Texas Commission on Environmental Quality Responses to Peer Review Report

April 5, 2012

Texas Commission on Environmental Quality
Toxicology Division
Chief Engineer's Office

Prepared by:

Toxicology Division
Chief Engineer's Office

Comments from Panel Members and the Public

The August 31, 2011 Report of a Letter Peer Review of the Texas Commission on Environmental Quality's (TCEQ) updates to its Guidelines to Develop Inhalation and Oral Cancer and Non-Cancer Toxicity Factors (TERA 2011) (hereafter referred to as the Peer Review Report) is available on the following TERA website: http://www.tera.org/peer/tceqesl/.

The Toxicology Division (TD) of the TCEQ has prepared responses to comments from panel members and the public. The TD would like to thank the peer review panel for their careful review of the updated inhalation and oral cancer and non-cancer methodology document and their constructive suggestions. After careful review and due consideration, the TD revised the Public Comment Draft of the Guidelines document to incorporate suggested revisions of the panel members and/or public which increased the scientific and technical merit of the Guidelines or clarity for agency staff. Panel member comments are addressed in the order they appear in the Peer Review Report. Similar comments that were recorded multiple times have generally been addressed only once in order to avoid needless repetition of the same issues. Comments from the public and other interested parties are then addressed.

Comments and responses are organized in the following manner to be consistent with the Peer Review Report:

- 2.0 Peer Reviewer Comments
  - 2.1 General Issues
  - 2.2 Specific Issues for Evaluating Mode Of Action (MOA)
  - 2.3 Specific Issues for Hazard Characterization and Exposure-Response Assessment Using Epidemiology Data
  - 2.4 Specific Issues for Acute ESLs
  - 2.5 Specific Issues for Oral Cancer and Noncancer
  - 2.6 Specific Issues for Inhalation Cancer and Noncancer
  - 2.7 Specific Issues for Vegetation
  - 2.8 Specific Issues Related to Odor
- 3.0 General/Editorial Comments
- Appendix B: Public Comments
  - B.1 Public Comment 1
  - B.2 Public Comment 2
  - B.3 Public Comment 3

2.0 Peer Reviewer Comments

2.1 General Issues

2.1.1

Does the guidance reference the most current, valid, and generally accepted federal or state guidance documents or key papers (Section 3.1)?

Reviewer 1

Yes it does. The exception is it ignores some items from the National Research Council. This will be discussed later.

Response: No response required.

Reviewer 2

It appears that the guidance documents and key papers cited by the TCEQ toxicity factors guidelines are the most current and generally accepted. I know of no more recent citations that should be used in lieu of those referenced by the guidelines.

Response: No response required.

Reviewer 3

The guidance document is well written and definitely uses the most current references.

Response: No response required.

Reviewer 4

In my opinion section 3.1 does reference the most important guidance documents that provide an overview of the risk assessment process. I recommend that the 1986 cancer guidelines and 1996 and 1999 draft cancer guidelines be relocated from section 3.1.1 to 3.1.6.

Response: TCEQ agrees and appropriate changes have been made.

Reviewer 5

Section 3.1 of the guidance provides the most current and generally accepted Federal and State guidance, and key papers. This section could be expanded to include the guidelines for derivation of AEGLs (these are referenced in this section and elsewhere in the document) because they constitute the most formalized body of procedures for addressing acute exposure levels, and provide some insights on how to address issues of insufficient data. Also, please add CalEPA to the OEHHA citation in this section (as cited in line 27 on page 109) and to the glossary.

Response: For the purpose of the Guidelines, the references provided for AEGL and CalEPA are appropriate.

Reviewer 6

Review of Section 3.12 pages 83-93.

Based on my review of the section, the generally accepted federal and state guidance documents, and key papers are included and adequately referenced. There will always be some specific papers, new concepts and journal articles that are not represented. This is minimal at this time with the major well accepted papers, guidance and concepts well represented and cited. Rating=4. As requested, to quantify relative strengths and weaknesses, I have used a positive 5 point system for relative strength (1 is very weak, 2 is weak, 3 is consistent with accepted practice and documentation, 4 is strong, 5 is very strong).

Response: No response required.

Reviewer 7

I assume you are referring to Section 3.1 and not Section 3.12. The list of federal and state documents appears to be complete. European documents based on REACH and WHO approaches are not included, nor should they be, since the title of the section refers only to federal and state documents. I note, however, that WHO documents and other international documents are often cited in the text. I think you might consider adding a section on international documents of interest.

Response: TCEQ agrees and has included a brief section on international documents of interest.

Reviewer 9

The appropriate government guidance documents are cited (although it seems more appropriate to list the RfC guidance with risk documents (as a key reference for ESL derivation), rather than "other guidance documents"). If the distinction is guidance or methods vs. guidelines, maybe the header for the risk guidelines could be broadened.

A recommended addition – from the reference list of the draft ESL document:

- *United States Environmental Protection Agency (USEPA 2011). Recommended Use of BW ¾ as the Default Method in Derivation of the Oral Reference Dose. EPA/100/R11/001. Office of the Science Advisor. Washington, D.C. http://www.epa.gov/raf/publications/pdfs/recommended-use-of-bw34.pdf*

Response: The suggested reference was added appropriately.

More importantly, there are a number of key reference works from ILSI and IPCS that should be listed – some of which (but not all) are cited elsewhere in the ESL draft guidance:

- *IPCS (International Programme on Chemical Safety) (2005). Chemical-specific adjustment factors for interspecies differences and human variability: guidance document for use of data in dose/concentration–response assessment. (http://whqlibdoc.who.int/publications/2005/9241546786_eng.pdf)*
- *IPCS (International Programme on Chemical Safety) (2006) IPCS framework for analysing the relevance of a cancer mode of action for humans and case studies. (http://www.who.int/ipcs/methods/harmonization/areas/cancer_mode.pdf)*
- *Boobis, AR; Doe, JE; Heinrich-Hirsch, B; Meek, ME; Munn, S; Ruchirawat, M; Schlater, J; Seed, J (2008). IPCS Framework for Analyzing the Relevance of a Noncancer Mode of Action for Humans, Critical Reviews in Toxicology 38:87-96.*

- *Meek, M; Bucher, J; Cohen, S; et al. (2003). A framework for human relevance analysis of information on carcinogenic modes of action. Critical Reviews in Toxicology 33:581-653.*

Response: TCEQ agrees and all appropriate changes were made. References were cited in the last paragraph of Section 7.2.

Reviewer 10

This is a very useful compilation of the discipline references. Care should be taken about using some of the older references and tempered with more recent knowledge as necessary.

Response: TCEQ agrees with the reviewer's comments.

Reviewer 11

The guidance documents listed in Sections 3.1.1 to 3.1.6 are identical to those listed on the USEPA's IRIS website and generally represent the most up to date EPA guidance pertaining to the development of toxicity factors. I would suggest adding the two most recent EPA Risk Assessment Forum publications Draft Guidance for Applying Quantitative Data to Develop Data-derived Extrapolation Factors for Interspecies and Intraspecies Extrapolation (USEPA 2011a) and Recommended Use of the Body Weight ¾ as the Default Method in Derivation of the Oral Reference Dose (USEPA 2011b). The authors discuss and cite the 2nd new document in Chapter 5, but should also include it here in Section 3.1. The ATSDR guidance document on the development of toxicity values is also included in Section 3.1 which is helpful.

Response: TCEQ agrees with the reviewer's comments.

Reviewer 12

As indicated in the General comments [see section 3.0 below], this has been accomplished to a degree beyond what would be expected. Importantly, these reports have been implemented in this guidance document with reasoning and justification that allows for correct and direct application.

Response: No response required.

Reviewer 14

Section 3.1 of the TCEQ Guidelines provides a list of federal and state guidance documents for developing toxicity risk factors that are current, valid, and generally accepted. This list is so good that I am going to refer to it for my future projects.

Response: No response required.

Reviewer 15

To the best of my knowledge, this guidance is complete and thorough, even exhaustive, in its coverage of relevant guidance on development of toxicity criteria available in the United States and Europe. It seems almost impossible that anything was left out, but I found no mention of probabilistic methods, which was surprising.

Response: No response required.

2.1.2

Are the procedures for addressing the differences between children's and adult's susceptibility to risk (Chapters 3 and 5) appropriate and consistent with accepted risk assessment methods? Procedures for addressing children's risk are found throughout the different sections of Chapters 3 and 5. In particular consider the following specific recommendations:

2.1.2.1 Part A.

The definition of child as conception to 18 years of age

Reviewer 1

It is appropriate.

Response: No response required.

Reviewer 2

Even though the ages of childhood are defined in Chapter 3 to be conception through age 18 years, there seems to be some instances where this does not apply. For example, when considering a mutagenic carcinogen in Chapter 5, the lifetime risk calculation shows age-adjustments to the URF only up to age 16. I am not saying that that is inappropriate, but it does appear to be inconsistent with the blanket statement in Chapter 3 that childhood lasts until age 18. In fact, because of the extent of the discussion of the various factors that affect lifestage-related differences, and the fact that their rate of "maturation" is different, I would have expected the guidelines to avoid a strict definition of "childhood" and recommend adjustments based on what is known about the lifestage differences of the processes relevant to the toxicity in question.

A related question is: how were the adjustment factors for the URF (10 for ages less than 2 years, 3 for ages 2 to 16; see end of p. 162) selected? There is only a brief reference to the 2005 Supplemental Guidance. This is in contrast to the extensive discussion of other adjustments and age-specific considerations given elsewhere in the guidance.

Response: The definition of children has been revised to "conception to maturation." In addition, the reference for the ADAFs has been specified in Chapter 5. Section 7.9.4 has also been modified; the reader is referred to the USEPA guidance for information relevant to the related question.

Reviewer 3

[Please see overall comments under the section 2.1.2.5 Part E below].

Response: No response required.

Reviewer 4

Section 3.2 provides a fairly balanced discussion of the various issues related to differences in child and adults and their response to toxicants. (Note: P. 39 l 13 uses toxins instead of toxicants). The definition of a child is acceptable and Table 3-1 identifies and characterizes the important life stages that should be considered.

Response: The text has been revised to "toxicants" instead of "toxins."

Reviewer 5

[Please see overall comments under 2.1.2.5 Part E below].

Response: No response required.

Reviewer 6

Acceptable; Rating=4. As requested, to quantify relative strengths and weaknesses, I have used a positive 5 point system for relative strength (1 is very weak, 2 is weak, 3 is consistent with accepted practice and documentation, 4 is strong, 5 is very strong).

Response: No response required.

Reviewer 7

I prefer the definition that includes from conception to "attainment of physical and sexual maturity." (as stated in the footnote on page 8 of Appendix C, the white paper).

Response: See response to Reviewer 2 above.

Reviewer 9

In general the recommendations are appropriate.

Response: No response required.

Reviewer 10

[Please see overall comments under 2.1.2.5 Part E below].

Response: No response required.

Reviewer 11

[No comments specific to this bullet].

Response: No response required.

Reviewer 12

Although it is not clear from the footnotes to Table 3-1 which EPA documents are being referred to here (2005 a, b, or c?; 2006 a or b?), the upper portion of the age definition at 18 should be better justified, perhaps by examining and referring to Table 3-1 in the US EPA's RfD/RfC revision document (http://www.epa.gov/raf/publications/pdfs/rfd-final.pd) which presents a species age comparison of specific life events and life-stage using terms of "young", "puberty" and "sexual maturity". The age of 18 falls within the range of puberty (12-21) but just outside the range of and sexual maturity (21-40). Therefore the figure of 18 years as an upper range is consistent with a number of sources and could be considered defensible.

The identification of the lower end of the range of zero (i.e., inclusive of infants) could also be justified. This is perhaps harder to accomplish than the upper limit as life-processes and overall development status are so markedly different between infants (30 days – 1 yr in Table 3-1 in US EPA, 2002) and the upper age limit. As relevant dosimetric and dynamic results accumulate, however, (as documented in

the references cited and from the figures in your white paper) it is becoming more and more apparent that inclusion of the infant age range is justified as the lower age limit for this definition.

Response: References in Table 3-1 were corrected to EPA 2006b and EPA 2005d. Additional references were added to justify the upper age limit of 18 years and others describing other upper age limits of adolescence.

Reviewer 14

The definition of childhood as conception to 18 years of age is widely used

Response: No response required.

Reviewer 15

[Please see overall comments under 2.1.2.5 Part E below].

Response: No response required.

2.1.2.2 Part B.

The definitions of susceptible, sensitive, and vulnerable

Reviewer 1

I see the definitions for susceptible and sensitive on page 38. These definitions are appropriate. I do not see the definition for vulnerable. It should also be on the same page.

Response: In general, the term susceptible is used rather than sensitive or vulnerable. The idea is to use one term and avoid confusion arising from multiple, nearly equivalent, terms. This section was revised and the word vulnerable is now discussed.

Reviewer 2

In general I thought the discussion of and proposed approaches to handle differences between children's and adult's susceptibility to risk to be satisfactory. This is not my area of expertise, so my comments may be a bit rudimentary.

Response: No response required.

Reviewer 3

[Please see overall comments under 2.1.2.5 Part E below].

Response: No response required.

Reviewer 4

I think there is a problem with the definitions of susceptible, sensitive, and vulnerable. I did not find a TCEQ definition for vulnerable in this section. The definitions used for susceptible and sensitive are not mutually exclusive so the use of the terms remains unclear.

Response: This section has been revised, and use of the terms susceptible and sensitive have been reviewed throughout the document to ensure they are consistently applied.

Reviewer 5

[Please see overall comments under 2.1.2.5 Part E below].

Response: No response required.

Reviewer 6

Acceptable; Rating=4. As requested, to quantify relative strengths and weaknesses, I have used a positive 5 point system for relative strength (1 is very weak, 2 is weak, 3 is consistent with accepted practice and documentation, 4 is strong, 5 is very strong).

Response: No response required.

Reviewer 7

I would reverse the definitions of sensitivity and susceptibility given on page 38, lines 5-10. Sensitivity should refer to innate biological differences while susceptibility should refer to tendency to get a disease or adverse effect (which will include both sensitivity and exposure). For example, a fair-skinned person may be sensitive to the sun but they are only susceptible to getting melanoma if they stay out in the sun for a long time. Webster's dictionary defines sensitive as being highly responsive to an agent, while susceptibility is defined as capable of being affected by something. Being affected by something (i.e., developing a disease or an adverse effect) must involve both sensitivity and a high degree of exposure. As you say on p.38, line 5, it is most important that you define what you mean and be consistent in the use of the terms. In reading the text, I noticed that in most places your use of "sensitivity" is consistent with my definition, but not yours. See page 39, line 26, and page 42, line 40, Table 3-2, page 51, line 16; page 52, line 19, for example.

I did not see where the term "vulnerable" was defined. However, i have no objection to the manner in which you have used the term.

Response: See responses to Reviewers 1 and 4 above.

Reviewer 9

In general the recommendations are appropriate.

Response: No response required.

Reviewer 10

[Please see overall comments under 2.1.2.5 Part E below].

Response: No response required.

Reviewer 11

I don't recall seeing a definition of "vulnerable" but the definitions of sensitive and susceptible sounded reasonable and are consistent with definitions elsewhere (e.g., EPA's IRIS glossary, Wikipedia).

Response: See response to Reviewer 1 above.

Reviewer 12

Though not totally in synchrony with all definitions this reviewer has seen, these terms are clearly defined by the authors and these definitions held to in the various occurrences throughout the text. Therefore they are acceptable.

Response: No response required.

Reviewer 15

[Please see overall comments under 2.1.2.5 Part E below].

Response: No response required.

2.1.2.3 Part C.

The identification of critical lifestages

Reviewer 1

Table 3-1 is an excellent description.

Response: No response required.

Reviewer 2

In general I thought the discussion of and proposed approaches to handle differences between children's and adult's susceptibility to risk to be satisfactory. This is not my area of expertise, so my comments may be a bit rudimentary.

Response: No response required.

Reviewer 3

[Please see overall comments under 2.1.2.5 Part E below].

Response: No response required.

Reviewer 5

[Please see overall comments under 2.1.2.5 Part E below].

Response: No response required.

Reviewer 6

Acceptable; Rating=4. As requested, to quantify relative strengths and weaknesses, I have used a positive 5 point system for relative strength (1 is very weak, 2 is weak, 3 is consistent with accepted practice and documentation, 4 is strong, 5 is very strong).

Response: No response required.

Reviewer 7

I found no problems with your identification of lifestages.

Response: No response required.

Reviewer 9

In general the recommendations are appropriate.

Response: No response required.

Reviewer 10

[Please see overall comments under 2.1.2.5 Part E below].

Response: No response required.

Reviewer 11

I don't have a problem with TCEQ's definition of childhood or critical lifestages as shown in Table 3-1. However, the text in Section 3.2 and references provided for Table 3-1 give the appearance that these definitions are the same as the EPA's. They are not. Both of the EPA guidance documents cited for Table 3-1 actually include ages 16 – 21 in the definition of childhood and define the critical lifestages just a little bit differently. For example, the EPA guidance considers the infant stage as birth to <12 months, while TCEQ considers the infant stage as birth to <3years. I think the way TCEQ has defined childhood and critical lifestages is reasonable, however, the document would benefit from a more transparent explanation of what was done and why (including how they differ from EPA). For editorial purposes, it would be helpful if references were correctly cited. For example Table 3-1 cites USEPA 2005 and 2006. Yet there are several USEPA 2005 and 2006 references provided in the Reference Section. Also, the authors might consider including the USEPA Guidance on Selecting Age Groups for Monitoring and Assessing Childhood Exposures to Environmental Contaminants for completeness (USEPA, 2005).

Response: Additional references were added to Table 3-1 to clarify age ranges for various life stages and EPA references were corrected.

Reviewer 12

As is indicated in my general commentary [see section 3.0 below] on consideration of children's issues in this report, there exists a clear bias towards considering children always and foremost. It would seem that wording suggesting that both development of relevant processes during youth and their diminution during advanced age would result in and define critical lifestages in a more equitable, realistic and unbiased manner.

[In response to a clarifying question] This reviewer contends that there exists a bias in the document towards considering children always and foremost as a susceptible lifestage/group before all others which could and often include gender susceptibility and other lifestages such as pregnancy and advanced age. Diminution of life processes in advanced age may well be a more toxicologically sensitive period than development of life processes during infancy. A manner in which this bias could be eliminated would be to add wording that is inclusive of these groups.

Response: TCEQ's toxicity factors are set to protect the most sensitive individuals in a population, including children. The specific sections in the guidance document addressing the differences between adults and children are designed to highlight one potentially sensitive subgroup of the population, not to exclude other sensitive or susceptible groups. Sections discussing differences between children and adults have been revised to clarify the intent of the Guidelines.

Reviewer 14

Table 3-1 provides a good summary of life stages.

Response: No response required.

Reviewer 15

[Please see overall comments under 2.1.2.5 Part E below].

Response: No response required.

2.1.2.4 Part D.

The discussion of toxicokinetic and toxicodynamic differences between children and adults

Reviewer 1

I found this to be a very interesting section, but of necessity too short. I think there are very few generalizations that can be made, and these differences must be demonstrated for each agent.

Response: No response required.

Reviewer 2

In general I thought the discussion of and proposed approaches to handle differences between children's and adult's susceptibility to risk to be satisfactory. This is not my area of expertise, so my comments may be a bit rudimentary

Response: No response required.

Reviewer 3

[Please see overall comments under 2.1.2.5 Part E below].

Response: No response required.

Reviewer 4

I think it is an important step forward to discuss toxicokinetics and toxicodynamic differences between children and adults. I suggest that TCEQ use these terms to help the reader understand the points you are making. I think of toxicokinetic differences to give a different internal dose at the target tissue from the same external exposure; I think of toxicodynamic differences as giving a different response at the same internal dose in the target tissue.

Response: No response required.

Reviewer 5

[Please see overall comments under 2.1.2.5 Part E below].

Response: No response required.

Reviewer 6

Acceptable; Rating=4. As requested, to quantify relative strengths and weaknesses, I have used a positive 5 point system for relative strength (1 is very weak, 2 is weak, 3 is consistent with accepted practice and documentation, 4 is strong, 5 is very strong).

Response: No response required.

Reviewer 7

I saw no problems with the discussion of adult/child differences in toxicokinetics and toxicodynamics.

Response: No response required.

Reviewer 9

In general the recommendations are appropriate.

Response: No response required.

Reviewer 10

[Please see overall comments under 2.1.2.5 Part E below].

Response: No response required.

Reviewer 11

[No comments specific to this bullet].

Response: No response required.

Reviewer 12

Some specific comments have been made. In general, though, the guidance follows the content and concepts in the accompanying white paper which is quite an acceptable effort in this area. This reviewer considers this particular text as balanced (i.e., largely unbiased) in that it points out that children are not always more sensitive than adults.

Response: No response required.

Reviewer 14

It would be useful if numerical default values were added to Section 3.2.2 for toxicodynamic differences between children and adults.

Response: Toxicokinetic and toxicodynamic differences between children and adults are discussed in detail in Chapter 3.

Reviewer 15

[Please see overall comments under 2.1.2.5 Part E below].

Response: No response required.

2.1.2.5 Part E.

Other comments related to child/adult differences in risk.

Reviewer 2

In general I thought the discussion of and proposed approaches to handle differences between children's and adult's susceptibility to risk to be satisfactory. This is not my area of expertise, so my comments may be a bit rudimentary.

Response: No response required.

Reviewer 3

I found the procedures and comments on the differences between children and adults contradictory in some places. On page 37, children are defined "from conception to 18 years of age. Later, it is stated that no age adjustments are made for people over 16. This should be consistent. As either number is somewhat arbitrary, I would use 18 as it appears most often. I also consider the fetus unique and would not include it in the children category. Exposures to the fetus occur as a secondary event following exposure to the mother. As such, the fetus is exposed to metabolites and exposure to the parent compound can vary from high to non-existent depending on the volatility, solubility, and reactivity of the substance. The descriptions on pages 37-42 is excellent and concludes that differences can result in children being either more or less sensitive than adults, therefore, uncertainty factors should be evaluated on a case by case basis (P. 38 L 29-33; P. 39 L 29-36; P. 41 L 1-14; P. 42 section 3.2.4; and P. 78 L15-30). However, on P. 86 L 34-41 application of default UFs are described. The assumption in this section is that children are more sensitive than adults. These comments negate much of the earlier discussion. The other points in section 2 were well covered.

Response: The word "potential" was used to describe the sensitivity of children compared to adults and UFs will be applied on a chemical-by-chemical basis. Thus, the UFs reflect the availability of data regarding the relative susceptibility of children compared to adults. In addition, the reference for the age adjustments specified in Chapter 5 was added. The upper age limit of 16 years is derived from an EPA guidance and is specific to carcinogens acting through a mutagenic MOA.

Reviewer 4

In my opinion, many of the examples in section 3.2.2, P41 L37 and continuing to page 42, do not deal with toxicodynamic differences.

Response: This section of the document has been revised accordingly.

Reviewer 5

Sections describing differences between children and adults susceptibility are well described and consistent with current risk assessment practice. The definition of children from conception to age 18 is consistent with Texas' practice (age 21 is considered the upper limit at the Federal level). So are the definitions of sensitive, susceptible, and vulnerable populations. The identification of life stages is also appropriate. Differences in toxicokinetics and toxicodynamics between children and adults are presented in sufficient detail. Addition of a table summarizing the key differences would help follow the discussion on these differences. Some discussion of the interplay between environment (e.g., nutritional status) and toxicant effects in children and adults may be warranted. For the pre-natal life stage, it would be useful to add some explicit discussion on primary and secondary effects on the fetus, and the guidelines to address them.

Response: Rather than adding a table to highlight key differences between toxicokinetic and toxicodynamic differences between adults and children, a definition of toxicodynamics has been added to clarify this section. The other suggestions are beyond the scope of this document and/or will be discussed as applicable in the chemical DSDs.

The guidelines for identifying key studies are well presented. In section 3.4.3.4, the authors may consider modifying the text in lines 6-7 to "lack of biological plausibility" instead of "causal relationships" because epidemiology studies are generally not suitable for establishing robust cause-effect associations.

"Population Exposure Studies" may be a better subtitle for section 3.4.3.6.1 than field studies.

Section 3.4 does not mention other sources of information (which is mentioned in other parts of the report), such as non-peer review reports of studies by private companies that may provide information not available elsewhere.

Response: The suggestions have been added accordingly.

Reviewer 6

Acceptable; Rating=4. As requested, to quantify relative strengths and weaknesses, I have used a positive 5 point system for relative strength (1 is very weak, 2 is weak, 3 is consistent with accepted practice and documentation, 4 is strong, 5 is very strong).

Response: No response required.

Reviewer 7

I am not sure I understand this question, but I found that you gave adequate references to documents that describe the potential differences between adults and children in response to toxic agents.

Response: No response required.

Reviewer 9

In general the recommendations are appropriate. WHO (2006) gives an extensive review of the developmental timing of various organ systems and is worth citing for that information.

Response: WHO (2006) is included as a reference and is cited several places throughout the document for reasons the reviewer pointed out.

Reviewer 10

This section is well written and provides a good introduction to the factors that must be considered when looking at children. If provides general principles that should be considered when appropriate and data are available without resorting to an inflexible dogmatic approach.

Response: No response required.

Reviewer 11

The discussion of what toxicokinetics is and how the differences between children and adults can be determined qualitatively or quantitatively was adequate. The discussion of toxicodynamics was not as clear. In the toxicodynamics section it would have been helpful to begin with a definition of toxicodynamics. For example, the determination and quantification of the sequence of events at the cellular and molecular levels leading to a toxic response to an environmental agent. Toxicodynamic variability within the human population is calculated as the relationship between concentrations or dose metric values producing the same level of the response in the general population and in susceptible groups or individuals. Human studies which demonstrate a given level of response at a lower

concentration in the susceptible population as compared to the general population are one example of how toxicodynamic differences can be identified. In vitro data (e.g., genetic polymorphisms) representing the two populations is another approach. The Guidance for Applying Quantitative Data to Develop Data-Derived Extrapolation Factors for Interspecies and Intraspecies Extrapolation (USEPA 2011a) may be helpful in providing more easily understood definitions and examples of toxicodynamics.

Response: See response to Reviewer 5 above.

Reviewer 12

The guidance on study identification is both thorough and thoughtful. No comments are offered.

Response: No response required.

Reviewer 14

Guidelines on how to identify studies that describe toxicokinetic and toxicodynamic between children and adults are adequately described in Section 3.4.3.

Response: No response required.

Reviewer 15

The recommended methods for assessing children's risk are consistent with the best methods available, which I believe to be those of USEPA and California EPA. Like all other topics addressed in this guidance document, the subject is covered in exhaustive detail. The only topic on which I take issue is the division of the period from birth to 18 years into so many time periods, each with somewhat different methods for estimating risk. In my opinion, the evidence for this many divisions is rather weak, even though the treatment in the guidance is consistent with published materials from USEPA and Cal/EPA. The definitions of susceptible, sensitive, and vulnerable are rather prescriptive, quite possibly allowing the risk assessing inadequate latitude for addressing novel situations.

Response: See response to Reviewer 2 in 2.1.2.1 Part A above. Also, language was revised to clarify the purposes of Table 3-1 which are to inform the TCEQ when assessing toxicity data and determining appropriate uncertainty factors.

2.1.3

Does the guidance clearly describe the approaches to be used by TCEQ, and provide supporting rationale, in situations where it employed different procedures than those recommended in referenced federal or state guidance documents?

Reviewer 1

Yes.

Response: No response required.

Reviewer 2

The approaches to be used by TCEQ are clearly described. The discussion of the reasons for deviating from other procedures was one of the highlights of these guidelines; I found the support for those deviations to be very well developed and presented.

Response: No response required.

Reviewer 3

While the guideline does an excellent job of describing the approaches TCEQ should take and clearly references and summarizes Federal Guidance; in the application of UFs, it appears that the composite UF for many chemicals will be very large as a consequence of multiplying the many (5) UFs together. In many cases in the document, the argument is presented that large UFs are often not required. However, these excellent descriptions could be negated by the application of multiple UFs. In the development of AEGL values, the committee often limited itself to two UFs, one for animals to man and the second to include sensitive members of the population. In a few cases a modifying factor was also used to account for a poor data base or using a LOAEL instead of a NOAEL. Using a range of 1, 3 or 10 for each of the two factors and 2 or 3 for the modifying factor, limits the total uncertainty to a maximum of 300. For chemicals with a reasonable data base (GLP multiple acute, 28 day and possibly other data), we would not use a modifying factor. However, we are developing acute exposure guidance levels. One of the most recent reviews on data extrapolation (ECETOC Guidance on Assessment Factors to Derive DNELs, Technical Report 110 page 17, 2010) recommends a factor of 3 for subacute to subchronic, 2 for subchronic to chronic and 6 for subacute to chronic, this was based on reviews of a large number of studies. The discussion on P. 65 l 12-29 on the BMDL is noteworthy.

Response: TCEQ agrees with these comments and has added that TCEQ will generally use a maximum total UF of 300 for the acute POD.

Reviewer 4

I believe TCEQ has provided its rationale when it advocates using different procedures from referenced federal or state guidance documents. Although I do not necessarily agree with TCEQ's choice, these differences are science policy decisions.

Response: No response required.

Reviewer 5

In general, the document does a good job indicating deviations from other procedures and providing the rationale. However, need to be highlighted by using italics or boldface.

Response: TCEQ has already addressed the reviewer's comment to the extent possible.

Reviewer 6

The guidance clearly describes the approaches to be used by TCEQ, and the supporting rationale. Rating=3. As requested, to quantify relative strengths and weaknesses, I have used a positive 5 point system for relative strength (1 is very weak, 2 is weak, 3 is consistent with accepted practice and documentation, 4 is strong, 5 is very strong).

Response: No response required.

Reviewer 7

The general answer to this charge question is yes, the guidance is clear. But I have noted a few places where I think the guidance can be made clearer.

I found a problem with one part of Table 3-2. In the row titled "seriousness of the endpoints," I do not find it acceptable to list "less serious" as the reason for low concern. The term "less serious" is not informative. I think you could at least list "irreversible effects" under the column for high concern, and then list "reversible effects" or "adaptive responses" under the column for low concern. I would think irritation is another effect that might be listed under low concern. Later in the document, more specific listings of less serious effects are given (see 3.7.1.4.2)

Response: TCEQ agrees and has modified the row titled "seriousness of the endpoints" of Table 3-2 accordingly.

I was not sure as to the strict definition of "toxicity factor," a term used frequently in the document. The term is used in the title of the document and is described in a general sense in Chapter 1. But I did not find where there was a list of terms included under oral or inhalation toxicity factors. Perhaps you could list the definition of toxicity factors in your glossary of terms at the beginning of the document (page 4).

Response: TCEQ agrees and has added a definition for toxicity factors.

Reviewer 9

In general the recommendations are appropriate.

Some specific comments:

- *I would recommend that the literature search strategy include review of existing risk values (as noted elsewhere in the document), particularly since TCEQ looks at other risk values for appropriateness, as part of its strategy for leveraging resources.*

Response: TCEQ already has a check list that includes a comprehensive review of all the major risk values as part of the literature.

- *P. 51, l 12: I would disagree with the framing of the burden of proof for human relevance. As shown in the IPCS framework, rather than animal effects "must be relevant," animal effects are excluded if they are shown not to be relevant (but are included if it is uncertain or the MOA is unknown). Further, since this analysis can be conducted only after the MOA is evaluated, it may make sense to present Figure 3-1 after the discussion of evaluation of MOA.*
- *Section 3.1.2.2.1 mentions USEPA severity grades, but does not provide a reference. Please include the reference.*
- *Section 5.5 in general seems focused on child-adult differences, rather than issues for UFs specific to chronic values.*

Response: Section 5.5 has been revised accordingly.

- *Section 5.5.1 is almost entirely focused on age-related differences in sensitivity (i.e., children's risk). This is only one of many contributors to human variability. Other factors affecting sensitivity should be mentioned, including toxicokinetic variability (including genetic polymorphisms and lifestyle-related factors such as CYP 2E1 induction from alcohol consumption), and background disease (e.g., COPD, asthma).*

Response: See response to Reviewer 12 in 2.1.2.3 Part C above.

- *It would be useful for Sections 3.1.2.3.2 and 5.5.2 to note the key question addressed by this UF with respect to data gaps: would additional studies identify a different critical effect? (Many of the questions listed in 3.1.2.3.2 can be thought of as questions that assist in addressing this over-reaching question.)*
- *The variations from EPA's guidance for the approach for the UF for LOAEL to NOAEL appear appropriate and reasonably well justified from the literature.*

Response: No response required.

- *The approach for the database UF in Table 5-2 appears to be a useful extension/enhancement of current practice. Several aspects of the table would benefit from further explanation:*
  1. *While the potential UFD values listed are consistent with current practice, the use of these numbers for the specified databases would be strengthened by citation of relevant supporting analyses:*
     - *Dourson, M.L., L. Knauf and J. Swartout. 1992. On reference dose (RfD) and its underlying toxicity database. Tox. Ind. Health 8(3): 171-189.*
     - *And a more recent follow up analysis, including chemicals other than pesticides at: http://www.tera.org/Peer/UFD/UFDWelcome.htm.*
  2. *Where ranges are given, is the intent to use only the bounds of the ranges (i.e., 1-3 is really 1 or 3?) If not, how would numbers be chosen within the range and how does TCEQ intend to ensure a consistent approach?*
  3. *It would be useful also to give some rationale as to the use of 6 as a further intermediate value. Presumable it was chosen as the next arithmetic intermediate between 3 and 10, but no rationale is given.*

Response: The database UF is chosen based on confidence in the information available for the chemical of interest as discussed in Chapter 5. The intent is to provide the flexibility based on best scientific judgment to use any number within the range of values in Table 5-2 (i.e., 1, 2, 3….10), rather than the bounds of the ranges.

Reviewer 10

See response to charge question 2.4.8.

Response: No response required.

Reviewer 11

For the most part, this document is utilizing the approach used by USEPA for the development of chronic oral toxicity values. There were some minor differences such as not including a modifying factor along with the uncertainty factors (Section 3.12.4), limiting cumulative uncertainty to 3000 (Section 3.13), and developing toxicity values for chemicals with limited toxicity data (Section 3.16). The rationale for these different approaches is clearly described and sounds reasonable.

Response: No response required.

Reviewer 12

The General Comment [see section 3.0 below] contains statements regarding the clarity and thoroughness in the manner in which this is accomplished in the guidance.

Response: No response required.

Reviewer 14

The TCEQ Guidance Document clearly describes the approaches to be used and provides documentation where procedures differ from federal or state guidelines, e.g., Section 3.12.2.2.

Response: No response required.

Reviewer 15

The guidance certainly does describe every conceivable method and provides well documented rationale for each. The draft document actually reads more like a textbook than a guidance document. I found very few deviations from Federal guidance, and none of major consequence.

Response: No response required.

2.1.4

Are the recommendations for determining adverse from non-adverse effects and the approach to determining if a response in a toxicology study is adverse or non-adverse appropriate and consistent with accepted risk assessment practices (Section 3.7.1)?

Reviewer 3

The description of the criteria for the differentiation of adverse from non-adverse effects is well presented and consistent with risk assessment procedures. Care should be taken to assure that these procedures are carried out in practice as well. The comments on P. 55 l 26-33 define an excellent approach.

Response: No response required.

Reviewer 4

I believe the practices summarized in Section 3.7.1 are consistent with accepted risk assessment practices. Although the section does not clearly state that TCEQ is adopting these practices, I assume that is the intent as there is no discussion of deviation from the practices summarized here.

Response: No response required.

Reviewer 5

The recommendations are appropriate and consistent with current practice.

Response: No response required.

Reviewer 6

The recommendations for determining adverse from non-adverse effects and the approach to determining if a response in a toxicology study is adverse or non-adverse appropriate is consistent with accepted risk assessment practices. Rating=3. As requested, to quantify relative strengths and

weaknesses, I have used a positive 5 point system for relative strength (1 is very weak, 2 is weak, 3 is consistent with accepted practice and documentation, 4 is strong, 5 is very strong).

Response: No response required.

Reviewer 7

In general, the answer is yes. But there is one sentence in Section 3.7.1.1 that is in error. The sentence on page 55, lines 33-36, mixes apples and oranges. Statistical significance is required to determine if an effect has actually occurred, i.e., that the observation is not due to chance alone. Once a statistically significant effect has been observed, then one can evaluate whether the effect is biologically significant. This is well illustrated in Figure 3-4. But it is not true that one can have a biologically significant response that is not statistically significant. However, one can have a statistically significant response that is not biologically significant. The article by Lewis et al., 2002 was making the latter point, not the former. (See Figure 3-4 and also see page 63 for an accurate description of the distinction between statistical significance and biological significance.)

Response: TCEQ agrees and has made the recommended revision to this section.

Reviewer 10

This section reflects the fact that judgment must be exercised to make this determination.

Response: No response required.

Reviewer 11

Although definitions of what constitutes an adverse effect are readily available, the application of that definition to a toxicological study is commonly subjective and problematic. Section 3.7.1 attempts to provide more definitive information on how to identify an adverse effect. The guidelines and recommendations provided in this section (e.g., biological vs. statistical significance, adaptive responses, precursor effects, etc.) are scientifically sound and consistent with toxicological practices.

Response: No response required.

Reviewer 12

A number of specific comments and specific suggestions have been made in this section. In particular this reviewer considers that addition of Schulte-type diagrams would be more suitable than that presently in this section.

Response: TCEQ agrees; a figure adapted from Julien *et al.* (2009) was added. This figure is a Schulte-type diagram and should address the reviewer's concern.

Reviewer 15

I found these sections to be quite thorough, but overly prescriptive. Every peer review panel on which I have sat has needed to address the issue of the borderline between adaptive and adverse effects. Regardless of any guidance TCEQ might publish, risk assessors and peer reviewers will bring their own prejudices to this topic. I believe this openness is appropriate for a topic so difficult and controversial.

Response: TCEQ will exercise best scientific judgment.

I found that the treatment of precursor effects is not adequate. Definitions of precursors are good, but the guidance needed is how to apply uncertainty factors when the critical effect is a precursors. TCEQ should offer some wisdom here in order to prevent another debacle like perchlorate. RfD guidance says a precursor may be selected as the critical effect only if it is the immediate precursor of the toxic effect. As systems biology is applied to toxicity criteria, our profession will be examining ever earlier combinations and permutations of biological phenomena. It is necessary to avoid the practice of "lowering the bar" by applying UFs to very earlier precursors.

Response: TCEQ agrees and has added clarifying language in this section.

2.1.5

Section 3.7.1.4 provides a list of effects classified as not adverse, less serious, transitional, or serious. Do you agree with this classification of effects? Would you move any effect to different category or add an effect to any category?

Reviewer 3

The listing of effects and the degree of seriousness appears to be complete and should serve as guidance to a far wider audience than TCEQ.

Response: No response required.

Reviewer 4

This section provides useful general guidance on the severity of adverse effects. I would add a disclaimer that this is not an exhaustive list, but is included to provide a general overview. In general I agree with the classification scheme presented. However, some listed as Serious Effects could be considered transitional depending on the degree of the response in the test species. I recommend adding some wording that TCEQ will consider the degree of the response when distinguishing between Transitional and Serious Effects.

Response: TCEQ agrees and has added a disclaimer to this section and text as recommended.

Reviewer 5

Effects categories are generally reasonable and consistent with ATSDR's definitions. Please note that the ATSDR guidance is still in draft form (i.e., do not cite or quote). This section should include some text describing the rationale for each category, and the limitations for use of these categories as understood by ATSDR.

Response: TCEQ has revised this section to refer to ATSDR (2007) for detailed description of the rationale for each category, and the limitations for use of these categories.

Reviewer 6

The list of effects classified as not adverse, less serious, transitional, or serious is exhaustive. Rating=4. As requested, to quantify relative strengths and weaknesses, I have used a positive 5 point system for relative strength (1 is very weak, 2 is weak, 3 is consistent with accepted practice and documentation, 4 is strong, 5 is very strong).

Response: No response required.

Reviewer 7

I have some questions. Under less serious effects, what is meant by moderate serum chemistry changes? What is considered to be moderate? This should be specified.

Response: Serum chemistry changes can indicate the level of effects for specific organ function based on clinical observations. For example, liver function tests (LFT) are a helpful screening tool, which are an effective modality to detect hepatic dysfunction. The level of clinical significance can be indicated by elevation of aminotransferases (SGOT, SGPT). Levels of aminotransferases increased 1-3, 3-20, and > 20 times the normal ranges are typically considered mild, moderate, and severe, respectively (Thapa and Walia (2007)). These examples have been added as appropriate.

In two places (page 60, lines 1 and 13) you list weight loss or gain "assuming normal food consumption." I do not know how one could have weight loss unless food consumption was also decreased (I have known people on diets who wish that were possible!). What do you mean by "normal food consumption"?

Response: This section was based on Tables 3-1 to 3-17of the *Guidance for the Preparation of a Twenty First Set Toxicological Profile* (ATSDR 2007). The ATSDR indicates that if a decrease in body weight is accompanied by decreased food consumption in a feeding study, then neither effect is considered an adverse effect. Chemically-induced effects on the thyroid (e.g., hyperthyroidism) or on homeostatic functions (e.g., normal metabolism), for example, could cause body weight changes.

Reviewer 10

The categories presented are reasonable. As always judgment must be exercised. As noted, irreversible effects increase the concern level. If the assessment is based upon a POD for relatively minor effects, consideration should be given to how close that POD is to doses that cause severe toxic effects or a false sense of security may result. This is especially true for compounds with steep dose response curves.

Response: TCEQ agrees and has added "As always judgment must be exercised" as part of a disclaimer.

Reviewer 11

As stated in Section 3.7.1.4 this listing was copied from the ATSDR guidance document. I'm sure arguments could be made for moving effects up or down on the list depending on one's idea of severity. For example, what is meant by moderate serum chemistry changes? If you are talking about statistically significant elevations from the control, but still within normal physiological ranges, I would consider this to be a non-adverse effect. However, I think the list provides a reasonable starting point for classifying the severity of an effect. This classification is then used in the quantification of toxicity factors. The USEPA guidelines for RfD development recognize that the severity of an adverse effect should be considered in the application of the uncertainty factor for LOAEL to NOAEL extrapolation. However, EPA does not provide specific examples such as those shown here in Section 3.7.1.4. Consequently, severity of effect is seldom taken into consideration in the development of values for the IRIS database. I am encouraged to see this issue presented in detail in the TCEQ document in a manner that can be applied in practice.

Response: No response required.

Reviewer 12

Many such classifications have been advanced and found almost immediately to be invalid. This classification, however, appears to be a bit more in alignment with actual pathophysiology findings commonly encountered in toxicity studies. Also, the classifications are further qualified somewhat regarding whether they are found in conjunction or absence of related effects. Yet another attribute is the lack of a formal categorization value or numeric score that inherently and erroneously implies an internal quantitative relationship (e.g., a severity rating of 4 is 2x that of a severity 2).

Response: As indicated by ATSDR (2007), the list is to provide guidance for the derivation of its minimal risk levels (MRLs). However, this is not an exhaustive list, but is included to provide a general overview and as always best scientific judgment must be exercised.

However, the actual application of such a gradation of adverse effects (less serious, transitional) in estimation of a BMR or Uncertainty Factor determination for anything other than descriptive purposes remains and is considered to be inherently problematic.

Response: TCEQ agrees. The Bullet "Since the benchmark response (BMR) (1, 5, or 10%) or critical effect size at 1 SD ($CES_{1\ SD}$) should represent a response level of no significant concern, the corresponding benchmark dose lower confidence limit (BMDL) should be comparable to the NOAEL rather than the LOAEL" has been deleted.

Regarding specific recommendations, this reviewer finds the placement of "moderate" clinical enzyme alterations under "less serious" to be quite vague and lacking a counterpart in the preceding non-adverse category.

Response: TCEQ agrees. As always best scientific judgment must be exercised. In addition to the level of enzyme changes (moderate or major), both statistical significance and relevance of enzyme changes to organ function(s) are required to classify whether specific enzyme alterations are non-adverse, less serious or severe. See response to Reviewer 7 above.

Other than deletion of the erroneous BMD/C comment under "non-adverse" effects, no other additions or deletions are suggested.

Response: No response required.

Reviewer 15

Where does TCEQ place behavioral effects and clinical observations?

Response: Mild behavioral effects have been added to the Less Serious Effects category, and moderate to severe behavioral effects have been added to the Sever Serious Effects category. As listed in the section on clinical effects of significant organ impairment, clinical observations are considered serious effects.

I take issue with the treatment of necrosis. This term means death of cells, usually visible histopathologically. This cannot be non-adverse, given TCEQ's earlier definition of adversity which include microscopically visible changes. These two sections should be brought into agreement. Also, pathologists customarily grade severity of necrosis with as many as five different descriptors: none, mild,

moderate, marked, or severe. In my opinion, the category which marks the border for loss of function of the organ is very likely to be a matter of disagreement among pathologists. This is another case where being overly prescriptive will only lead to arguments.

Response: TCEQ agrees and added mild to moderate and moderate to severe necrosis to the less and severe serious effect categories as appropriate.

2.1.6

In Chapter 3, are the general approaches described for identifying critical effect, selecting a key study, or defining the point of departure (including the hierarchy for selecting a point of departure) consistent with accepted risk assessment methodologies?

Reviewer 3

The general approaches for defining critical effects; selecting key studies; and PODs are well presented and consistent with current risk assessment procedures.

Response: No response required.

Reviewer 4

I agree that the general approaches described in Chapter 3 are consistent with accepted risk assessment practices.

Response: No response required.

Reviewer 5

The approaches described for consistent with current practice.

Response: No response required.

Reviewer 6

There are many general approaches in the literature and guidelines by regulatory organizations, for identifying critical effect, selecting a key study, or defining the point of departure. The selections and explanations provided in the document are well suited for the applications presented, and consistent with generally accepted risk assessment methodologies. Rating=4. As requested, to quantify relative strengths and weaknesses, I have used a positive 5 point system for relative strength (1 is very weak, 2 is weak, 3 is consistent with accepted practice and documentation, 4 is strong, 5 is very strong).

Response: No response required.

Reviewer 7

Yes, I had no problems with the approach described. I will note a few places where I think the wording needs to be changed.

- *On page 63, line 43, it is stated the "the terms BMD and BMDL will be used to describe both oral and inhalation exposures." My question is why do that? And I did not notice that you actually did use those terms for both oral and inhalation exposures. So I recommend leaving that sentence out or explain why you need it.*

Response: TCEQ agrees and the text has been revised accordingly.

- *On Figure 3-7, I did not understand the use of the terms protective and predictive. They appear to be modifiers of the term "exposure" but on second thought I think they refer to the general process. A footnote is needed.*

Response: Figure 3-7 (Figure 3-5 in the revised draft) was taken from USEPA (1994) and will not be revised as suggested.

- *On page 69, line 11, I was not sure what was meant by saying that if experimental data were used to "fit" the model, additional uncertainty would be added to the model. I would think that observed data would add more certainty to a model. Perhaps I misunderstand. An example might help.*

Response: The text has been revised for clarification.

- *On page 70, lines 1-10, I found it surprising that 2 of the three models listed are for inhaled particles, even though the preceding pages were mainly a discussion of inhaled gases. The discussion of the factors affecting the deposition of inhaled particles, particularly discussion of the effect of particle size, comes in a later section (3.10.3). Even in that section, there was very little discussion of the importance of particle size in inhalation toxicology and I found no discussion of nano particles and their deposition rates and sites.*

Response: At the present time, TCEQ does not characterize for nano particles in particulate matter. Also, at this time, the ambient air monitoring networks do not have the capability to monitor for nano particles. Therefore, the TCEQ Guidelines do not have a discussion of nano particles, their deposition rates and sites, etc. No changes were made.

- *Table 3-4, is taken directly from the Hanna et al. (2001) paper and is written in an engineering language that may not be familiar to all. For the definitions of the three types of gases, I prefer what is given in the abstract of the paper:*
  - *Category 1: Gases that are highly soluble and/or reactive, absorbing primarily in the extrathoracic airways.*
  - *Category 2: Gases that are moderately soluble and/or reactive absorbing throughout the airways;*
  - *Category 3: Gases that have low water solubility and are lipid soluble such that they are primarily absorbed in the pulmonary region and likely to act systemically.*

Response: The table was taken directly from Hanna *et al.* (2001) and therefore will not be revised as suggested.

- *Figure 3-10 comes from the White Paper (Appendix C) where the equations mentioned in the figure are presented. However, neither the figure nor the equations are referenced here. Thus the reader does not know where to find them.*

Response: A reference has been added to Figure 3-10 (Figure 3-8 in the revised draft) as suggested.

- *On page 85, line 41, what is the "remaining factor" referred to here?*

Response: The sentence containing the "remaining factor" has been deleted.

- *Page 100, lines 30-33: This sentence needs to be written more clearly.*

Response: This sentence has been modified accordingly.

Reviewer 10

Yes.

Response: No response required.

Reviewer 11

Yes. The methodologies described for developing chronic oral cancer and non-cancer toxicity values are essentially the same as the USEPA methodology for developing RfDs and oral slope factors.

Response: No response required.

Reviewer 12

These procedures are mostly like others this reviewer is familiar with, although they are a bit more detailed and seem to have the potential to be especially useful for those less than familiar with risk assessment processes.

Response: No response required.

Reviewer 15

I agree with the authors' treatment of this important subject.

Response: No response required.

2.1.7

Should the critical effect be selected before or after uncertainty factors are applied (Section 3.11)?

Reviewer 3

Section 3.1.1 presents all key considerations for the determination of the POD. As the UFs will be determined based on the severity of the effect selected and the MOA, the data analysis, and hence the POD, must be determined before selection of appropriate UFs.

Response: TCEQ agrees with these comments.

Reviewer 4

TCEQ does not mention in Section 3.11 that it is considering an option. My recommendation is to select the critical effect first and then apply appropriate uncertainty factors.

Response: TCEQ agrees with these comments.

Reviewer 5

The critical effect should be selected before applying uncertainty factors.

Response: TCEQ agrees with these comments.

Reviewer 6

The critical effect should be selected prior to the application of uncertainty factors. Rating=2. As requested, to quantify relative strengths and weaknesses, I have used a positive 5 point system for relative strength (1 is very weak, 2 is weak, 3 is consistent with accepted practice and documentation, 4 is strong, 5 is very strong). The use of uncertainty factors in the risk assessment process is useful because data are not generally available to indicate how humans react to an exposure, and to protect more sensitive members of the general population. The value and use of uncertainty factors has improved greatly with more research, data and understanding. This improved understanding and detailed evaluation of data bases has led to improvements that allow for the incorporation of more scientific data into the dose-response assessment of non-cancer toxicity. The goal of risk assessment is to be able to describe the risk, or lack of risk, for various exposures. The state of the science has now evolved so that every effort should be made to use all of the available scientific information in establishing appropriate UFs. Selection of the critical effect prior to the application of uncertainty factors will improve the selection of the uncertainty factor by focusing the risk assessor (and the subsequent data selections) on the critical effect to be evaluated. The section was rated as 2 because I did not see that the method described the recommended selection of the critical effect prior to the application of uncertainty factors.

Response: No response required.

Reviewer 7

The critical effect should be selected before uncertainty factors are applied.

Response: No response required.

Reviewer 10

The critical effect should be selected before applying uncertainty factors. Values determined with UF application are derived values, not data. They reflect uncertainty. That is a pervasive problem in regulatory toxicology. All too often derived values take on a life of their own such that it is sometimes impossible to determine what the data were. To accomplish your mission to derive "safe" exposure levels you have to apply the UFs. However, the critical effects should be selected based upon the analysis of the DATA and not a value that incorporates degrees of uncertainty that will differ depending on who did the evaluation. As you note in your document, if the critical effect is a minor one but close in dose to an exposure causing a serious effect, you might want to incorporate another factor to ensure safety.

Response: No response required.

Reviewer 11

The typical sequence of events in the derivation of a threshold toxicity value is to identify the critical effect and effect level first, and then apply uncertainty factors to derive the RfD.

Response: No response required.

Reviewer 12

Selection of the critical effect after UF application would most always result in the effect about which there is the most uncertainty. The critical effect, by most definitions, is that which would avoid all other adverse effects. Therefore UF should always be applied after the critical effect has been chosen.

Response: No response required.

Reviewer 15

When methods for the RfD and RfC were first devised in the 1980s, selection of the critical effect and study were always done prior to the application of UFs. In my opinion, TCEQ has got it right when they recommend completing the entire calculation before making their selection(s). I have seen no clear guidance on this subject from USEPA or elsewhere; perhaps the treatment in this draft document can enter common use.

Response: As recommended by the majority of reviewers, TCEQ chooses to identify the critical effect prior to applying uncertainty factors to derive toxicity values.

2.1.8

Is the general information including cutoffs and definitions on physical/chemical properties described in Section 3.4.1 complete and appropriate?

Reviewer 3

The information in the physical properties section is well presented and generally appropriate. However, in looking at the criteria for PM vs. vapor, one must consider overall toxicity. A substance with a vp of 0.01 mm Hg @ 25◦C has a saturated vapor concentration of 13 ppm @ 25◦C. Many substances can be toxic at that level. Therefore, while this is good guidance, a caveat that cautions the Risk Assessor to consider the toxicity of the substance as well as the vp should be included in the document.

Response: TCEQ agrees and changes have been made accordingly. While physical and chemical characteristics of a chemical can influence the nature of its toxic effect, the toxicity of a chemical and the vapor pressure of the chemical will be an important consideration when conducting a chemical assessment.

Reviewer 4

The information outlined on physical/chemical properties seem appropriate to consider in a risk assessment.

Response: No response required.

Reviewer 5

In general, this information is correct and appropriate, including the cut-offs. A caveat should be added to the effect that the cut-offs represent approximations (or rules of thumb, as indicated in the subtitle for Log Kow).

(Editorial: replace "is" for "it" in line 10 of page 47)

Response: TCEQ agrees and appropriate changes have been made accordingly.

Reviewer 6

The general information on physical/chemical properties is complete and appropriate. Rating=3. As requested, to quantify relative strengths and weaknesses, I have used a positive 5 point system for

relative strength (1 is very weak, 2 is weak, 3 is consistent with accepted practice and documentation, 4 is strong, 5 is very strong).

Response: No response required.

Reviewer 7

Yes, except the point is not clearly made that the physical state of a chemical (solid, liquid or gas) depends not only on the temperature, but also on the atmospheric pressure. This becomes important for people living at altitude.

Response: No response required.

Reviewer 10

This section provides an excellent synopsis of the type of analysis that is currently used in the Premanufacture Notice (PMN) program of EPA. It also discusses the consequences of differing values on absorption and penetration into the lung. It is a useful surrogate when data on absorption are not available.

Response: No response required.

Reviewer 11

I thought Section 3.4.1 provided some good examples of chemical and physical properties which may impact what we know about the toxicity of a chemical. Of course, the chemical and physical properties of importance will vary depending on the chemical and may include different or more properties than those shown here. For example, molecular weight and Henry's Law constant are useful properties when evaluating volatile organic compounds. So rather than seeing Section 3.4.1 as an all-inclusive list, I interpreted it as an example of some chemical/physical properties and how they influence toxicity.

Response: No response required.

Reviewer 12

Several specific comments have been made here, one regarding the inclusion of vapor pressure and estimation of maximum allowable air concentrations. The explanation of Kow values and bioaccumulation is excellent.

Response: No response required.

Reviewer 15

I have just one comment on this topic. The discussion of the importance of boiling point, vapor pressure, and Henry's Law should be presented earlier in this section, because these properties are so very crucial to predicting inhalation toxicity and understanding inhalation exposure.

Response: TCEQ is satisfied with the placement of this topic in the Guidelines. No changes have been made.

2.1.9

When a free standing NOAEL is used as the POD, are exposure duration adjustments in Section 3.9.3 appropriate?

Reviewer 3

The application of n=1 to a free standing NOEL in going to longer exposure levels is appropriate. In going to short exposure durations, ten Berge et al. have shown that n=3 is a reasonable approximation. I therefore do not agree with the recommendation on P. 73 l 11-13. The slope of the line when n=3 is very nearly flat. However, it is important to look at data from other studies on that material. Many fluorinated hydrocarbons have free standing NOELs. The upper limit for repeat exposure testing is 50,000 ppm. Above that, there is a risk of inducing effects due to anoxia since the oxygen level is lower (e.g. at 100,000 ppm or 10% the oxygen level is reduced by 10% from 21% to 19%). However, with added oxygen in acute exposure studies, animals can survive exposures of 400,000 ppm (40%) and greater. In this case it is appropriate to extrapolate to a shorter time. I would not recommend n=1, but either n=2 or n=3 could be appropriate.

Response: There is no information on the steepness of the dose-response curve for a free-standing NOAEL. TCEQ does agree that in some cases it may be appropriate to extrapolate to shorter exposure duration by applying a n value lower than 3 if data shows that such an adjustment is scientifically defensible. As commented by most of the reviewers, this section is considered appropriate.

Reviewer 4

This section appears to apply only to inhalation exposures.

Response: The TCEQ agrees with reviewer's comment.

Reviewer 5

Yes. Adjustment from a shorter to a longer duration is justified because it is more conservative, while adjustment from the longer to shorter duration is not appropriate unless there is evidence to support it.

Response: No response required.

Reviewer 6

The choice of free standing NOAEL as the POD is appropriate. Rating=3. As requested, to quantify relative strengths and weaknesses, I have used a positive 5 point system for relative strength (1 is very weak, 2 is weak, 3 is consistent with accepted practice and documentation, 4 is strong, 5 is very strong).

Response: No response required.

Reviewer 7

I consider the exposure adjustments to be valid.

Response: No response required.

Reviewer 10

The response to this question cannot just focus on Section 3.9.3 because the concepts there are developed in earlier parts of Section 3.9. The response therefore will discuss all of Section 3.9.

- *Section 3.9.1.*
  - *i. The most important comment here is that these curves are valid for frankly toxic effect exposure levels. The AEGL Program assumes the duration scaling for lethality is generally*

applicable to disabling effects. This relationship is probably reasonable since the two levels are often within a factor of 3 of each other. However, there are two factors at variance with the TCEQ mission. 1) The first is that the AEGL Program is deriving levels that are thresholds for frank toxicity. At these levels, homeostatic processes have often been totally overwhelmed. Conversely TCEQ is deriving "safe" levels of exposure. At "safe" levels homeostasis is generally operating so the shape of the curve for minor effects tends to flatten out – e.g. the value of n approaches infinity. The point is that the value of n that is observed for lethality will not be valid for whatever endpoint used to derive "safe" levels. Thus, even if a value of n is known for a chemical, it will be based upon the lethality endpoint and will not be valid for lesser severity effects. For example, in the AEGL Program the AEGL-1 value (threshold for discomfort) is often held constant in value from 10 minutes to 8 hours of exposure. In this case the value of n is infinity. For acute effects for minimal toxicity, this is probably a more valid approach. This will be amplified in the discussion in the next session. 2) The second point is that even if the value of n is known, this only exists within a limited duration of exposure. For example, n for HCl is probably valid in the 10 minute to about 2 hour range. After about 2 hours, homeostatic processes start to become relevant and the value of n will increase with increasing duration of exposure. However, the value of n for phosgene is probably valid in the 10 minute to 8 or 24 hour range before homeostasis kicks in. Note that these comments are for frankly toxic effects, not the kind of effects that TCEQ will be modeling.

Response: TCEQ agrees with these comments. However, this section will not be revised based on these comments because the Guidelines already address the issues discussed by this reviewer. For example, the following quotes may be found in this section:

- "However, certain health effects such as irritation, narcosis, or asphyxia may be more dependent on concentration than duration so exposure concentrations are not adjusted for these health effects."
- "…where "n" represents a chemical- and endpoint-specific exponent."
- "The experimental data are deemed to be adequate if the different exposure durations of the studies are similar to the desired exposure duration; the studies evaluate the appropriate health effect endpoint; and the quality and quantity of the data are adequate."
- "Exposure duration adjustments using this procedure are only conducted over a limited time extrapolation."

ii. *To derive n, the document refers to using the ten Berge model. That is certainly valid. However, there will sometimes only be LC50 data available. In that case a least squares analysis of the LC50 data is also reasonable (NRC, 2001, Section 2.7). Keep in mind that this is simply an empirical approach that is valid only within a chemical specific time frame.*

Response: TCEQ agrees with these comments. This section has been revised to state that if LC_{50} data are not amenable for modeling using the ten Berge model, a least squares analysis of the LC_{50} data may be used (NRC 2001).

iii. *The discussion about the significance of raising C to the nth power and t to the mth power adds little and could be deleted. In fact all of the equations discussed can generate the same curve. There is no way to determine the values and significance of n and m from lethality data. This was discussed in NRC, 2001 and has been excerpted as follows: "2.7.5.3 Curve Fitting and Statistical Testing of the Generated Curve Once the health-effect endpoint and data points describing the exposure concentration-duration relationship have been selected, the values are plotted and fit to a mathematical equation from which the AEGL values are developed. There may be issues regarding the placement of the exponential function in the equation describing the concentration-duration relationship (e.g., $C^n \times t = k$ vs $C \times t^m = k_2$ vs $C^x \times t^y = k_3$). It is clear that the exposure concentration-duration relationship for a given chemical is directly related to its pharmacokinetic and pharmacodynamic properties. Hence, the use and proper placement of an exponent or exponents to describe these properties quantitatively is highly complex and not completely understood for all materials of concern. The quantitative description of actual empirical data of the concentration duration relationship can be expressed by any of a number of linear regression equations. In the assessment of empirical data reported by ten Berge et al. (1986), these workers quantified the exposure concentration-duration relationship by varying the concentration to the nth power. Since raising c or t or both to a power can be used to define quantitatively the same relationship or slope of the curve and to be consistent with data and information presented in the peer-reviewed scientific literature, the equation Cn × t = k is used for extrapolation. It must be emphasized that the relationship between C and t is an empirical fit of the log transformed data to a line. No conclusions about specific biologic mechanisms of action can be drawn from this relationship."*

Response: While TCEQ agrees, this section provides important information to the reader about the complexity of the relationship between concentration and duration. Despite data not typically being available to derive a chemical-specific n and m value, which is stated in the text, this discussion provides fundamental knowledge about the basis of the calculations used to conduct duration adjustments. No changes were made.

- *Section 3.9.3. The response to this question is same whether a value of n can be derived or not.*

i. *When you have data or validated models that allow you to quantitatively relate exposure to blood or deposition values as a measure of time then you should use that relationship.*

Response: TCEQ agrees. However, the Guidelines will not be revised based on this comment because the Guidelines already address the issues discussed by this reviewer. If a model is available, it will be used. This section provides default guidelines for when a model is <u>not</u> available.

ii. *It is not valid to use a value of n derived for lethal effects for the mild effects used to derive "safe" values. The biological endpoints and duration response relationships are different. That being said, one is left with using a conservative approach.*

Response: TCEQ agrees. This section will not be revised based on this comment because the Guidelines clearly state that values of "n" are endpoint specific. However, if death occurred due to respiratory failure, and the critical effects at lower concentrations are respiratory effects, the "n" value derived

using lethality data will be deemed to be acceptable (but not minor respiratory irritation, which are considered to be concentration dependent).

1. *Extrapolation from short to long durations of exposure. For severe effects the low end of the range of n values is 1 and for effects of lesser severity the value of n is greater than 1. Therefore the approach of using an n of 1 when extrapolating from short to long duration exposures is the most conservative approach to use.*

Response*:* TCEQ agrees with this comment.

2. *Extrapolation from long to short durations of exposure. The convention used in Chapter 4 uses an n of 3 when extrapolating from long to short durations of exposure and n is not known. That is generally a valid approach for lethal effects. For effects of lesser severity, the value of n approaches infinity as discussed above. The extrapolation to shorter durations of exposure using an n of 3 is actually not conservative since n should be higher for minor effects. When extrapolating to short durations of exposure the concentration should be held constant to take the most conservative approach, e.g. n is infinity.*

Response: TCEQ prefers to use data most similar to the exposure duration of interest (e.g., 1-hour) to limit the temporal range over which extrapolation may be performed. If exposure duration adjustment is needed and adequate experimental data at different exposure durations are available, a chemical- and endpoint-specific "n" value can be derived. Otherwise, the TCEQ considers using a default "n" value of 3 for extrapolation from long to short durations of exposure as consistent with other agencies. For minor respiratory irritation or critical effects where the MOA is stimulation of the trigeminal nerve, the concentration is held constant.

3. *The approach discussed in 1. and 2. above should be used in general unless you have decent blood/deposition data. Modeling lesser effects to derive n with ten Berge model is problematic. The model assumes dichotomous data. However, when modeling lesser effects, the effect is not simply yes or no. It is really yes of varying severities and no. The yes responses are actually a continuum and not dichotomous because they are not identical.*

Response: TCEQ agrees. This section will not be revised based this comment because if the data are not amenable to mathematical modeling, default procedures will be conducted. If there are sufficient blood/deposition data, then data will be used, not default procedures.

4. *Chronic data should also be examined to decide if the extrapolation is too conservative and the numbers adjusted accordingly.*

Response: TCEQ agrees and has added this comment to this section. TCEQ intends to use chemical-specific data to make adjustments, if available. However, if data is not available, then default procedures will be used.

5. *At the end, there are approaches that will have to be used as defaults when data are limited. However, there will never be a substitute for the exercise of professional*

judgment. You must always ask, are the derived values reasonable when the entire body of knowledge on the chemical is examined?

Response: TCEQ agrees with these comments.

Reviewer 11

I don't believe the adjustment proposed in Section 3.9.3 applies to chronic oral toxicity values.

Response: TCEQ agrees with this comment.

Reviewer 12

Yes, especially with regard from longer to shorter durations where extrapolations even under n=3 curve sharply upwards to higher and higher exposure concentrations.

Response: TCEQ agrees with this comment.

Reviewer 15

I agree with the method recommended by TCEQ. However, the 4-page discussion of Haber's Law in Section 3.9.1 is excessive.

Response: No response required.

2.1.10

Is the approach for identifying Inhalation Effect Levels consistent with accepted risk assessment methodology (Section 3.14).

Reviewer 3

The approach described in section 3.14 for identifying inhalation no adverse effect levels is appropriate and consistent with current practices.

Response: No response required.

Reviewer 5

The approaches described in 3.14 are consistent with current practice.

Response: No response required.

Reviewer 6

There is no one standard risk assessment methodology, rather there are many accepted risk assessment methodologies adapted to each unique relevant scenario. TCEQ's selection of risk assessment methodology is well adapted, explained and appropriate for the applications described. Rating=3. As requested, to quantify relative strengths and weaknesses, I have used a positive 5 point system for relative strength (1 is very weak, 2 is weak, 3 is consistent with accepted practice and documentation, 4 is strong, 5 is very strong).

Response: No response required.

Reviewer 7

I thought this was an interesting discussion of predicting the central estimate of the lowest exposure that can be expected to cause an adverse response in humans. It does not include use of UFs, so is not what is usually done in risk assessment methodology. It is for informational purposes only (page 95, line 26) and for comparison with "safe" levels determined with the use of UFs for risk assessment. The general tone of this section reminds me of when I served on a committee to advise the Navy on the toxicity of airborne substances on submarines. The commanders of the submarines were not interested in "safe" levels for a general population. They wanted to know when they had to take emergency measures and surface to vent the air on the submarine. In that case, the population at risk was somewhat homogenous, i.e., they were healthy young men. But in Section 3.14, the general population is of concern, including among others the elderly, the young and the highly sensitive. So it is much harder to estimate a practical, "real" inhalation effects level for such a diverse group. There will be a great deal of uncertainty associated with such an estimate. There is a good discussion in this section about those uncertainties. But there is one paragraph (Section 3.14.2) in which the authors claim that their approach will allow them to predict an inhalation effects level with "a reasonable degree of certainty" (page 96, line 24). Using duration adjustments and UFs is said to add uncertainty (page 96, lines 21-22). This gives a false impression of precision in the calculation of the inhalation effects levels that is not warranted. The effects level will depend on who is being exposed. It would be more reasonable to give a range of effects levels, considering the diverse population of concern. The approach is apparently intended to give a "central tendency" value, but this should not be confused with certainty. As long as these calculations are restricted to the DSDs and all the caveats that are discussed in Section 3.14 are clearly given, then I have no problem with the exercise. However, it must be made clear that such estimates have a great deal of uncertainty.

Response: TCEQ disagrees with this comment. TCEQ does not claim that the approach will always allow prediction of an inhalation effects level with "a reasonable degree of certainty," TCEQ merely states "a reasonable degree of certainty" as a desired goal or attribute and to the extent possible, attempts to identify procedures consistent with furthering the purpose of, "identifying a level where with a reasonable degree of certainty, a response in some individuals may be expected." Furthermore, TCEQ does not claim that a given level will, or can, account for any potential intrahuman variability in sensitivity beyond that captured by the dose-response data made basis for the value (i.e., the value cannot directly account for truly sensitive subpopulations, if any, if relevant human dose-response data are not available). Ultimately, the degree of certainty/uncertainty associated with the expectation of a human response occurring in some individuals will depend upon the data available. As acknowledged in the comment, the TCEQ provides a good discussion of the associated uncertainties. The TCEQ understands that effects levels may vary among individuals for a given chemical, and if the data allow, the TCEQ will provide a range of effects levels considering population diversity. If dose-response data are only available for a group that is in fact less sensitive, then any resulting effects level would be even more certain to increase the likelihood of an effect in a general population that includes more sensitive subgroups. The goal of the TCEQ is to reduce the uncertainty, to the extent possible, associated with the expectation of a human response occurring in some individuals at an air concentration based on the dose-response data available, which may not include the potentially most sensitive subpopulations.

However, when determining the lower end of actual effects levels, available data are a limiting factor. Consistent with the comment, the TCEQ will discuss the caveats and uncertainties associated with these values in the DSDs.

Reviewer 10

Typically this type of assessment is not done except sometimes when incidence levels are calculated for risk benefit determinations. Thus it is difficult to answer the question. Is this determination for susceptible populations? If so that should be stated.

Response: TCEQ is interested in this determination for the general population and susceptible populations in particular, as data allow.

Section 3.14.1 uses the most sensitive species. This may not always be appropriate. For irritants, the rat is probably the better model than the mouse. The most appropriate species should be used, not the most susceptible. You are using an HEC determination that has problems. See my general comment 5 on dosimetry corrections.

Response: Text has been added to this section regarding appropriate animal model considerations, and consistent with TCEQ procedures, the most defensible established dosimetric adjustments will be performed.

Section 3.14.2. Since you are not using UFs that implies the effects predicted will occur at lower doses in susceptible individual. If that is what is meant then state it.

Response: UFs are not used in this evaluation as they introduce uncertainty about the expectation of a human response occurring in some individuals based on the dose-response data. Not using UFs in-and-of itself does not imply that significant intrahuman variability exists regarding sensitivity to the critical effect for a given chemical, much less that effects will in fact occur at lower concentrations in some susceptible individuals. For example, in some cases, the dose-response data used may include susceptible individuals or populations such that the intrahuman UF would have been 1. No changes were made.

Section 3.14.3. There are a lot of caveats in this section so I am not sure what will be calculated. One gets the impression that the problems are so great these values might not be calculated. Consideration might be given to computing the 1 in 10 cancer rate. It would be a novel exercise of predicting cancer rates in the experimental range. If these calculations are performed TCEQ might consider staying away from predicting cancer from single exposures. The uncertainties with lifetime predictions are difficult enough without adding single exposure estimates.

Response: TCEQ will keep these considerations in mind.

Reviewer 12

This section attempts to apply what this reviewer refers to as probabilistic considerations of PODs (both N/LOAEL and BMD/Cs) and what they may mean in human populations. Most "accepted" risk assessment methodologies do not usually provide the reader with this aspect, thus this approach should be considered not only as acceptable, but an attribute of the guidance.

Response: This section was developed in response to public inquiries regarding likely effects levels. Along with the derived health-protective values (e.g., ReVs, ESLs), the TCEQ believes these levels will help provide the public with a more complete picture of the conservativeness inherent in the protection of public health and environmental quality in our heavily-industrialized state, and agrees that this additional approach is an attribute to TCEQ guidance.

2.2 Specific Issues for Evaluating Mode Of Action (MOA)

2.2.1

TCEQ follows current U.S. EPA cancer guidelines for hazard identification, mode-of-action analysis and dose-response assessment for developing inhalation URF and oral SFo values (Chapter 5). Is this appropriate, or would other guidance be more appropriate? Is TCEQ interpreting and applying the guidance correctly? If not, what changes to the methods would you suggest?

Reviewer 1

In general, this document follows USEPA cancer guidelines. However, I believe that the field of Toxicology is undergoing a development that is not captured in this document. I refer to the 2007 National Research Council (NRC) document "Toxicity Testing in the 21st Century: A Vision and a Strategy" and some other documents relating to this issue (NRC, 2007, Models in Environmental Regulatory Decision Making; NRC, 2007, Applications of Toxicogenomic Technologies to Predictive Toxicology and Risk Assessment; NRC, 2009, Science and Decisions; Advancing Risk Assessment, Washington, DC: National Academies Press). Newer methods of molecular biology, bioinformatics, and computational toxicology based on human biology are going to be much more informative about MOA in the future. This will involve a broad range of doses and will focus on perturbations of critical cellular responses. This type of pathway analysis is a move away from high dose to low dose extrapolation. It allows one to distinguish between perturbations in pathways that are adaptive responses and perturbations of pathways leading to morbidity/toxicity.

Response: TCEQ acknowledges the importance of the referenced documents for the future of toxicology testing and regulatory chemical risk assessment. The TCEQ Guidelines will be updated periodically to incorporate appropriate procedures which have been developed and incorporated into the U.S. EPA cancer guidelines, or developed and well-established within the scientific regulatory community for chemical risk assessment. In the interim, TCEQ is free to use considerations based on the referenced documents (e.g., toxicogenomics data), as appropriate and applicable, to aid in informing chemical toxicity assessments (e.g., MOA).

Using some of these newer approaches will eliminate some of the confusion about adverse effects. I found lines 12-16 on page 58 to be vague and unintelligible. What is a reserve loss?

Response: TCEQ agrees that the referenced lines regarding reserve loss are confusing. The sentences have been removed.

Section 3.7.1.4 is also confusing. What is the purpose of distinguishing between non-adverse, less serious, transitional, and serious effects? Why aren't changes in gene expression or protein levels that are adaptive responses listed under non-adverse effects? What level of necrosis is less serious? Why does this section not include a discussion of apoptosis?

Response: TCEQ did not revise the Guidelines based on these comments. The purpose of distinguishing between non-adverse, less serious, transitional, and serious effects is to help determine the level of severity of the critical effect, which informs the selection of the appropriate LOAEL to NOAEL uncertainty factor (if necessary). This section was based on Tables 3-1 to 3-17of the *Guidance for the Preparation of a Twenty First Set Toxicological Profile* (ATSDR 2007). The TCEQ modified the ATSDR list.

However, this is not an exhaustive list, but is included to provide a general overview for the severity of adverse effects; as always best scientific judgment should be exercised.

What is a "major serum chemistry change"?

Response: See response to Reviewer 7 in 2.1.5 above.

By the way, SFo does not appear in the Glossary or list of Acronyms and Abbreviations.

Response: SFo has been added to the list of Acronyms and Abbreviations.

Reviewer 2

Mode of action evaluations are not my area of expertise, but I know of no current better options for hazard identification, mode-of-action analysis and dose-response assessment. It appears to me that TCEQ is interpreting the guidance correctly.

Response: No response required.

Reviewer 9

The EPA guidance presented is appropriate, but our understanding of the guidance and how to apply it is further enhanced by the publications and case studies cited under Question 1, and I would recommend that additional emphasis be placed on those publications.

Response: See response to Reviewer 1 above.

Reviewer 14

It is most appropriate that the TCEQ follows current U.S. EPA cancer guidelines for hazard identification, mode-of-action analysis and dose-response assessment for developing inhalation URF and oral SFo values. Since SF in toxicological literature refers to a "Safety Factor", it would be useful to list SFo in the Acronyms and Abbreviations as a "Slope Factor". The TCEQ is using the U.S.EPA cancer guidelines correctly.

Response: The SFo has been added to the list of Acronyms and Abbreviations.

2.2.2

Chapter 3 (Section 3.5) and Chapter 5 (Section 5.7) describe the general process that TCEQ will follow to conduct an MOA analysis. Is this process complete, accurate, and consistent with accepted risk assessment practices?

Reviewer 1

Section 3.5 hardly deals with MOA at all. Most of Section 5.7 does not deal with MOA analysis either. Section 5.7.2 is titled "MOA" but refers to the 2005 Cancer Guidelines. The criteria needed to call something a Mutagenic Carcinogen is in Section 5.7.4.1, but it would make more sense to have it earlier.

In general, this discussion follows ideas in USEPA (2007). However, there are some problems. The definition of McCarroll et al. (2010) of mutagenicity is not accepted by many experts (including those who reviewed USEPA, 2007). The key element of mutagenicity must be heritability (see USEPA, 2007). As currently practiced, mutagenicity assays depend on heritability (clones are counted), but chromosome

aberration assays do not. In fact, most chromosome aberration assays do not measure cytotoxicity properly, and so the aberrations seen are in dead cells that will never replicate [see discussion in Klein et al., Toxicol. Appl. Pharmacol. 222:289-297, 2007]. There are many examples of agents that cause chromosome aberrations in the absence of mutagenicity.

Response: TCEQ agrees that a key element of mutagenicity is heritability, as referred to in the Draft Guidelines. Thus, the definition of mutagenicity was revised to make this clear. The comments on chromosome aberrations do not call for any particular revisions, and thus are simply duly noted for future reference and review of chromosome aberration results.

Page 156, line 22 lists WOE approach #1 as "Does the carcinogen have genotoxic and/or mutagenic potential?" This should be changed to "Does the carcinogen show mutagenic activity at relevant concentration?" (i.e., at concentrations that do not result in high toxicity). It is NOT enough that an agent is "genotoxic". This is much too vague a category. Assays for "genotoxicity" (DNA adducts, strand breaks, chromosome aberrations, micronuclei, SCE, UDS, etc.) are useful as biomarkers of exposure and in hazard identification. They do not predict mutagenicity, which is the key event in a mutagenic MOA.

Response: TCEQ agrees and has made revisions consistent with these comments. In particular, it is emphasized that genotoxicity alone is not sufficient evidence of a mutagenic MOA.

Line 35 states "Within the context of these definitions, agents that are mutagenic are also genotoxic…." This is not necessarily true. An agent can cause mutations by altering nucleotide pools and by interfering with proteins that are important in DNA replication, repair, apoptosis, and cell cycle control. For example, if mismatch repair is blocked either by enzyme inhibition or by effects on gene expression, a mutator phenotype is produced. Indirect mutagenicity can be defined as interactions with non-DNA targets leading to mutagenic effects. It is expected that indirectly mutagenic agents should have a threshold concentration below which there is no effect, due to the fact that non-DNA targets exist in many copies in the cell, unlike DNA (Kirsch-Voldars et al., Mutat. Res. 540:153-163, 2003). In the European Union, considerations of indirect genotoxic mechanisms have led to new appreciation of thresholds in risk assessment (Pratt and Baron, Toxicol. Lett. 140-141:53-62, 2003).

Response: TCEQ appreciates the noted exception to the statement and has added language to reflect the possibility that some chemicals may not be directly genotoxic but still be mutagenic through an indirect mechanism, which may have a threshold.

Page 157 continues based on these errors, and should be rewritten or deleted. Table 5-3 is incorrect in a number of instances. Aneuploidy is rarely evaluated in gene mutation assays, and usually does not result from DNA damage. It also usually has thresholds (see Elhajouji et al. Mutagenesis 26(1), 199–204, 2011). Neither are chromosome aberrations assessed in mutagenesis assays. It is sometimes assumed that so-called "small colonies" in the MLA assay are due to chromosome aberrations, but this is not usually proven. Many chromosome aberrations are lethal. In any case, as mentioned above, this type of assay is not based on heritability. The Comet assay also detects alkali-labile sites if run under alkaline conditions. Other endpoints of interest for a MOA include gene amplification and epigenetic effects (which are often mistaken for mutagenesis).

Response: TCEQ has revised the section consistent with these comments and an emphasis on heritability.

Reviewer 2

The general process for applying MOA analysis appears to be accurate and consistent with accepted risk assessment practices. I would recommend that TCEQ consider the ideas in Moore et al. (2008: Analysis of in vivo mutation data can inform cancer risk assessment. Regulatory Toxicology and Pharmacology 51:151) and Allen et al. (2005: Dose-response modeling of in vivo genotoxicity data for use in risk assessment: some approaches illustrated by an analysis of acrylamide. Reg. Tox. Pharm. 41:6-27) as an approach for more completely evaluating the consistency of purported genotoxic responses with the endpoints of ultimate regulatory concern (health-related adverse effects). Such approaches may have merit for deciding whether mutagenic or genotoxic insults are responsible for the health effects of concern and whether they should be regulated on the basis of a linear low-dose extrapolation.

Response: The two referenced articles have been added to the text.

Reviewer 9

Section 3.5 gives a nice discussion of how MOA is applied for multiple decisions in risk assessment. It would be useful to more directly tie the analytical thinking in evaluating MOA to the modified Hill criteria, and explain either here or elsewhere in particular the importance of dose-response and temporal comparisons between the key events and the apical endpoint. (This is important because the thinking is a bit different from the simple application of these ideas for evaluation of causality – the key point here is to compare the key event data with those for the apical endpoint). In addition, reference to case studies (e.g., the Meek et al., 2003 reference cited above [under charge question 1]) is useful for illustrating how the concepts are applied. It may also be useful to cite work of Rhomberg regarding identifying "high-stakes" hypotheses resulting from the hypothesized MOA and using that to test the MOA hypothesis. (He uses a different term that I can't recall, but the idea is developing hypotheses that are not easy to meet, and that rigorously test the MOA.)

Response: The two references (Meek et al. 2008, Allen et al. 2005) suggested by Reviewer 2 have been added to the text to emphasize the importance of dose-response and temporal comparisons between the key events (i.e., mutgenicity/genotoxicity) and the apical endpoint (i.e., cancer). A discussion of how the modified Hill criteria relate to the types of evidence relevant to evaluating the carcinogenic MOA is beyond the scope of the document.

Reviewer 14

Section 3.5 is very brief and inadequate. It needs to be stated that an MOA analysis will be discussed in detail in Section 5.7. Section 5.7 provides an excellent discussion of the TCEQ MOA process that is complete, accurate, and consistent with accepted risk assessment practices.

Response: TCEQ has made the recommended revision to this section.

2.2.3

Are the definitions of genotoxicity and mutagenicity used in the guidelines and adopted from McCarroll et al. (2010) appropriate in the context of discussing a mutagenic MOA evaluation? Are all

major and relevant considerations (or key steps) included and put into appropriate context in weight of evidence approach for chemical carcinogenicity via mutagenic MOA? If possible, please help identify endpoints already included in the guidelines as genotoxic or mutagenic and provide other relevant input.

Reviewer 1

I already wrote my comments on McCarroll et al. [under charge question 2.3.2].

The important considerations for WOE (p. 158) are well stated.

There is no discussion about whether it is always justified to assume that mutagenic carcinogens never have thresholds. Section 5.7.3.3 gives a few examples of carcinogens that have thresholds because a key event is cytotoxicity. As discussed above, indirect mutagenicity is expected to have a threshold concentration below which there is no effect, due to the fact that non-DNA targets exist in many copies in the cell, unlike DNA (Kirsch-Voldars et al, 2003a). It is also the case that some directly mutagenic agents have a threshold for clastogenesis (reviewed in Elhajouji et al. Mutagenesis 26(1), 199–204, 2011). Some DNA repair mechanisms are robust at low concentrations, and become saturated as the concentration increases. This leads to a threshold (e.g. for alkylating agents that form O6 methylguanine adducts (see discussion in Slikker et al., TAAP 201:203-225, 2004; and review by Jenkins et al., Toxicology 278:305-310, 2010). Data is appearing showing thresholds for tumorigenesis of mutagenic agents (e.g. dibenzo[a,l]pyrene in trout, Bailey et al., Chem Res Toxicol. 22:1264–1276, 2009).

Response: See response to Reviewer 1 in 2.2.2 above. Additionally, text has been added regarding the possibility of a threshold for mutagenic/genotoxic carcinogens, referencing the Bailey *et al.* (2009) and Jenkins *et al.* (2010) studies.

Reviewer 2

The definitions used by TCEQ are adequate for the purposes of the TCEQ guidelines. It appears that all the relevant key steps and major considerations are considered for the weight of evidence approach. See comment above about considering methods of Allen et al. and Moore et al. [under charge question 2.3.2] as they pertain to the weight of evidence for chemical carcinogenicity (i.e., relevance of genotoxic or mutagenic responses to the ultimate caner response).

Response: See response to Reviewer 2 in 2.2.2 above.

Reviewer 9

The definitions are appropriate, and the key steps are included.

Response: No response required.

Reviewer 14

Section 5.7.4.1 provides an excellent discussion of the definitions of genotoxicity and mutagenicity that are appropriate for a mutagenic MOA evaluation. The major and relevant considerations are included and put into an appropriate context in a weight of evidence approach for chemical carcinogenicity via a mutagenic MOA.

Response: No response required.

2.2.4

Please comment on any issues relate to mode of action analysis that have not already been addressed.

Reviewer 1

Chemicals can work in a variety of ways that could lead to cancer. For example, take benzo(a)pyrene (BaP), a model compound that is metabolized to genotoxic intermediate(s). Recently, it was reported that challenge of HeLa cells with BaP activates Long Interspersed Nuclear Element-1 (LINE-1), a mobile element within the mammalian genome. This is accomplished epigenetically (Teneng et al., Epigenetics 6:335-367, 2011). Mobilization of LINE elements is one mechanism for genetic rearrangements, a potential MOA. Another example: The carcinogen inorganic arsenic has numerous possible MOAs, is not significantly mutagenic, and yet is treated as a mutagenic carcinogen (Klein et al., Toxicol. Appl. Pharmacol. 222:289-297, 2007).

Response: TCEQ agrees with these comments and Chapter 5 now acknowledges the possibility of multiple MOAs.

Reviewer 2

Although not part of one of the charge questions, I feel the need to comment on one aspect of the discussion in Chapter 3 concerning the determination of adversity (P. 55 ff., Section 3.7.1). There are several comments and implications in this discussion that appear to be inappropriate or incorrect.

- *For example, lines 13-14 of P. 55 seem to imply that adaptive or incidental changes are ones that occur "merely by chance" and be unrelated to exposure. However, some adaptive or incidental changes are related to exposure; it is the fact that one somehow (more appropriately) determined that they are not adverse that allows an assessor to "ignore" such endpoints even though they do demonstrate a dose-related behavior. Conversely, effects that truly are adverse can occur "merely by chance;" in fact, many adverse effects may occur merely by chance after exposure to a toxicant when the effect of that toxicant is unrelated to that endpoint (i.e., it is affecting other systems or organs). It is one of the tasks of a risk assessment to separate out the truly adverse effects that are related to the exposure in question from those that are not related to exposure.*

Response: TCEQ has made revisions to Chapter 3 to address this comment.

- *The consideration of statistical significance of dose-related changes in relation to the determination of adversity (bottom of P. 55, top of P. 56) is misplaced. Adversity should be able to be decided on a chemical-independent basis; the chemical-specific decisions are related to the magnitude or level of the change in the adverse effects, e.g., estimation of a BMD or some similar value, as a function of the exposure level. It is not the case that EPA has "provided guidance … to determine whether the benchmark response and corresponding 95% confidence level of the benchmark dose is adverse or not" (P. 56. lines 16-18). The BMR defaults recommended by EPA say nothing about adversity; in fact, they assume that the endpoints for which BMDs are derived are relevant for a risk assessment (including considerations of adversity) and have been chosen based on expectation about what levels of response are generally*

consistent with responses close to the observable range and which can be used for PODs with some comfort that derived RfDs (or a similar estimate) that are going to be safe enough.

Response: TCEQ agrees with these comments and has removed text regarding the consideration of statistical significance of dose-related changes in relation to the determination of adversity.

- *The adversity question should be evaluated separately. Figure 3-4 is, in a sense, upside down: first rule out those effects that are not adverse, then for those that are, determine if they demonstrate a dose-related behavior. If some endpoint does not show a dose-related change, then it is not correct to say that that endpoint is not adverse, only that it is not an effect that is induced by the chemical in question.*

Response: TCEQ believes that the description of the criteria for the differentiation of adverse from non-adverse effects is sufficient to assure that these procedures are carried out in practice. Nevertheless, this figure has been removed from this section.

- *The use of statistical significance (at the 0.05 level or any other level) is not appropriate for determining adversity (as in the bullet lists under Section 3.7.1.4.1). It is well known that sample size affects statistical significance even if the magnitude of the difference between treated and control groups does not change. Moreover, the reference to the BMDL being comparable to a NOAEL (P. 59, l 28-29) has no place in this list of criteria for determining severity. this bullet item is merely a restatement of one objectives that BMD(L) estimation was designed to satisfy, i.e., that it provide a more robust, consistent and stable estimate that could replace the problematic NOAELs that used to be relied upon for POD determination.*

Response: See response Reviewer 12 in 2.1.5 above.

- *Related to the above comments about the BMD and BMR, the "definition" in Table 3-3 for the BMR is not correct. The BMR is not an adverse effect; it is a specific (user-selected) level of change in an effect (probably an adverse effect) that has some specific characteristics (e.g., represents a 10% increase in risk, or a change of 1 standard deviation, etc.). Moreover, the reference in that table to a change in the range of 5 to 10% is unnecessarily restrictive and inappropriate for some types of response measures (e.g., continuous endpoints).*

Response: The definitions of BMD and BMR in Table 3-3 have been revised.

- *The discussion at the end of section 3.7.2 (P. 63, l 12 ff.) is consistent with my comments above, i.e., that statistical significance is not an appropriate criterion for judging adversity. More importantly for the guidance document, those comments quoted from Lewis et al. (2002) are not consistent with what was written earlier in the guidance, i.e., those items that I have commented on above.*

Response: TCEQ agrees with this comment and the discussion and comments quoted from Lewis *et al.* (2007) have been moved to the previous section.

- *And finally, do not make the claim that BMD model can identify a threshold (as is done on P. 63, l 33-34). The real advantages of the BMD approach over the NOAEL approach are sufficient to support its use, as TCEQ correctly does. [By the way, the percentiles given on P. 66, l 7 cannot be*

correct: the normal distribution is symmetric, so it probably should be "below the 5th percentile or above the 95th percentile."]

Response: TCEQ has made appropriate revisions pursuant to these comments (i.e., text was revised to indicate that BMD modeling can be used to identify a POD comparable to a NOAEL, instead of a threshold, and 95th percentile was corrected to 98th percentile).

Reviewer 9

P. 161: Cell proliferation is a key event (and one that is occurs in any carcinogenic process), not a MOA. While the term may have been used that way by Swenberg et al., it is better not to perpetuate the terminology.

Response: TCEQ carefully considered this comment, but has not made revisions as "cytotoxicity-induced cell proliferation" captures the two MOA key events commonly referred to as a carcinogenic MOA.

The approach for comparing dose-response and temporal relationships between mutagenic events in the target tissue and a tumor was addressed in concept by Moore et al. 2008:

- *Moore, M.M., R.H. Heflich, L.T. Haber, B.C. Allen, A.M. Shipp, R.L. Kodell. 2008. Analysis of in vivo mutation data can inform cancer risk assessment. Reg. Toxicol. Pharmacol. 51(2):151-61.*

Response: This reference was added to discussion in the text.

It may be useful to supplement the equations on P. 163 with text explaining the concept – that the ADAF is applied for only a relatively small portion of the lifetime. Thus, the impact of the ADAF on the risk estimate is a change of only 60%, even though the adjustment for early years is a factor of 10.

Response: The relevant time intervals applicable to the ADAFs are already given, but some revisions were made to make the early-life exposure period affected (0-16 years) even more clear.

Reviewer 14

None.

Response: No response required.

2.3 Specific Issues for Hazard Characterization and Exposure-Response Assessment Using Epidemiology Data

2.3.1

The major topics covered in Chapter 7 are listed below. Are there additional important topics that should have been discussed in these guidelines? Should any of these topics have been covered in more depth (keeping in mind that the Guidelines mainly present summary information with references to key documents that contain additional information)?

7.4 Endpoint Selection
7.5 Exposure Characterization
7.6 Exposure Metric
7.7 Dose-Response Models
7.8 Quantitative Cancer Exposure-Response Characterizations
7.9 Excess Risk Calculations for the General Population
7.10 Determination of URFs and SFo Values from Dose-Response Modeling
7.11 Meta-Analyses
7.12 Reality Checks
7.13 Uncertainty Analysis

Reviewer 1

7.4 Endpoint Selection: Why is MOA discussed here? Epidemiologists use disease mortality or incidence. I see no other examples here. MOA would come into play only if one believes that an endpoint other than disease (but key to it) could be used as a surrogate. Is that what is being proposed here? If so, you need a lot more discussion with many examples and justifications.

Response: TCEQ agrees that epidemiologists use disease mortality or incidence and are not suggesting "that an endpoint other than disease (but key to it) could be used as a surrogate." The reason for bringing up the MOA is given/emphasized in the third paragraph in 7.4; namely, "There are limitations to assessments of toxicity endpoints that involve multiple MOAs." and "Thus, the TCEQ takes special care when the toxicity endpoint involves grouping of responses with possibly multiple target tissues, mechanisms of action, MOAs, and severities." No changes were made.

7.5 Exposure Characterization: I have no problem with this.

Response: No response necessary.

7.6 Exposure Metric: I have no problem with this, except that I would expand on the subject of exposure metric relevance (p. 17, lines 12-14.).

Response: It is difficult to expand on this subject without being overly restrictive. However, the general nature is captured in the current text "When several exposure metrics are available to characterize exposure-response, the TCEQ will evaluate the relevance of each exposure metric being considered for selection in terms of the specified toxic response, the chemical, and its mechanism(s) and mode(s) of action." No changes were made.

7.7 Dose-Response Models: This section is out of date with regard to thresholds, as I discussed concerning the previous Chapters.

Response: TCEQ agrees that thresholds have not been thoroughly discussed thus far. However, at this point in time, we have only expanded the last sentence in the second paragraph of 7.7.7.1. TCEQ will consider a more thorough discussion of thresholds in future revisions of the Guidelines.

Section 7.7.1, line 7: "genotoxic" should be replaced by "mutagenic".

Response: In this section, the original Crump and Allen article in 1985 used the word "genotoxic" which is why it is used it here. However, for greater clarity text was added to the next paragraph.

I do not understand the point about "those acting at a site where cancers occur spontaneously". Why does this make it biologically plausible to use a linear model?

Response: Text was added for clarification.

I assume this means linear but with a positive Y intercept.

Response: Yes.

I also do not understand the statement #3 (line 12). Suppose more than 1 model fits the data? How do you choose? Philosophy?

Response: Although goodness-of-fit and choice among competing models with approximately the same goodness-of-fit is a worthwhile discussion, here the choice is based on, (1) "In the absence of mechanistic information about the carcinogen and the health endpoint, the linear exposure-response model is the most parsimonious and simplest polynomial that should be used to fit epidemiology data" and (2) although some authors have used more sophisticated exposure-response models for epidemiology data, there has not yet been any statistical evidence showing any superiority of these models over the linear model in describing the relationship between exposure or dose and cancer endpoints.

7.8 Quantitative Cancer Exposure-Response Characterizations: OK assuming linear D/R

Response: No response necessary.

7.9 Excess Risk Calculations for the General Population: I have no problem with this.

Response: No response necessary.

7.10 Determination of URFs and SFo Values from Dose-Response Modeling: For linearity, this is reasonable. Section 7.10.2 deals with non-linear Models. More needs to be said about MOA and determining whether response is really linear.

Response: TCEQ will consider this for future revision of the Guidelines.

7.11 Meta-Analyses: I have no problem with this, assuming a linear D/R. However, looking at Appendix B, where this assumption was made, it seems to have been made in contradiction to the data. Figure B.4 (Lubin et al. data) appears to have a threshold up to about 6000 µg.m^3-years. Figure B.6 (the Jones et al.

data) shows a clear threshold up to at least 400 µg.m^3-years. It is not possible to see the lower doses from the other studies. Do they have thresholds?

Response: The Lubin *et al.* and Jones *et al.* data do seem to have a threshold if the higher SMRs at the lowest non-zero exposure group are ignored. Section 7.7.7.1, however, indicates that in the absence of mechanistic or biological data that justify a model other than linear, the TCEQ uses linear dose-response models. There is no mechanistic or biological data related to exposures to arsenic that warrants a specific dose-response shape.

7.12 Reality Checks: We need more of this. How about applying it to actual data vs. philosophy, as in Appendix B? Why not look at the data to see if it shows a threshold?

Response: The purpose of Appendix B is to illustrate how to carry out a meta-analysis. The purpose of the Guidelines is to indicate and illustrate, when necessary and when possible, different aspects that TCEQ needs to consider when conducting a risk assessment. Application of risk assessments to specific compounds belong in the development support documents developed by TCEQ for specific chemicals.

7.13 Uncertainty Analysis: I have no problem with this.

Response: No response necessary.

Reviewer 2

Given the fact that the guidelines cannot hope to cover every case or anticipate every circumstance, and the fact that epidemiology studies vary tremendously in design and extent of reporting, it is my opinion that the discussions of these topics have been sufficient to describe the approaches or "thought processes" that TCEQ is likely to apply in the instances where epidemiologic data are available for a risk assessment. Perhaps some discussion of how (in general) to deal with study designs that are not the typical occupational cohort, retrospective follow-up designs would be useful; I am thinking of case-control and cross-sectional studies in particular, designs that have their own challenges for use in risk assessment.

Response: The "important topics" in the last paragraph of Section 7.2 was expanded to acknowledge this issue.

Reviewer 9

The modeling text addresses covariates, but it would be useful to address bias and confounding (using those terms) qualitatively in the context of general evaluation of studies.

Response: Confounding was added to the list of things considered in the general evaluation of studies.

Reviewer 14

Perhaps, a Section should be added on how results from animal studies may be used to support or clarify results from epidemiological studies.

Response: Text, as well as several references to key papers, has been added to Section 7.12.

2.3.2

Should the discussion of endpoint Selection (Section 7.4) include any additional pitfalls associated with using certain types of endpoints for risk assessment?

Reviewer 1

Such as what? See my comments for 7.4 above [In section 2.3.1 above].

Response: No response required.

Reviewer 2

I can think of no additional pitfalls that need to be listed (although I must comment that the discussion on P. 12, l 16 – 32, does not really appear to relate to the issues of endpoint selection). And note the typo on line 18: 'MOA' rather than 'MOE.'

Response: The reasons for the discussion are, (1) "There are limitations to assessments of toxicity endpoints that involve multiple MOAs" and (2) "the TCEQ takes special care when the toxicity endpoint involves grouping of responses with possibly multiple target tissues, mechanisms of action, MOAs, and severities." The typo has been corrected.

Reviewer 9

The issue of mortality vs. incidence data should be noted. Mortality is a good surrogate for incidence for tumors (or other endpoints) that have low survival rate. However, mortality is a poor measure of cancer incidence for some tumors with high survival rates (e.g., skin cancer, and some other types). (This issue is addressed within the context of modeling and quantitative estimates on P. 32, but it would be useful to address the generic qualitative issue and implications in this section).

Response: Mortality vs. incidence is raised in the last bullet at the end of Section 7.4. (As noted, mortality vs. incidence is discussed in 7.7 Dose-Response Models and 7.8 Quantitative Cancer Exposure-Response Characterizations.) Two sentences were added to the last bullet at the end of Section 7.4.

Reviewer 14

Some discussion of potential confounding factors, e.g., cigarette smoking, should be considered in the selection of endpoints.

Response: A bullet addressing this was added at the end of Section 7.4.

2.3.3

Should the discussion of Exposure Characterization (Section 7.5), include additional issues (e.g., temporality, measurements, models, reasonableness of underlying modeling assumptions, exposure estimation errors, grouped versus individual exposure values, and biomonitoring) associated with using certain types of exposure characterizations for risk assessment? Should the discussion of industrial hygiene measurements be expanded and, if so, how?

Reviewer 1

I have no problem with this.

Response: No response required.

Reviewer 2

I think all the relevant considerations are listed in Section 7.5. It might be worth pointing out that the relative importance of these factors will vary from case to case.

Response: Text was added to the text preceding the list of factors.

I think the guidelines are correct in emphasizing that exposure assessment is very often the most difficult and error-filled aspect of an epidemiologic risk assessment. (However, on P. 14, l 22, I would not say that all exposure estimates 'involve errors' as if that was the intention; I think I would say that all exposure estimates have errors associated with them, for the reasons listed.)

Response: The text was modified accordingly.

A brief sentence or two describing why the 95th percentile of a distribution is likely to be over-estimated by the sample values would be valuable; the reasoning there (say, in contrast to what is happening with the mean or median) would be illustrative of the types of concerns that need to be considered. I find that, often, statistical insights are more difficult to grasp for many risk assessors than are principles of other aspects of risk assessment.

Response: Many people find that statistical insights are more difficult to grasp for many risk assessors than are principles of other aspects of risk assessment. This one is no different, and maybe harder than most. A simple example illustrates the overestimation bias in the higher sample percentiles when the sample data contain errors:

> If the true value of a variable under investigation is always equal to 200, and the error impacting the sample value is a multiplicative error, which is equally likely to be -10%, 0%, or +10%, then the sample values would be equally likely to be 180, 200, or 220. The expected value of the higher percentiles is nearly 220. Thus, the bias in the higher percentiles is nearly +20.

For clarity, an explanatory reference was added and text in the second-to-the-last sentence in the 5-th paragraph in Section 7.5 was expanded.

Reviewer 9

It's confusing to address biomonitoring and exposure measurement techniques with so little transition (P. 13 at the bottom), although the topics are related. For the text on biomonitoring, it would be useful to cite the ACGIH Biological exposure indices (BEI) documentation. This documentation provides very useful information on many topics related to the potential BEIs for each chemical covered, including specificity of the biomarker, and whether the biomarker reflects short-term or chronic exposure – all of which should be considered in determining whether biomonitoring data can be used to reflect exposure. I think this information is more useful for exposure characterization than the Hays work (as nice as that work is), which focuses more on how to interpret biological monitoring data relative to an RfD or RfC, rather than the appropriateness of specific biomarkers.

Response: Text was added at the end of the 3rd paragraph in 7.5.

The advantage of biomonitoring data in integrating exposure from multiple routes should be noted (since the observed effect may reflect both the occupational exposure and exposure from other sources – and the inhalation equivalent needs to be determined for this integrated exposure).

Response: This is a complex issue. The biological response is a function of the integrated exposure from all sources and pathways. However, only some of these sources and pathways can be regulated. Thus, the risk and exposure related to that component needs to be identified. Therefore, the "advantage" of the integration associated with biomonitoring data is not always clear. This is at least partially why the text "Ideally, the exposure concentrations in an exposure history are based on accurate analytical measurements (e.g., biomonitoring where the relationship between the exposure concentration and the biomonitored value is known)" included the comment "(e.g., biomonitoring where the relationship between the exposure concentration and the biomonitored value is known)".

The text on P. 15 (at the end of Section 7.5) on biomonitoring addresses the topic more completely (and should be integrated with the text on P. 13), but the issues noted earlier in this paragraph apply. In particular, the statement that biomonitoring may be of limited value is overly strong (and one I would disagree with, particularly as the field develops over the next 5-10 years). Methods are being developed to address the issues noted, and for many chemicals, appropriate biomarkers of long-term exposure have been identified; one needs to ensure that the study used an appropriate biomarker.

Response: Text on P. 15 (at the end of Section 7.5) on biomonitoring has been moved and modified to follow the text on P.13 related to biomonitoring.

With regard to the IH measurements, it would be useful to mention the preference for personal sampling over area sampling.

Response: This is covered in the first sentence "with personal measurements being better than area measurements". No changes were made.

In the discussion of databases for exposure models, it would be useful to note the different types of data addressed by different databases (e.g., food vs. occupational). It would also be useful to address how TCEQ staff would use the exposure databases in their assessments, given the lack of connection to effects. (The utility of these databases and models for ESL development is not clear to me, unless the TCEQ staff would be doing a de novo epidemiology assessment – but my focus area isn't exposure.)

Response: A discussion of the databases underlying the different exposure models mentioned in 7.5 is contained in the references cited and is too voluminous to be incorporated explicitly in the text.

The text on categorization on P. 15 can note that confidence in the result is increased if the author conducts a sensitivity analysis of the impact on the ultimate result of various cutpoints for the categorization.

Response: A sentence has been added at the end of the last paragraph in 7.5.

Reviewer 14

Additional discussion not needed.

Response: No response required.

2.3.4

Should the guidelines have discussed additional exposure metrics (Section 7.6) or discussed additional issues related to cumulative exposure metrics (e.g., cumulative ppm-years)? Is the discussion of the

advantages and disadvantages of using transformed exposure values as the exposure metric for exposure-response modeling complete? Are there additional issues and/or references relating to including more than one exposure variable (e.g., a second characterization of the exposure to the primary chemical, exposures to other chemicals) that should have been cited?

Reviewer 1

I have no problem with this [referring to all portions of this charge question].

Response: No response required.

Reviewer 2

I liked the discussion of Section 7.6. I think the advantages and disadvantages of the cumulative dose metrics have been adequately presented for a summary-type document. The presentation (P. 16, l 38) of the general form ($(C-C_0)^n$ x $(T-T_0)^m$) is a valuable inclusion. It might be enhanced further still if it could be presented, in equation form, in conjunction with the weighting that was mentioned in the paragraph starting on line 3 of that page. That would show the full range of cumulative dose metrics, a class of metrics that I think almost always includes the best options for dose metric selection (as opposed to duration or intensity alone, even though those two options are included in the class shown by $(C-C_0)^n$ x $(T-T_0)^m$ with suitable selection of n or m to be zero).

Response: The following sentence was added: "Such metrics can also be weighted as discussed above."

The one issue that I have with this section is in the paragraph stating on P. 17, l 15. It is not clear how the differences between the modeling scenario and the inference scenario are relevant to the consideration or determination of an appropriate dose metric. In fact, the whole purpose of a dose metric is to allow a rational extrapolation from one exposure scenario to another. It is not the differences in the scenarios that define the metric, but rather the metric that defines the differences in the scenarios (from a risk perspective). That is why different dose metrics matter; different metrics have different risk implications, even when starting from a given modeled exposure scenario (i.e., from one epidemiologic study). In essence, what I am saying is that, by and large, the choices related to dose metric should be made independently of the differences in the modeled and inference scenarios; to the extent that a priori, biologically motivated decisions can be made, the stronger the support will be for the dose metric(s) ultimately used in the risk assessment.

Response: TCEQ agrees with this comment and the text has been modified accordingly.

Reviewer 9

The importance of tying the exposure metric to the chemical's MOA should be noted. The text alludes to the point with the note that the cumulative exposure metric assumes that cumulative exposure is more biologically relevant than other measures, but the idea should be developed more completely. This then suggests some other useful metrics to note – e.g., peak exposure, number of events (or duration of) exposure above some level (e.g., high intensity tasks).

Response: Text has been inserted as the fourth paragraph from the end of 7.6.

Reviewer 14

Exposure metrics is a complex issue. The material on Page 16, lines 37-42, is important and warrants more discussion.

Response: TCEQ agrees this is important; therefore, text was modified and an additional reference (USEPA 1998) was added.

The discussion of transformed exposure values for the exposure metric for exposure-response modeling is adequate. I am not aware of additional citations relating to including more than one exposure variable.

Response: No response required.

2.3.5

Is the discussion of Dose-Response Models (Section 7.7) used in epidemiology studies complete or should have additional classes of dose-response models been discussed? Does the document accurately and completely present the issues/pitfalls associated with dose-response modeling based on reported summary characterizations (e.g., SMRs, RRs, ORs)?

Reviewer 1

I have no problem with this [referring to all portions of this charge question].

Response: No response required.

Reviewer 2

I believe the guidelines do present a fairly complete discussion of the modeling approaches available (and routinely used) for an epidemiologically based risk assessment. I have a few specific comments about this section.

- *The claim (P. 17, l 29-34) that TCEQ will use a model that conforms to the assumption of linearity at low doses has implications about the dose-metric as well as the dose-response model. In particular, that position means that TCEQ would never use a metric of the form $(C-C_0)^n \times (T-T_0)^m$ with C_0 greater than zero, a form that was discussed in the previous section. For such a metric, no matter what the model, the resulting curve would be non-linear; in fact it would have a threshold below C_0.*

Response: The intent here is to be linear in the exposure metric not necessarily linear in the administered dose or exposure concentration. Clarifying text was added at the end of the paragraph.

- *Starting with Section 7.7.4, and in many places following (in Section 7.7 and beyond) there is major confusion and misleading verbiage concerning the adjustment for background hazard rate and its implications for risk assessment practice. I agree that one may need to adjust for differences in the background hazard. But at the stage of the modeling, the target population (what is elsewhere referred to as the inference population or inference situation) has no bearing on that adjustment at all.*

Response: TCEQ agrees.

What is important for the modeling is that the members of the epidemiologic study population may differ with respect to background hazard from the reference population. This is the population that is used to determine what would be expected in the study's population if they had not been exposed, and it has nothing to do with the population for which TCEQ ultimately wants to estimate risk. As an example, if a Chinese worker cohort is studied, the study authors would most like have picked a Chinese reference population as the basis for estimating what mortality patterns those workers would have experienced had they not been exposed; that population would share many common characteristics and behaviors with the workers and would be judged "most similar" to the workers with respect to mortality patterns. But, the workers will still differ (perhaps in many and unknown ways) from the more general population, so it is appropriate (even important) to allow the background hazard rate to differ, by estimating a model parameter that adjusts the rate(s) associated with the reference population to be more relevant to the worker subpopulation.

Response: TCEQ agrees.

So, the equations on P. 19 are correct. Although, to call λ_0 the "cohort's estimated background hazard rate" (P. 19, l 26) does not seem quite right: it is the rate suggested by the reference population, probably looked up and so not estimated at all, in the context of the dose-response modeling under consideration. Rather, it is the product of λ_0 and α that is the estimated (because α is estimated) cohort background hazard rate.

Response: TCEQ agrees, cohort's was changed to reference population's.

The bottom line here (and in additional discussions referenced in my comments below), is that the target (or inference) population has no bearing at all at this stage of the analysis. Whether the ultimate population of concern for TCEQ was all of Texas or the male population of Timbuktu, the dose-response modeling discussed in this section would be exactly the same.

Response: TCEQ agrees.

The guidance could possibly comment on the use of Texas population rates with the model parameters (as is done on lines 30-34 of P. 19), but that is done in later sections (where they are more appropriate and relevant to the step of the risk assessment in question) and I think it just adds more confusion to that already introduced by mention of the target population on line 20 of P. 19. This same confusion is present on lines 29-34 of P. 21; while technically correct (background hazard for the inference population is used in the risk calculations) it is out of place in a discussion of "Parametric Dose-Response Models."

Response: Text was modified for clarification.

- *P. 22, l 2: the dose-response model need not be limited to polynomial functions. Models with non-integer powers on dose are perfectly good models (that also have the linear model as a special case).*

Response: Although non-integer powers are possible and have sometimes been used, our experience is that fractional powers less than 1.0 have frequently lead to models that are not biologically justified.

The text says that "Generally, the dose-response model can be assumed to be a polynomial function of the dose metric" rather than saying that they have to be polynomial models.

- *P. 22, l 23: It would be good to define here exactly what the rate ratio is: the ratio of the risk at a particular dose to that in the absence of the exposure. Otherwise, the denominator for this equation appears without context and without the less-informed reader knowing why that particular ratio and its linear form are of interest.*

Response: The following sentence was inserted: "The rate ratio compares the rate of events at a dose to the corresponding rate when the dose is zero."

- *P. 24, l 10-11: This statement that splines should not be used in place of biological or mechanistic interpretations is not particularly "damning." The same could be said about any of the other models discussed in this section.*

Response: TCEQ agrees.

Reviewer 9

The text appears to be complete and accurate.

Response: No response required.

Reviewer 14

A good 10-page discussion of Dose-Response Models used in epidemiological studies is provided in Section 7.7. Issues associated with dose-response modeling based on reported summary characterizations are presented in Section 7.7.

Response: No response required.

2.3.6

The guidelines recommend using best estimates (e.g., maximum likelihood estimates) as the basis for comparing the exposure-response of two chemicals (Section 7.8). Is this recommendation consistent with accepted risk assessment practices? Is the recommendation clearly explained and well supported by the available references? Has the document completely discussed the issues/pitfalls/references associated with using bounds (e.g., 95% LCL, 95% UCL) as the basis for points of departure (PODs)?

Reviewer 1

I have no expertise in this area.

Response: No response required.

Reviewer 2

I think the document is a little weak in regards to the use of the MLE vs. the bounds. I would agree with the contention that for comparative purposes, the MLE estimates are most appropriate. However, the justification (e.g., see p. 29, lines 9-16) is not strongly presented. Why, and in what way, is the LEC "much less responsive to the observed epidemiology data?"

Response: A new sentence and reference were added for clarification.

Furthermore, with respect to a POD, I think it could be argued strongly that the LEC is a better basis for health-protective extrapolation to risk estimates of interest. The extent to which the LEC may be "determined more by study designs and statistical assumptions than the observed data" is debatable, but even if that were absolutely true (in relation to the EC), do not the designs of the studies and the assumptions used to analyze them need to be reflected in the estimates from which reference values are to be derived?

Response: The text does indicate that TCEQ current policy is to report both best estimates and bounds. EPA's standard practice is to use the LEC as a POD – arguing that this is most health protective. These Guidelines do not provide the space to fully discuss this. While EPA's practice may be the most health protective looking at only one specific case, there is a question as to whether it leads to the greatest health protection overall in the sense that risk at the resulting regulatory acceptable air concentration (LEC) could be zero (de minimis) and it may not result in the most health protective use of limited resources (e.g., divert regulatory focus such that the greatest real risk reduction is not achieved).

How else do we distinguish between two studies (one "good/large" and the other "bad/small") that give the same MLE estimate, unless we resort to the arbitrary and "data-ignoring" application of uncertainty factors?

Response: These factors may be more appropriate for evaluating uncertainty and the confidence to place in the estimate rather than changing the estimate itself. In addition, in analyses where studies are being weighted on the basis of inverse variance, the MLEs based on the "good/large" study would be given more weight because of their smaller variance while the MLEs based on the "bad/small" would be given less weight because of their larger variance of the estimated MLEs.

Other comments on this section are as follows:

- *The statement on lines 4-8 of P. 28 is not true as written and is subject to misinterpretation. While it may be the case that the inference scenario is adjusted or assumed to be the same as the modeled scenario with respect to something like days/week or hours/day for exposure, there is no need to assume or adjust for other aspects of the exposure scenario differences. In fact, the life-table approach discussed elsewhere allows one to extrapolate (with some assumptions) from any exposure scenario that allows estimation of the model parameters to any other scenario of interest. Thus, it is common practice to model occupational exposure scenarios (with, typically, exposure starting only in adulthood and ending some time before death) and then use the modeling results to predict risks associated with constant lifetime exposures without adjusting the one to match the other, in terms of years of exposure, timing of exposure, etc. This is the power of the life-table methodology that TCEQ is recommending.*

Response: TCEQ agrees with what the life-table methodology is intended to do. In an attempt to better clarify what was intended, the text was modified.

- *Comments in this section (e.g., P. 29, l 23) and elsewhere about the EC being within the observed range are not as straightforward as in the case of an animal experiment. Recall that the EC_x*

values are calculated on the basis of a life-table calculation, for a specific exposure scenario of interest (say constant lifetime exposure) that is not the same as the scenarios from which the model estimates were derived. It is not so clear or obvious what "within the observed range" means for such EC values compared to the data used to derive them. It may not even be possible to determine what the "observed range" of lifetime (or up to age 70) risks for the study population members are; by definition many of them would have been observed or follow-up for only some portion of their lifetimes. I suggest that here, and elsewhere in these guidelines, this reliance on estimating values in the observable range is a constraint that is not only needless but difficult or impossible to ascertain.

Response: What was intended was that exposure corresponding to the EC be in the range of the observed exposures, not that the risk corresponding to the EC be in the range of study risks. For example, if the EC was 10 ppm-years, there would be several members of the study cohort that experienced 10 ppm-years, and that 10 ppm-years would neither be much greater or much less than what was experienced by several members of the study cohort. In the preceding response, it was clarified that exposures from the inference situation are converted to exposures in the exposure-response model. The implicit assumption is that equivalent cumulative exposures reflect the same level of risk, regardless of how those exposures where cumulated. In other words, for a specific exposure scenario of interest (e.g., lifetime exposure) and an exposure metric (e.g., cumulative exposure) the relevant variable is cumulative exposure and not the time-profile of exposure concentrations leading to that cumulative exposure. This is the underlying standard assumption when an estimated dose-response model is used to estimate risks in a different situation. In order to better clarify this, the text was modified.

- *The distinction on the top of P. 30 (l 1) about some models not including time or age is a straw man not worthy of inclusion. The multistage model referenced there is surely not an epidemiological problem, since it is but one of the models fit to bioassay data. The "multistage" (two-stage) models that are used in epidemiological analyses absolutely do include age and/or time.*

Response: In order to clarify, the text was modified.

Reviewer 9

This recommendation is appropriate and generally clear. It would be useful to explain a bit more the statement "determined more by study design" – explaining how the confidence limits reflect study design.

Response: An example was added for clarification.

Reviewer 14

The TCEQ Guidelines recommend using best estimates as the basis for comparing the exposure-response of two chemicals. This recommendation is explained clearly on Page 29, lines 4-16, and is consistent with accepted risk assessment practices as supported by appropriate references. Confidence limits are necessarily employed in developing a point of departure for low dose extrapolation in order to allow for

uncertainties in the process. The document adequately discusses issues in the calculation of confidence limits.

Response: No response required.

2.3.7

Is the discussion of excess risk calculations (Section 7.9) complete? Would you recommend adding discussion of unidentified issues/pitfalls related to the calculation of excess risks for a specified general population (e.g., the Texas population) based on the dose-response modeling for the population (e.g., workers) in the epidemiology study?

Reviewer 1

I have no problem with this [referring to all portions of this charge question].

Response: No response required.

Reviewer 2

In general, the discussion in this section is complete. The emphasis on the life-table approach is warranted and appropriate; it is the best option for deriving and comparing life-time risks from various exposures. I would quibble a bit with the statement that the method is "computationally intensive" (P. 31, l 22); since it can be coded and set up in a spreadsheet with documentation sufficient to allow almost anyone to do the computations for almost any exposure scenario of interest, it can hardly be considered computationally intensive. I do agree with the conclusion that approximate methods are not required and should not be used.

Response: Text was changed from "are computationally intensive" to "used to be computationally intensive".

I do not believe that there are any particular unidentified pitfalls associated with the calculations for a specific inference population (e.g., Texans). However, on a related issue, here is the section to discuss and emphasize the fact that the background rates for the inference population of interest (Texas) are what will be used. So, perhaps the discussion of the importance of getting and using appropriate background hazard estimates could be emphasized a little more here. See my comments above about the earlier sections and the confusion caused by inserting comments about the inference population when discussing the modeling; all of that previous discussion should be edited and presented here to avoid that confusion and to make sure that the reader (and importantly TCEQ practitioners) understand how the inference population-specific information (with respect to background rates of all-cause mortality and endpoint-specific mortality) are used in the appropriate manner at this phase of the risk calculations and to highlight the key assumptions that allow one to do that (i.e., that the dose-related parameters of the model estimated from the epidemiology study population are the same as those for the inference population).

Response: As indicated earlier, the text in Sections 7.8 and 7.9 has been clarified.

Other comments on this section:

- *P. 32, l 6-8: Is it necessary or appropriate to constrain the choice of terminal age for the risk calculation based on the assumption of exposure in the inference population at older ages? The inference population is rather like a hypothetical population: exposed at a constant level for their lifetimes. As such, the "fact" of exposure at the older ages is merely a matter of what the risk assessor defines the inference population to be.*

Response: There are two points here. First, it may be unreasonable to substantially extend an estimated exposure-response model beyond its foundation. For example, extending the effect of exposure estimated on the basis of data for ages 20 to 65 to ages substantially beyond age 65. Second, it seems unreasonable to calculate risks using false assumptions about the population that the inference population is trying to represent.

- *P. 33, l 12-14: This parenthetical statement is not needed and not true. The only adjusting that takes place is related to the adjusting of the exposure scenario; the estimated model parameters are not adjusted at all; the first paragraph of this subsection makes it clear that the dose metric should be exactly the same as used for the model fitting; therefore any adjustments to the model parameters would be wrong, since their values are a function of the metric used in their estimation.*

Response: The intent was not that the model be adjusted but that the exposure calculation be adjusted. Given that the exposure calculation and the transformation from the inference scenario back to the estimation scenario (i.e., the epidemiology study scenario) have been extensively discussed, the best thing (to avoid confusion) is to delete the parenthetical comment. Thus, the parenthetical statement has been deleted.

- *P. 33, l 21-27: Does TCEQ not adjust for the number of weeks of exposure per year? I have seen a factor of 52/50 used to convert from occupational settings, where it is assumed that the workers went to their jobs 50 out of the 52 weeks in a year.*

Response: No, TCEQ does not adjust for the number of weeks, but rather adjusts for occupational exposure being for 5 days per week and environmental exposure being 7 days per week. This means that occupational exposure is approximately 71% of the year. There are other candidate adjustments. Some other people have further adjusted by 50/52 weeks which would result in approximately 69%, or, equivalently, adjusted for 250/365 days per year. None of these adjustments are necessarily right or wrong, and none make much difference. Some people have argued that both environmental and occupational ought to be 50 weeks per year because both typically involve 2 weeks of vacation.

- *P. 33, l 36-37: it would be nice to have at least a brief (one or two sentence) description here of how the ADAF life-table adjustment of Sielken and Valdez-Flores differs from what EPA did.*

Response: A detailed explanation is given in the cited reference (Sielken and Valdez-Flores (2009a)); however, a simple informative one-line explanation is not really possible. The text has been modified to indicate that it was a mathematical failure (i.e., a failure to follow the mathematical procedure rather than some other more general philosophical or biological failure).

Reviewer 9

The text is generally clear and complete. However, it is a bit confusing to have the discussion of the slope (β) followed by discussion of the POD and slope calculated by drawing a line from that point. It would be useful to clarify the relationship between the two. If one is using β to determine the slope, why is a separate URF calculated?

Response: The reason for the discussion is that the slope β (MLE and bounds) is used to calculate the POD. However, the risk slope (e.g., URF = 0.10 / LEC10) is not β. Clarifying text was added.

It would be useful to provide some additional context/rationale to the text on dosimetric adjustments, perhaps referring back to where these are discussed in the remainder of the guidelines.

Response: The text was changed to refer back to 7.6 and 7.8.

It would be useful to explain further what the issue was with EPA's application of the ADAF.

Response: Please see last response to Reviewer 2 above.

Reviewer 14

The discussion of excess risk calculations in Section 7.9 appears to be adequate.

Response: No response required.

2.3.8

Do the Guidelines clearly explain how TCEQ will determine whether to use linear or nonlinear dose-response models to develop URFs and SFo values (Section 7.10)? Are the recommended practices consistent with accepted risk assessment practices and completely supported by the available references? Should the document include any additional issues/pitfalls/references related to the choice between linear and nonlinear extrapolation below points of departure (PODs)?

Reviewer 1

I have discussed this issue already. I suggest that this decision be based on actual data rather than on philosophy.

Response: No response required.

Reviewer 2

This section does not explain how TCEQ will determine whether to use a linear or nonlinear model. In fact, that determination is best left to other chapters of the document (Chapters 3 and 5). What is important here is that this section describes how the URFs and SFo values will be derived once that determination has been made (which is not entirely, nor perhaps predominantly, an epidemiological determination). In that regard, this section is adequate. As I have commented above, I would argue in favor of using lower bounds on EC estimates as the basis for URFs and SFo's, but the decision by TCEQ to present URF(MLE) and URF(95% UCL) values is a good one. Also as mentioned above, the insistence on PODs within the range of observations is problematic for epidemiologically derived PODs. [If there is intended to be a distinction between URF(95% UCL) and the URF(LEC), it is obscured by the definitions on

lines 18-21 of P. 35: those two sentences are identical except for the fact that one has URF(95% UCL) and the other has URF(LEC).]

Response: TCEQ does "address the potential" but does not necessarily go to the extreme of ruling it out. Note that determining the MOA is not linear at low doses (the first part of the EPA sentence and the TCEQ text commented on) implicitly requires consideration of mutagenicity or another activity creating a low-dose linear MOA (the second part of the EPA sentence). Furthermore, simply demonstrating mutagenicity (or often simply genotoxicity) is possible in certain cells and under certain experimental conditions, which may not be relevant to biological conditions in the target tissue of interest, is not tantamount to demonstrating a mutagenic MOA is operable in potential target tissues at environmentally-relevant doses.

Reviewer 9

The text is clear. However, the recommendation to focus on the MLE, rather than the upper bound on the unit risk appears to be different from the recommendation in Section 5.7 of the TCEQ guidance regarding considerations for determining whether to use the central tendency or bound (e.g., the EC or LEC). Unlike the list of considerations in Section 5.7, chapter 7 appears to place strong preference on use of the central tendency.

The EPA guidelines state that "A nonlinear approach should be selected when there are sufficient data to ascertain the mode of action and conclude that it is not linear at low doses and (emphasis in the original) the agent does not demonstrate mutagenic or other activity consistent with linearity at low doses." It appears that the TCEQ guidelines are stating that it is sufficient to have MOA information indicating a nonlinear dose-response relationship, but that one does not need to also rule out mutagenic or other activity consistent with low-dose linearity. Is this correct? While the EPA approach may be overly conservative in some ways, I would think that it is necessary and appropriate to address the potential for linearity or linear components at low doses.

Response: There is no intended distinction between URF (95% UCL) and URF (LEC). If both notations were to be used, then URF (95% UCL) = URF (LEC). In order to make things less confusing, text has been changed to URF (95% UCL) and not URF (LEC).

Reviewer 14

The TCEQ Guidelines clearly explain whether to use linear or nonlinear dose-response models to develop URFs and SFo values in Section 7.10. The recommended procedure is consistent with accepted risk assessment practice and supported by available references. No additional discussion is needed.

Response: No response required.

2.3.9

TCEQ provides recommendations for how to determine the appropriate choice of the risk level (e.g., 1-in-a-thousand, 1-in-ten-thousand, 1-in-a-hundred-thousand, 1-in-a-million) for estimating the point of departure using linear low dose extrapolation (Section 7.10). Are these recommendations complete and consistent with accepted risk assessment practices? Please suggest any additional issues/pitfalls/references that should be added to the document.

Reviewer 1

This is a political question, not a scientific one.

Response: No response required.

Reviewer 2

I have commented on the problems associated with the choice of a POD close to or within the range of the observations. As a general rule, I do not think this is necessary, and, as argued above, I think the determination of whether a particular POD is or is not in the range of the observations is not necessarily a simple determination.

Response: See response to Reviewer 2 in 2.3.6 above.

Therefore, the recommendations for when to use a particular risk level (1-in-a-thousand vs. 1-in-a-hundred-thousand) are not likely to be achievable. Frankly, for any epidemiology study (or for any analysts using an epidemiology study) to claim that 1-in-a-hundred-thousand lifetime risk is "within the range of the study" is ludicrous. That applies, for all epidemiology studies of which I am aware, even for risk levels of 1-in-ten-thousand. So, I do not see how one could ever, realistically, be considering risk levels less than about 1-in-a-thousand as the basis for PODs. I would argue that TCEQ could simply use that level as the default for all epidemiologically based risk assessments (and present other PODs to the extent that the analysts can rationalize them).

Response: There is a wide range of risks experienced by the individuals in an epidemiology study. Some individuals may have only background risks (zero excess risk). Other individuals, as they progress through time (age), may have very small excess risks at early ages and progress to much larger and more readily observable risks by the end of the study.

In general, it is not advisable to a priori establish one specific risk level for all PODs (all chemicals, etc.). In particular, it is not advisable to use an excess risk level to define a POD if the corresponding exposure level is beyond (above or below) the exposures experienced by the individuals in an epidemiology cohort. Using too high (or low) a risk level that result in PODs beyond the observed range of exposures increases the uncertainty in the estimates and is statistically inappropriate.

Reviewer 9

The recommendation to use a POD that is in the range of the observed data, but at the low end of that range is appropriate. However, given the text notes that in some cases epidemiology data are available in the range of low risk levels (1/100,000 or 1 in a million), it would also be useful to address the case where extrapolation is not needed, because the data can be used directly to estimate risk.

Response: Sometimes the dose corresponding to a low risk level can be estimated directly from the data. This was the case for ethylene oxide and butadiene: "… or in some cases as low as 1/100,000 or 1/1,000,000 (e.g., ethylene oxide (Valdez-Flores and Sielken 2010) and butadiene (Sielken and Valdez-Flores 2011))." Although the policy is to use the POD to characterize the low-dose risk, it is possible in some cases to use the estimated exposure-response model to estimate the risk corresponding to a specified low dose. When to do the latter instead of the former is a possible subject for a future revision of the Guidelines.

P. 35, l 20 is a bit confusing as worded – particularly because the URF isn't a concentration – it's a risk per unit concentration.

Response: See response to Reviewer 9 in 2.3.8 above.

While I agree that the MLE is preferred for risk-risk comparisons, the rationale is not clearly presented. Can't different dose-response data (for the same study design) also apply to the MLE (different curves going through the same point)? It may be clearer to add explicitly that the sample size is a key determinant of the difference between the MLE and lower bound, and thus can distort comparisons between chemicals.

Response: Text in Section 7.8 has been changed to include sample size in this context.

It's not clear what is intended by the statement that the bounds on the URF may not be the best estimates for risk management. While I agree that this is at least partially a policy decision, most risk management for EPA for carcinogens is based on the bounds – when risk management includes such issues as limits on amount of exposure, cleanup goals, etc.

Response: It is true that in the past most regulatory risk management for EPA for carcinogens has been based on the bounds. The text has been modified.

Reviewer 14

The TCEQ Guidelines provide complete recommendations in Section 7.10 for selecting a risk level for the point of departure that are consistent with accepted risk assessment practices.

Response: No response required.

2.3.10

Is the discussion of the general approaches to quantitative meta-analyses (Section 7.1), including discussion of combining risk measures, slope estimates, or data sets, complete? Please suggest any unidentified issues/pitfalls related to quantitative meta-analyses that should have been discussed.

Reviewer 1

It seems fine.

Response: No response required.

Reviewer 2

The descriptions and discussions of meta-analytical approaches included in the guidelines appear to be complete, at least insofar as can be accomplished with a summary. However, it is not clear to me that the many and varied problems associated with conduct of a good meta-analysis have been given sufficient weight in the document. Meta-analysis is a very difficult exercise to complete adequately; perhaps that should be made clearer in the document so that readers/regulators do not expect meta-analysis to be the norm rather than the exception.

Response: Meta-analyses are a complex topic and can only be incompletely addressed in a few pages. A sentence was added to the second sentence in Section 7.11.

What I found to be somewhat misleading were the parts of the discussion that referenced combining URFs. I would not expect URFs to be the product of any standard epidemiology studies; as noted elsewhere in the document, the URF is calculated using the life-table approach for a specific target or inference population. Therefore, the discussion on P. 39 (l 30-37) and P. 40 (l 1-23) that talks about combining estimates of URFs is basically meaningless. How or when would you ever expect to see an epidemiology study (let alone more than one) with URFs calculated for the inference population of interest to TCEQ (i.e., the Texas population)?

Response: TCEQ agrees, epidemiology studies rarely publish URFs and it is likely that there will not be many URFs published and based on Texas background rates. However, there are some studies that publish dose-response models and corresponding parameter estimates that can be used to derive URFs based on life table analyses using specific population background risks. Thus, URFs can be derived using the life-table approach from several publications with different exposure metrics, different models, etc., and these URFs can then be combined using the methodology discussed in Section 7.11.3

The discussion of the combination of slope estimates has more likelihood of being useful. But, as noted, for such a combination to be undertaken, the same model form and dose-metric would need to have been used in order for those slope parameters to be comparable. In fact, as implied by the statement on lines 31-32 of P. 39, one might be interested in combining results where the modeling was done with a linear model in one case and a polynomial model in another case.

Response: TCEQ agrees, this is one of the "limitations" of combining the slope estimates. It is doable if the data to fit models that satisfy all the conditions are available. However, your comment "as implied by the statement on lines 31-32 of P. 39, one might be interested in combining results where the modeling was done with a linear model in one case and a polynomial model in another case" points to the usefulness of 7.11.3 of combining URFs that was discussed above in reference to your previous comment.

The resolution of this problem is at least hinted at near the close of section 7.11.4 (P. 41, l 1-4), where it is stated that the combining of TCEQ-derived risk estimates could be a possibility. I would strongly encourage that point of view, and suggest that it be given a more prominent and thorough discussion. If that approach is followed, regardless of the models, dose metrics, or study populations, one can derive, using the study-specific estimates, an EC_x for each study (using the inference population of choice) and combine the EC_x values using meta-analytical techniques. This requires that each study publish or otherwise provide all of the information necessary for an EC_x calculation via a life-table analysis, but if that is not the case for a study, that study would not be useful in the (quantitative) risk assessment to begin with.

Response: In this section, TCEQ's performance of a meta-analysis combining URFs that TCEQ determines from individual epidemiology studies is being discussed. An individual epidemiology study would never be expected to contain that type of meta-analysis. Of course, there are several examples of meta-analyses of different types in the literature.

Reviewer 9

I am not aware of any additional discussion that is needed.

Response: No response required.

Reviewer 14

The TCEQ Guidelines provide an excellent 5-page discussion of quantitative meta-analyses.

Response: No response required.

2.3.11

In Section 7.12 (Reality Checks), TCEQ discusses the steps that can be used to at least partially evaluate the reasonableness of dose-response modeling assumptions and resulting estimates and bounds. Is this discussion complete, or are there additional approaches to reality checks that the TCEQ should be aware of? Can you recommend additional guidance concerning how the TCEQ should incorporate/utilize reality checks in its Development Support Documents (DSDs)?

Reviewer 1

If the final result suggests that the only safe concentration is below ambient levels, there is a problem (see analysis of chromium 6 in drinking water). Again, I would like to stress that real data should trump philosophy when deciding on thresholds.

Response: The text was modified accordingly.

Reviewer 2

I know of no other general approaches to reality checking. It is important, as emphasized throughout the document, that this be considered case by case.

Response: TCEQ agrees, clarifying text was added to the end of the first sentence in this section.

It is important to note in the discussion on lines 5-13, P. 42, about cross validation using other studies, that to be able to compare URFs derived from one study to results from another study, the exposure scenario (including exposure levels) of that other study will have to be known; this may be problematic, or else that study might well have been used for its own URF estimation. And, really, it is not the URF that should be compared, but rather a risk estimate relevant to the population comprising that other study. The risk estimated to be associated with some 'typical' exposure level from that other study would be a more appropriate basis for the cross-study reality check.

Response: TCEQ agrees, hence the emphasis in the last line of this paragraph on "predict the number of responses in the other population and the reasonableness of this prediction evaluated." No changes made.

Moreover, the discussion on P. 42, lines 14-24, insofar as it relates to attainable levels, is not a reality check of the data and their analysis. It may very well be the case that the risk predicted for some very small level of exposure, beyond the current state-of-the-science to attain, is indeed higher than one might want to accept. That kind of comparison does not mean that the modeling is wrong or in some way biased. Rather, it suggests that under current circumstances, the attainable levels may be associated with levels of risk that are 'too high."

Response: True.

I would recommend excluding the discussion in this paragraph to the extent that it implies that predictions that are "too low" from a management perspective should be considered unrealistic.

Response: Clarifying text was added.

Reviewer 9

This section is generally complete and well-written. Another possible reality check applies for rare (or relatively rare) tumors. One can calculate the expected response in the population based on the estimated URF and estimated exposure (or high end estimate of exposure). This estimated incidence for a population can then be compared with the reported incidence in a registry such as SEER. If the estimated incidence of the cancer from the one chemical source is substantially higher than all reported cancers of that type, this suggests that the risk has been overestimated.

Response: This idea was incorporated into Section 7.12. The phrase "or high end estimate of exposure" was not used so that the reality check is on slope (URF) and not on the exposure estimate. If the estimated number of responses was too large because the estimated exposure was too large, then that would not be the desired check.

Reviewer 14

The discussion of Reality Checks in Section 7.12 is adequate.

Response: No response required.

2.3.12

Is the discussion of uncertainty analysis (Section 7.13) complete? Please suggest and additional general uncertainty analyses that the TCEQ should consider or any additional guidance concerning how the TCEQ should incorporate/utilize quantitative uncertainty analyses in its Development Support Documents (DSDs)?

Reviewer 1

Not my area of expertise.

Response: No response required.

Reviewer 2

This section is correct in suggesting that any analysis that excludes an uncertainty analysis should be considered inadequate. But here, perhaps more than in earlier sections, the overview summary given here appears to be a bit cursory. This section is nothing more than a list of the items that TCEQ has included in past assessments. In fact, that list included items that may or may not be important at all, let alone for an uncertainty analysis, for some other chemicals.

I think this section would have been better served by having a more overarching list of potential sources of uncertainty. More importantly, the guidelines provide no guidance or suggestions on how various sources of uncertainty should be addressed or what the results of an uncertainty analysis ought to include. So, while the document is correct to emphasize the importance of uncertainty analysis, as a set of guidelines it appears to fall short.

Response: Uncertainty analysis itself is a big topic, as is how that analysis should impact risk characterization and risk management. A clarifying sentence was added before the list of examples.

Reviewer 9

I am not aware of any additions needed for the uncertainty analysis discussion.

Response: No response required.

Reviewer 14

The discussion of Uncertainty Analysis in Section 7.13 is adequate.

Response: No response required.

2.3.13

Chapter 7 of the revised Guidelines includes the following appendices. Please comment on the usefulness of these appendices for illustrating the principals discussed in Chapter 7.

- **Appendix A: Linear Multiplicative Relative Risk Models**
- **Appendix B: Example of a Meta-Analysis of Arsenic Cancer Dose Response Models Based on Published Summary Data**

Reviewer 1

Appendix A: Linear Multiplicative Relative Risk Models: Probably useful for those in the field of risk modeling.

Appendix B: Example of a Meta-Analysis of Arsenic Cancer Dose Response Models Based on Published Summary Data: Not sure. I enjoyed looking at it because of my interest in arsenic. I don't know enough about meta-analysis to know if this is a good example. I had problems with the assumptions.

Response: The audience for these appendices as well as the entire Guidelines is TCEQ and those served by TCEQ. There seems to be strong support for Appendix A and moderate support for Appendix B. Because Appendix A has been judged as useful, it has been retained and the length considered as only a minor concern. The Arsenic DSD has been published (TCEQ 2012), so Appendix B has been replaced with a reference to the appropriate section of the Arsenic DSD discussing meta-analysis procedures.

Reviewer 2

Appendix A is a very good summary of the ways in which data are summarized and analyzed using the multiplicative relative risk model. However, it is not clear who the audience is supposed to be. I fear that non-statisticians will get lost or scared off by the equations, while statisticians will know all of this material already.

The example in Appendix B also provides an interesting example of a meta-analysis. The detail of the calculation of the weighting and the alternatives for parameter combination or combination of the data should be very useful to TCEQ practitioners and reviewers of TCEQ assessments.

Response: See response to reviewer 1 above.

Reviewer 9

These appendices are beyond my area of expertise to evaluate, although it does appear useful to include an example of a meta-analysis.

Response: See response to reviewer 1 above.

Reviewer 14

It is extremely useful to have Appendix A (Linear Multiplicative Relative Risk Models) available in the TCEQ Guidelines. It is suggested that Appendix B (Example of a Meta-Analysis of Arsenic Cancer Dose Response Models Based on Published Summary Data) be removed from these Guidelines. These Guidelines are already quite lengthy. Appendix B is a very well-written and important meta-analysis of arsenic cancer risk. This information should be presented in a separate document and/or publication.

Response: See response to reviewer 1 above.

2.3.14

Please comment on any other issues related to using epidemiology data for hazard and exposure-response characterization that have not already been addressed.

Reviewer 1

I have no further comments.

Response: No response required.

Reviewer 2

I have addressed all of my concerns in response to the questions above.

Response: No response required.

Reviewer 9

- *P. 10: The objectives explicitly state that this text focuses on the application of epidemiology data for carcinogenic response, but no rationale is given for why the text does not also address applications for noncancer endpoints. I would recommend continuing the focus on cancer, but also explicitly addressing noncancer endpoints. Similar principles would apply for many of the topics addressed (including modeling of the data), and so also including noncancer endpoints would not require substantial revision. Several RfCs are based on epidemiology data (e.g., carbon disulfide, toluene) and so analysis of epidemiology data for noncancer endpoints is likely to be relevant to ESL development.*

Response: These Guidelines were meant to be for carcinogenic endpoints. While many of the points made regarding cancer endpoints also apply to noncancer endpoints, expanding this chapter to include noncancer endpoints is a topic for future Guideline revisions.

- *P. 12, P. 17, and elsewhere: It is not clear what is intended by the term "common mechanism of action; common MOA." The term frequently is encountered in the context of mixtures assessment (meaning more than one chemical that acts via the same MOA), but that does not seem to be the intent here. If "common" is intended as a well-studied or frequently-encountered*

MOA, those terms should be used instead of "common," in light of the ambiguity. An even better way of approaching the issue would be to say that these models are most appropriate when the MOA for the chemical has been identified – this can then apply even if the TCEQ assessor had done the MOA evaluation, even if the MOA occurs less frequently.

Response: TCEQ agrees, "common" usually refers to multiple substances acting by the same (common) mechanism of action or at least MOA. That was not the intent here. Text was modified accordingly.

The intent of the example of lymphohematopoeietic cancers is not clear, since the default approach for those would be to assume low-dose linearity, as discussed elsewhere in the text.

Response: For this grouping, there is no defensible alternative to the default, and that default really is unlikely to be biologically relevant (other than as a default) for many in the group.

- *Also – while some within EPA continue to use the term "multiple MOAs," several leaders in the field have emphasized that the chemical acts via one MOA, which may have multiple components (e.g., mutagenic and non-mutagenic components).*

Response: See response to P. 12, P. 17 comment above.

- *P. 13, l 13: I would say that the exposure measurement method is important (not "can be.")*

Response: Changed the text to "is".

- *P. 16, l 14: does this refer to a specific evaluation that was conducted, or is this a general statement that should be in the present tense?*

Response: Changed text to present tense in the last two sentences of this paragraph.

- *P. 16: Note that the 2005 EPA guidelines should be referred to as the Carcinogen Risk Assessment guidelines.*

Response: Changed text accordingly in several places.

- *P. 17, l 29-30: Is the intent here that there is rarely information to justify a chemical-specific dose-response model? One usually can justify the dose response models chosen based on science policy.*

Response: The intent here is that there is rarely sufficient information to justify a specific detailed model for that chemical. Text was modified for clarification.

- *P. 17, l 31-35: Here and elsewhere in this section, it is important to distinguish between linearity in the model fit to the data (which is based on data), and linearity in extrapolation from a point of departure to low doses.*

Response: Text was modified accordingly in the second paragraph of 7.7 and similarly at the end of 7.7.1.

Reviewer 14

Chapter 3, Section 3.7, Page 66, line7. Change 95th to 98th percentile.

Response: Text was modified accordingly.

2.4 Specific Issues for Acute ESLs

2.4.1

Is the list of sources for published acute inhalation toxicity factors complete? (Section 4.1).

Reviewer 3

The AIHA's Workplace Environmental Exposure Level guidelines should be added to Table 4-1 under occupational guideline levels. There are approximately 150 WEELs and with only 5 exceptions, they do not repeat any chemicals for which there are TLVs

Response: The AIHA's WEELs have been incorporated appropriately.

Reviewer 5

The list is complete for acute levels published by US organizations. The AGEL values are published by the National Advisory Committee for Acute Exposure Guideline Levels for Hazardous Substances (NAC/AEGL Committee), not NAC/COT as indicated in Table 4-1. It would be useful to provide a link to the documents where the various levels are defined, or to provide a list of definitions as an appendix so that readers can place the ESLs in context.

Response: Changes have been made accordingly.

Reviewer 10

This is a reasonable compilation. George Woodall and OECD are working on documents related to determination of safe levels of exposure from single exposure scenarios. This is an ongoing activity but should be considered in the future when it becomes available. Table 4.1 you might add that AEGL, ERPG, and TEEL values are thresholds for detection, disability and death to emphasize they are not safe levels of exposure.

Response: No response required.

2.4.2

Are dosimetric adjustments for inhalation studies appropriate? (Section 3.9, 4.2 and 4.3).

Reviewer 3

The dosimetric adjustments used in sections 3.9, 4.2 and 4.3 represent the current approaches used by risk assessors. The discussion on the ten Berge modification to Haber's rule is still the most current approach and the default to longer time periods is the same procedure used by the AEGL committee. However, in going to shorter durations I repeat my comments from point 2.1.9 above: In going to short exposure durations, ten Berge et al. have shown that n=3 is a reasonable approximation as it represents a value close to the upper bound. Where n=3 the slope of the line is very nearly flat. However, it is important to look at data from other studies on that material. Many fluorinated hydrocarbons have free standing NOELs. The upper limit for repeat exposure testing is 50,000 ppm. Above that, there is a risk of inducing effects due to anoxia since the oxygen level is lower due to dilution by the fluorocarbon (e.g. at 100,000 ppm or 10% the oxygen level is reduced by 10% from 21% to 19%.) However, with added oxygen in acute exposure studies, animals can survive exposures of 400,000 ppm (40%) and greater. In these

cases it is appropriate to extrapolate to shorter time periods. I would not recommend n=1, but either n=2 or n=3 could be appropriate. Use of PB/PK model extrapolations can yield more precise results, but this data is not often available.

Response: See response to Reviewer 3 in 2.1.9 above. Also, as stated in the Guidelines, TCEQ considers all available data for a given chemical to perform exposure duration adjustments.

Reviewer 5

The sections dealing with dosimetric and duration adjustments are very well described, appropriate, and consistent with current understanding. This is a major improvement over past TCEQ practice.

Response: No response required.

Reviewer 10

- *The question should be phrased in terms of duration adjustments. "Dosimetry" has a different connotation.*

Response: No response required.

- *Please see the response to charge question 2.1.9 above. That response covers most of the issues in Section 4.2.*

Response: See response to Reviewer 10 in 2.1.9 above.

- *For chronic studies EPA makes the Haber's rule conversion from a number of hours exposure per day for 5 days to 24 hours exposure for 7 days. This approach will give at least protective values.*

Response: No response required.

- *Section 4.3. Derivation of 24 hour values. There is some discussion here about comparing derived values to 24 hour TWA samples. This type of assessment should be done in a chemical specific manner. For example, you may have a low 24 hour TWA for sulfur dioxide that is deemed "safe". However, if that amount is emitted during a 30 minute period, the concentration might be enough to trigger an asthmatic attack – an event that requires few breaths. The impact of short duration peaks should be considered where possible.*

Response: TCEQ agrees. The Guidelines recommend that derivation of a 24-h value should be done in a data-driven, chemical-specific manner. Data are evaluated to determine, when possible, whether toxicity is concentration-dependent, duration-dependent, or both. Thus, if available data indicate that exposure to a chemical at peak concentration for a short-duration would have the potential to cause an adverse health effect, such data would be used to inform the risk assessment process and the derivation of the 24-h ReV or AMCV. At the same time, the resulting 24-hour values are derived to represent a fairly constant level of exposure for comparison to 24-hour monitored values for which information on peaks is lacking. This is considered in the evaluation of air monitoring data (e.g., fairly constant area source(s) or highly variable point emissions source). No revisions were necessary.

- *Sections 4.3.1, 4.3.2, 4.3.3 and 4.3.4 are very well written. They provide guidelines to be considered when the judgment is made whether to derive 24 hour values. Rather than a horse*

blinder formulaic approach, they recommend key toxicology issues to be considered. The issues discussed give the writer guidance for most common issues and the flexibility needed with novel chemical/data issues. A novel approach that has much merit is reference to Figure 3-14 (Exposure-Reference Value Arrays) and the information such graphical displays can give to inform the decision about developing 24 hour values. This is an important application the consideration of the entire weight of evidence when developing values.

Response: No response required.

- *Section 4.3.4.1. The concepts discussed in this section are reasonable approaches even though values of n derived from frankly toxic effects may not be scientifically applicable to the value of n for "safe" levels of exposure. The risk assessor has to derive a number with limited data and the approaches discussed are reasonable compromises for a difficult process. When they err, it is on the conservative side. As discussed in the response to question 9 above, keep it simple and use C^n x t = k and remove the discussion of raising C and t to the m and n power. These are empirical observations. Without mechanistic data, assigning toxicological significance to them is unnecessarily complicating matters.*

Response: See response to Reviewer 10 in 2.19 above.

- *Sections 4.3.4.1 through 4.3.6 are very well written and provide the necessary guidance and flexibility to derive the most scientifically defensible values. Some of the earlier discussions on the derivation of n should be modified because it is sometimes confusing and too detailed for the science it provides. My response to charge question 9 covers this in more detail.*

Response: See response to Reviewer 10 in 2.1.9 above.

- *The use of developmental toxicity to derive acute values is a difficult one. Reference should be made to two publications that bear on this matter. 1) RIVM Report 301900004/2003. Van Raaij and Janssen. The relevance of developmental toxicity endpoints for acute limit setting. 2) Davis et al. 2009. The role of developmental toxicity studies in acute exposure assessments: Analysis of single-day vs. multiple-day exposure regimens. Regulatory Toxicology and Pharmacology 54 (2009) 134–142.*

Response: The suggested references have been incorporated.

2.4.3

Have all of the appropriate uncertainty factors been considered (Sections 3.12 and 4.5)? Would you make recommendations for a different approach to uncertainty factors? In particular, focus on the following issues that have been updated in the revised guidelines:

- **Accounting for differences between children and adult risk susceptibility in choice of UFH and UFD**
- **Choice of different values (other than 1, 3, or 10) for UFL**
- **Choice of different values (other than 1, 3, or 10) for UFD**

Reviewer 3

In my opinion, the application of too many uncertainty factors can invalidate the risk assessment. Uncertainty factors are combined by multiplying them together. Thus two factors of 10 each become one factor of 100. This implies that they are independent of one another which is not true. The TCEQ ESL document, and many other documents, suggest using 5 UFs. The AEGL program used 2 plus a modifying factor. The two were for animals to man and from the typical man to the sensitive members of the population. While we did use values of 1, 3 and 10, we probably used 3 more often than either 1 or 10 for one of the two parameters. For example, if we had multiple acute exposure studies with a tight grouping of LC_{50} values we would use 3 to go from the animal data to man. If the studies were in multiple species and were within a factor of 3, we would use a UF of 1 from the most relevant species. For either a weak data set or an LOAEL or both, we would use a modifying factor of 2 or 3. This would also factor into the animal to man extrapolation. Thus our typical UF would be 10 to 30 and the maximum would be 300. For a chronic value, if the POD was a 28 day study, one might add a UF of 6 so the maximum would be 1800.

Response: TCEQ agrees and has added that TCEQ generally uses a maximum total UF of 300 and 3,000 to the respective acute and chronic POD. The reasonableness of all risk assessment decisions has been discussed in this section.

Reviewer 5

All the uncertainty factors have been appropriately considered and the selection of factors to account for adult/child differences, as well as the approach used to derive factors, appear justified. The comparison of the TCEQ factors with those from other Agencies shows relative good consistency.

Response: No response required.

Reviewer 10

- *Accounting for differences between children and adult risk susceptibility in choice of UF_H and UF_D*
  - *The approach is reasonable and not formulaic. Presumably the default UF_H will be used to address susceptibility of children unless chemical specific data indicate otherwise.*
- *Choice of different values (other than 1, 3, or 10) for UF_L*
  - *The approach described in Section 3.12 is reasonable and provides useful guidelines about when to apply UF values of less than 10.*
- *Choice of different values (other than 1, 3, or 10) for UF_D*
  - *The approach taken is reasoned. However, see comment below about excessive multiplication of UFs.*

Response: See response to Reviewer 3 above.

2.4.4

Is the definition of a minimum database for developing an acute toxicity factor adequate (Table 4-2; Sections 4.4)? In particular, is the discussion of having adequate studies to evaluate children risk differences complete?

Reviewer 3

The information in Table 4-2 for a minimum data base FOR ACUTE 1-hr ReV reads like something created by a bureaucrat to destroy all of the hard work that went into this document. I do NOT agree with the information in even one box. Based on my 25 years as founding chair of both the ERPG and AEGL committees, for an acute risk assessment of high quality, what is needed are 3-4 GLP acute inhalation studies preferably in different species and for different exposure times plus a GLP 28 day inhalation study. It would be helpful to have a developmental toxicity study and 2-4 mutagenicity studies. Since exposures will be acute and infrequent, one does not have to consider carcinogenicity or chronic toxicity. I realize that I have entered strong comments regarding this question. In my opinion, this one table truly could destroy the value of the entire document. In fact, the information requested in this Table is far more than would be needed to conduct a chronic risk assessment on the vast majority of chemicals in commerce.

[In response to a clarifying question] Regarding my comment about the information in Table 4-2, and Section 4.4 on pages 124-126, the first sentence of Section 4.4 does an excellent job of summarizing what is needed for an acute ReV. "The minimum toxicological database required for the development of an acute 1-hr ReV is a well-conducted acute or subacute inhalation bioassay that evaluates a comprehensive array of end-points, including an adequate evaluation of POE respiratory tract effects, and establishes an unequivocal NOAEL and or LOAEL." I would add that with this information and an understanding of the chemical and physical properties one could develop a robust ReV. However, when one goes on to read Table 4-2, reference is made to two bioassays, two developmental toxicity studies and concludes that if one only had a single inhalation bioassay the confidence in the data base is LOW. A bioassay has almost nothing to do with a risk assessment of an acute (i.e. single short term) exposure. Throughout the document, there were excellent discussions on how to utilize available data. But, the list in Table 4-2 opens the door to apply multiple safety factors that are not needed and could drive an otherwise valuable risk assessment into something so conservative as to be not only of little use, but harmful because it could lead to unnecessary protective actions (e.g. evacuation or refusal to grant a permit to a plant when there was no significant risk). I have participated in acute risk assessments on over 300 chemicals. In only one case was carcinogenicity even considered as a possible factor. The discussion on page 125 lines 17 to 25 does an excellent job of discussing the typical short falls of inhalation studies, but the key question is "Does the study provide enough information to support an ACUTE risk assessment?" If it does, that may be all the information needed to derive accurate values that will be protective without being too conservative. The reason for my concern is that on both the AEGL and ERP committees we have spent a great deal of time considering the consequences of being too conservative and trying to be as precise as the data will allow us to be. Both committees typically use total uncertainty factors of 30 or less and rarely would they use a value greater than 100. Looking at the "Minimum Database for an Acute 1-hr ReV ..." one could easily envision a typical uncertainty factor of 1000 which would be very over-conservative.

Response: It is difficult to identify any specific revisions the reviewer is trying to suggest and the basis of some of the comments. For example, Table 4-2 does not consider carcinogenicity or chronic toxicity and may have been misinterpreted by the reviewer. In option 4, Table 4-2 does accommodate the situation the reviewer points out; the absolute minimum database for development of an acute 1-hr ReV is a well

conducted study. Table 4-2 is not a prescriptive table, nor is it meant to be the sole consideration in selecting a UF_D; rather it is guidance for TCEQ toxicologists to aid in determining a potential starting point for a UF_D, which calls for use of best scientific judgment. Text has been added to clarify that use of this table alone cannot adequately represent the completeness of the overall database for a given chemical as this would be a significant oversimplification of the scientific judgment necessary for selection of the UF_D.

Reviewer 5

Table 4-2 and the accompanying discussion do a good job in defining criteria for what constitutes a minimum database for developing acute toxicity factors. The discussion about having adequate studies for estimating children's risk is dispersed in several sections of the guidelines and appears fairly complete. Section 4.4 does not have a specific subsection for discussion of adequacy regarding adequacy of children's studies.

Response: No response required.

Reviewer 10

The minimum database is complete and the database UF application will certainly give protective values.

[Response to clarifying question] I suspect an adequate response lies somewhere between reviewer 10 and 3. The sentence in the second paragraph of this section "Confidence in toxicological databases will vary depending on how much is known about each chemical's MOA and the quality of the experimental study (Section 3.12.3)" should be the operative for this consideration. Use of proscriptive factors can lead to a formulaic approach that results in unrealistically high UFs. For example, the criteria in footnote would seem to exclude consideration of studies that would be useful in the assessment. The judgment of the assessor is paramount. There are many studies in peer review journals that are of low quality. Conversely there are reports that may not have formally adhered to "good laboratory practices" that should be considered. Give the reviewer some wiggle room. There is nothing wrong with requiring the reviewer to justify their conclusions but a rigid formulaic approach will result in many values unrealistically, if safely, low. In another section you emphasize that overly low values do not serve the public. If formally applied, the rules in Table 4-2 will do just that.

It is interesting to note that Dr. Dourson did an excellent analysis a number of years ago of what you miss when certain data elements are missing. Perhaps that should be examined when making these database uncertainty decisions. Guidelines rather than rules should be considered.

Response: See response to Reviewer 3 above.

2.4.5

The revised guidance discusses several approaches for developing an acute generic ESL for chemicals with limited data. For each approach, please discuss whether the approach is applied correctly and appropriate for developing acute generic ESL in the absence of data.

- **Route-to-route extrapolation (Section 3.16 and Subsection 3.16.1)**
- **Relative potency (Section 3.16 and Subsection 3.16.2)**
- **Use of the Threshold of Regulation and N-L Ratio Approach (Section 4.6)**

- **In the revised guidelines, TCEQ is proposing to primarily use the N-L ratio approach rather than the Threshold of Concern method to develop generic acute ESLs (Section 4.6). Is this reasonable given the literature cited?**

Reviewer 3

Alternative approaches for acute generic ESL values

a. *The development of oral to inhalation ESL represents an improvement over what has commonly been done in the past. Application of PBPK modeling is an excellent approach if adequate data exists. The criteria for not doing the extrapolation are good.*
b. *The approach described for relative potency is fine.*
c. *The default ESL described in section 4.6.1 of 2 µg/m^3 appears overly conservative. The N-L ratio approach appears to represent a state-of-the-art approach.*
d. *I support using the N-L as a replacement for the TOC. It should produce values of greater reliability*

Response: No response required.

Reviewer 5

It is difficult to make a blanket statement on the suitability of the route-route extrapolation approach. Section 3.16 presents the caveats inherent in adopting this approach fairly well, so it is expected that application of these caveats will be reflected in their application to specific chemicals. Likewise, the relative potency approach, which has been used for establishing some acute exposure levels, appears reasonable when suitable data are lacking. The N-L Ratio approach is based on statistical considerations and thus more questionable, although not clearly unreasonable. Lacking data or a good indicator of reliability when utilizing a single approach, the alternative would be to derive the generic ESLs using more than one of these approaches (if possible) to determine if the estimates are reasonably close.

Response: No response required.

[*Response to clarifying question****]*** *The answer to these questions meant to convey that the various approaches are well described and appropriate for developing acute generic ESLs in the absence of suitable data. The comment on using more than one approach referred more specifically to the adoption of the N-L Ratio method in the context of the results from comparisons with TOC discussed in the document. Results from these comparisons add confidence to the N-L Ratio method but the number and range of compounds used in the comparison is limited. Therefore, it cannot be assumed that the conclusions regarding the suitability of N-L ratio approach is generalizable to the universe of chemicals for which generic ESLs need to be derived. The document is not clear about evaluation of this approach for reasonable value grounding for compounds that could differ in structure and properties from those already included in the prior comparisons. One way of doing this would be to derive the generic ESL by more than one feasible approach, including the N-L ratio.*

Response: This section has been revised to clarify when the TOC and N-L Ratio approach will be conducted.

Reviewer 10

- *Route-to-route extrapolation (Section 3.16 and Subsection 3.16.1)*
  - *The discussion is reasonable. Pragmatically one will rarely find a chemical with a data set necessary to overcome the many difficulties with this approach. In the AEGL program we only did this extrapolation for one chemical, methanol – a case study where many people were poisoned and detailed blood levels were available. That being said, TCEQ has a charge to develop values and must do so. The alternative to this type of extrapolation is the default approach discussed in section 4.6. This is a tough judgment call that must be made. With those caveats, the discussion in Section 3.16 is appropriate.*

Response: No response required.

- *Relative potency (Section 3.16 and Subsection 3.16.2)*
  - *This is appropriate.*

Response: No response required.

- *Use of the Threshold of Regulation and N-L Ratio Approach (Section 4.6)*
  - *This response is to this question and the one below. I don't feel qualified to comment on this approach since I am not familiar with the literature. A few minor notes.*
    - *The default cited is 2 ug/m^3. How was this figure derived?*

Response: The default short-term ESL value of 2 µg/m^3 is based on TCEQ's policy when conducting effects evaluations for air permits, which is based on Grant et al. (2007). The TCEQ considers a chemical's emission *de minimis* if the emission rate is equal to or less than 0.04 lb per hour, provided that the short-term ESL is ≤ 2 µg/m^3. The default ESL is conservative. The TCEQ believes that the short-term ESLs for chemicals with limited data, if developed, are expected to be higher than 2 µg/m^3. Highly toxic chemicals (e.g., metals) would likely have sufficient toxicity data available for developing either interim or final ESLs, which, for some chemicals, already have ESLs lower than 2 µg/m^3.

    - *Section 4.6.2 implies the TOC approach will not be used and cites Grant as support for the TOC approach. Next it says Phillips demonstrated the TOC approach was overly conservative and the N-L approach is discussed. However, Section 4.6.2.2 uses a discussion of Grant to support the N-L approach while in the earlier section it looked like Grant was associated with the TOC approach. I am a little confused by this discussion but as indicated, I am not familiar with the literature.*

Response: This section has been modified appropriately for clarity.

2.4.6

In the revised guidelines, (Section 4.3) TCEQ has presented methods for developing 24-hour ReVs and 24-hour Air Monitoring Comparison Values (AMCVs) for use in evaluating ambient air monitoring data. Are the proposed approaches appropriate and consistent with accepted risk assessment practices? Can you make any recommendations to improve these methods?

Reviewer 3

It is not appropriate to compare air sampling data (AMCVs) at one interval with ReVs at a different time value. The fact that one approach is "conservative" and another not "conservative" is not relevant. The

approach is wrong. Either the sampling interval should be adjusted to match the time for the ReV or the ReV should be adjusted (e.g. using the ten Berge calculation) to match the sampling time.

Response: It is not clear what statement or sentence the reviewer is referring to in this comment. However, the TCEQ agrees that it is not appropriate to compare AMCVs at one interval with air monitoring data collected at a different time value. This issue is the key justification for derivation of 24-h ReVs for use in the TCEQ air monitoring program. The approach that is selected to derive a 24-h AMCV will be based on best available science, assessment procedures, and professional judgment with the goal of generating AMCVs specific for the duration over which an air sample was collected. As the Guidelines discuss, there are occasions where the use of the ten Berge calculation may result in the generation of an unrealistic 24-h ReV (i.e., one that is lower than a chronic ReV derived from a high quality study). Thus, the TCEQ suggests that all relevant data and professional judgment be used when deriving a scientifically-defensible 24-h ReV from a study with a significantly different exposure duration (e.g., 30 min or 1-h) to avoid the generation of misleading regulatory values.

Reviewer 5

These are different exposure and risk levels than those promoted by other agencies, including TCEQ in the past. The proposed approach appears consistent with derivation of other ReVs, with the appropriate caveats. It would have been useful to see an example presented as part of the document. I do not have additional recommendations.

Response: TCEQ agrees. However, since this is a guidance document, case studies will not be included. Instead, TCEQ intends to generate a white paper or a peer-reviewed manuscript detailing the approach using examples; the goal is having this approach peer-reviewed. No changes were made.

Reviewer 10

See response to charge question 2.4.2 above.

Response: See response to Reviewer 10 in 2.4.2 above.

2.4.7

For the chronic ReV, the total UF is capped at 3000 for 5 areas of UFs. For the acute ReV, there are only 4 areas of UFs. Should a lower cap be used for the acute ReV (i.e., 1000 or lower)?

Reviewer 3

As noted in the [section 2.4.3 response], I would not recommend a combined UF of more than 300 for acute exposures providing that there are at least 2 GLP acute studies and one repeat exposure study. Within the AEGL program it was rare that we would be above 100 and typically we were at or below 30.

Response: See response to Reviewer 3 in 2.1.3 above.

Reviewer 5

No, unless a justifiable rationale is provided.

Response: See response to Reviewer 3 in 2.1.3 above.

Reviewer 10

This is a reasonable approach to what is generally accepted in the hazard assessment community. Of course one can question how relevant the data are if you have to apply a thousand fold uncertainty factor. A number of years ago a group looked at the multiplication of worst case factors of 10 times 10 etc. Their reflection was that this gave an unrealistically high resultant UF. I think they came out with something like 10 x 10 is about 67. This would be an important concept to pursue but is beyond the scope of this document. The UFs discussed will at least give health protective values.

Response: No response required.

2.4.8

Are there other issues specific to developing acute inhalation toxicity factors that have not been adequately addressed in the document?

Reviewer 3

My final comments are that this is an excellent review of the risk assessment process. I am concerned about the application of these procedures. In several situations (e.g. children vs. adults, minimum acute data base, application of UFs) there is a move towards overly conservative risk assessments. The goal of a risk assessment is to be as precise as possible. If it is not protective, people may be injured. If it is too conservative, it loses credibility and may be unnecessarily costly; also if used for emergency planning, it could create unnecessary actions such as evacuations and result in accidents that otherwise would not have occurred such as happened with hurricane Rita.

Response: No response required.

Reviewer 5

None that I can identify at this time.

Response: No response required.

Reviewer 10

Although there are a lot of comments above, I would like to close by saying this is in general a well researched and written document. It provides the scientific basis for the determinations being made and guidelines to guide the writer in assessing safe levels of exposure. The emphasis is on a refreshing use of developing the values based upon the guidelines and an intelligent use of professional judgment.

Response: No response required.

2.5 Specific Issues for Oral Cancer and Noncancer

2.5.1

Is the list of sources for published chronic oral toxicity factors (both RfD and SFo values) complete? (Chapter 5, Section 5.1).

Reviewer 4

The list in Section 5.1 is complete. I would caution against using HEAST Toxicity Factors as these values have not been updated since 1997. The Provisional Peer-Reviewed Toxicity Values source is preferable as it is being updated.

Response: The TCEQ agrees with these comments.

Reviewer 11

The sources of toxicity values presented in Section 5.1 provide a good representation of the state, national, and international databases available. There are other state, national and international toxicity databases available such as the USEPA's Pesticide Registration database (http://www.epa.gov/pesticides/reregistration/status_page_o.htm), the state of New Jersey (http://www.nj.gov/dep/standards/all%20tox%20factors.pdf), the European Chemicals Agency (http://apps.echa.europa.eu/registered/registered-sub.aspx#search), etc. A complete listing would not be possible or practical here, but the major ones are shown.

Response: The TCEQ agrees with these comments.

Reviewer 15

I find this list to be complete, although I see no mention of USFDA' pronouncements on regulating ethical pharmaceuticals.

Response: As the TCEQ does not typically prepare toxicity factor derivation support documents (DSDs) for pharmaceuticals, no revisions were made.

2.5.2

Are dosimetric adjustments for oral cancer and noncancer appropriate? (Section 3.9, 5.2.2 and 5.3)

- **Adjustment of human and animal data to mg/kg-day**
- **Interspecies scaling for cancer and non-cancer effects**

Reviewer 4

TCEQ has provided its hierarchy of procedures for adjusting data from oral exposure in animals to humans in Section 5.3. This scheme is consistent with accepted risk assessment practices. As a default TCEQ has elected to use $BW^{3/4}$ for this adjustment without the use of the UF_A of 3 as advocated by EPA. This is clearly explained and identified as a science policy choice.

Response: No response required.

Reviewer 11

The dosimetric adjustments for converting the doses administered in a toxicological study to both a continuous exposure and a human equivalent dose are consistent with the methodologies used by the USEPA in their derivation of cancer and non-cancer toxicity values. The authors do a good job of discussing the most recent EPA guidance on the use of body weight ¾ as a default procedure (for non-cancer assessments) for scaling from animal to human doses (as opposed to the older default uncertainty factor of 10 for interspecies variability).

Response: No response required.

Reviewer 15

I agree with TCEQ's selections for dosimetric adjustments. The treatment of using $BW^{3/4}$ is clearly excessive. The subject requires treatment, but not page after page. Also, TCEQ might find that the classic publication of Guyton (1947) should really be the backbone of any such discussion.

Response: The TCEQ believes that electing to use $BW^{3/4}$ without the use of the additional UF_A of 3, when a UF_A of 3 is advocated by EPA, requires some discussion. As TCEQ references the publications relied upon, and to TCEQ's understanding Guyton (1947) is more relevant to inhalation, Guyton (1947) was not added as a reference.

2.5.3

Have all of the appropriate uncertainty factors been considered [Sections 3.12 and 5.5]? Would you make recommendations for a different approach to uncertainty factors? In particular, focus on the following issues that have been updated in the revised guidelines:

- **Accounting for differences between children and adult risk susceptibility in choice of UF_H and UF_D**
- **Choice of different values (other than 1, 3, or 10) for UF_L**
- **Choice of different values (other than 1, 3, or 10) for UF_D**
- **UF_{Sub}**

Reviewer 4

I would suggest that only UF_H be used to account for any potential differences in life stages (children versus adults). The discussion in Section 5.5.3 seems to be saying that both could be used. Does this mean that TCEQ could use a total UF for a difference in potential toxicity to children of 100?

Response: The TCEQ believes both UFs are related to protecting children and wants to maintain the flexibility of being able to address potential child/adult differences under the most appropriate UF, depending on available data (e.g., the lack of a developmental or reproductive study versus expected differences in metabolism). In most cases, the TCEQ would not expect the combined UF to exceed 30 based on these considerations.

Section 3.12.2.1 provides TCEQ's scheme for UF_L. This is clearly explained, but references USEPA grades of adversity. I am not aware of any EPA grading system of this type and no reference is supplied.

Response: The TCEQ points out that OEHHA (2008)is actually the reference for grades of severity.

A scheme for UF_D is presented in Table 5-2. However, this table does not supply any discussion of how values of 1-3, 3-6, 3-10 will be selected and does not clearly identify these values as being adopted by TCEQ.

Response: The range of UF_D values in Table 5-2 are simply *potential* values (not adopted values) within which the final value may be generally expected to fall. This section indicates that Table 5-2 is not the sole consideration for selecting a UF_D value (i.e., would represent an oversimplification of the scientific judgment necessary), which will be determined case-by-case based on best scientific judgment considering available data, chemical-specific concerns, and data gaps.

Similarly Section 5.5.4 provides no definitive guidance on how a UF_{Sub} different from 1 or 10 will be selected.

Response: This section contains many important considerations (e.g., TK, TD, study type, critical effect, study duration), but is not intended to be prescriptive or deterministic in nature regarding selecting a UF_{Sub} value, which will be determined case-by-case based on best scientific judgment considering available data and chemical-specific concerns (e.g., study duration, TK, TD).

Reviewer 11

The TCEQ uses an uncertainty factor approach which is very similar to the approach used by the USEPA. They use five areas of uncertainty UF_H (intraspecies), UF_A (interspecies), UF_{sub} (subchronic to chronic), UF_L (LOAEL to NOAEL), and UF_D (database deficiency). The issue of whether chronic toxicity values derived from adult humans or animals is protective of children is frequently raised. There are three important factors to consider when discussing this issue. The first, as TCEQ emphasizes in Section 3.2, is that because of the toxicokinetic and toxicodynamic differences between children and adults, sometimes children may be more sensitive to the effects of contaminants than adults, sometimes they may be the same and sometimes they may be less sensitive. Therefore intraspecies variability (UF_H) should be evaluated on a chemical by chemical basis and not by the application of an across the board safety factor. The 2nd consideration is that the overall intraspecies uncertainty factor has been shown to be protective of intraspecies variability in humans, including children (Section 5.5.1). The 3rd consideration is the UF_D or database uncertainty factor. The database UF is intended to account for the potential for deriving an underprotective RfD/RfC as a result of an incomplete characterization of the chemical's toxicity. Typically this factor is applied when a developmental study or two generation reproductive study is missing. As such, the additional 10-fold factor for infants recommended by the National Research Council and called for in the 1996 Food Quality Protection Act is similar to the database UF (USEPA 2002). After careful consideration of these three factors, I believe the authors have correctly chosen not to include an additional safety factor for the protection of children, but to evaluate the situation on a chemical by chemical basis.

Response: No response required.

The authors provide guidelines and suggestion for uncertainty factors less than 10 for the UF_L, UF_D and UF_{sub}. This is a scientifically sound concept and I find the text to be very useful. The USEPA guidance documents for the derivation of RfDs discuss the concept of uncertainty factors less than 10 for these particular factors but does so in a very general manner. Very seldom are severity of effect, dose spacing

between LOAEL and NOAEL, bioaccumulation, etc., taken into consideration in the development of the IRIS RfDs and, as a result, the UF_L, $UF_{sub,}$ and UF_D are seldom reduced from 10. I appreciate the use of specific examples which guide a user in applying an uncertainty factor or 1, 3 or 10.

Response: No response required.

In conclusion, I believe that all the appropriate uncertainty factors have been included. TCEQ chose not to include the modifying factor used in EPA's RfD/ RfC methodology, which is rarely used anyway. I also believe that the TCEQ document advances the state of risk assessment and toxicity derivation by moving away from the robotic application of default uncertainty factors of 10 and specifically listing toxicological and chemical/physical criteria to consider when applying quantitative uncertainty factors less than 10.

Response: No response required.

Reviewer 15

The treatment of UFs for childhood exposure is adequate, although, as mentioned above, I believe the cited guidance from USEPA and CalEPA makes too many subdivisions of the period from birth to 18 years of age. Hence, TCEQ offers an excessive and overly prescriptive number of possible adjustments.

Response: The Guidance is intended to contain important considerations and not to be prescriptive in nature (and will not be used as such) regarding selection of a UF value, which will be determined case-by-case based on best scientific judgment considering available data, chemical-specific concerns, etc. The Guidance should be viewed as, at most, simply providing examples of *potential* values (not adopted values) within which the final value may be generally expected to fall.

2.5.4

Is the definition of a minimum database for developing a chronic toxicity factor adequate? In particular, is the discussion of having adequate studies to evaluate children risk differences (Table 5-2; Sections 5.4 and 5.5) complete?

Reviewer 4

I believe the definition of the minimum database is adequate. See comment above about using UF_H and UF_D for evaluate differences in risk to children.

Response: See response to Reviewer 4 in 2.5.3 above.

Reviewer 11

In Section 5.4, the authors recommend using a similar definition of minimum database for high, medium, and low confidence to that used by the USEPA in their derivation of non-cancer toxicity values. To consider differences in toxicity between children and adults, a two generation reproductive study or study that evaluates reproductive endpoints is recommended. Typically, if a two generation reproductive study or developmental study is missing, a database uncertainty factor of 10 is applied. The authors might consider adding text to that effect. The authors go on to state that cancer or non-cancer toxicity values may be developed even though the minimum database requirements are not met. This is a policy call and up to the program developing and using these toxicity values. As long as the strengths/ weaknesses and confidence in the toxicity value are clearly explained and understood, this is appropriate.

Response: The TCEQ does not want to imply that a UF_D of 10 should be the general rule if a two generation reproductive study or developmental study is missing as chemical-specific information may not justify a value of 10. Hence, no revisions were made.

Reviewer 15

This treatment is adequate, although I have heard [experts in the field] expound on the relative lack of importance of a 2-generation reproduction study when an adequate chronic toxicity study is available.

Response: No response required.

2.5.5

The revised guidance discusses several approaches for developing an RfD for chemicals with limited data. For each approach, please discuss whether the approach is applied correctly and appropriate for developing chronic RfDs in the absence of data.

- **Route-to-route extrapolation (Section 3.16 and Subsection 3.16.1)**
- **Relative potency (Section 3.16 and Subsection 3.16.2)**
- **Use of LD_{50} (Section 5.6)**

Reviewer 4

It is always a difficult task to deal with a chemical with limited data in a regulatory setting. I think TCEQ has identified an appropriate hierarchy of methods to use in this situation. I would, however, use the LD_{50} approach only as a last resort.

Response: No response required.

Reviewer 11

The use of route to route extrapolation to develop a toxicity value is oftentimes useful in risk assessment when there is no data available by the route of exposure desired. There are physiological and toxicological criteria which should be evaluated and considered prior to undertaking such an operation (e.g., portal of entry effects, first pass effects, etc.). These are clearly presented in Section 3.16 along with the equations for calculating the conversion. The application of relative potency factors to an index chemical is also useful in risk assessment when limited toxicity data are available for structurally similar chemicals. Dioxins, PCBs, and PAHs are good examples. The approach and equations described in Section 3.16.2 are consistent with relative potency approaches used in risk assessment practice. I'm not familiar with the NOAEL to LD_{50} ratio approach discussed in Section 5.6 and am not familiar enough with comparisons between acute and chronic toxicity to comment on the accuracy of this methodology. As mentioned previously, the use of methods with limited data and low confidence to develop toxicity values is strictly a policy call. It is important though, that the low confidence in the value be transparent and understood by those using the value.

Response: No response required.

Reviewer 15

For route-to-route extrapolation, I recommend including strong language recommending against using this approach for metals. Although TCEQ states that route-to-route extrapolation should not be used

when POE effects are present, which is usually the case with metals, I recommend a statement countering OEHHA's practice of naming a substance carcinogenic by all routes if it is carcinogenic by one route.

Response: Metals, irritants, and sensitizers have been added as examples of chemicals for which route-specific effects may be expected to occur in Chapter 3. In Chapter 5, text has been added indicating that the carcinogenic descriptor for one exposure route should not be viewed as automatically relevant to another route, as there may be important route-specific considerations (e.g., significant differences in absorption or response at the POE). For example, when a chemical produces tumors only at the POE, the descriptor generally applies only to that exposure route (unless the MOA is also relevant to other routes).

2.5.6

Are there other issues specific to developing chronic oral toxicity factors that have not been adequately addressed in the document?

Reviewer 4

I believe the document adequately addresses the issues related to the development of oral toxicity reference values.

Response: No response required.

Reviewer 11

Earlier in Section 3 I think I saw reference to the use of a margin of exposure (MOE) approach. Yet it was never really discussed in this document. I think it would have been useful to see a discussion of this alternate risk assessment method which doesn't rely on the development of an RfD or oral slope factor but uses site specific exposures and the LOAEL from toxicological studies instead. Different decision criteria are needed, obviously, but the method may allow more flexibility in risk decision making, especially for chemicals with limited data.

Response: While MOE is a useful methodology for some chemical exposures (e.g., dioxins), the TCEQ generally prefers the conservativeness of the RfD/SFo approach and anticipates rarely (if at all) using the MOE approach for the regulatory setting of environmental media comparison values (although helpful for putting risk into perspective when comparison value exceedances occur). Thus, no additional discussion/text was added.

Reviewer 15

This draft guidance is not just comprehensive, it is encyclopedic. Except for those few issues I raise above, I find virtually nothing missing. Apparently, Texas law and its Health and Safety Code have a great deal to say about airborne pollutants; hence, it is not surprising that the document gives greater weight to inhalation than other routes. This is not necessarily a weakness, however, because the shorter treatment of the oral route is still quite complete. The dermal route receives little treatment. The authors might want to consider expanding their discussion of the wisdom of USEPA recommendations in RAGS Part E to use predictive equations for dermal absorption and absorbed dose.

Response: The TCEQ rules which utilize oral toxicity factors (e.g., Texas Risk Reduction Program (TRRP) rule) already contain appropriate equations to evaluate the dermal route. Such equations concern the appropriate application of toxicity factors in the context of remediation baseline risk assessment and are beyond the scope of a document, which simply seeks to provide guidance on development of the toxicity factors themselves.

2.6 Specific Issues for Inhalation Cancer and Noncancer

2.6.1

Is the list of sources for published chronic inhalation toxicity factors (both ReV (or in some cases RfC) and URF values) complete? (Chapter 5, Section 5.1).

Reviewer 6

The list of sources for published chronic inhalation toxicity factors is adequate. Rating=3 [is consistent with accepted practice and documentation].

Response: No response required.

Reviewer 7

The only other source of chronic toxicity values that I know is the Committee of Toxicology of the National Research Council. They have volumes on "Continuous Exposure Guidance Levels for Selected Submarine Contaminants." The submarine tours are 90 days and the documents are peer-reviewed. The population of concern is limited to healthy young sailors.

Response: TCEQ develops toxicity factors for ambient air. The submarine conditions apply to indoor air.

Reviewer 12

In the view of this reviewer, yes.

Response: No response required.

2.6.2

Are dosimetric adjustments for inhalation cancer and noncancer appropriate? (Section 3.9 and 3.10 and 5.2.1).

Reviewer 6

Dosimetric adjustments for inhalation cancer and noncancer are subject to individual selection, interpretation and judgment. The adjustments are reasonable. Rating=3 [is consistent with accepted practice and documentation].

Response: No response required.

Reviewer 7

The only area I find missing is any discussion of adjustments for nanoparticles. Otherwise, the adjustments appear to be appropriate.

Response: See response to Reviewer 7 in 2.1.6 above.

Reviewer 12

This question covers sections of the guidance dealing with duration adjustment (5.2.1) and (3.9) and more formal dosimetric procedures involved in animal to human extrapolation (Section 3.10).

The duration adjustments for both acute durations (both shorter to longer and longer to shorter durations) and chronic durations (i.e., hourly to daily to weekly) are consistent with the most recent and

reasoned approaches available. These approaches are inclusive of the concept of C^n x T as well as new ten Berge model in the US EPA's BMD model suite. A specific comment has been made to add two additional references regarding these type of duration adjustments.

The dosimetric adjustments advanced here are the basic default versions of the RfC Methodology and, as such, are acceptable but conceptually incomplete. Specific commentary has been offered to explain and address this issue for both gas and particle dosimetry.

Regarding to that portion of the query regarding the query cancer and noncancer inhalation dosimetry it needs to first be stated that with the US EPA, there has been no issue regarding differences between the inhalation dosimetry for cancer and noncancer as both use the RfC Methodology. This reviewer finds few statements regarding this issue in the current guidance document. It is noted that the term used to indicate application of inhalation dosimetry to a point of departure, POD_{HEC}, is used in a few instances in these sections (e.g., in the route-to-route extrapolations). It thus seems that this guidance presumes no difference in performing dosimetry for cancer and noncancer endpoints, but does not explicitly state so. Thus, the authors should consider this issue and add clear statements to address this dosimetry issue, possibly citing the USEPA's procedures.

Response: As the comment indicates, there has been no issue regarding differences between the inhalation dosimetry for cancer and noncancer with the USEPA as both use the RfC Methodology. TCEQ interprets similarity in cancer and noncancer inhalation dosimetry as standard practice as opposed to an issue which needs to be explicitly addressed. Thus, after due consideration, no changes were made.

2.6.3

Have all of the appropriate uncertainty factors been considered [Sections 3.12 and 5.5]? Would you make recommendations for a different approach to uncertainty factors? In particular, focus on the following issues that have been updated in the revised guidelines:

- **Accounting for differences between children and adult risk susceptibility in choice of UFH and UFD**
- **Choice of different values (other than 1, 3, or 10) for UFL**
- **Choice of different values (other than 1, 3, or 10) for UFD**
- **UFSub**

Reviewer 6

The inclusion of the updated uncertainty factors greatly enhances the applicability and usefulness of the document. Rating=4 [strong].

Response: No response required.

Reviewer 7

This discussion is strong in that the reader is repeatedly reminded that the use of UFs must be done on a case by case basis and scientific judgment is required. I have already made a recommendation to improve Table 3-2 to be more specific about what constitutes a low concern in an endpoint.

Response: No response required.

Reviewer 12

These UF are fairly standard and as such are acceptable.

Response: No response required.

2.6.3.1 Part A.

Accounting for differences between children and adult risk susceptibility in choice of UF_H and UF_D

Reviewer 6

[See above comment under 2.6.3].

Response: No response required.

Reviewer 7

The discussion of this issue was thorough and I saw no need to change it.

Response: No response required.

Reviewer 12

This reviewer has already commented extensively upon the manner in which children's issues are addressed in this guidance. If this reviewer's general comments are addressed in their fullest form, this section would likely be renamed and the specific mention of children replaced with wording such as "any known subpopulation with an identifiable basis for being susceptible". Again, as most differences in child-adult susceptibility fall into the range of 1-2, the differences would be then stated to be subsumed in the UF_H of 10. If a particular instance was identified in children that exceeded this level, then that subpopulation would be identified and the extra UF applied, as is proposed in this Section.

Response: These sections have been revised in response to other peer review comments regarding child adult differences.

2.6.3.2 Part B.

Choice of different values (other than 1, 3, or 10) for UF_L

Reviewer 6

[See above comment under 2.7.3].

Response: No response required.

Reviewer 7

I agree with the choices of different values for UF_L.

Response: No response required.

Reviewer 12

Severity is used consistently as a modifier of assessment evaluations throughout this guidance including fraction of UF in Table 3-5. The usual basis for fractionating UF has been to account for distinct discrete components of uncertainty such as PD & PK in both UF_H & UF_A. The basis for using severity for fractionating UF_L is not component- based but is dose-response. It is a rather blindly applied factor used

to move down the dose-response continuum to a place of greater subjective comfort. This reviewer thus considers this use to be a subjective and misplaced activity that belongs in dose-response, not in uncertainty factor assignment. Nonetheless, a clear acknowledgement of the underlying assumptions and subjectivity of this use would be helpful, i.e., such as that using severity in this determination is based on the assumption that more or less (or no) severe effects (and also LOAEL or NOAEL) are likely to be further up or down the continuum of dose-response. It should also be noted that at least the EPA does not perform a dose-response assessment when only an FEL is available. This reviewer does not recall encountering such a caveat in this guidance.

Response: The Guidelines were not revised based on these comments. TCEQ agrees that the basis for using severity for fractionating UF_L is not component-based but is dose-response. It is used to account for the uncertainty in the dose-response assessment when a LOAEL is the POD for a toxicity assessment (i.e., used to move the dose-response continuum to a place of greater subjective comfort) to convert the LOAEL-based POD to a value analogous to a NOAEL. Chapter 3 already addresses these issues and the reader is referred to USEPA (2002a) and ECETOC (2003) which provide additional discussion:

> "It is important to consider the slope of the dose-response curve in the range of the LOAEL in making the determination to reduce the size of the UFL (USEPA 2002a). Several studies have reported that the LOAELs rarely exceed the NOAEL by more than about 5-6 fold and are typically close to a value of 3 based on the ratios of LOAELs to NOAELs for a range of different chemicals and different study durations (subacute, subchronic, and chronic) (ECETOC 2003), although these findings are dependent on the dose spacing chosen in toxicity studies."

The EPA does not perform a dose-response assessment when only an FEL is available, but the TCEQ is required to develop inhalation toxicity factors even though there are limited data. The problem formulation information in Chapter 1, which the peer reviewers were not asked to review, clearly states that "Because of the comprehensiveness of the language in the Texas Health and Safety Code, ESLs are developed for as many air contaminants as possible, even for chemicals with limited toxicity data." The TCEQ uses conservative procedures for developing toxicity factors for chemicals with limited toxicity information. These procedures are generally accepted by the scientific community as being conservative and are based on scientific analyses.

2.6.3.3 Part C.

Choice of different values (other than 1, 3, or 10) for UF_D

Reviewer 6

[See above comment under 2.7.3].

Response: No response required.

Reviewer 7

I agree with the choices of different values for UF_D.

Response: No response required.

Reviewer 12

The basis for fractionation of this UF is data on life-stage components and, as such, is acceptable and reasonable.

Response: No response required.

2.6.3.4 Part D.

UF_{Sub}

Reviewer 6

[See above comment under 2.7.3].

Response: No response required.

Reviewer 7

I agree with the discussion of the use of UF_{Sub}.

Response: No response required.

Reviewer 12

Although severity is also listed as a modifier for this UF, it is not as prominent as the duration as is clearly explained in Section 5.5.4. No alterations are thus suggested.

Response: No response required.

2.6.4

Is the definition of a minimum database for developing a chronic toxicity factor adequate? In particular, is the discussion of having adequate studies to evaluate children risk differences (Table 5-2; Sections 5.4 and 5.5) complete?

Reviewer 6

The explanation needs improvement. It is unclear. The amount of information available to evaluate children's risk differences is insufficient. However, until more information is available, it is understandable that assumptions and allowances must be made. The choices described by TCEQ appear to be consistently appropriate and conservative. Having this information described and applied consistently is helpful. Rating=3 [is consistent with accepted practice and documentation].

Response: This section has been revised to clarify that reproductive effects (not child-adult differences) are evaluated by reproductive studies and those considering reproductive endpoints.

Reviewer 7

I think the discussion of the minimum database required for developing a chronic toxicity value is well done in Section 5.4. However, I do not agree with the statement on page 143, line 32 and going over to page 144, lines 1-2. A two-generation reproductive test is a good measure of the effects on reproduction, but is certainly not necessarily a valid way to measure "chronic child adult differences." I suggest you rethink that sentence.

The discussion of the use of uncertainty factors to account for child-adult differences in Section 5.5 is quite good.

Response: See response to Reviewer 6 above.

Reviewer 12

Based on the authors' decision to implement this reviewers comment offered in the general comments on children's risk as used in this guidance, this query could be somewhat moot. Even in the event that the comment would not be implemented this reviewer would, of course, endorse the manner of implementation shown here, as long as the justification for doing so reflects relevant facts.

Response: The Guidelines have been revised in response to Reviewer 12's other comments regarding child-adult differences.

2.6.5

The revised guidance discusses several approaches for developing ReVs for chemicals with limited data. For each approach, please discuss whether the approach is applied correctly and appropriate for developing chronic ReVs in the absence of data.

2.6.5.1 Part A.

Route-to-route extrapolation (Section 3.16 and Subsection 3.16.1)

Reviewer 6

The approaches for developing ReVs for chemicals with limited data are consistent with current literature and some regulatory applications. However, caution should be executed regarding the applicability of individual selections in light of missing data. Individual selections can create different ReVs. It is important to confirm that the selections result in an appropriate ReV that is applicable to the scenario evaluated.

Response: TCEQ will consider the merits of the alternatives available and use best scientific judgment in determining the most appropriate methodology for the final ReV value.

Reviewer 7

This discussion on pages 99-102 is especially well done (a strength). However, there is one incomplete sentence (page 100, lines 12-13 and one very complex sentence (page 100, lines 30-33). I have read the latter sentence 10 times and I still do not know what the meaning is (a weakness).

Response: These sentences have been revised for clarity.

Reviewer 12

Use of validated PBPK models are, of course, the preferred manner in which such extrapolation should be attempted, either for inhalation to oral or oral to inhalation. The basic default approaches given here are simplistic and their use will count as null what may be extremely relevant kinetic differences such as major differences in absorption as well as unanticipated toxicity, such as to POE tissues. Nonetheless, that the default dose is considered to be 100% of that present (e.g., air concentration x breathing rate) in the inhalation to oral makes this the definitive worst–case scenario in regards to exposure. The

extrapolation from oral to inhalation however, has the potential to result in unrealistic results and further specific commentary has been offered concerning this issue. That the authors acknowledge and put forth these limitations and liabilities make this approach more defensible than not.

Response: No response required.

2.6.5.2 Part B.

Relative potency (Section 3.16 and Subsection 3.16.2)

Reviewer 6

[see above comment under 2.6.5.1 Part A].

Response: See response to Reviewer 6 in 2.6.5.1 Part A above.

Reviewer 7

I thought this section was exceptionally well- written and have no changes to recommend. I also liked the section of sensitization (3.17). This is an important type of toxicity that is sometimes ignored.

Response: No response required.

Reviewer 12

This practice, of course, has quite limited application as is acknowledged in this section. The approach, as described here in all of its limitations, has the potential to result in derivation of toxicity factors that are of use to the risk assessment community.

Response: No response required.

2.6.6

Are there other issues specific to developing chronic inhalation toxicity factors that have not been adequately addressed in the document?

Reviewer 6

There are a wide range of issues relating to the development of chronic inhalation toxicity factors. However, I am not aware of any other issues regarding chronic inhalation toxicity factors that are reasonable and appropriate, that have not been addressed in the document. Rating=4 [strong].

Response: No response required.

Reviewer 7

My only comments would be that the toxicity of chronic exposure to particles could be expanded. The sections on gases are much more extensive than the sections on particles, at least in what I was assigned to review.

Response: No response required.

Reviewer 12

Not at this time.

Response: No response required.

2.7 Specific Issues for Vegetation

2.7.1

Are the overall approaches outlined for developing vegetative ESLs adequate?

Reviewer 8

The overall approaches seem reasonable. The availability and quality of the information on vegetative effects will determine whether establishing an ESL can even be considered. The obtained levels leading to moderate effects will actually determine whether the vegetative ESL is higher than any human ESL, and thus, whether the human ESL will be protective and obviate the need for setting a vegetative ESL.

Response: No response required.

Reviewer 13

Threshold values for effects on vegetation are not clearly defined in the literature, even for Criteria Pollutants such as ozone (probably the best studied). These thresholds depend on species and on environmental conditions of relative humidity, temperature, and water availability. These confounding factors should at least be mentioned.

Response: TCEQ agrees and a brief description of the confounding factors has been added.

2.7.2

Are other available alternate approaches more appropriate?

Reviewer 8

Not known to this reviewer.

Response: No response required.

Reviewer 13

Given the paucity of data this protocol is unlikely to be applied to very many compounds. The procedures are very limited but probably appropriate.

Response: No response required.

2.7.3

Are the criteria for deciding whether or not to develop a vegetation ESL appropriate?

Reviewer 8

Yes.

Response: No response required.

Reviewer 13

No. I did not identify these criteria in the document.

[In response to a clarifying question by TCEQ] I have again reviewed the relevant section of the Guidelines, in order to respond to the [clarifying] question. I confirm my earlier assessment.

The entirety of Section 2.3 is very, very brief considering the wide distribution of vegetation and criticality of ecosystem services. The only relevant criteria I can identify are page 34, lines 13-17. This wordy and internally redundant sentence indicates that "if hazard identification and dose-response assessment of a chemical's vegetation effects indicates that the observed effects concentration is close to or below levels of…..human…concern", then ESLs are set. It would seem that the first priority would be to identify whether an ecosystem, canopy, or even single plant is present in a threatened location, as a key part of deciding whether to move forward with setting ESLs. This would then likely focus the search for relevant plant response thresholds. Then would come the current focus of the section, i.e. deciding if the threshold is below that for direct human impacts.

In my opinion, this section should be clear why, when and where this effort would be undertaken. Otherwise, a single search for available thresholds for any plant anywhere for the chemical species in question would be applied in a single sentence to any vegetation in any situation. This could be done without regard for the vegetation that is present (contradicting the concept of using Texas plants first), or the exposure mode that represents the risk. For example, there may be salt pans that are devoid of vegetation for which no meaningful risk can be defined, even if the chemical is known to harm vegetation.

Response: A vegetation ESL will be developed if adequate vegetation toxicity data are available, regardless of whether the observed effects level is higher, close to, or below its odor threshold or health effect levels. The derived vegetation ESLs (in general this is the LOEL for the most sensitive plant species) will be applied to air permits on a case-by-case basis (e.g., if the predicted concentrations from a proposed facility occur in an area where the most sensitive plant species is unlikely to exist, then the effect caused by a specific chemical is not a concern).

2.7.4

Is it appropriate to focus vegetation ESL development on plant species native to or grown in the state of Texas?

Reviewer 8

Yes, provided that the relevant experts continue to track the diversity of plant species that are grown or become established in Texas.

Response: No response required.

Reviewer 13

It is reasonable to focus on such plants, and particularly on these plants growing under the environmental conditions of Texas. However, much more data will be available on plants not growing in Texas, or on similar/same plants growing under non-Texan environmental conditions. This protocol would be improved considerably with development of a method for determination of similarity indices (waxy leaves, deciduous, annual, etc.) along with genetic relationships (congeners, same family, etc.), that would allow any and all data available for the toxic material to be applied in some way to the regulatory situation in Texas.

Response: The Guidelines will not be revised based on these comments (i.e., the current methodology will be maintained). If a new method is developed in the scientific peer-reviewed literature consistent with that suggested by the reviewer, the TCEQ will consider adopting it. Vegetation ESLs will be developed based on the critical level identified in the most sensitive species. The vegetation ESLs will be applied for evaluating vegetation effects only in areas where the studied plants are grown.

2.7.5

Is it appropriate to base the vegetative ESL on the LOEL in the most sensitive species as opposed to the NOAEL in the most sensitive species?

Reviewer 8

Yes, given that only moderate, not mild, effects will be considered.

Response: No response required.

Reviewer 13

This does seem consistent with odor ESL determination, although it clearly does not provide complete protection for the most sensitive species. As these data will be based on discrete experimental concentrations, a better approach could be to derive a linear or curvilinear regression relating effect to concentration. The appropriate ESL from such a relationship would be just below the LOEL, rather than the highest NOAEL that happened to be tested.

Response: TCEQ disagrees with the reviewer. The Vegetation ESL for both acute and chronic scenarios will be determined from the lowest LOEL among several species. The term LOEL is similar to the threshold concept and will replace the term "Threshold effects" used in our previous guidelines.

2.7.6

Is it appropriate to base the vegetative ESL on relatively moderate adverse effects plant damage as opposed to milder vegetative effects?

Reviewer 8

Yes, unless data exist or become available to suggest that mild effects progress to moderate effects under conditions of chronic exposure.

Response: This comment has been incorporated accordingly.

Reviewer 13

In practice it will be very difficult to distinguish these classes of damage. The example given, "dry sepal injury" is unclear. Is this damage to sepals after they are dessicated (unlikely to happen) or damage that is observed after the sepals are dried (clearly subjective, depending on when observation occurs)?

Response: TCEQ agrees and has added "observed after the sepals are dried" after "dry sepal injury."

Any visible injury, organ malformation, or stunted growth should provide an adequate indication of effect. Experience indicates that visible damage does not always correlate with reduced productivity.

Response: No response required.

2.7.7

Are there other issues specific to developing vegetative ESLs that have not been adequately addressed in the document?

Reviewer 8

Not to this reviewer's knowledge.

Response: No response required.

Reviewer 13

The reference to Air Pollution Control Association, 1970 should be updated. It has been revised as Flagler et al., 1998, providing photographs which are very useful for identi fying effects of criteria air pollutants on plants.

Response: The reference Flagler *et al.* (1998) has been added.

2.8 Specific Issues Related to Odor

2.8.1

Are the overall approaches outlined for developing acute odor ESLs (Figure 2-1) adequate?

Reviewer 8

The first stage of the approach for developing acute odor ESLs is quite acceptable and I commend the decision of the TCEQ to move from accepting the lowest odor detection threshold in the literature, irrespective of how or when it was generated, to a more scientifically valid threshold based on data quality. However, this is where the approach both begins and ends and I must voice my disagreement with the notion that the ODT (however it is determined) is a concentration that would pose even the slightest risk to the health and welfare of humans.

According to the guidelines in Chapter 2 (Welfare-based ESLs), short-term ESLs are based on data concerning …..'the potential for odors to cause indirect health effects or nuisance…. This rationale is based on considerable data showing that exposure to "persistent or recurrent exposures to ***strong*** *odors may cause indirect health effects such as headache or nausea in some individuals". An odor concentration that meets this definition must be readily detectable, allow recognition of odor quality (which rarely occurs at the ODT), and be judged by some proportion of the population to be of moderate to* ***strong*** *intensity. The 50% odor detection threshold, however rigorously determined, does not appear to fit this definition, and is certainly an extremely conservative overestimate of odor potency, even in the laboratory.*

Response: TCEQ agrees that the 50% odor detection threshold is conservative and would not pose the slightest risk to the health and welfare of humans. "Strong odors" usually refers to exposure to levels > 3-5 times the 50% odor detection or > the recognition threshold. Clarification has been added.

By definition, the odor detection threshold (as is true for any absolute sensory threshold) represents the limits of performance in the detection of an odor. Very little information is conveyed to the perceiver at their individual threshold. The 50% odor detection threshold for any panel is the concentration which is discriminable from a clean air blank for half of those individuals at a 50% detection probability. By its very nature, that odor is just on the cusp of detectability. In my 18 years of experience collecting odor thresholds, I have never encountered a threshold concentration for a group of individuals that would be described as a distinct odor, much less a strong or annoying odor.

Response: No response required.

Moreover, it is well-recognized that the absolute detection threshold for 50% of the population that is obtained under laboratory conditions (focused attention, forced-choice comparison with clean air, zero background odors) would almost never be a concentration that would rise to the level of awareness in the field. For example, studies in which subjects have either been specifically directed to detect the odor or not, have found that there is a four-fold increased detection threshold for the undirected test as compared to the directed test [1]. This would be on top of the need for an odor concentration to be at least three times the ODT for its quality to be recognized. Thus, in the opinion of this reviewer, the TCEQ

has made substantial progress in the procedure to establish ESLs for odor, but much more needs to be done to be able to set a scientifically-defensible guideline for odor nuisance in a community.

***Response**:* TCEQ agrees. However, the Guidelines were not revised because odor-ESLs are comparison levels, not ambient air standards, the simple exceedance of which does not establish an odor nuisance condition (e.g., the TCEQ has procedures for the investigation of odors). Thus, consistent with the TCEQ's use of these screening values, they are generally set at levels well below their recognition thresholds (i.e., set at 50% detection thresholds).

Reviewer 13

The overall procedure is appropriate and logical, the values that have the most scientific merit of those derived from level I values. In these cases it may make more sense to utilize the lowest published values, unless these can be discredited, rather than the geometric mean. The logic of the geometric mean is more regulatory than scientific.

Response: The Guidelines were not revised based on this comment. Using the geometric mean of reported Level 1 and/or Level 2 thresholds instead of the lowest odor threshold to set the odor-based ESL is consistent with procedures recommended by the NAC/AEGL Committee. Specifically, when several studies reporting the odor threshold of a given chemical are available, which are of similar quality in terms of methodologies used to collect the data, it is the practice of the TCEQ to take a geometric mean of those data rather than choose the lowest odor threshold with the goal of utilizing an odor threshold value representative of all available high-quality data rather than defaulting to a lowest value.

This subject is addressed again (section 2.2.3.1), although this paragraph is extremely difficult to interpret. The use of "acceptable" odor threshold from the 2006 document is not carried through later usage of the term "odor threshold". It is therefore unclear whether "acceptable" is an intentional distinction, or what its intent may be. The complex sentence (p. 27, l. 7) is very difficult to interpret, appears illogical, and provides little support for the use of the geometric mean.

Response: Clarifying text has been added.

2.8.2

Are other available alternate approaches more appropriate?

Reviewer 8

In the opinion of this reviewer, there exists a far superior method for determining the concentration of odor that is likely to alert individuals to its' presence and that in some circumstances may elicit community concern or become a nuisance. This is the method proposed in RIVM report 609200001/2009, entitled "Assessment of odour annoyance in chemical emergency management'"[2]. The study reports a methodology that is intended to derive an airborne chemical concentration producing a distinct odour perception in more than half of an exposed, distracted population that would qualify as significant odour awareness (designed as the Level of Distinct Odour Awareness or LOA).

Using the RIVM classification for the three levels of odor detection thresholds, the TCEQ process begins with the selection or calculation of an appropriate odour threshold. However, at that point, the two methods for setting a guideline value diverge. The TCEQ ESL is simply based on this threshold (either the geometric mean of the compiled odor thresholds, an extrapolated threshold based on structurally-related chemicals or the lowest odor threshold). In contrast, the RIVM process uses the threshold as the starting point and then derives a distinct odor level using adjustment factors based on the slope of the psychophysical intensity (dose-response) function, and field conditions. The distinct odor perception in the field (LOA) derived from this will be ~16 times the laboratory-based ODT.

In addition to this approach, there are other well-established methods to establish enforceable odor guidelines. A number of US states have adopted regulations that allow them to quantify an odor nuisance level using a field olfactometer. McGinley et al have developed an instrument that allows field odor detection in a quantifiable manner (Nasal Ranger™)[3]. Although there is variation between the nuisance levels between some states using this methodology, the odor nuisance level is never set at the ODT. Rather the level is set at a specified dilution ratio where a unit of odorous air is diluted by a certain number of units of clean air. This refers to the amount of clean air that needs to be mixed with odorous air for the odor to be undetectable. The most common dilution to threshold ratio used to set nuisance regulations is at or greater than a 7:1 dilution to threshold (d/t).

Response: See response to Reviewer 8 in 2.8.1 above.

Reviewer 13

An alternative to the geometric mean proposed above, in those cases for which data are available, would be to recalculate the 50% threshold for all participants, rather than the mean of 50% thresholds derived from the various studies.

Response: The Guidelines were not revised based on this comment because the original data for all participants may not be available for recalculation and the existing procedure is deemed to provide satisfactory results.

Because "code A" (AIHA; p. 25, l. 17) is a rather specific set of guidelines, these could be described briefly in this document.

Response: The specific description for code A (critique acceptance) has been added to the Level 3 quality level for odor thresholds in Chapter 2.

2.8.3

Have the definitions of the quality levels (Levels 1, 2, and 3) for odor threshold values been described correctly? Have these quality levels been appropriately used in the development of odor ESLs?

Reviewer 8

Yes, the definitions of the quality levels for ODT values are correctly described. This is a significant advancement in the guidelines toward the use of the most scientifically valid ODT values. This change acknowledges the lack of methodological precision for thresholds obtained more than several decades ago and the recent move toward standardization and harmonization across olfactometry laboratories on odor measurement practices.

Response: No response required.

Reviewer 13

It is not clear in Chapter 2 whether a Detection or Recognition Threshold is being applied.

Response: Chapter 2 has been revised to specify that the lowest acceptable odor detection threshold value is applied. Otherwise the odor recognition threshold is applied when detection thresholds are not reported. For data categorized at level 1 or 2, the geometric mean of these values is used. For data categorized as level 3, the lowest odor threshold value is used.

In section 2.2.3, is not clear why the use of extrapolated surrogate odor ESL's is refined to air permit reviews only. This may have regulatory significance but does not seem to be scientifically justified.

Response: The text has been revised to not limit the use of extrapolated surrogate odor-ESLs to air permit reviews only.

The calculation of level II odor detection thresholds (section 2.2.4) relies on the n-butanol standard, which seems logical. However the specific error(s) for which this standard compensates should be described.

Response: The Guidelines were not revised base on this comment as a more detailed discussion is available in the TCEQ odor specific guidelines and publication.

2.8.4

Is the use of the geometric mean of all equivalent odor threshold values (for Level 1 and Level 2 data) an appropriate approach for developing the odor ESLs?

Reviewer 8

The use of the geometric mean of all equivalent odor threshold values meeting the criteria for Level 1 (best) or Level 2 (next preferred) data is a far better representation of the population odor detection threshold than simply selecting the lowest obtained value, due to the log-normal distribution for odor thresholds [4]) One minor concern, however, is the potential for an outlier among the compiled ODT values to inordinately affect the calculated mean value. Although most of the Level 1 and 2 values are likely to have less variation than values derived prior to 1990, a rule for exclusion of an outlier, particularly when there is good agreement among the rest of the ODT values, might be necessary.

Response: TCEQ agrees with this comment and has added the rule of exclusion.

Reviewer 13

Given the log normal distribution, this appears justified.

Response: No response required.

2.8.5

The TCEQ has determined that if nuisance odors have actually been reported at concentrations lower than the geometric mean, then lowest odor threshold value will be used for the odor ESL. Is this decision appropriate?

Reviewer 8

Although this reviewer has serious doubts that a concentration lower than the established odor ESL (as determined by the proposed guidelines), would generate a nuisance complaint based on a discernible odor, the implementation of this rule needs to be better defined. What process and constituency would form the basis for validating the presence of a nuisance odor and how would it be measured in the field to compare with the current ESL? Nevertheless, if it turns out that the odor ESL is set at a level that consistently is judged to be a nuisance odor and an independent process has been used to confirm the nuisance level, then it would be reasonable to adopt a lower value. Those unusual circumstances, however, would call into question the quality of the original data set (or the presence of a significant outlier) and one might question whether even adopting the lowest value from the available data would be advisable.

Response: TCEQ staff must follow the Odor Complaint Investigation Procedures (TCEQ 2007) (http://www.tceq.texas.gov/complaints/protocols/odor_protopdf.html/at_download/file) to validate the presence of a nuisance odor (e.g., by using the FIDO Chart to characterize odor's frequency, intensity, duration, and offensiveness). The reference for the TCEQ procedures has been added to the Guidelines.

Reviewer 13

The setting of an ESL at a lower value than the geometric mean it will occur if valid reports exist of odor at a lower concentration. Unless these field reports are exceptionally well documented, this leads to a level of subjective interpretation that could prove troublesome. At a minimum, some indication of what constitutes a valid field report should be included.

Response: See response to Reviewer 8's comment above.

It is not clear (p. 27, l. 15) whether these example classes are materials that have been treated in this way, or classes of materials that might be candidates for such treatment.

Response: These examples are materials that either have been treated or may be potential candidates if identified in the future.

2.8.6

Is it appropriate to adopt the acute odor ESL for longer exposure durations?

Reviewer 8

It is entirely appropriate to adopt the acute odor ESL for longer exposure durations; the intensity of odor is dependent on the concentration, and there is no cumulative intensity as a function of exposure time. Indeed, the opposite is true: continued exposure to an odor will results in a decreased intensity for that odor, a well-recognized and validated process known as olfactory adaptation or 'odor fatigue' [5]*. Thus there is no need to establish a chronic ESL for odor, as the most intense odor will be experienced acutely.*

Response: No response required.

Reviewer 13

This is probably appropriate, however in cases where the airborne compound is toxic, clearly the longer exposure even at threshold levels of odor detection will pose a larger hazard. In section 2.2.5 it should be stated explicitly that the one-hour ESL would apply even if the 24-hour concentration is below the odor-based AMCV.

Response: TCEQ agrees and has revised the section accordingly.

2.8.7

Is the process for developing generic odor ESLs in the absence of chemical-specific data appropriate (see Figure 2-4)?

Reviewer 8

While the scarcity of credible and reliable ODTs for many chemicals is lamentable, it is also well-recognized that in some cases guidelines must be established on the basis of no acceptable data. The generic process set forth to determine an ODT under these circumstances is defensible, and can be used subject to the minor considerations I have listed below.

Response: No response required.

Reviewer 13

[No specific comment here]

Response: No response required.

2.8.7.1 Part A.

Use of data on correlation between chain length and odor to generate odor ESL equations that will estimate an odor ESL in absence of good odor data (see Table 2-1).

Reviewer 8

Unfortunately, the draft of the consigned review and analysis that supports the implementation of this process was not available for dissemination at this time. However, this reviewer was able to obtain and review the excel sheets showing the data analysis for the various chemical groups. This review was convincing that the process represents a scientifically-defensible approach for determining thresholds for compounds where good correlations exist between chain length and ODTs. However, it is important to recognize that when few threshold values are available within a group, there can be significant effects from one or more datapoints. In particular, I am referring to the graph in Figure 2-4 on the aldehyde group. The correlations between ODT and chain length for this group are the lowest among all of the 'good correlations' (R^2=.46) and if the data point for formaldehyde is removed, this correlation falls below the criterion for a 'good correlation' (R^2=.20). Thus, estimates of a compound's odor threshold using the generated equation may be flawed.

Response: Some data points were found to be missing from the aldehyde group and were subsequently added. The data have been reevaluated using correlation and statistical analysis (this was done for all chemical groups). The correlation between carbon chain and odor thresholds is moderate, yet statistically significant for aldehydes and the equation will be utilized to estimate odor thresholds.

Reviewer 13

These procedures seem appropriate, and are a good example of the use of related data in a semi-quantitative fashion. The presentation is unclear (p. 28, l. 33), where "$r^2 > 0.3$" may be a typographical error. In any case the r^2 values are not informative without either sample size or probability of a significant regression. This information should be added.

Response: TCEQ agrees. All data have been reevaluated wherein chemical groups with statistically significant ($p<0.05$) correlation between carbon chain number and odor threshold are used to predict odor thresholds. The R^2 values are used to determine how well data fit the trendline, which describes the relationship between carbon chain number and odor threshold.

2.8.7.2 Part B.

For chemicals that are from groups with poor correlation between odor threshold and carbon chain length, the geometric mean of the OT50 of that group is used as the generic odor-based ESL (see Table 2-2).

Reviewer 8

Given the numerous compounds for which no reliable ODT exists, it is likely that there will be a need to derive a generic odor threshold for a compound where data from structurally-related compounds exists, but for which there is poor association between physico-chemical parameters and odor. In this case, the use of the geometric mean of the group is a reasonable way to develop the ODT, provided the issues regarding significant outliers (such as formaldehyde data point, discussed above) can be managed in a scientifically-defensible way.

Response: No response required.

Reviewer 13

More detail should be provided on what would lead to "anticipation" of an unpleasant odor. As this anticipation could lead to regulatory action, this procedure suffers from an excessive element of subjective judgment (p. 31, l. 5).

Response: "Unpleasant odors" has been replaced with "similar types of odor." Exposures to an odorous chemical at levels above its detection threshold may be discernible regardless of whether it is pleasant or unpleasant.

2.8.8

Are there other issues specific to developing odor ESLs that have not been adequately addressed in the document?

Reviewer 8

On the basis of nuisance odor, the decision to adopt a single odor ESL for both children and adults is reasonable and defensible. Among very young children, there is a tendency toward a <u>greater</u> acceptance of odors that may be aversive to adults. By age 3,however, these differences have largely disappeared and children appear to have similar odor preferences and aversions as their parents or guardians[6]. Thus there is no scientific reason to treat children as a sensitive sub-population on the basis of odor impact.

Response: No response required.

Reviewer 13

The important subject of indirect health effects is mentioned (p. 22, l. 4), but appears already to have been removed from consideration in chapter 1. This is an important subject that should be considered, particularly for airborne toxics that may pose a human health risk through bioconcentration despite being below the 1-hour concentration threshold in air.

Response: As stated in Chapter 1, the THSC authorizes the prevention and remedy of air pollution based on direct effects. Additionally, as noted by the commenter, TCEQ does consider indirect health effects such as nausea and/or headache due to strong persistent and unpleasant odors. The chronic toxicity assessment addresses any effects due to any build-up of dose of a chemical in the human body over time due to long-term inhalation exposure.

The decision to remove tabulation of odor thresholds if they are above the health based ESL (p. 31, l. 15) should be reconsidered. There may be instances when the two types of information may need to be applied differently so that both will be required. In any case, they are independent scientific findings. In an extreme case, the lower health threshold might eventually be changed or discredited, leaving the odor ESL as the remaining scientifically justified threshold. Both should be reported.

Response: TCEQ agrees and the paragraph has been deleted.

3.0 General/Editorial Comments

Reviewer 4 (Typographical errors)

- *P. 137 l 6: the units should be mg/day.*

Response: The suggested revision has been made.

- *P. 148 l 20 to 26: the units for the 6.7 x 10-6 factor (1/day?) need to be included to make the units of the RfD correct.*

Response: The suggested revision has been made.

Reviewer 5

- *Some of the diagrams are too small to read, so they may need to be displayed by themselves on a separate page.*
- *It would be useful to number the equations (chapter#/equation # in order of appearance).*
- *Some of the symbols in the equations are used to represent different variables. The authors may consider having a uniform use of symbols and variables throughout the text.*

Response: The equations in the Guidelines were created by other entities and TCEQ is using them as they were intended to be used, without modification. The symbols associated with an equation are defined when that equation is introduced. No changes were made based on this comment.

- *Although outside of the charge questions above, I have a concern about the carcinogenic ESLs for children. Specifically, estimates of lifetime risk appear not to consider the time window in which a specific cancer may occur, which could be related to an earlier exposure also during childhood. This is not discussed in the document.*

Response: Regulatory risk assessment involves calculating excess risk on a lifetime basis. The issue of a cancer potentially being related to an earlier childhood exposure is addressed in the document through the application of ADAFs for chemicals with a mutagenic MOA for carcinogenesis. This accounts for early life susceptibility and makes the unit risk factor more conservative, increasing estimates of lifetime risk accordingly and driving down the air concentration corresponding to a particular risk level.

- *On page 29, graph C (aldehydes): the slope of the line seems to be driven by just one point between 100 and 1000. Reconsider if this point should be included.*

Response: The slope of the trendline for aldehydes is not purely driven by one point. In analyzing the relationship between odor threshold and carbon chain number, TCEQ takes into consideration that it is possible that a single point could drive the shape of a given curve. To investigate whether this was the case for aldehydes, the curve was generated both with and without the aforementioned point, formaldehyde. Removal of this point did reduce the R^2 values (R^2= 0.52 before and 0.41 after), demonstrating that curve fit is somewhat dependent upon this data point. However, Spearman's rho (ρ= -0.77 before and -0.72 after) and the subsequent p-values (p=0.0003 before and 0.004 after) reveal that this point is not the only driver for correlation or statistical significance between odor threshold and carbon chain number for aldehydes. Rather, the addition of this point provides a more data-rich, better

fitting curve and as such, the formaldehyde data was retained and no changes were made to that particular graph in response to this comment.

Reviewer 7

- *P. 47, l 18: Need to close the parenthesis after the word "polymers"*

Response: The suggested revision has been made.

- *P. 67, l 17: This sentence would be clearer if the parentheses around "as opposed to particles" were replaced with commas.*

Response: The suggested revision has been made.

- *Figure 3-9: The lowest point on the "n=3" line should be a diamond, not a closed circle.*

Response: The error has been corrected with a new graph.

- *P. 74, l 23:"physic-" should be "physico-"*

Response: The suggested revision has been made.

- *P. 92, l 21: The close parenthesis sign after "3" should be removed.*

Response: The suggested change could not be located in the document, no change was made.

- *P. 99, l 12 and 18: "as-needed" should be hyphenated*

Response: The suggested revision has been made.

- *P. 100, l 12-13: This is an incomplete sentence.*

Response: The suggested revision has been made.

Reviewer 8

Reference List

1. *Amoore JE. The perception of hydrogen sulfide odor in relation to setting an ambient standard. El Cerrito, CA: Olfacto-Labs; 1985. Report No.: ARV contr. a4-046-33.*

Response: TCEQ has chosen not to add this reference.

2. *Ruijten MWvDR, Van Harreveld AP. Asssment of odour annoyance in chemical emergency management. The Netherlands: National Institute for Public Health and the Environment; 2009. Report No.: 609200001.*

Response: TCEQ has already cited this reference.

3. *McGinley MA, McGinley CM. Comparison of field olfactometers in a controlled chamber using hydrogen sulfide as the test odorant. Water Sci Technol. 2004;50(4):75-82.*

Response: TCEQ has chosen not to add this reference.

4. *Cometto-Muniz JE, Abraham MH. Human olfactory detection of homologous n-alcohols measured via concentration-response functions. Pharmacol Biochem Behav. 2008 May;89(3):279-291.*

Response: TCEQ has already cited this reference.

5. *Dalton P, Wysocki CJ. The nature and duration of adaptation following long-term exposure to odors. Percep Psychophys. 1996;58(5):781-792.*

Response: TCEQ has chosen not to add this reference.

6. *Schmidt HJ, Beauchamp GK. Adult-like odor preferences and aversions in three-year-old children. Child Dev. 1988;59:1136-1143.*

Response: TCEQ has chosen not to add this reference.

Reviewer 10

1. *This is a very useful compilation of the discipline references. Care should be taken about using some of the older references and tempered with more recent knowledge as necessary. See discussion below.*

Response: No response required.

2. *Include also ATSDR toxicology profiles. They are an excellent source. Also the TSCA Section 8(e) submissions (http://www.epa.gov/opptintr/tsca8e/) – this is a good source for unpublished studies.*

Response: In being consistent with the common procedures to derive toxicity factors, TCEQ reviewed numerous federal and state guidance documents and scientific articles to develop these guidelines. ATSDR is among those specifically referenced for the development of this document. Although TSCA Section 8(e) is not specifically mentioned, TCEQ considers all relevant data on a chemical by chemical basis.

3. *One of the charge questions should have been on the use of benchmark dose methodology. The discussion Section 3.7.3 is well done. As I recall it reflects what is in the EPA documentation. EPA will be coming out with a revision to the guidance document in the near future. TCEQ can contact Karen Hogan (NCEA, Washington, DC) for the schedule. There are at least two factors that should be considered:*
   a. *The good thing about the EPA BMD software is that it is publicly available and easy to use. One of the bad things about the EPA BMD software is that it is easy to use. There should be some discussion of the types of judgment that should be applied before using BMD modeling because there are times when it is simply inappropriate. Discussion of some of the points in "4. Minimum Data Set 1 for Calculating a BMD" in USEPA 2000 would be appropriate. The AEGL Program (NRC, 2001) states "Because of uncertainties that may be associated with extrapolations beyond the experimental data, the estimated values are compared with the empirical data. Estimated values that conflict with empirical data will generally not be used." Without this consideration, the user can fall into the trap of applying the methodology by rote and obtain numbers that have no relation to reality.*

Response: Two suggestions have been provided here. One request is that a discussion of the types of judgment that should be applied before using BMD modeling be included. The Guidelines already provide this information in this section and a reference to USEPA 1995 and 2000b is provided. Therefore the Guidelines were not revised based on this suggestion. The second request was to include a statement concerning the uncertainties associated with extrapolation beyond the experimental data when using BMC modeling. The statement from the AEGL program has been included in the Guidelines.

b. *One thing the AEGL SOP does comment on is "Because of uncertainties that may be associated with extrapolations beyond the experimental data, the estimated values are compared with the empirical data. Estimated values that conflict with empirical data will generally not be used."*

Response: No response required.

4. *Point of trivia. The reference to Haber's rule (P. 70-l 35) should be: Haber, F. 1924. Zur Geschichte des Gaskrieges. P. 76-92 in Fünf Vorträge aus den Jahren 1920-1923. Berlin: Springer-Verlag.*

Response: The suggested reference has been added.

5. *Section 3.10 on dosimetry adjustments from animal to human.*
   a. *Some general comment on your discussion as it relates to USEPA, 1994 follow:*
      i. *For Category 1 gases:*
         1. *Category-1 gases in the ET region. The current methodology in the RfC document assumes that the deposition is identical throughout the ET region. This results in an HEC of about 1/5 for the equivalent concentration. The work of Kimbell (you cite Kimbell et al. 1993) on formaldehyde and others on other gases clearly demonstrated with both modelling, and data, that the HEC was effectively 1 at physiologically reasonable concentrations. See also "2.4 DOSIMETRY CORRECTIONS FROM ANIMAL TO HUMAN EXPOSURES" in NRC, 2001 for a discussion about how the AEGL Program addresses this issue.*

Response: TCEQ is familiar with how the AEGL program addresses this issue and will generally use the dosimetric adjustment equation given in this section for Category 1 gases. However, TCEQ will also consider the scientific defensibility of using alternate dosimetric adjustments when models have been applied on a chemical-specific basis (e.g., formaldehyde), and utilize best scientific judgment in determining the most appropriate dosimetric adjustment equation or model.

         2. *If the pulmonary region is the target, a default HEC calculation using minute volume and surface area of animal and human will give HEC values that are about 3X higher in the human. This is anti-conservative and certainly greater than the differences between human adult and child. For many chemicals this difference may be true and it may take a 3X greater concentration to deliver the same dose to the human lung than an animal. As always if you have quantitative data or validated models, they should be used. In the PMN Program at EPA we use the default minute volume to surface area alluded to above unless we have something better. I just want to add a caution that while you are looking at differences between adult and child, there may be other matters that contribute a greater difference. While we may have to use that default I would hope the science can improve in this area. It is certainly one we should always be mindful of when default approaches are applied.*

Response: No response required.

         3. *For Category 2 gases: Fortunately you picked up on this being reconsidered by EPA. It is my opinion that for systemic effects the formula presented is incorrect. It looks like the modeler mistakenly inverted two parameters. If you reverse the parameters*

in the equation you get more reasonable numbers. I didn't take the time to locate where the error occurred but something is wrong.

Response: No response required.

4. *For Category 3 gases: This one actually seems to work and using the blood air partition coefficients is a good surrogate.*

Response: No response required.

b. *An alternative approach can be found in the AEGL SOP (NRC, 2001 Section 2.4) that goes into more detail on these discussions. For AEGLs we don't make dosimetry corrections between animals and humans unless we have data or a good rationale.*

Response: See response to Reviewer 10, comment point number 5, above.

c. *Another reference you might add is EPA/600/R-09/072. 2009. STATUS REPORT: Advances in Inhalation Dosimetry of Gases and Vapors with Portal of Entry Effects in the Upper Respiratory Tract. It can be found at http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=212131*

Response: The suggested reference has been added.

d. *I talked with George Woodall and he said that EPA is currently using a body weight to a power conversion for inhalation assessments. You might want to check with the EPA/NCEA people on that.*

Response: We thank the reviewer for this comment.

e. *Part of the rambling nature of this section is the discomfort I feel when we do these corrections. There are a lot of uncertainties that should be considered if possible and when the defaults are used, they should be used with some degree of humility. If you revise this section you could discuss some of the uncertainties involved so the reader doesn't get the impression that we know more than we really do.*

Response: The chemical-specific development support documents (DSDs) will address uncertainties in the discussion sections.

6. *Section 3.12 Uncertainty factors.*
   a. *This seems like a balanced presentation of what is done by the different agencies at this time. It is appropriate that this section acknowledges that defaults are not always used and decisions will be made based upon the chemical specific data.*

Response: TCEQ agrees and has revised the section accordingly.

7. *Section 3.18 Rounding*
   a. *Mathematically rounding 13,563 to 13,000 is mathematically incorrect. It should be 14,000. If the value is 13,500 exactly rounding down is a legitimate policy decision.*

Response: The rounding procedures are discussed in this section. No response required.

Reviewer 11 (References)

- *USEPA (2002). A Review of the Reference Dose and Reference Concentration Processes. EPA/630/P-02/002F. http://www.epa.gov/raf/publications/pdfs/rfd-final.pdf.*

Response: TCEQ has already cited this reference.

- *USEPA (2005). Guidance on Selecting Age Groups for Monitoring and Assessing Childhood Exposures to Environmental Contaminants. EPA/630/P-03/003F. November 2005. http://www.epa.gov/raf/publications/pdfs/AGEGROUPS.PDF.*

Response: The suggested reference has been added.

- *USEPA (2011a). Guidance for Applying Quantitative Data to Develop Data-derived Extrapolation Factors for Interspecies and Intraspecies Extrapolation. External Review Draft. EPA/100/J-11/001. May 2011. http://www.epa.gov/raf/files/ddef-external-review-draft05-11-11.pdf.*

Response: Text has been added to Chapter 3 indicating that, when finalized, the draft USEPA (2011) guidance document on data-derived extrapolation factors (DDEFs) provides an approach fairly similar to IPCS (2005) and may be useful in deriving interspecies or intraspecies DDEFs for either toxicokinetics or toxicodynamics.

- *USEPA (2011b). Recommended Use of Body Weight ¾ as the Default Method in Derivation of the Oral Reference Dose. EPA/100/R11/0001. http://www.epa.gov/raf/publications/interspecies-extrapolation.htm.*

Response: See response to Reviewer 9 in 2.1.1 above.

Reviewer 12 (general comments)

Chapters 1-6

The authors of this report are to be commended for the thoroughness, accuracy and usefulness instilled into this report. This reviewer was impressed that, for example, even reports that have not been adequately implemented by the originating Agency, such as EPA's application of BW allometry for oral studies, use of data arrays, use of ten Berge models and C^n x T, have been cited, discussed accurately and implementation thoughtfully proposed. The point-by-point analysis and examples of the Supplemental Guidance and the in-depth checklists for conducting WOE for mutagenic origins of cancers are, in my judgment, more helpful and clearer than those offered by the originating agency. This reviewer further opines that risk assessors utilizing this guidance will finish with products that are more transparent and more internally consistent than they would be by attempting to apply much of the existing EPA guidances.

Response: No response required.

The only issue considered major that this reviewer has with the guidance regards the manner in which children's issues and dosimetry are handled. The reason that the NAS report on this issue (Bruckner JV. Regul Toxicol Pharmacol. 2000 Jun;31(3):280-5) is not even mentioned in this report is not clear although one of the overall NAS evaluations, that the current use of uncertainty factors adequately addresses risk to children, is clear and prominent in the guidance. A principal reason for NAS putting this conclusion forward is the realization and documentation that risks to children fall largely in the range of 1-to 2-fold that of adult values. Other clearly stated directives of the NAS report included (1) using PBPK models to continue and refine evaluation of child risk and (2) to continue collection of data needed to support these models. Your white paper report captures most of this work that can be considered generated by the NAS directives including the models of Pelekis et al. (2001), Price et al. (2003), Sarangapani et al., (2003), Clewell et al. (2004) and data sets collated for this purpose (e.g., Ginsberg et

al., 2002). The guidance itself even mentions some other relevant child/adult modeling efforts (Garcia et al., 2009). The findings of all these modeling efforts using these collated data support the initial NAS evaluation of child to adult differences being in the range 1-2-fold that of adults and, as such, would be accommodated by current commonly used (10-fold) uncertainty factors. This finding is essentially that of the white paper accompanying the guidance as an appendix (see Figures therein). Also at issue with this reviewer is the explicit consideration of children as a susceptible subgroup to the apparent exclusion of all other susceptible subgroups. This is quite apparent in those sections addressing uncertainty factors. The several proposals throughout the noncancer sections for differential applications in consideration of children are to this reviewer in contrast and at odds with the consideration made for cancer risk in children. As accurately reviewed in your guidance, the Supplemental Guidance proposes a single quantitative modification to risk factors in response to the documentation that children are more susceptible to cancer in a degree that is considered by the modification. The equitable and supportable parallel for noncancer effects in children would be to acknowledge that this degree of difference has been determined to be already incorporated in the assessment procedures, i.e., the UFs. If this comment is accepted and considered for implementation, there would be a spectrum of choices as to how it could be accomplished. These include, on the minimal side, that text be added to cite and explain the NAS findings in sections deemed appropriate whereas a broader scale may be major alteration or even exclusion of some sections including those for certain uncertainty factors.

Response: The Bruckner reference has been cited in several places in the document. Other sensitive subgroups have been discussed specifically in Chapters 1 and 3.

Another but less pressing issue than the one above for this reviewer is the guidance's reliance and prominence of MOA information in the various schema and procedures proposed herein. Toxicity factors are primarily dose-response evaluations, not MOA evaluations. This reviewer appreciates MOA information and how these data allow for an assessment to be based more scientifically on events proximal in time and place to the target tissues. This reviewer is also keenly aware that MOA information is relatively rare and more often than not confusing due to the situation of its generation (e.g., in vivo vs. in vitro, extremely high doses, objective of the experimenter). Indeed it has been established with several toxic agents that different MOAs function at different dose levels. The manner in which these have been discovered is, of course, through observations made in dose-response information which, again, demonstrates the primacy of dose-response information. Thus it is considered advisable for the authors to establish and make clear the primacy of dose-response information in deriving toxicity factors through a careful and thoughtful reading of the guidance, especially those sections in which MOA is so prominently featured. Another thought to aid in these sections where MOA is prominent is for the authors to undertake the addition of text and synthesis that relates SAR and MOA information as per how SAR is utilized in the sections on relative potency factors.

Response: Pursuant to other more specific comments by this reviewer regarding the emphasis placed on MOA, Chapter 3 and the "Document Description and Intended Use" section have been revised with more emphasis on dose-response. Similarly, text has been revised throughout the document to clarify this subject.

Chapter7

Much of the materials in this chapter were outside the expertise of this reviewer. It was noted that the dosimetry proposed to apply to occupational studies (Section 7.9.3) was consistent with that used by US EPA. Section 7.10 appears to be a complete and thorough implementation of both the cancer guidelines (US EPA, 2005) and the quantitative BMD/C models. The sections on meta-analysis were also outside this reviewer's area of expertise.

Response: No response required.

Reviewer 12 (editorial comments)

- *xiii, MPPD = multiple -**path** particle dosimetry. Please perform a universal search to correct this error as it appears several times throughout the guidance.*

Response: The suggested revision has been made.

- *I, l 20. Despite all that follows in this document regarding MOA, the basis of any derived value has to be dose-response. This is present but obscured in this document as consideration is given for deriving toxicity values in the absence of MOA information but not in the absence of dose-response information. Toxicity factors are dose-response evaluations and cannot be derived without dose-response information. Please consider changing MOA here to "dose-response information" and rewording the section to reflect dose response as being the primary and foremost information required for toxicity factor derivation.*

Response: TCEQ believes that it goes without saying that information on dose-response, the cornerstone of toxicology, is of paramount importance to dose-response assessments and the development of toxicity factors. To greater emphasize the importance of dose-response data the referenced sentence has been revised.

- *P. 3, l 13 ff. This description ignores nonlinearity as it occurs at high doses/exposures. Please acknowledge this and further specify the nonlinear dose-response discussed here, probably most accurately in terms of becoming parallel to the dose-axis at lower and lower doses such that there is dose at which no effect occurs. This concept then can also be used to more specifically identify and define both "threshold" and "nonthreshold" here and below. Nonthreshold would be simply and strongly impressed on the reader as effects potentially occurring at any dose.*

Response: The Guidelines were not revised based on this comment. This comment is referring to the beginning of the definitions of nonlinear dose response, linear dose response, and threshold. The preceding paragraph states "The low-dose region is generally defined as the dose range below the experimental doses. Nonlinear dose response, linear dose response, and threshold are defined as follows." Therefore, the definitions are applied to the low-dose region, below the doses where effects occur. The concept of threshold is defined adequately in these definitions, as is the concept of nonthreshold.

- *P. 4, l 9 ff. Please consider amplifying this paragraph. Typically further distinction is required to differentiate an acute exposure from just a single exposure in a repeated dose exposure regime. For EPA an acute exposure is considered as a single exposure of 24 hours or less in duration followed by an observation period, usually 24 to 96 hours, in which to observe any poisonous effect produced from that single exposure.*

Response: The Guidelines were not revised based on this comment. The definition in Chapter 1 for an acute ReV is an introductory definition– acute ReVs are derived for evaluation of a 1-h exposure duration, although they are to be protective for intermittent exposure. Therefore, the definition of an acute ReV is different from the definition of an acute exposure used by EPA (i.e., a different problem formulation). Chapter 3 goes into greater detail on the definition of acute exposures and the types of acute and subacute studies that are used to develop acute ReVs. Chapters 3 and 4 discuss the procedures for adjusting the duration of experimental studies to be representative of a 1-h exposure.

- *P. 4, l 23 ff. The POD is a vitally important concept of how assessment is done that is blindly introduced here. Please consider the IRIS definition of a POD as a standalone paragraph, or as a footnote here or as an amplified definition in your glossary as well as further reference to your Section 3.7 below: "**Point of Departure:** The dose-response point that marks the beginning of a low-dose extrapolation. This point can be the lower bound on dose for an estimated incidence or a change in response level from a dose-response model (BMD), or a NOAEL or LOAEL for an observed incidence, or change in level of response" (from IRIS glossary).*

Response: TCEQ agrees and has revised the definition of POD in this section.

- *P. 4, l 32 ff. This document could be more inclusive of the population and also avoid becoming biased and defensive regarding the children's issues by just adding "… pregnant women and the elderly" here as you do further down in the document.*

Response: This section has been revised to include pregnant women and the elderly as additional examples of sensitive individuals.

- *P. 5, l 2-3. This statement is too simplistic and can and will be misused. Please consider adding phrasing such as "…in a population known to be sensitive through data and/or mode-of-action or other chemical specific information".*

Response: TCEQ agrees and has added the suggested phrase to this statement.

- *P. 5, l 11. The nondefined and simplistic use of the descriptor "excess" for risk here is not acceptable. That this term addresses inclusion of background rates needs to be clear. An option for this could be to eliminate it as an adjective in the first sentence and then follow with a short descriptive sentence about excess risk and what it entails as is made clear in the additional materials and definitions at the US EPA BMD web site.*

Response: The section has been changed to include a brief definition of excess risk.

- *P. 5, l 17. Consider adding here "statistically-derived" upper-bound estimate.*

Response: The suggested revision has been made.

- *P. 38, l 11 ff. This counterpoint regarding lung alveoli does not follow. For example, fewer alveoli being exposed to more particulate-containing air both indicate a potential for higher local doses to these few alveoli (adults would be able to distribute and lower the local dose to their more numerous alveoli at a decreased ventilation rate). I would therefore eliminate this statement.*

Response: This sentence has been deleted.

- *l 21-23. The exhaustive and excellent PBPK-based study of gas dosimetry in adults in children of Sarangapani should be included here.*

- *Sarangapani, R.; Gentry, P. R.; Covington, T. R.; Teeguarden, J. G.; Clewell HJ 3rd (2003). Evaluation of the potential impact of age- and gender-specific lung morphology and ventilation rate on the dosimetry of vapors. Inhal Toxicol, 15: 987-1016.*

Response: The suggested reference has been added.

- *l 17. ~~and~~ **although** the significance*

Response: The suggested revision has been made.

- *l 24. The formal response of the NAS to the FQPA should be cited here:*
  - *Bruckner, J. V. (2000). Differences in sensitivity of children and adults to chemical toxicity: The NAS panel report. Regul Toxicol Pharmacol, 31: 280-285.*
- *l 33. The last statement should be followed by both the Guzelian and Henry (1992) and the Bruckner (2000) reference.*

Response: The suggested revision has been made.

- *P. 41, l 15 ff. The notable finding of Ginsberg et al. (2002) of higher clearance in 0.5 to 12 years children is not mentioned here. Suggested text addition; " … results showed that, for those chemicals with clearance data (27 substrates), premature to 2-6 months of age infants showed significantly lower clearance (P<0.01) whereas 6 mo to 12-yr-old children had significantly higher clearance (P<0.0001) than adults." It should also be pointed out early in the text that these are all oral exposures of high dose therapeutic agents with short (hrs) t 1/2s..*

Response The suggested text has been added to Chapter 3.

- *P. 42, l 8. … between adults and children **and between laboratory rodents and humans***

Response: This sentence was revised to delete the text, "between adults and children."

- *P. 47, Section 3.4.1.6 The authors should consider adding the significance of this value to risk assessment in that the vapor pressure defines the maximum air concentration the particular substance is able to obtain.*

Response: TCEQ agrees and has added the suggested phrase.

- *P. 49, l 11ff. This discussion here of various factors and causality seems to be misplaced (more appropriate in occupational exposures) and misguided especially in regard to laboratory animal studies. Dose-response seems to be discounted here as just one of many equal factors, such that if a dose-response relationship is seen but the MOA is not known a case could be made for discounting or dismissal of the effect. This section should stress clearly the primacy of dose-response (which incidentally this reviewer considers more concrete and substantive than Hill's "biological gradient") which is typically the purpose of the study and always the point of this assessment.*

Response: The text has been revised to include more emphasis on dose-response.

P. 49, Section 3.4.3.4. This section and those immediately following need to be considered more closely by an epidemiologist as it is confused and erroneous in the manner in which "epidemiological" is discussed. The two principal observational epidemological study types are case-control and cohort of which occupational studies being those where the cohort is certain workers.

Response: TCEQ agrees and this section has been rewritten to address the concern of the reviewer. Additionally, a reference to Adami *et al.* 2011, describing the use of epidemiology studies in the weight of evidence framework, has been added to this section.

- *P. 49, Section 3.4.3.7. Define here WOE (weight of evidence).*

Response: WOE has been defined earlier in the document.

- *P. 52, l 5. "proceed … with the risk assessment" (vice characterization).*

Response: The Guidelines were not revised based on this comment. The legend for Figure 3-1 is a direct quote from Boobis *et al.* (2006), and is reproduced with permission from Informa Healthcare.

- *Figure 3-2 legend. Please add mode of action (MOA) to the legend.*

Response: The Guidelines were not revised based on this comment. Mode of Action (MOA) is described in the previous figure (Figure 3-1).

- *P. 54. l 7-14. Please emphasize your dichotomy of external-internal dose measures here by clearly designating the first measure as "external" or "applied" as opposed to the rest which are internal.*

Response: A sentence has been added to the beginning of the first paragraph following the referenced bulleted list for clarification of this issue.

- *P. 56, Figure 3-3. This reviewer considers this figure from the Lewis review as vague and unclear with respect to the what it attempts to clarify here (i.e., adversity). Therefore it is strongly recommended that Figure 2-1 from RfC Methodogy (P. 2-4; or any similar version of the "Schulte" diagrams) be substituted here to graphically demonstrate the overall process of producing an effect as a continuum from nonadverse to adverse with all of the difficulties of designating adversity that follow.*

Response: TCEQ agrees and has removed this figure and related statement from this section.

- *P. 56, l 16-18. Please revise or eliminate these lines. BMD plays no roll in determining adversity. The BMD and the BMDL are only the statistical solutions to the BMR (or CES as is ponted out below) which is input by the modeler as the level/amount/degree that the modeler considers as potentially adverse.*

Response: See response to Reviewer 2, second bullet, in 2.2.4 above.

- *P. 63, l 31. Please define BMDL here in terms of the BMD.*

Response: BMDL has been defined in Table 3-3.

- *P. 66, l 19. The models referred to here are those that refine internal dosmetry, i.e. dose to the target tissue) and can have nohing to do with defining the "pathogenic process".*

Response: TCEQ agrees and the related statement has been deleted.

- *P. 67, l9. Please add "chemical or agent partitioning " to flow and metabolism here.*

Response: The suggested revision has been made.

- *P. 67. Section 3.8.1. PBPK models have many uses and applications. Please modify this section to make it clear that the discussed use of PBPK models here is for interspecies (i.e. animal to human) dosimetric extrapolation.*

Response: The section has not been revised because it is only listed as an example of application of PBPK modeling and does not serve as a complete discussion on this topic.

- *P. 67, l 20. …delivery of dose, (add)* ***in particular for effects occuring systemically.***

Response: The suggested revision has been made.

- *P. 67, l 21. "… prime importanc(add)* ***in achieving equilibrium concentrations in the body.***

Response: The suggested revision has been made.

- *P. 70, l 5 path (not pass).*

Response: The suggested revision has been made.

- *P. 70, Section 3.9 - The following references are proposed to be added to this section.*
  - *Woodall, G.M., Gift, J.S., and Foureman, G.L.. 2009. Chapter 112, Empirical methods and default approaches in consideration of exposure duration in dose-response relationships, pp.2743-2757 In "General and Applied Toxicology, 3rd Edition, Vol 5, Wiley*
  - *Rhomberg, L.R. 2009. Uptake kinetics, specis differences, and the determination of equivalent combinations of air concentration and exposure duration for assessment of inhalation toxicity. Human and Ecological Risk Assessment 15, 1099-1145.*

Response: TCEQ has added the Woodall et al. (2009) reference, as suggested. However, at this time the Rhomberg (2009) reference has not been added. TCEQ would like to review this article in-depth. If, after a more in-depth review of the article, TCEQ decides to add this reference to the Guidelines, a White Paper will be written for incorporation.

- *P. 74 ff, Sections 3.10.1 & 3.10.2. This section for gases reflects accurately the procedures in the RfC methodology as they are used generally in practice. It is, however, not complete and probably needs to be so in light of this document's repeated reference to (overall) mass transfer coefficients (Kg) in various resiratory tract models (above Sections) where they are dismissed.*

Response: The Guidelines incorporate information contained in numerous other documents by reference and cannot serve as a complete discussion on this and other subjects. Readers are referred to the RfC Methodology and related documents (Hanna *et al.* 2001) for more detailed information. The referenced sections contain sufficient information for TCEQ to conduct chemical toxicity assessments.

- *The RfC Methodology gives clear indications to perform dosimetry in RT regions that is inclusive of fractional penetration (fp) which requires regional specific Kg values for calculation of gas uptake in these regions. And when these Kgs are not available, the RfC Methodology presents equations that still allow for consideraton of fractional penetration. For example, Equations 4-21 and 4-22 (page 4-49) in RfC Methods for dosimetry to the TB region considers the fraction absorbed in the ET region via the right hand portions of the equations.*
- *These complete equations are typically not used in practice, even within the EPA. However, Kgs for this purpose have been discussed (e.g., Hanna et al., 2001) and are now becoming available (e.g., Asgharian, B., Price, O. T., Schroeter, J. D., Kimbell, J. S., Jones, L., & Singal, M. (2011). Derivation of Mass Transfer Coefficients for Transient Uptake and Tissue Disposition of Soluble and Reactive Vapors in Lung Airways. Ann Biomed Eng. doi: 10.1007/s10439-011-0274-9; Madasu, S. (2007). Gas uptake in a three-generation model geometry during steady expiration: Comparison of*

axisymmetric and three-dimensional models. Inhal Toxicol, 19: 199-210.) Inclusion of these concepts also serve to demonstrate the completeness of the RfC Dosimetry models.

- *Therefore a brief paragraph or two added to this section acknowledging and generally explaining fp in gas dosimetry is recommended to make this section more complete and accurate.*
- *Also, this overall discussion also applies to the following particle dosimetry section (3.10.2) which does not explain or address the deposition fraction (DF). Explaining DF allows the reader and practitioner the realization that the particle dosimetry programs actually correct for the particle dose taken up by RT regions anterior to the affected or target region, much in the same manner that the regional specific Kg values are intended to correct for gas taken up by anterior RT regions. Therefore, it is recommended that brief explanation be made of DF (from the RfC methodology) and added to Section 3.10.2.*

Response: Clarifying text has been added to this section.

- *P. 82, Legend of Figure 3-12. Please add that the NOAEL in the figure is the POD. Remove (erase with "paint" function?) the "x MF" as the EPA no longer uses MF (modifying factors). UF = composite of uncertainty factors. The x – axis should read "air" concentration.*

Response: The Guidelines were not revised based on this comment. Figure 3-12 (Figure 3-10 in the revised draft) is taken from USEPA (1994) and the figure legend indicates TCEQ does not use a MF.

- *P. 88, l 25. Cite the source of this "grading" (RfC Methodology, Table 4-3). These values are actually an ordinal rank, not a grade and should be so noted. Also see Charge Question commentary of liabilities of numerical/categorical assignment of severity.*

Response: This section was revised to use the word "rank" instead of "grade" and to indicate the source of the ranking (i.e., Table 4-3 of USEPA 1994).

- *P. 90, l 24 ff. The reference of Bruckner (2000) that reflected the NAS findings regarding the FQPA, may be used here:*
  - *Bruckner, J. V. (2000). Differences in sensitivity of children and adults to chemical toxicity: The NAS panel report. Regul Toxicol Pharmacol, 31: 280-285.*

Response: The suggested reference has been added to this section.

- *P. 101, l 29 ff. Such a procedure is highly problematic for all the reasons already given. In addition, this procedure has the potential for highly misleading and nonrealistic results in the oral to inhalation direction if the volatility of the compound being extrapolated is not considered. For the factors used in this procedure (ignoring comparative absorption information which this reviewer has seen maybe once, for cadmium, in 25 years of assessment experience) the air concentration will always be about a factor of 3 of the oral dose. So an oral POD of 10 mg/kg would yield a value from extrapolation of around 30 mg/m3. If the vapor pressure of the compound, for example, would support a maximum air concentration of < 1 mg/m3 then the procedure would yield clearly irrational results. The same situation may well happen when utilizing oral slope factors in such an extrapolation scenario. It is therefore recommended that text be added that physical characteristics (especially vapor pressure) be considered in the feasibility of performing generic/default route-to-route extrapolation procedures, especially via the oral to inhalation path.*

Response: TCEQ will evaluate the relevant considerations given in the Guidelines prior to conducting route-to-route extrapolation. A clarifying sentence was added to the end of the last paragraph in this section.

- *P. 107, l3 ff. The statements made here regarding the LLNA appear to be oversimplified and not current. That the cytokine profiles listed here can be used diagnostically has been contradicted in the literature (e.g., Selgrade M., et al., Inconsistencies between cytokine profiles, antibody responses, and respiratory hyperresponsiveness following dermal exposure to isocyanates. Toxicol Sci. 2006 Nov; 94(1):108-17.) The WHO guidelines would perhaps be a better and more current source to employ as a citation here (http://www.who.int/entity/ipcs/methods/harmonization/areas/immunotoxicity_review/en/index.html).*

Response: TCEQ agrees and more contemporary references have been added to the document.

- *P. 107, Section 3.18. The policy consequences of this manner rounding toxicity values should at least be mentioned. Rounding up values for risk (e.g., oSF and IUR) and rounding up threshold values imply diametrically opposed attitudes for overall public health. A point should perhaps be made that these rounding procedures are based on arithmetic principals with policy implications being of minor importance especially in reference to other conservative procedures in the overall risk assessment process such as UF that may vary over much greater ranges.*

Response: Rounding of the final values to the nearest integer to obtain two significant figures is an arithmetically accepted method. The final values derived in the DSD are conservative values, and to ensure a uniform and transparent approach TCEQ rounds the final values to two significant figures.

- *P. 113. Excellent accounting of current C x T procedures.*

Response: No response required.

- *P. 114, Sections 4.2.1., 4.2.2. It would be beneficial to note that this basic procedure in use of exponents to duration adjustment is used both by the AEGL committee (NRC, 2001. Standing Operating Procedures for Developing Acute Exposure Guideline Levels for Hazardous Chemicals, National Academy Press, Washington, DC, pp. 92-110) and has been advocated by EPA scientists (Woodall, G.M., Gift, J.S., and Foureman, G.L.. 2009. Chapter 112, Empirical methods and default approaches in consideration of exposure duration in dose-response relationships, pp.2743-2757 In "General and Applied Toxicology, 3rd Edition, Vol 5, Wiley).*

Response: TCEQ acknowledges this supporting material but has chosen not to reference it.

- *P. 116, Section 4.2.4.2, l 2-8. The manner in which this paragraph is ended leads to vagueness and misleading thoughts about the facts stated. One interpretation that may more or less identified in the paragraph is that this high rate of malformation in humans remains to be explained or "linked" directly to environmental agents. Text needs to be added to aid the reader in avoiding this interpretation. A more realistic and important interpretation is for the reader to realize the range of sensitivity for developmental bioassays to detect effects that occur at a level of 3% in the human population.*

Response: Information on background rates of developmental effects was not necessary for the Guidelines and was removed.

- *P. 121, Figure 4.2. It seems correct and desirable to perform the dose-response analysis as near the actual recorded/designated exposures as possible, before any manipulation or adjustment has taken place as is done here. There seems to be no allowance for dosimetric adjustment in this diagram.*

Response: TCEQ agrees. Figure 4.2 has been modified as recommended through addition of a step to account for dosimetric adjustment.

- *P. 128, Section 4.6. Of all the places that MOA could play a role, it would seem to be especially relevant here in evaluating the validity of the decision to apply the default ESL of 2 µg/m^3. I am certain that the process would unfold differently for one compound with SAR characteristics of an irritant and another with SAR suggesting estrogenic activity. Please consider adding text to this section regarding the potential evaluative use of MOA.*

Response: TCEQ has determined that MOA considerations cannot be incorporated into this section.

- *P. 133, Section 5.1.3. That the ATSDR does not derive cancer factors is also often considered a deficency and may be mentoned here also.*

Response: The suggested sentence has been added in this section.

- *P. 134, l 9. It should also be mentioned here that the PPRTVs also derive subchronic noncancer values for both oral and inhalation routes of exposure. Also, the chronic PPRTV values derived are only those that IRIS does not have. PPTRVs are revised every 5 years.*

Response: The appropriate information has been added in this section.

- *P. 139, Section 5.3.2. The auhors are to be commended for incorporation of the most recent advances in this area.*

Response: No response required.

- *P. 147, Section 5.6, the N-L Ratio approach. This approach has the attributes of being simple to perform and being free of extra procedures that lead to a false sense of precision.*

Response: No response required.

- *P. 151, Section 5.7.3. The modified polynomial cancer model in EPA's BMD modeling suite also provides a slope factor based on the POD determined by the cancer model. It is this slope factor that may then be further adjusted by opinions entering from the Supplemental Guidance. From the explanation given here, BEIR IV seems to be an alternative process, complete with a life-table analysis, which would seem to automatically accommodate those life-stage-based sensitivity issues in the Supplemental Guidance . Could further explanation be added here to clarify for the reader that this is or is not the case?*

Response: Revised text was added to make this more clear in the relevant section.

- *P. 154, l33 ff. The example given here of vinyl acetate is relevant and excellent. However, text should be added here to indicate the need for professional judgment and firm criteria as to what constitutes a precursor effect. These effects are in themselves uncertain both qualitatively, (often*

being linked to the ultimate event by speculative MOA events), and quantitatively as few precursor effects develop into ultimate events.

Response: Revised text has been added for clarification.

Reviewer 15

A Final Word: The authors really should speak with a book publisher about using this guidance as the basis for assembling a textbook to train risk assessors. Only two main topics are lacking: (1) probabilistic risk assessment; and (2) some treatment of newer methods not yet in anyone's guidance, such as methods recommended in Science and Decisions (NRC, 2008) and approaches to include informatics and systems biology into risk assessment.

Response: No response required.

Appendix B: Public Comments

B.1 Public Comment 1

Elena Craft, PhD
Environmental Defense Fund

Executive Summary

Environmental Defense Fund and Air Alliance Houston submit these comments on several proposed revisions to the Effects Screening Levels (ESLs). The proposed revisions were developed by the Texas Commission on Environmental Quality (TCEQ) and Sielken & Associates Consulting, Inc. to ensure that the TCEQ is using the most up-to-date and scientifically-sound methods and that the agency stays on the leading edge of regulatory toxicology and risk assessment.

While we appreciate the effort on the part of the agency to further refine screening level guidelines, we feel that the proposed revisions are inadequate in achieving the goals for which they were intended.

In response to other issues raised within these revisions, we recommend the Agency adopt the following actions:

- *Enforce the ESLs as standards as opposed to screening level guidelines;*
- *Institute a notice and comment rulemaking on adoption of the Protocol for Notification and Work Group Functions for Evaluating Potential and Active Air Pollution Watch List (APWL) Areas1;*
- *Take immediate action in addressing hydrogen sulfide and other contaminants which have state health standards;*
- *Adopt more aggressive strategies to remediate risks from air pollutants and work to better understand and develop community level risk from a multi-pollutant standpoint as opposed to focusing solely on individual pollutants;*

In short, if the agency's goal is to stay on the leading edge of risk assessment, then the agency should develop risk screening levels commensurate with a more thorough examination of the factors that can result in increased cumulative risk from exposure to air toxics as a whole. These factors have been discussed extensively in the recent peer reviewed literature2,3,4,5 and include, but are not limited to:

[1] http://www.tceq.state.tx.us/assets/public/implementation/tox/apwl/protocol/draft.pdf

[2] Morello-Frosch R, Zuk M, Jerrett M, Shamasunder B, Kyle AD. (2011). Understanding the cumulative impacts of inequalities in environmental health: implications for policy. Health Aff (Millwood). 2011 May;30(5):879-87.

[3] Clougherty JE, Rossi CA, Lawrence J, Long MS, Diaz EA, Lim RH, et al. 2010. Chronic Social Stress and Susceptibility to Concentrated Ambient Fine Particles in Rats. Environ Health Perspect 118:769-775. doi:10.1289/ehp.0901631

[4] Searing DA, Rabinovitch N. (2011). Environmental pollution and lung effects in children. Curr Opin Pediatr. 2011 Jun;23(3):314-8. doi: 10.1097/MOP.0b013e3283461926.

- *Life stage*
- *Exposure to multiple pollutants*
- *Genetic polymorphisms*
- *Low socio-economic status*
- *Pre-existing diseases*
- *Race/ethnicity*
- *Obesity*
- *Stress*

While the agency asserts that the ESLs encapsulate the total risks from air pollution, we believe the agency falls short of providing the public health protections afforded under the Clean Air Act as well as the TCAA. Exceedance of chronic ESLs for a decade or longer6, as well as demonstration of increase in birth defects in some heavily polluted areas in the state that may be related to environmental exposures, for instance, suggest that the agency is not fulfilling its responsibility to meet the current health threats of air pollution.

Response: No additional discussion/text was added. The Guidelines consider lifestage, genetic polymorphisms (toxicokinetics), pre-existing disease (sensitive subpopulations), and other available and relevant data (intrahuman variability, known race/ethic differences in sensitivity, etc.) in deriving toxicity values. Discussion of the remainder of the aforementioned issues is beyond the scope and purpose of this document.

TCEQ's Statutory Obligations under the Texas Clean Air Act (TCAA)

The TCAA (Chapter 382 of the Texas Health and Safety Code (THSC)) specifically mandates that the TCEQ to conduct air permit reviews and protect public health from the harmful effects of air pollution. From personal communications with TCEQ7, the definition of "condition of air pollution," is as follows:

"The Texas Clean Air Act (TCAA) defines "air pollution" as "the presence in the atmosphere of one or more air contaminants or combination of air contaminants in such a concentration and of such duration that: a) are or may tend to be injurious to or to adversely affect human health or welfare, animal life, vegetation, or property; or b) interfere with the normal use or enjoyment of animal life, vegetation, or property." 382.003(3). The breadth of the definition allows for applicability to either an episodic event or the ambient quality of a regional area.

For example, the language in the definition of "air pollution" is identical to the definition of "nuisance" in TCEQ rules (101.4), and nuisances are generally considered transient in nature. Similarly, the term "air pollution" appears in the enforcement section 382.085(a) which provides: Except as authorized by a commission rule or order, a person may not cause, suffer, allow, or permit the emission of any air contaminant or the performance of any activity that causes or contributes to, or that will cause or

[5] Postma DS, Kerkhof M, Boezen HM, Koppelman GH. (2011). Asthma and chronic obstructive pulmonary disease: common genes, common environments? Am J Respir Crit Care Med. 2011 Jun 15;183(12):1588-94. Epub 2011 Feb 4.

[6] http://tceq.com/assets/public/implementation/tox/apwl/annual_report/2009.pdf

[7] Booker Harrison, TCEQ Office of Legal Services. Email communication August 17, 2009.

contribute to, air pollution. Finally, the term "air pollution" could also be used in a more general context to apply to a non-attainment area since the criteria have some relationship to the NAAQS.

The concept of "condition of air pollution," while not specifically defined in the TCAA has generally been used to refer to an episodic event such as an upset event that results in a shelter in place order, or a monitored exceedance of a NAAQS or ESL. The phrase "condition of air pollution" does appear in the TCAA, specifically in section 382.055(e) related to renewals ("...avoid a condition of air pollution..."), and Water Code section 5.514(a) related an order issued for air emergencies ("if the commission finds that a generalized condition of air pollution exists..."). The phrase also appears in chapter 118 rules related to control of air pollution episodes."

Given such broad authority afforded under the TCCA, it would appear that the TCEQ has considerable legislative authority to address air pollution in a more holistic and aggressive way.

Response: No additional discussion/text was added. The intent of the guidance is to document the scientific procedures employed in deriving toxicity factors. Discussion of the aforementioned issues is beyond the scope and purpose of this document.

General Comments of the Guidelines to Develop Inhalation and Oral Cancer and Non-Cancer Toxicity Factors

The margin of exposure approach on page 5, section 1.2.3 scenario does not detail under what instances TCEQ would use the margin of exposure. The commenters do not support the use of a margin of exposure approach without scientific evidence of the necessity of using the approach and the opportunity for public comment.

Response: No additional discussion/text was added. While MOE is a useful methodology for some chemical exposures (e.g., dioxins), the TCEQ generally prefers the conservativeness of the standard toxicity factor (RfD, RfC, SFo, URF) approach and anticipates rarely (if at all) using the MOE approach for the regulatory setting of environmental media comparison values (although helpful for putting risk into perspective when comparison value exceedances occur). As with any other Development Support Document (DSD), any DSD incorporating a MOE approach would justify the approach used and be posted for public comment.

The commenters feel that statements made on pages 8 and 9, sections (1.4.2 and 1.4.3) of the draft revisions are misleading. The comments make reference to the least protective end of EPA's acceptable level of risk and never mention the more protective end of the range. A more accurate assessment of TCEQ's risk management goal of 1 x 10-5 value is that it lies midway between EPA's acceptable level of risk; as such, it should not be described as overly conservative.

Response: TCEQ agrees and has revised the statement in this section accordingly.

The commenters do not support the exemption of air toxics from the ESL list as described on page 15 of the draft revisions without adequate opportunity for public comment.

Response: No additional discussion/text was added. TCEQ strongly encourages interested parties (e.g., industry trade associations, individual companies, environmental groups, academia) to submit technical information to aid in the categorization of substances as exempt or not. Substances may be added or

removed from the exempt list if scientific evidence or regulatory experience dictates a change in status. TCEQ will give the public an opportunity to comment whenever substances are added or removed from the exempt list.

Section 2.2.3.1 discusses setting the odor-based ESL: "Considering that most standardized methods as those listed in Section 2.2.2 7 are the compatible and preferred methods for measuring odor, it would be more reasonable to use the geometric mean of odor thresholds rather than selecting the lowest one." It is unclear why selection of the geometric mean would be the default in setting odor-based ESLs instead of using the lowest odor threshold, especially when the agency states in the same section that "the TCEQ may set the acuteESLodor values at the lowest acceptable reported odor thresholds rather than the geometric mean if available data indicate that TCEQ air mobile monitoring staff members and/or field investigators have reported odors at measured levels at or lower than the geometric mean. Examples of such odorous air contaminants include, but are not limited to, styrene, alkyl amines, reduced sulfur compounds, or other sharp/pungent odorous compounds." The commenters recommend that the lowest odor threshold be the default in setting odor-based ESLs.

Response: TCEQ derives the $^{acute}ESL_{odor}$ value based on a system that allows comparison of values collected by methodologies of similar quality and sensitivity. Selection of the lowest available Level 1 or 2 odor detection threshold as the $^{acute}ESL_{odor}$ is inappropriate because such an approach would entail disregarding data of equal validity. By taking the geometric mean of various odor detection thresholds, the $^{acute}ESL_{odor}$ is the most representative value based on the collective data concerning odor detection thresholds in human subjects. For Level 3 odor detection thresholds, the lowest odor detection threshold is used as the basis of the ESL_{odor} because the quality of these data is considered to be lower. Importantly, this approach is conservative. In the peer-review process for this document, the reviewer assigned to evaluate the odor regulations section of the TCEQ guidance document expressed concern that use of a 50% odor detection threshold was an "extremely conservative overestimate of odor potency." The reviewer went on to describe the 50% odor detection threshold as a value where "odor is just on the cusp of detectability." The TCEQ chooses to use the 50% odor detection threshold as the conservative basis of the $^{acute}ESL_{odor}$ because most people are exposed to chemicals as mixtures rather than single chemical emissions. Since a number of studies indicate that chemicals with similar structure may act additively or synergistically in regard to odor perception, the TCEQ uses the odor detection threshold as the basis of the $^{acute}ESL_{odor}$ to be as conservative as is scientifically defensible in its approach to odor regulations.

In Chapter 3 of the ESL revisions, the commenters believe that the TCEQ should expand the federal and state guidance documents to include risk values established in other countries. At a minimum, TCEQ should examine risk assessment factor values developed by agencies such as Health Canada[8], and the National Institute of Public Health and the Environment[9] in the European Union.

Response: No additional discussion/text was added. References to guidance produced by international organizations as well as a brief acknowledgement and reference to organizations from other countries

[8] http://www.hc-sc.gc.ca/ewh-semt/pubs/contamsite/part-partie_ii/trvs-vtr-eng.php
[9] http://rivm.openrepository.com/rivm/bitstream/10029/9662/1/711701025.pdf

that conduct toxicity assessments have been included in the updates to the Guidelines (i.e., Health Canada, and the European Centre for Ecotoxicology and Toxicology of Chemicals), although no attempt has been made to produce an exhaustive list.

While TCEQ touts the benefits o f the air monitoring network system in the state in section 1.1.1 of these revisions, the commenters believe that additional information regarding the system also be shared. Specifically, we believe that the shortcomings of the monitoring system and occasional unreliability of the system should be mentioned as monitors have failed during critical times during air pollution events. For instance, the recent mass upset events at numerous facilities in Texas City caused by unexpected power outages exposed the paucity of the established air toxics monitoring system to accurately characterize community exposure.[10]

Response: No additional discussion/text was added. This document is primarily written as guidance for TCEQ staff, documenting the processes used to develop different toxicity values for any interested person with training in inhalation and oral toxicology and risk assessment. Evaluation of the advantages and potential shortcomings of the state-wide monitoring network is beyond the scope and purpose of this document.

EDF and Air Alliance Houston support the use of exposure response or reference arrays (as presented in Section 3.15) to aid in the development ESLs.

Response: No response required.

In reference to section 5.5.3, commenters support the use of uncertainty factor greater than 10 to account for child-adult differences in susceptibility. At this time, commenters do not support reducing UF_H values based on toxicodynamic differences in children as discussed in 5.5.1.

Response: No response is required for the first part of the comment. The UF_H will not be reduced consistently based on differences between children and adults, but will be evaluated on the data available for each chemical. This section has been clarified.

General Comments on the Hazard Characterization and Exposure Response Assessment Using Epidemiology Data

Commenters recommend a statement explaining how the Consulting firm of Sielken and Associates Consultants, Inc of Bryan, TX was chosen to develop the framework for the Hazard Characterization and Exposure Response Assessment. Given that the overwhelming majority of the firm's clients are companies those most directly impacted by the regulations that are inherent in meeting health protective standards set by the TCEQ, a public statement of disclosure regarding any conflict of or competing interest would have been warranted.

Response: In December, 2008, A Solicitation for Specialized Toxicological Analyses (582-9-80187) was issued for Advanced Mathematical Modeling and Statistical Analyses and Review of Epidemiological Analyses and Dose-Response Modeling. Standard TCEQ contracting and purchasing protocols were followed.

[10] http://www.cnbc.com/id/42760237/Power_Cuts_Hit_BP_Marathon_Valero_Texas_Plants

Several proposals were received and independently reviewed. Sielken & Associates Consultants, Inc (the contractor) was awarded the umbrella contract. When TCEQ requires the Contractor's technical expertise, a work order is issued to the Contractor detailing the scope of work, background, and other necessary information. A work plan outlining the work, deadlines, and deliverables is submitted to the TCEQ by the Contractor. Each work plan contains a Conflict of Interest Certification. For Work Order 582-9-80187-05 (Preparation of an Epidemiology Section including Dose-Response Modeling to be Included in the Cancer Section of the updated ESL guidelines Participation in a Letter Peer Review of the Epidemiology Cancer Section), the Contractor certified "that there is no conflict of interest between key personnel and the services required by the TCEQ as outlined in the work order. I understand that I am under a continuing obligation to provide this information."

It is difficult in a field as specialized as epidemiological statistical analyses and modeling to avoid an appearance of conflict of interest because there are a small number of experts in this field. These experts have diverse clientele. Additionally, these statistical methods are essentially standard and common in the field. Chapter 7 was prepared by the Contractor but was thoroughly reviewed by TCEQ before it was peer-reviewed by other scientific experts. Chapter 7 will be updated based on peer-review comments, then the revised Guidelines as well as the response to comments will be posted for an additional public comment period. The peer-review process has been transparent. We are early in the review process and the public is invited to review the Guidelines to make sure the process is objective and to determine if the report was influenced by an actual conflict of interest.

Commenters have concern over statements in section 7.9 Excess Risk Calculations for the General Public. Systematic adoption or incorporation of an inference population, such as the general population of Texas for instance, without careful scrutiny, may mask health impacts on that inference population if care is not taken to avoid such a "type-1" error. Additionally, whittling down the n in a study population under guise of selection of an inference population will necessarily result in a study with less statistical power, possibly resulting in masking a true effect.

Response: TCEQ focuses the inference population to be the Texas population (rather than the US population) because TCEQ's objective is to protect the population of Texas.

The inference population does not have any effect on the "type-1" error or the statistical power of the model to detect a response. The inference population is used in the calculation of excess risks but is not used in the exposure-response model estimation based on the epidemiology data.

While recognizing that risk assessment development is a dynamic process, commenters are concerned that the meta-analysis discussed in section 7.11, and the reality check discussion in section 7.12 suggest that "cherry-picking" of data is acceptable when analyzing multiple studies. Commenters would like to see a scientific validation system in determining criteria that would warrant of specific exclusion of studies from a risk analysis – "reality check" or other vague descriptions. For instance, in section 7.11.2: "Reproducibility of results. Epidemiology studies selected for inclusion in a meta-analysis should include enough information to corroborate or reproduce the results used for the meta-analysis. Studies that only include summary data without enough data to support the reported results should be seriously considered for exclusion from the meta-analysis," commenters are uncertain whether a new study whose findings have not been replicated would be automatic grounds for exclusion from a risk analysis.

Response: Meta analyses are designed to combine results rather than to exclude results. In order to combine studies, those studies have to meet some a priori qualifications in order for the meta analysis to be scientifically defensible.

The Guidelines state in Section 7.11.2 that "Although different criteria for selection of individual studies should be tailored to the problem at hand, the criteria should follow some pre-established guidelines to be valid. The criteria should be developed a priori, before a meta-analysis is started, to avoid selection bias." These criteria, being selected a priori, are precisely designed to avoid "cherry picking" acceptable data when analyzing multiple studies.

The other concern of reproducibility of results refers to the study having sufficient data published "to corroborate or reproduce the results" presented in the study report/article/paper. The "reproducibility" here refers to the ability to reproduce calculations and not studies. The reproducibility in the context of the Guidelines for meta analysis does not refer to collecting data from other epidemiology studies and obtaining similar results. With respect to being able to reproduce calculations, studies that report only, for example, a risk of 0.01 should be questioned if there is no other information indicating where the 0.01 came from (how it was calculated) and what data are there to support such a risk estimate. We have changed the word "results" to "calculations" in the chapter text corresponding to the quote.

Conclusion

EDF and Air Alliance Houston also believe that, given the scientific nature of the ESL program and the complexity of setting individual guidelines for such a vast array of chemicals, there is a natural barrier to full and meaningful public participation in a process which has enormous implications for public and community health. We recommend that the TCEQ institute a formal and fully independent scientific review board consisting of stakeholder representatives from all sectors but predominated by independent academics and scientists with expertise in this field.

Thank you for the opportunity to comment.

Response: TCEQ believes the creation of such a scientific review board is unnecessary for several reasons:

- The revised Guidelines underwent an independent scientific peer review organized by Toxicology Excellence for Risk Assessment (TERA), an organization internationally-renowned for conducting peer reviews. TERA assembled a panel of independent distinguished experts from around the country to conduct an external scientific peer review of the Guideline revisions to make sure that the TCEQ incorporated the latest, most up-to-date scientific methods and models and to ensure that the methods are scientifically defensible. The TCEQ's methods are of the highest scientific quality and have been well received in the scientific and regulatory communities.
- The TCEQ does not have the resources to fund or institute a formal and fully independent scientific review board consisting of stakeholder representatives from all sectors but predominated by independent academics and scientists with expertise as suggested.

- The TCEQ seeks to ensure a transparent process and welcomes public comments on the technical aspects of DSDs, which are naturally commensurate with the degree of technical and scientific expertise. A public comment period which allows ample opportunity for input from the public is incorporated into the process and anyone wishing to provide comments is invited to do so. We invite participation in the TCEQ Toxicology Announcement Listserv. To join, send a blank email to: join-tox@listserv.tceq.texas.gov.

B.2 Public Comment 2

Laurie Haws, PhD, DABT
ToxStrategies, Inc.

The Guidelines are timely and well organized, are accompanied by robust, comprehensive discussion, and rely on peer-reviewed, scientific approaches and information. The Guidelines are broad in scope, as well as provide a detailed discussion on each topic. The information presented represents both standard and state-of-the-science approaches in toxicology and risk assessment. The Guidelines also contain detailed evaluations of issues of particular importance to toxicity factor development, including differences in responses to toxicant exposures between adults and children, threshold vs. nonthreshold mode of action, benchmark dose and physiologically-based pharmacokinetic modeling, and the latest advancements in determining odor thresholds.

Overall, ToxStrategies is very supportive of the Guidelines. In an effort to further enhance the scientific robustness of the document, the following comments and suggestions are offered, along with more detailed discussions on selected topics:

General Comments re: Clarification and Simplification of the Document

The title of the Guidelines suggests that the content is limited to the development of toxicity factors, when in fact the Guidelines cover both the development of not only toxicity factors (e.g., RfDs, SFos, etc.) but also environmental media evaluation levels for air (ESLs and AMCVs). This should be reflected in the title to make it clear to the reader.

Response: As the primary purpose of the Guidelines is to discuss the methods used to derive toxicity factors and not to address the development of evaluation levels for all relevant media (e.g., soil, groundwater), TCEQ has revised the document title to: TCEQ Guidelines to Develop Toxicity Factors.

The term "ReV" should be eliminated and replaced with the term RfC. The scientific community understands what an RfC is, therefore there is no need to invent a new term to describe what is essentially an RfC. The term "ReV" is not explicitly defined in the TRRP rule but rather was developed to distinguish between those ESLs that were historically used in air permitting and those that were explicitly developed solely for use in TRRP, in cases when published RfC values were not available for other sources in the hierarchy of toxicity values specified in TRRP.

Response: The TCEQ methodology is somewhat divergent from the EPA RfC methodology. Therefore, TCEQ chose to use the term "ReV" as a unique identifier to distinguish a TCEQ-derived value from other regulatory values derived using other methods.

We also recommend that the document be broken up into the following two distinct sections: 1) the first section describing the development of the actual toxicity values themselves (subdivided in turn into a section for air toxicity values – URFs and RfCs, and a section for oral toxicity values – SFos and RfDs) and, 2) the second section describing development of the air comparison values. Additionally, it should be made clear in both the air and oral toxicity values subsections that if a carcinogen is determined to act through a threshold mode of action, then a cancer-based RfD or cancer-based RfC will be developed rather than a SFo or URF, respectively. As currently written it is very difficult to read and understand

Chapter 1 of the document with the back and forth discussions between the two primary categories of values (i.e., toxicity values vs. air comparison values) and the different terms to describe the different values.

Response: The Guidelines will not be revised based on this comment. The Guidelines are primarily written for TCEQ staff, who prefer to keep the document layout similar to the original version for ease of reference.

Further, as just alluded to, as written, the Guidelines appear to be more complicated than necessary and could be greatly simplified by eliminating some of redundant and confusing terms currently being used to describe the various "toxicity values" (e.g., ReVs, AMCVs, RfDs, ESLs, URFs, SFos, etc.) and by using standard terminology to describe the toxicity factors (i.e., SFos and IURs for non-threshold-based carcinogens and RfDs and RfCs for non-carcinogens and carcinogens determined to act through a threshold MOA). The terms ESLs and AMCVs could then be employed to describe the air comparison values used to assess air permit application and air monitoring data, respectively.

Response: TCEQ has chosen to keep terminology consistent with its historical usage, as alteration of current terminology may give rise to additional confusion.

Finally, the terms "threshold" and "non-threshold" should only be used in the context of describing the development of the actual toxicity values themselves.

Response: TCEQ is uncertain about what specific contexts the reviewer finds usages of these terms to be objectionable.

Clarification of direct dermal effects: Chapter 1 (p.1, lines 23 and 24)

The text states that direct effects are those that result from direct inhalation and dermal exposures to chemicals in air. This section is the only instance in the Guidelines that discusses dermal exposure as a direct pathway along with inhalation. Sensitization is discussed in Section 3.17, though only respiratory sensitizers are mentioned. Since dermal exposures are not discussed elsewhere in the Guidelines with regard to direct effects, we recommend that the reference to direct dermal effects be omitted.

Response: Direct dermal effects should be mentioned because they are potential critical effects. However, as this is considered unlikely, TCEQ has not developed procedures for derivation of dermal-based air toxicity factors for incorporation into the Guidelines. If needed, these procedures will be developed and incorporated into the document at a later time.

Removal of ReV adjustment: Chapter 1

The Guidelines discuss cumulative and aggregate exposures and indicate that the Agency uses a HQ of 0.3 to calculate ESLs for chemicals with a threshold dose-response assessment (p. 8, lines 8 through 21). In contrast, ReVs are calculated using a HQ of 1. Further, the TCEQ uses a target cancer risk level of 1 x 10^{-5} to calculate ESLs for individual chemicals (e.g., carcinogens) with a nonthreshold dose-response assessment. However, on the previous page of the Guidelines (page 7), the TCEQ states that a HQ of 1 and a risk level of 1 x 10^{-5} correspond to no significant risk levels for both threshold and nonthreshold chemicals, respectively. As currently written, these two sections appear to be contradictory.

Response: TCEQ does not interpret these sections as contradictory. The difference observed by the reviewer is related to the fact that ESLs are adjusted for use in the air permit program to evaluate facilities often located in areas of dense industrial activity (i.e., there may be multiple sources of a chemical). Thus, the HQ of 0.3 is applied to account for this consideration and TCEQ does not interpret this as contradicting the policy that an HQ of 1 and target risk level of 1×10^{-5} are no significant risk levels.

Regarding cumulative and aggregate exposures, the Guidelines state that ESLs are not further adjusted for chemicals with nonthreshold dose-response (1×10^{-5} risk level is used) since few chemicals with a nonthreshold dose-response assessment are routinely permitted in Texas for a given facility, and this risk management goal is ten times lower than the 1×10^{-4} level, which is defined by USEPA as an acceptable level of risk. This statement presupposes that the person already knows whether each and every chemical for which he or she will be developing toxicity factors acts through a threshold or non-threshold mode of action (MOA) before they have even begun conducting the assessments. As such, we recommend that this text be deleted.

Response: To address this comment, the sentence has been revised as follows: "Further adjustment of this no significant excess risk level is not necessary since few chemicals with a known or assumed nonthreshold dose-response are routinely permitted in Texas for a given facility and the risk management goal of 1×10^{-5} is ten times lower than the 1×10^{-4} level, defined by USEPA as an acceptable level of risk (USEPA 2000e)."

We recommend that the 70% adjustment to the ReV to calculate the acute and chronic ESLs (adjustment decreasing the HQ from 1 to 0.3) be removed, as this adjustment was deemed unnecessary for nonthreshold AMCVs and ESLs as well as threshold AMCVs (i.e., ReVs).

Response: The Guidelines will not be revised based on this comment. AMCVs are used for review of air monitoring data. There are significant differences between values used for air permitting and air monitoring, so different adjustments are required as discussed in the Guidelines as well as in the fact sheet and AMCV document at http://www.tceq.texas.gov/toxicology/regmemo/AirMain.html#compare. Adjustments to nonthreshold AMCVs (and long-term ESLs) are not required. TCEQ uses an excess risk management goal of 1×10^{-5} to calculate nonthreshold AMCVs and long-term ESLs for individual chemicals (e.g., carcinogens) with a nonthreshold dose-response assessment. The risk management goal of 1×10^{-5} is ten times lower than the 1×10^{-4} level, defined by USEPA as an acceptable level of risk (USEPA 2000e).

Further, no explanation is provided as to the selection of 0.3 as the HQ for threshold-based ESLs. The current application of an adjustment to only the threshold ESL seems inconsistent and the value appears to be chosen somewhat arbitrarily. There is no scientific basis for applying a cumulative adjustment factor solely to chemicals that act through a threshold MOA. There is also no scientific basis for applying a cumulative adjustment factor to toxicity values used in permitting vs. in evaluation of air monitoring data.

Response: The selection of 0.3 as the HQ for threshold-based ESLs is a risk management decision. In addition, these risk management goals were approved by the Commissioners and Executive Director of

the TCEQ. The reason the adjustment is deemed necessary for chemicals with a threshold MOA for air permit reviews is stated in the Guidelines: "During the air permit review process, emissions of one chemical from one site are evaluated, not emissions from multiple sites or multiple chemicals (i.e., chemicals are evaluated on a chemical-by-chemical basis). In consideration of cumulative and aggregate exposure, the TCEQ uses an HQ of 0.3 to calculate acute and chronic health-based ESLs for chemicals with a threshold dose-response assessment".

Since not all chemicals act on the same target organ, not all chemical mixtures are additive or synergistic (some are antagonistic), and different chemicals have different pharmacokinetics, it is recommended that a HQ of 1 and risk level of 1 x 10^{-5} be used for both ESLs and AMCVs, as these levels are considered to be protective of cumulative and aggregate exposure and are in keeping with the approaches used by other regulatory agencies to develop conservative environmental media-specific screening values.

Response: TCEQ understands the idea implied by the reviewer. However, these risk management goals are policy-based and, as stated in the response to the previous comment, these risk management goals were approved by the Commissioners and Executive Director of the TCEQ.

Clarify data sufficiency criteria for development of toxicity values: Chapter 1 (p. 9-10)

On page 9, lines 28-29, the Guidelines state that ESLs are developed for all chemicals, even if they have limited toxicity information. However, on the next page (p. 10, lines 20 and 21) it is stated that if adequate data are not available and route-to-route extrapolation or a surrogate chemical approach is not defensible, a health-based chronic ReV or ESL will not be developed. Text presented on these two pages appears to be contradictory. It is recommended that the text be revised to clearly indicate whether ESLs are in fact developed for all chemicals.

Response: To address this comment, the sentence has been revised as follows: "Acute ESLs are developed for all chemicals, even if they have limited toxicity information, although data may not be available to develop chronic ESLs."

Modify Figure 1-2 to reflect text: Chapter 1

Figure 1-2 accompanies the discussion on developing long-term ESLs. However, the Figure would be more consistent with the discussion if it were modified to include the condition where a mutagenic MOA has been determined and the ESL becomes 6 x 10^{-6}/URF (p. 11, lines 15-18).

Response: The figure has been revised accordingly.

Clarification of review status: Chapter 1 (p. 21)

The Guidelines state, "this regulatory guidance document did undergo external scientific peer review and public comment" (lines 24 and 25). It is assumed that the TCEQ is referring to the original 2006 ESL development guidance (RG-442). This should be clarified and a citation added.

Response: The referenced statement, "this regulatory guidance document did undergo external scientific peer review and public comment", refers to the Guidelines currently under review. When the Guidelines are finalized the referenced statement will be applicable.

Distinguish approach for developing odor thresholds: Chapter 2

The discussion regarding the TCEQ's current approach for setting the $^{acute}ESL_{odor}$ for Level 1 and 2 odor thresholds is somewhat confusing. The first paragraph states that the agency sets the $^{acute}ESL_{odor}$ at the lowest acceptable odor threshold (p. 27 lines 2 through 17). The remaining text in the paragraph then discusses the merits of the NAC/AEGL Committee's approach of using the geometric mean of all Level 1 or 2 odor thresholds, instead of the lowest value. Then the following paragraph implies that the TCEQ only sets the $^{acute}ESL_{odor}$ at the lowest acceptable reported odor threshold, instead of the geometric mean, if certain criteria exist. It is recommended that the language in these two paragraphs be clarified with respect to the approach(s) used in setting the odor-based ESLs.

Response: This section will not be revised based on this comment. The second paragraph of this section indicates the exception for not using the geometric mean for chemicals that have been identified with odor nuisance incidence in the field. For such odorous chemicals, in order to protect against odor potential, TCEQ uses the lowest odor level as the odor-ESL.

Consider the 2011 EPA Guidelines on Extrapolation: Chapter 3

In Section 3.1 of the TCEQ Guidelines, the Agency cites the 2007 USEPA draft guidance document titled Framework for Determining a Mutagenic Mode of Action for Carcinogenesis. USEPA has since issued new draft guidelines that address the issue of extrapolation (Guidance for Applying Quantitative Data to Develop Data-Derived Extrapolation Factors for Interspecies and Intraspecies Extrapolation, 2011). It may be worthwhile to consider whether this new EPA guidance document would be beneficial in providing support for approaches used to develop toxicity factors.

Response: See response to Reviewer 11 (comment on USEPA (2011a)) in 3.0 above.

Use the terms "susceptible" and "sensitivity" consistently: Chapter 3

In Section 3.2, the terminology for susceptibility and sensitivity appear to be used inconsistently. The following definition provided in the Guidelines implies that sensitivity involves susceptibility and differences in exposure:

"The TCEQ defines susceptible as a capacity characterized by biological factors that can modify the effect of a specific exposure, leading to a higher health risk at a given exposure level (Hines et al. 2010). The word sensitive describes the capacity for higher risk due to combined effect of susceptibility and differences in exposure (Hines et al. 9 2010)."

However, in the next sentence, the TCEQ appears to be using the term "susceptibility" in a different context (p.38, lines 11-12):

"A number of physiologic and metabolic factors, toxicodynamics, and diet and behavior patterns influence the increased or decreased susceptibility of children (Snodgrass 1992)."

It is the behavior patterns of children that lead to their differences in exposure relative to adults, thus it would seem that sensitivity should be used in lines 11-12. It is recommended that the Guidelines be carefully reviewed to ensure these terms are used consistently.

Response: For children, exposure via inhalation is expected to be higher on a body weight basis not just due to behavior such as outdoor play or exercise, which adults also participate in, but also (more

importantly) biological/physiological requirements (e.g., infants and children have a higher respiratory rate per unit body weight). Biological/physiological factors are included in both definitions of susceptibility. Nevertheless, in an attempt to promote clarity regarding the relevant considerations, the definition sentences used by the TCEQ were revised and other language added to this section.

Consider revisiting selection and/or definition of uncertainty factors: Chapter 3

First, it is recommended that Sections 3.12.1.1 and Sections 3.12.1.2 include discussion of data-derived extrapolation factors (DDEFs), consistent with new EPA guidance discussed in a previous comment.

Response: See response to Reviewer 11 (comment on USEPA (2011a)) in 3.0 above.

Second, clarification regarding the application and definition of the UF_H would be beneficial. Parts of Section 3.12.2.1 are not consistent in this regard:

"Factors of 10 have been commonly applied by default to account for interspecies and intraspecies sources of variability. The factor of 10 is considered to protect the majority of the human population including children and the elderly. Renwick (1993) proposed that…"

"Based on scientific data and an evaluation conducted on a chemical-by-chemical basis, the UF_H may need to be greater than 10 in order to adequately protect children."

The statements support the use of DDEFs if data suggest that children are likely more susceptible, rather than increasing the UF_H.

Response: This section has been revised in response to this comment and other comments received.

Third, many potential values for UFs are offered, though the implementation of such may be cumbersome. In Section 3.12.2.1, the following values are recommended for UF_L: 1, 2, 3, 6, and 10. By allowing for five potential values to describe this UF, it conveys a false sense of accuracy in the UF_L. Further, five values add a level of ambiguity that external (i.e., non-TCEQ) risk assessors would be unable to reproduce without additional justification and/or clarification on implementing such. Thus, it is suggested that the TCEQ use values of 1, 3, or 10 for UFs, and justification is provided for the selection of such.

Response: The Guidelines will not be revised based on this comment. As indicated in Chapter 3, these default values are based on the criteria for determining the severity of adverse effects and on Table B-1 in Appendix B and may be replaced by more specific values where appropriate data are available (OEHHA 2008). In addition, the Guidelines are written for TCEQ toxicologist risk assessors who will provide justification for the value of UF_L. Toxicity assessments often differ between agencies based on scientific judgment, and non-TCEQ risk assessors will have to provide justification for their own UF values and assessments based on their own guidance.

Fourth, in Section 3.13, the discussion related to capping the composite UF at 3000 would be improved with additional rationale and justification. The rationale for such a cap in IRIS is that USEPA does not conduct assessments for chemicals that have insufficient data. UFs used in PPTRVs are in fact higher than 3000. If TCEQ has a scenario in which UFs > 3000 were applied, then the Agency should acknowledge such and consider the values as provisional.

Response: This section will not be revised based on this comment. When default factors of 10 are used to account for each area of uncertainty (e.g., interspecies, intraspecies, LOAEL to NOAEL, incomplete database, subchronic to chronic), the product of the UFs can be as high as 10,000-100,000, which may result in an overly conservative toxicity value (Swartout et al. 1998). The justification for capping the composite UF at 3000 is given in this section. There is no need to designate certain values as provisional based on the total UF value as it would serve no real use and regardless, toxicity factors can be updated as new data, which reduce uncertainty, become available.

Clarify study duration definitions as they relate to UFs: Chapter 3

In Section 3.3, it is stated that chronic studies are defined as those that are longer than 3 months, but typically 2 years. However, more specific guidance regarding the relationship between study duration and the application of uncertainty factors would be useful. For example, would one apply a subchronic-to-chronic uncertainty factor if a study is 6 to 12 months?

Response: According to the USEPA IRIS, repeated exposure by the oral, dermal, or inhalation route for more than approximately 10% of the life span in humans (more than approximately 90 days to 2 years in typically used laboratory animal species) is considered chronic exposure (the IRIS definition for chronic exposure has been added to this section). Thus, there is no need to apply a UF_{sub} for animal exposure of 6 to 12 months.

Consider revising approach to account for uncertainty prior to study selection: Chapter 3

In Section 3.11, it is stated that the critical effect (POD_{HED}) should be selected prior to extrapolation. However, each study has uncertainties inherent to the study design as well as differences in extrapolation based on the underlying data. By basing selection on the POD_{HED} instead of the RfD, these aspects of the studies are ignored during selection of the critical effect. It is typical EPA practice to base selection on an array of RfD values. This approach should be considered by the TCEQ – especially as data-derived extrapolation factors (DDEFs, USEPA 2011) may be applicable to some endpoints and not others.

Response: Chapter 3 indicates that the critical effect is selected based on human equivalent exposure after considering each key study. Thus, as implied by the first line of the comment, the comment is only relevant to whether the critical effect should be selected before or after the application of UFs. TCEQ agrees with all members of the external expert peer review panel, who unanimously indicated that the critical effect should be selected before uncertainty factors are applied. Otherwise, selection of the critical effect after UF application would most always result in the effect about which there is the most overall uncertainty, instead of being driven by dose-response data (and any appropriate animal-to-human dosimetric adjustments).

Clarify appropriate application of approaches to identify toxicity: Chapter 4 (p. 129-130)

The N-L ratio approach is discussed as an approach for developing generic ESLs in the Guidelines. This approach appears to be the preferred approach, based on text stating that a study by Phillips et al. (2011) demonstrated that the N-L ratio approach was more predictive of toxicity, whereas the TOC approach was overly conservative (p. 129 lines 10-13) (and therefore the N-L ratio was the only approach discussed). However, on the next page (p. 130) and in Figure 4-4, the TOC approach is discussed as part

of the criteria for selection of acute lethality data. Please clarify this text regarding use of the TOC approach in developing generic ESLs.

Response: See response to Reviewer 10 in 2.4.5 above.

We commend TCEQ on the approach for applying bodyweight scaling: Chapter 5

In Section 5.3.2, it is noted that TCEQ will apply a UF_A value of 1 when $BW^{3/4}$ scaling is used. The standard EPA approach is to use a value of 3 to account for potential differences in toxicodynamics. However, since empirical studies cited by the TCEQ authors note that $BW^{3/4}$ scaling predicts toxicity, which is the result of both pharmacokinetic and pharmacodynamic processes, we agree with TCEQ that the UF_A is therefore unnecessary. Moreover, USEPA uses $BW^{3/4}$ for dose adjustments in cancer risk assessment without application of additional uncertainty factors for pharmacodynamics. Thus, we commend TCEQ for implementing a risk assessment approach that reflects the state-of-the-science, as documented in the peer-reviewed literature, on this topic.

Response: No response required.

Provide additional guidance on application of study-specific bodyweights: Chapter 5

Toward the end of Section 5.3.2, it is implied that study-specific bodyweight should be used in the dosimetric adjustment factor (DAF), yet a table of default species bodyweights and DAFs from USEPA (2002a) is provided. Additional guidance is needed on this topic. It would be particularly useful if the TCEQ would address two specific issues associated with the use of study-specify bodyweight. First, the human bodyweight (70 kg) in the $BW^{3/4}$ DAF calculation is simply a representative bodyweight. Thus, rationale for applying a representative bodyweight for humans, but not for the species used in the toxicity study, would be helpful. Second, if the POD is based on a BMDL, it is unclear which bodyweight would be selected for the study-specific bodyweight (e.g., the average bodyweight from control group, highest treatment group, or group closes to the BMDL, etc.). It would seem that default bodyweight values, or DAF values reported in Table 5-1, are sufficient for this extrapolation.

Response: The rationale for applying a representative bodyweight for humans but not for the species used in the toxicity study is that study-specific information is always preferable to default values and often available for animal studies, while a representative bodyweight for humans must be used for the extrapolation as there is no other option. While default animal bodyweights are sufficient, study-specific data are preferable. Study-specific bodyweights will be selected based on relevant scientific considerations for the particular study (e.g., POD dose group, dose-group specific bodyweights). No changes were made.

Recognition of threshold-based MOA

We commend TCEQ for recognizing that different approaches should be implemented for compounds that have demonstrated a threshold-based response. Examples are as follows:

*"**Threshold**: The dose or exposure below which no deleterious effect is expected to occur. In addition to noncancer effects, this may also apply to cancer effects for some chemicals (e.g., formaldehyde-induced respiratory tract cancers, dioxin)."*

"However, some carcinogenic effects, such as formaldehyde-induced respiratory tract cancers (TCEQ 2008) and possibly cancers from other chemicals (e.g., dioxins), are understood to exhibit a threshold dose-response."

Response: No response required.

Ensure the Guidelines are focused on science and not policy: Chapter 6

Chapter 6 outlines the TCEQ's approach for assessment of chemical groups and mixtures and includes a discussion of polyhalogenated aromatic hydrocarbons, dioxin-like compounds, and product formulations. Much of the text in these sections reflects policy decisions applicable to TRRP. Given that the Guidelines represent a scientific document focused on providing a general approach for the development of toxicity values and air comparison vales, we recommend that Sections 6.2 and 6.3 in particular be deleted. The discussions in these two sections would be more well-suited for a DSD on each of these specific classes of compounds. Further, discussions regarding the development of cleanup levels should be reserved for TRRP rather than included in a document providing general guidelines on the development of toxicity values. With respect to the dioxin-like compounds in particular, given the strong position that the TCEQ appears to be taking with respect to MOA, dioxin-like chemicals and formaldehyde should be among the first chemicals considered for development of threshold-based toxicity values that are protective of cancer.

Response: Although some values developed under the Guidelines are used in TRRP (e.g., chronic ReVs, URFs), to limit unnecessary references to TRRP approaches and increase focus on deriving toxicity values using Guideline methodology, TRRP-oriented language was deleted from these sections.

Additional discussion on study findings: Chapter 7

Text in Section 7.9.4 is unclear with respect to the findings of Sielken and Valdez-Flores (2009a). The discussion on this topic would be improved if the findings from this study were clarified.

Response: See response to Reviewer 2 in 2.3.7, P. 33, l 36-37 comment, above.

Editorial comments:

- *p. 19, Table 1-5: It seems the Short-Term Health box should contain text specifying "the lowest value of" the available values, as do the other types of ESLs and AMCVs.*

Response: Table 1-5 will not be revised based on this comment. There are significant differences between health effects reviews for air permit applications (which use ESLs) and ambient air monitoring data (which use AMCVs). The term "AMCV" is a collective term that refers to all odor-, vegetation-, and health-based values used in reviewing ambient air monitoring data.

To assess the potential for measured concentrations of specific chemicals to cause health effects, health-based AMCVs are based on ReVs and/or URFs (for carcinogens). To assess the potential for measured concentrations of specific chemicals to cause welfare effects (i.e., nuisance odors, vegetation effects), AMCVs are based on odor- or vegetation-based ESLs. Both health and welfare values are used in the review of ambient air monitoring data.

Prior to the peer-reviewed 2006 Guidelines, ESLs were developed using occupational values divided by safety factors to obtain a short- and long-term ESL. These ESLs, which have not undergone the Guidelines process, are termed interim ESLs. For the health-based AMCV, if the Guidelines have not yet

been used to develop a health-based ReV, the interim ESL is the AMCV. If the peer-reviewed 2006 or updated/current Guidelines have been used to develop a health-based ReV, the health-based ReV is the AMCV, unless a ReV could not be developed, in which case the generic ESL is the AMCV. Please refer to the Uses of ESLs and AMCVs Document, located at http://www.tceq.texas.gov/toxicology/AirToxics.html for a more detailed explanation of the differences.

- *p. 29, Figure 2-2: The labels for boxes C and D (alkenes and aldehydes) are reversed.*

Response: TCEQ has revised the labels for boxes C and D accordingly.

- *p. 29, line 13: In this instance, the word "toxins" should be changed to "toxicant" or another appropriate term.*

Response: It appears this comment is actually in reference to p. 39, line 13 rather than p. 29, line 13. See response to Reviewer 4 in 2.1.2.1 Part A above.

- *p. 72, line 21: "Values" should be changed to "value".*

Response: This section was deleted from the Guidelines.

- *p. 125, line 22: "NOEAL" should be changed to "NOAEL".*

Response: TCEQ has revised the text accordingly.

- *p. 159, line 27: "carcinogenic" should be changed to "carcinogen".*

Response: TCEQ has revised the text accordingly.

- *Throughout the document, there are references to U.S. EPA (2006) that should be specified as 2006a or 2006b.*

Response: TCEQ has revised the text accordingly.

- *Anderson (2002) is not listed in the References.*

Response: TCEQ has added this reference to the Guidelines.

- *In the References, Anderson et al. (2005) should be Andersen et al. (2005).*

Response: TCEQ has revised the text accordingly.

B.3 Public Comment 3

Christina Wisdom
Texas Chemical Council

On behalf of the Texas Chemical Council (TCC), thank you for the opportunity to submit comments on the Texas Commission on Environmental Quality (TCEQ) draft document, Revisions to Guidelines to Develop Effect Screening Levels, Reference Values, and Unit Risk Factors (RG-442), which is dated June 2011.

TCC is a statewide trade association representing over 70 chemical manufacturers with more than 200 Texas facilities. The Texas chemical industry has invested more than $50 billion in physical assets in the state, pays over $1 billion annually in state and local taxes and over $20 billion in federal income taxes. TCC's members provide approximately 70,000 direct jobs and over 400,000 indirect jobs to Texans across the state. TCC member companies manufacture products that improve the quality of life for all Americans and millions of people around the world.

TCC appreciates the opportunity to comment on the development of inhalation and oral toxicity factors. TCC understands the importance of ESLs in providing the TCEQ with guidance to protect human health and welfare regarding its authority for air permitting and air monitoring. Air quality is also important to the regulated community, particularly to members of TCC. TCC also understands the importance of oral exposure factors to help TCEQ guidance for their remediation guidance efforts. The regulated community appreciates consistent and scientifically robust methods for the development of these toxicity factors.

In general, TCC views the draft document favorably. TCC supports the thorough scientific process by which chemicals will be evaluated for inhalation (ESLs) and oral exposure values. A few comments on the document are listed below. In addition, for reference, we have also attached a copy of TCC comments to the prior TCEQ ESL guidance draft document (comments dated June 2005).

Response: The comments dated June 2005 regarding the 2006 Guidelines were previously addressed. Responses to those comments can be found at this location http://www.tceq.texas.gov/assets/public/implementation/tox/response/2006 prp.pdf.

TCC Comments on the Guidelines to Develop Inhalation and Oral Cancer and Non-Cancer Toxicity Factors Effects Screening Level (ESL) Methodology draft document (draft dated June 2011):

- *On page 33 regarding odor-based ESLs, the impact of the use of correlations with respect to chemical groups is not clear. The use of professional judgment is indicated, but it is with respect to the use of correlations when there is limited data. What is not clear is what is meant by 'limited' data, as opposed to no data. TCC requests clarification as to whether even limited data should have preference over correlations that in some cases are weak.*

Response: TCEQ agrees and clarification has been added accordingly.

- *On page 94, the table on uncertainty factors used by different health organizations includes the recommendations of the European Centre for Ecotoxicology and Toxicology of Chemicals (ECETOC). Note there is a more recent document by ECETOC that should be considered: ECETOC Technical Report No. 110, October 2010; Guidance on assessment factors to derive a DNEL.*

Response: A reference to the ECETOC Technical Report No. 110, October 2010 has been placed in the Guidelines.

- *Beginning on page 155 and ending on page 163, there is a discussion of adjustment factors for early life exposure to carcinogens. After going through the U.S. Environmental Protection Agency (EPA) 2005 supplemental guidance, the TCEQ document refers to the alternative method of the BIER IV life table analysis approach. This alternative approach seems to be favored when mentioned in Chapter 7 (page 33) regarding adjustments for early-age exposures. The charge questions for chapter 7 seem to request comment on the issue of risk assessment approaches for children versus adults. In this regard, TCC favors the BIER IV life table analysis approach when it is applicable.*

Response: No response required.

- *In Chapter 7, hazard characterization and exposure response assessment using epidemiology data, there is a discussion on page 41 regarding meta-analysis. TCC requests clarification as to whether meta-analysis of individual data is also referred to as a "pooled analysis."*

Response: Pooled analysis is a form of a meta-analysis not accounting for any sources of heterogeneity in the data. As indicated in the Guidelines a meta-analysis should "…control for all possible covariates and sources of heterogeneity." Bravata and Olkin (2001) make a distinction between a pooled analysis and a well-conducted meta-analysis when they state that, "In simple pooling, data are combined without being weighted. Therefore, the analysis is performed as if the data were derived from a single sample. This kind of analysis ignores characteristics of the subgroups or individual studies being pooled and can yield spurious or counterintuitive results. In meta-analysis, data from subgroups or individual studies are weighted first, then combined, thereby avoiding some of the problems of simple pooling."

Bravata, D. M. and I. Olkin. 2001. Simple pooling versus combining in meta-analysis. Eval Health Prof;24:218–230.

By offering the following comments, TCC hopes to provide perspectives to enhance the regulatory process and assist in establishing a science-based and health protective method for evaluating chemicals. TCC also hopes to continue this dialogue as the TCEQ subsequently seeks comment on its recommendation of inhalation and oral values for individual compounds.

Again, the TCC appreciates the opportunity to comment on this important document and looks forward to future discussions with the TCEQ.

Attachment: Copy of TCC Comments to the prior TCEQ ESL guidance draft document (comments dated June 2005)

Response: As stated in the first response in B.3, the below comments (1.0 to 16.0) are regarding the 2006 Guidelines and were previously addressed. Responses to those comments can be found at this location http://www.tceq.texas.gov/assets/public/implementation/tox/response/2006 prp.pdf. No response required for comments B.3 1.0 through 16.0 below.

1.0 TCC Supports Explicit Recognition That Exposures Above an ESL Are Not Necessarily Indicative of Adverse Health Impacts.

The Review Draft notes that ESLs are designed to be protective of the general public, including sensitive subgroups such as children, the elderly, pregnant women, and persons with preexisting health conditions, and that exposure to a chemical in air at or below the ESL is not likely to cause adverse health effects. (pp. 2, 5) Further, because of the protective assumptions built into the ESL derivation process, the Review Draft recognizes that if exposures exceed an ESL, "adverse effects would not necessarily be expected to result." (p. 2) Stated differently, "Since UFs are incorporated to address data gaps and other uncertainties, exceeding the ReV does not automatically indicate an adverse health impact." (p. 5) TCC believes that similar statements can be made about the derivation of unit risk factors (URFs) for carcinogens. URFs are explicitly designed to be "upper bound" estimates of levels of risk. Because URFs are upper bound values, actual or true human health risks will most certainly be lower than calculated upper bound risk levels. (See p. 5). To illustrate the relationship between upper bound estimates of risk and actual risk, some have used an analogy to batting averages in baseball; while the upper bound batting average in baseball is "1000," in reality, no batter will ever reach, or likely even come close to, that average.

TCC believes these are important points and supports their explicit recognition in the ESL methodology document. Further, TCC believes that these points should be reiterated when documentation is generated for each new ESL value, so that the public understands the health protective, "upper bound" nature of individual ESLs. It is important for the public to understand that exposure to chemicals in air at levels equal to or even above the ESL is not indicative of any imminent human health effects and they should not assume that any exposure above an ESL indicates that adverse effects on human health have or will result.

As has been seen in recent years in statements in the popular press and in some public policy arenas, the meaning of screening values is often misunderstood; it is typically assumed that ambient air levels above ESLs are known to pose significant health risks, and further that exposures above URFs show that excess cancers are definitely occurring as a result of those exposures.

TCC believes that an objective and transparent presentation of ESLs, with explicit recognition that true risks are likely to be below estimated upper bound risks, would help prevent such misunderstanding and misapplication of ESLs. This might be most efficiently and effectively accomplished by including a statement at the beginning of each ESL assessment document that defines the ESL explicitly as a level of a chemical in air that, even if exceeded, would not necessarily be expected to result in adverse human health effects.

2.0 TCC Does Not Support Use of a Target Hazard Quotient of 0.1

The Review Draft indicates that the TCEQ Toxicology Section (TS) will use a target hazard quotient (HQ) of 0.1 to calculate screening levels for noncarcinogenic constituents (p. 6), but does not explain why the TS will use a HQ other than the more typical value of 1.0. TCC believes the target HQ for noncarcinogenic effects should not be less than 1.0. By definition, a HQ is the ratio of the expected exposure level divided by the acute or chronic toxicity value, where a value of 1 indicates that exposure has not exceeded the toxicity value. It is well-established that a hazard quotient less than 1 indicates that there is little

likelihood of risk. Since TCEQ stated in the Review Draft that the ESLs or reference values it derives represent values that are likely to be without "an appreciable risk of adverse effects" (See p. 5), it appears inconsistent for the TS to then assume that an extra 10-fold adjustment is needed when that reference value is used to calculate the HQ.

Moreover, consistent with the preceding point in these comments, TCC believes that ReVs and similar chronic values (such as EPA RfCs) are sufficiently health protective to be applied without adjustment. It is important in this context to note the many protective layers built into the chronic ReV derivation process. After application of uncertainty factors and a duration adjustment for discontinuous exposure, a chronic ReV often will be set at a level 1000-fold or more below the NOAEL in the most sensitive animal study. Further, the LOAEL may be based on relatively mild or minor effects. In any event, since the chronic ReV is designed to be protective of the general population, including sensitive subgroups, TCC does not believe there is any scientific justification for discounting the value by an additional factor of 10. Rather, the ReV should be applied as is.

In addition, the Risk Assessment And Management Commission (Commission) established by the 1990 amendments to the federal Clean Air Act (CAA) to provide EPA guidance on implementing the risk-based provisions of the CAA recommended that EPA should only conduct further risk assessment and analysis of source categories where the screening level noncancer hazard quotient exceeds 10.0. The Commission arrived at this target HQ following Congress' mandate that it "make a full investigation of the policy implications and appropriate uses of risk assessment and risk management in regulatory programs under various Federal laws to prevent cancer and other chronic human health effects which may result from exposure to hazardous substances."[11] The Commission also noted that categories of sources with a HQ of less than 10.0, based on a screening level risk assessment, should be ranked as only "medium" priority and recommended that if a more detailed risk assessment (reducing or eliminating the conservative assumptions associated with the screening level assessment) yields non-cancer HQs less than 1.0, then no further action should be required.[12] The Commission's recommendations therefore support no further regulatory action if a HQ is less than 10.0 based on a screening level analysis, and no further action of any kind if the HQ is less than 1.0 following more refined analysis. TCC notes that EPA applied a HQ of 1.0 in its first residual risk rule under the CAA.[13]

3.0 TCC Would Not Support Use of a Risk Management Goal for Carcinogenic Effects More Restrictive Than 1 x 10^{-5}

The Review Draft indicates that the TCEQ TS will use a risk management goal of 1 x 10^{-5} excess lifetime theoretical cancer risk in calculating screening levels for individual carcinogens. (p. 6). Given the health protective nature of the methodology used to derive URFs, TCC agrees with this approach and believes it to be fully protective of human health. Further, TCC believes that it would not be scientifically justified to use a more conservative target risk management level for known, probable or suspected carcinogens. As already noted above, the scientific reality is that actual risks are most certainly well below upper bound

[11] CAAA of 1990, P.L. 101-549, § 303(a).
[12] Presidential/Congressional Commission on Risk Assessment and Risk Management, Risk Assessment and Risk Management in Regulatory Decision Making (1997), Vol. 2, pp. 110-11 and Figure 7.1.
[13] National Emission Standards for Coke Oven Batteries; Final Rule. 70 Fed. Reg. 19992 (April 15, 2005).

calculations of theoretical risks, such that when the risk management goal of 1 x 10^{-5} is met, actual risks are likely to be much lower than that risk management goal, and could be zero.

4.0 TCC Supports Peer Review and Public Comment on Proposed ESL Values

The ESL development process (described at p. 12) includes a number of positive features, including the annual listing of chemicals under consideration, the solicitation of relevant data from interested parties, and publication of proposed Development Support Documents for public comment. However, TCC believes the process may not be sufficient in all cases. In particular, the specified 30-day comment period may be insufficient for some data rich substances, particularly where difficult risk assessment choices are made. Also, in many cases, an independent, external peer review may be appropriate, and where that occurs, TCC believes TCEQ should invite public comment on the peer review report and make those comments available to the peer reviewers. The derivation of ESLs for certain chemicals will be difficult, with no two reviews being based on the same quality and types of data. Therefore, the need to allow for longer comments periods may be dictated by the complexity of the ESL derivation process on a case-by-case basis.

5.0 TCC Supports the Use of Data over Default Assumptions

TCC believes that default assumptions should be replaced with chemical-specific information whenever possible, as is discussed at pages 26-27 of the Review Draft. Chemical-specific acute or chronic (noncancer and cancer-based) ESLs should be based on the best available science, with the use of default assumptions only when necessary. Default uncertainty factors should be replaced with scientifically defensible values for acute and chronic ESLs, and default assumptions regarding carcinogens for the establishment of unit risk factors (URFs) should be replaced with values based on data whenever possible. TCC is encouraged that TCEQ mentions the ability to add mode of action data, consistent with the final cancer risk assessment guidelines of the EPA.

6.0 Assignment of Confidence Levels to ESLs Should Not Imply That Values Might Be Under-Protective

The Review Draft describes a process for assigning confidence levels to the key study used to derive an ESL, to the database as a whole, and to the final ESL. (See p. 20) The process resembles the process used by EPA for assigning confidence levels to inhalation reference concentrations. TCC is concerned that this process could give stakeholders an unduly negative impression of ESL values by implying that the values are equally likely to be too high (under-protective) or too low (overly protective). In reality, with the use of uncertainty factors as described in the Review Draft, ESLs based on non-cancer effects are very protective, and confidence in the health protective nature of such ESLs (versus their precision) should be high. Similarly, with the many conservative (health protective) assumptions built into the process for deriving upper bound URFs, confidence in the health protective nature of ESL values based on cancer also should be high. Accordingly, TCC urges TCEQ not to give a confusing or unduly negative impression of the protective nature of ESL values through the assignment of confidence levels. Indeed, while it may be appropriate to assign a confidence level to the key study used to derive an ESL, it is not clear that assignment of a confidence level to an ESL value itself is a very meaningful or helpful exercise. Assignment of any type of confidence level or quality factor to the ESL will likely lead to public misinterpretation of the complex process that was undertaken. If such quality factors are used, it will be

critical to succinctly describe the reasoning behind such choices. As was discussed above in the context of the conservative nature of the ESL development process, the public must be able to understand that, even with the uncertainties that may go into any individual calculation or derivation, the resulting ESL is a conservative, upper bound value that will be protective of public health.

7.0 Application of a 10-Fold Uncertainty Factor to a BMCL Level is not Generally Accepted

TCEQ discusses (p. 26) converting a $BMCL_{10}$ into a NOAEL by applying an UF of up to 10. The Review Draft supports this approach by stating that a $BMCL_{10}$ can be considered equivalent to a LOAEL. TCC disagrees with this approach. By definition, a $BMCL_{10}$ is the lower bound on the 95% confidence limit of the 10% response level. The lower confidence limit (LCL) approach assures that the level of response is not exceeded. Although a $BMCL_{10}$ is not distinctly a NOAEL value, neither is it a LOAEL. Studies have shown that the $BMCL_{10}$ is actually some value between the two levels, and is often essentially equivalent to the NOAEL (Fowles et al. 1999[14] as cited in EPA 2000[15]). Additionally, LCLs are sensitive to changes in the number of non-responding dose groups whereas maximum likelihood estimates (MLEs) are insensitive (Gephart et al., 2000).[16] Therefore, there is no scientific support for applying a default 10-fold factor to a $BMDL_{10}$ when a $BMDL_{05}$ is not available. To improve the proposed ESL methodology, we suggest using the $BMDL_{10}$ (LCL on 10% response level) without an additional and arbitrary 10-fold factor. Alternatively, we suggest using the MLE at the 10% response level with the additional factor. Since the first approach avoids use of an additional arbitrary factor, it appears to be the most scientifically valid approach. TCC asks that the TCEQ re-visit this issue.

8.0 Selection of Uncertainty Factors For Acute ESLs Should Be Made Carefully

Beyond the mention above of the need to replace default uncertainty factors with alternative factors that are based on data, there is a need to better explain uncertainty factors for acute ESLs. For example, the Review Draft (p. 27) describes UFs for converting lethal effects levels to ESLs. In that section, use of a UF of more than 10,000 is discussed. However, since an acute ESL is meant to be a value that is representative of an exposure period that is very short, much shorter in most cases than the exposure period of an animal study, the animal data are by design exaggerating exposure. This approach, combined with the a 10-fold factor to account for lack of identification of a NOAEL, and then another 100-fold factor to account for both interspecies and intraspecies differences in response, is clearly overly conservative. It is known that there is overlap in the uncertainty accounted for by the various UFs employed, i.e., both interspecies and intraspecies factors account for age differences in response. Therefore, applying such large UFs to acute ESLs (factors of 10,000, as stated at p. 27) appears to be problematic and not scientifically justified. This is especially true for any situation where the TCEQ may contemplate application of an additional UF for an incomplete database when deriving an acute ESL.

---

[14] Fowles, J.R.; Alexeeff, G.V.; Dodge, D. (1999). The Use of Benchmark Dose Methodology with Acute Inhalation Lethality Data. Regul. Toxicol. Pharmacol. 29(3):262-278.

[15] United States Environmental Protection Agency. (USEPA 2000). Benchmark Dose Technical Guidance Document. EPA/630/R-00/001. Risk Assessment Forum. Washington, D.C.

[16] Gephart, L.A.., W.F. Salminen, M.J. Nicolich, and M. Pelekis (2001). Evaluation of Subchronic Toxicity Data Using the Benchmark Dose Approach. *Regulatory Toxicology and Pharmacology* 33, 37-59.

The TCC suggests TCEQ instead consider adopting some of the procedures and factors for deriving Acute Exposure Guideline Levels (AEGLs) as referenced in the National Research Council's AEGL Standing Operating Procedures (NRC, 2001).[17] *With this approach, in many cases, the 10X default UFs are replaced by 3X UFs. The rationale for using these UFs, as described in the AEGL approach, could be referenced or included in the TCEQ ESL Methodology document. In addition, in cases where an incomplete database UF is required, the TCC suggests that TCEQ use a factor in the range of 2-3X, as is commonly used by the AEGL Committee. In some cases, depending on the nature of the endpoint being addressed and the underlying data (e.g., sensory irritation reported in naïve human subjects), an uncertainty factor approach may not be necessary at all to derive a value that is sufficiently protective of human health.*

9.0 When Estimating NOAELs from LOAELs, TCEQ Typically Should Apply a Factor of 3X

TCC suggests replacing the default factor of 10X for estimating NOAELs from LOAELs with a 3X factor. As described by ECETOC (2003)[18]*, in published studies where the ratios of LOAELs to NOAELs were compared for different chemicals and different study durations (subacute, subchronic, and chronic) the LOAEL rarely exceeded the NOAEL by more than about 5-6 fold. As shown in the table below, the central ratio is typically close to a value of 3.*

Ratio of LOAEL/NOAEL

Study Type	*Mean of LOAEL/NOAEL*	*Reference and comments*
Subchronic (n = 27) *Chronic (n = 25)*	*3.02 range (2-3, one at 5)* *3.8 (range 2-4, two at 10)*	*Dourson and Stara, 1983.*[19]
Chronic (n = 175)	*4.5 +/- 1.7 (95th% = 11)*	*Kramer et al., 1996*[20] *- oral*
Chronic (n = 7)	*5.7 (95th% = 11)*	*Kramer et al., 1996*[21] *- inhalation*
Subacute (n = 95)	*3.5 +/- 1.8 (95th% = 9)*	*Pieters et al., 1998*[22] *- subacute*
Subchronic (n = 226)	*4.3 +/- 2.2 (95th% = 16)*	*Pieters et al., 1998*[23] *- oral*
Subchronic (n = 23) *Chronic (n = 23)*	*91% ≤ 6* *87% ≤ 5*	*Kadry et al., 1995*[24] *Based on 6 chlorinated compounds*
Subchronic (n = 18) *Chronic (n = 18)*	*< 3* *≅ 3.5*	*Beck et al., 1993**[25]

[17] NRC (2001). Standing Operating Procedures for Developing Acute Exposure Guideline Levels for Hazardous Chemicals. National Academy Press.
[18] ECETOC (2003). Derivation of Assessment Factors for Human Health Risk Assessment. Technical Report No. 86. ISSN00773-6347-86. Brussels, February 2003.
[19] Dourson ML, Stara JF. 1983. Regulatory history and experimental support of uncertainty (safety) factors. Reg Toxicol Pharmacol 3: 224-238.
[20] Kramer HJ, van den Ham WA, Slob W, Pieters MN. 1996. Conversion factors estimating indicative chronic no-observed-adverse-effect levels from short-term toxicity data. Reg Toxicol Pharmacol 23: 249-255.
[21] *Id.*
[22] Pieters MN, Kramer HJ, Slob W. 1998. Evaluation of the uncertainty factor for subchronic-to-chronic extrapolation : statistical analysis of toxicity data. Reg Toxicol Pharmacol 27: 108-111.
[23] *Id.*
[24] Kadry AM, Sknowronski GA, Abdel-Rahman MS. 1995. Evaluation of the use of uncertainty factors in deriving RfDs for some chlorinated compounds. J Toxicol Env Hlth 45: 83-95.
[25] Beck BD, Conolly RB, Dourson ML, Guth D, Hattis D, Kimmel C, Lewis SC. 1993. Improvements in quantitative noncancer risk assessment. Fund Appl Toxicol 20: 1-14.

Developmental (n = 246)	*2, 3 or 4 with equal frequencies*	*Faustman et al., 1994*[26]

** Beck et al. considered the results obtained in their analyses for subchronic and chronic ratios an over-estimation*

In addition, the LOAEL/NOAEL ratio is highly dependent on the spacing between the doses chosen by the investigator. Since in recent study designs doses are typically spaced such that they increase by a factor of 2 to 4 between dose levels, it is logical to conclude that the ratio data support a value of 3 as a default. This would not preclude using a higher value, if warranted by data on a specific chemical. As already described in the Review Draft, the default value can replaced by a higher or lower value based on the data for a specific chemical.

10.0 EPA IRIS Values Should Be Used With Caution

The Review Draft indicates that published chronic toxicity factors, such as EPA's Integrated Risk Information System (IRIS) values, may be used when certain indicia of reliability are met. (p. 28). The published chronic toxicity factor must be based on a well-conducted scientific study, evaluation of the body of scientific literature (including studies made available after the date of publication) must indicate that the published toxicity factor is health-protective, and the risk assessment procedures used to derive the published value must be similar to those described in the Review Draft. These are reasonable points. However, TCC urges TCEQ to exercise caution when considering using published toxicity values, including EPA's IRIS database, as a source of existing toxicity factors.

IRIS values are frequently out of date. Additionally, though EPA has recently modified its IRIS document development process to allow greater stakeholder involvement and limited peer review, very few values have been established through the new process. Most published IRIS values did not go through a rigorous or transparent peer review, and where peer review was conducted, peer reviewers typically were not given access to stakeholder comments (as they are now). Further, the IRIS document development process is not a rulemaking process, and in some cases very little public input was permitted. As a result of all these considerations, EPA has recognized on many occasions that IRIS values should not be given conclusive effect in subsequent rulemaking proceedings. To the contrary, directives issued by Drs. Henry Longest and William Farland in 1993[27] *and John Seitz in 1994,*[28] *and a Federal Register notice signed by former EPA Administrator Whitman on September 7, 2001,*[29] *expressly provide that IRIS values are not entitled to conclusive weight in any rulemaking, and program offices must consider all credible and relevant information that is submitted in any rulemaking proceeding.*

[26] Faustman EM, Allen BC, Kavlock RJ, Kimmel CA. 1994. Dose-response assessment for developmental toxicity. 1. Characterization of database and determination of no observed adverse effect levels. Fund Appl Toxicol 23: 478-486.

[27] William H. Farland, Director, Office of Health and Environmental Assessment, and Henry L. Longest, II, Director, Office of Emergency and Remedial Response, Use of IRIS Values In Superfund Risk Assessment, OSWER Directive #9285.7-16 (December 21, 1993).

[28] John S. Seitz, Director, Office of Air Quality Planning and Standards, Guidance On Use of Integrated Risk Information System (IRIS) Values (August 26, 1994).

[29] 66 Fed. Reg. 46928, 46929 (Sept. 7, 2001) (reflecting settlement of legal action brought under the Safe Drinking Water Act).

Even the most recently adopted IRIS values should be used with caution, as TCEQ may not necessarily agree with each and every decision embedded in each RfC or cancer potency derivation (See, .e.g., discussion in Point 13.0, below). Further, an assessment that may seem scientifically reasonable on its face may seem less justifiable when the comments of stakeholders are considered. Unless TCEQ consults the full record underlying the derivation of an IRIS value, it may not fully appreciate the risk assessment choices that were made by EPA, and the alternative approaches that might have been considered.

TCC believes in most cases it would be more appropriate for TCEQ to use IRIS as a source of information – a starting place for analysis – rather than as a source of finished values. While the existence of IRIS might be very helpful to TCEQ efforts to develop ESLs, reliance on IRIS values should not be a substitute for exercise of sound scientific judgment by TS scientists based on the most recent data available. Furthermore, in its evaluation of the most recent data available, TCEQ should not exclude the consideration of high quality, GLP studies simply because they have not yet appeared in the peer reviewed literature.

11.0 TCEQ Should Not Rely on Toxicity Values Developed By Other States

In the hierarchy of preferred sources for published toxicity values (p. 28), RELs developed by California's Office of Environmental Health Hazard Assessment are listed third. Also, when addressing the derivation of acute toxicity values (p. 21), the Review Draft lists Cal EPA RELs in the first tier of its hierarchy of database sources. TCC believes TS should not rely on values or documentation developed by other states, except perhaps as a source of information. TCC believes there is too great a risk that reliance on values developed by other states will introduce inconsistencies in approach and inconsistent treatment of compounds that toxicologically are more alike than different. Values developed by other states also may reflect policy choices that are not readily transparent. TS scientists should exercise their own best judgment, and should not abdicate that responsibility to scientists in other states.

12.0 HEAST Values Should Not Be Used

The hierarchy of preferred sources for published toxicity values (p. 29) also lists EPA's Health Effects Assessment Summary Tables (HEAST) database as a source of chronic toxicity factors. As the Review Draft notes, these values are provisional only and have not undergone even intra-agency peer review. HEAST clearly should not be considered a source for toxicity values. At most, the database might be considered a source for scientific references, but TS should use HEAST with caution even for that limited purpose because the HEAST database is not being kept current.

13.0 Adjustments for Discontinuous Exposure Should Be Determined To Be Scientifically Appropriate In Each Case Where They Are Made, And An UF Of 10 Should Automatically Be Applied In Every Case Where a Chronic Study Is Lacking

The Review Draft's description of duration adjustments (p. 30) appears similar to EPA's approach to this issue in IRIS. TCC understands the scientific basis for making duration adjustments, but believes that such adjustments should not be made automatically in every case. In particular, for materials with a very short half-life in blood, a duration adjustment may not be scientifically appropriate. Accordingly, TCC believes the ESL methodology should describe how duration adjustments are made, but also indicate that TS scientists will consider in each case whether a duration adjustment is scientifically appropriate.

Similarly, TCC believes TCEQ should not automatically add a UF of 10 whenever a chronic study has not been conducted (as has been done in many cases in IRIS). Such a practice creates an incentive for the conduct of chronic studies even where such testing is not scientifically justified (such as where the National Toxicology Program or EPA under the Toxic Substances Control Act has determined that chronic toxicity testing is not warranted). Where a substance has a very short half-life in blood and subchronic testing has not demonstrated significant systemic toxicity, chronic toxicity testing may not be scientifically justified, and it also may not be scientifically appropriate to add any UF for lack of a chronic study. This issue, like many others raised in the Review Draft, should be addressed on a case-by-case basis using best scientific judgment. In each case, TCEQ should determine that the UFs applied are scientifically appropriate for the substance under consideration.

14.0 TCC Believes That When Human Data Are Used, the Point of Departure Typically Should Be Based on the MLE, Not the UCL.

The Review Draft states, "Typically, the lower 95% confidence limit on the lowest concentration that can be supported for modeling by the data is used as the POD." (p. 35) No distinction is made between human and animal data. This approach appears consistent with the recently published updated EPA cancer risk assessment guidelines, which also do not distinguish between values derived from animal data and those derived from human data.[30] However, historically, EPA has used the MLE (maximum likelihood estimate) when deriving unit risk factors (URFs) from human data, and the new cancer risk assessment guidelines do not offer any scientific rationale for changing that policy, nor is it even clear from the new cancer risk assessment guidelines that this issue was addressed during the external peer review of the new cancer risk assessment guidelines.

TCC believes that use of the MLE is scientifically appropriate in most cases for URFs derived from human data, and that the upper confidence limit (UCL) should be used with human data only where substance-specific justification is presented.

More generally, when making this decision, TCEQ should recognize that other health protective assumptions (such as low dose linearity) are already built into the URF derivation process. Historically, linear low dose extrapolation has been thought to be sufficiently conservative to be protective of potentially susceptible subpopulations. Thus, the question posed is this: in a given risk assessment based on human data, is it necessary to use the UCL as the POD in addition to other health protective assumptions to derive a URF value that is protective of human health? In most risk assessments based on human data, TCC believes the MLE will be scientifically sufficient and health protective.

15.0 TCEQ Should Be Cautious in Its Use of an Additional Factor to Account for Susceptibility Due to Early-Life Exposure to Carcinogens

The EPA has only recently finalized its guidance regarding early-life exposures to carcinogens,[31] and therefore its application has been limited. Moreover, TCC believes that the current cancer assessment methodologies, which do not adjust for early-life susceptibilities, protect both adults and children and

[30] *See* EPA, Guidelines for Carcinogen Risk Assessment. EPA/630/P-03/001F, (March 2005), pp. 3-16 – 3-18.
[31] EPA, Supplemental Guidance for Assessing Susceptibility from Early-Life Exposure to Carcinogens. EPA/630/R-03/003F, (March 2005).

that there is inadequate scientific evidence that the current methods are not suitably health protective. In fact, the data on early-life susceptibility to cancer have shown that the mechanisms of carcinogenicity that operate in adults also operate in children, and that to the extent that children may be more, less or equally sensitive to some substances, current cancer assessment methodology is sufficiently conservative to protect children. As a result, chemicals should be looked at on a case-by-case basis to determine whether there is justification for separate consideration of early-life exposure and cancer risk. In fact, the question should not be whether children are more or less sensitive but rather whether children are protected. Additionally, when considering a risk management goal of 1 x 10^{-5} excess lifetime theoretical cancer risk in calculating screening levels for individual carcinogens, an additional factor accounting for early-life susceptibility may not be necessary. Moreover, any consideration of adjustments for early-life susceptibility should be done recognizing that other factors may already have been included in the cancer risk calculations, such as the use of the UCL versus MLE (as discussed above) and the selection of dose response models.

Ultimately, TCEQ should evaluate chemicals individually and determine whether an early-life susceptibility factor is necessary by considering such scientific issues as: a) whether there are reasons to not apply this factor (e.g., the chemical is metabolized to the active metabolite by a pathway that is underdeveloped in early-life stages); and b) whether susceptibility has already been accounted for by other risk assessment choices, and the like.

16.0 TCEQ Should Make Sure That Its Risk Assessment Choices Are Scientifically Reasonable As A Whole.

Finally, TCC believes TCEQ should have a step at the end of the ReV and URF derivation process where it assesses the reasonableness of its approach as a whole. While it is appropriate that each risk assessment choice in a given risk assessment be health protective, the collective impact of too many conservative risk assessment choices can be the derivation of a chronic ReV or URF that is scientifically unreasonable as a whole. Thus, in the first instance, TCEQ should determine that each risk assessment choice is scientifically appropriate considered alone, but TCEQ also should determine that each risk assessment choice is reasonable in light of all other risk assessment choices that have been made, and TCEQ also should make sure that the aggregate impact of all risk assessment decisions is not overly conservative, but in fact produces a risk value that is scientifically appropriate, consistent with the underlying data, and reasonable. Further, TCEQ should ask the peer reviewers to go through the same decision logic, to make the final product is scientifically reasonable.