TERA

INDEPENDENT
NON-PROFIT
SCIENCE
FOR PUBLIC HEALTH
PROTECTION

Report of Letter Peer Review of the TCEQ's Isoprene – Section 4.2 Carcinogenic Potential – Development Support Document

Texas Commission on Environmental Quality (TCEQ)

July 1, 2013

Submitted by:

Toxicology Excellence for Risk Assessment
Contact: PatriciaNance (nance@tera.org)
www.TERA.org

This page left intentionally blank.

Note

This report was compiled by scientists of Toxicology Excellence for Risk Assessment (TERA). The peer reviewers served as individuals, representing their own personal scientific opinions. They did not represent their companies, agencies, funding organizations, or other entities with which they are associated. Their opinions should not be construed to represent the opinions of their employers or those with whom they are affiliated.

This page left intentionally blank.

TABLE OF CONTENTS

1. INTRODUCTION..... 1

2. PEER REVIEWER RESPONSES TO CHARGE QUESTIONS ON THE GENERAL STRENGTHS AND WEAKNESSES OF PROCEDURES..... 5

3. PEER REVIEWER RESPONSES TO CHARGE QUESTIONS ON CANCER ASSESSMENT AND UNIT RISK FACTOR (URF)..... 8

4. REFERENCES.....34

APPENDIX A - CHARGE QUESTIONS AND INSTRUCTIONS FOR PEER REVIEWERS.....36

APPENDIX B - CHARGE QUESTIONS AND INSTRUCTIONS FOR PEER REVIEWERS.....41

APPENDIX C – PUBLIC COMMENTS.....47

This page left intentionally blank.

1. Introduction

This report summarizes external peer review comments on Section 4.2 Carcinogenic Potential of the Development Support Document for Isoprene. Toxicology Excellence for Risk Assessment (TERA) organized and conducted an independent external scientific and technical peer review of this document for the Texas Commission on Environmental Quality (TCEQ). The goal of the peer review was to have a group of qualified external experts conduct a thorough and meaningful assessment of the document and provide an independent evaluation of the robustness of the science and whether the conclusions are supported by the body of evidence.

The Toxicology Division of the Texas Commission on Environmental Quality (TCEQ) has prepared a draft Development Support Document (DSD) that outlines the hazard assessment and dose-response processes used to derive health-protective Effects Screening Levels (ESLs) and Reference Values (ReV) for isoprene. The draft DSD includes Section 4.2, which documents the derivation of an inhalation unit risk factor (URF) and air concentrations corresponding to the policy-based 1 in 100,000 excess risk level based on lung cancer mortality. These toxicity values are used in the evaluation of air permit applications and ambient air data and were developed using RG-442 TCEQ Guidelines to Develop Toxicity Factors (TCEQ 2012). The TCEQ guidelines can be found at http://www.tceq.texas.gov/publications/rg/rg-442.html.

1.1 Peer Review Organization

TERA was responsible for managing all aspects of the peer review process, including selection of the reviewers, evaluation of potential conflicts of interest of candidate reviewers, development of the charge questions, distribution of the assessment document, collection and review of each expert's written comments, and compilation of all comments into a single report (this report).

Alliance for Risk Assessment (ARA) –This peer review is a project under the Alliance for Risk Assessment (ARA). ARA is a collaboration of organizations that fosters the development of technical chemical risk assessment products and services, through a team effort of specialists and organizations dedicated to protecting public health by improving the process and efficiency of risk assessment, and to increasing the capacity for developing risk values to meet growing demand. All ARA projects are vetted by a Steering Committee comprised of federal and state government, academic, and NGO perspectives, to promote scientific relevance and avoid

duplication of effort. As an ARA project, this project was led by an independent, nonprofit organization, performed in an open and transparent manner, and the results will be made publicly available at www.allianceforrisk.org.

Selection of Reviewers. TERA reviewed the draft assessment document and in consultation with TCEQ identified the types of expertise needed for the peer review. These included familiarity with isoprene toxicology literature, benchmark dose modeling, inhalation dosimetry, biostatistics/biomathematics, carcinogenic toxicology and mechanism of action, and risk assessment. TERA developed a list of potential reviewers that TERA judged to be qualified. This list was shared with TCEQ in order for the scientific authority to identify any reviewers who may have a potential conflict of interest or those who they thought unqualified. From the final cleared list, TERA independently selected four reviewers who collectively covered the needed areas of expertise to provide a high-quality peer review of the assessment.

TERA discussed the situations and conditions that may be considered potential conflicts of interest (COI) for the peer reviewers with TCEQ, and developed a COI questionnaire to screen all candidates. TERA's conflict of interest policy is found at http://www.tera.org/peer/COI.html. After reviewing credentials and COI information, TERA selected a group of reviewers that provide a balance of appropriate expertise and perspectives for this peer review. To maintain the independence of the peer review, the experts have had no direct contact with TCEQ. The expert peer reviewers for this assessment are listed below. Their affiliations are provided for identification purposes only. Appendix A contains short biographical sketches of the experts and results of the conflict of interest screening.

- Lynne Haber, Ph.D., D.A.B.T. – Toxicologist, Associate Director of Science, Toxicology Excellence for Risk Assessment, Cincinnati, OH, USA
- Richard Hertzberg, Ph.D. – Biomathematician, Adjunct Professor, Department of Environmental Health, Emory University, Atlanta, GA, USA
- Bruce Allen, M.S. – Biostatistician, Private consultant, Allen Consulting, Chapel Hill, NC, USA
- Jerry Rice, Ph.D. – Toxicologist, Distinguished Professor of Oncology, Member of the Cancer Prevention and Control Program, Lombardi Comprehensive Cancer Center, Georgetown University Medical Center, Washington, DC, USA

Development of Charge. A key aspect of a successful peer review is a comprehensive list of objective questions to frame the reviewers' comments and ensure that the reviewers are focused on the most important issues. TERA drafted a list of questions and issues for this "charge" to the peer reviewers and sent a draft of the charge to TCEQ for comment and input. . TERA, as the independent organizer of the peer review, considered TCEQ's input on the charge questions, but was responsible for the final content and wording of the charge. The charge questions focus on the adequacy, quality and relevance of the data and information and whether the conclusions reached are supported by the data. Focused and open-ended questions were used to provide reviewers with the opportunity to identify and discuss all the issues they felt were important. A copy of the charge and instructions for reviewers is found in Appendix B.

Reviewers' Comments. Reviewers were allotted several weeks to review the assessment document and submit comments to TERA. TERA compiled reviewers' comments by charge question, randomly assigning each reviewer a reviewer number that was used throughout the report. The assigned reviewer number is meant to keep each reviewer's specific comments anonymous, although the names and affiliations of the reviewers are provided. TERA staff screened the experts' comments for completeness and clarity, and TCEQ was given the opportunity to review the peer reviews' comments and submit to TERA clarifying questions for the reviewers.

Request for Public Comments. TERA posted information about the peer review on a publically-accessible web page and provided the opportunity for members of the public to submit comments. One public comment was received and has been included in this report (see Appendix C).

TCEQ Clarifying Questions. The TCEQ staff reviewed the reviewers' comments and submitted only one clarifying question, which was sent to the reviewer and their comments were revised. Below is the clarifying question:

- Reviewer 2, Question 1: The reviewer has answered this question with a note on the review process, which is not relevant to the DSD. It is not clear whether or not the reviewer thought the approaches used by TCEQ were clearly described or not.

The comments were compiled into this comprehensive report entitled, *Report of Letter Peer Review of TCEQ'sIsoprene - Section 4.2 Carcinogenic Potential - Development Support Document.*

2. Peer Reviewer Responses to Charge Questions on the General Strengths and Weaknesses of Procedures

2.1 Does the draft DSD clearly describe the approaches used by TCEQ to develop the URF? (Charge Question 1)

2.1.1 Reviewer 1

For the most part, the approaches are straight-forward and/or described clearly. There were some small details that could be explained further or that seem contradictory. For example, on p. 6, lines 14-16, it states that the diepoxide metabolite is formed by the minor epoxide intermediate. However, Figure 1 (p. 7) shows that the diepoxide is formed from both monoepoxide intermediates. As another example, on line 2 of p. 17, the DSD states that the "most relevant" dosimetry classification for isoprene was Category 3 gas, after initially stating it was a Category 2. Why was Category 3 chosen over Category 1? The abbreviation RGDR is introduced on p. 17 (line 11) for no reason; the document could just reference the ratio of partition coefficients without the added jargon. Similarly, line 5 of p. 17 described $H_{b/g}$ as a ratio of the partition coefficient. That is not true; it is the partition coefficient itself (though, of course, the definition of the partition coefficient does involve ratios of concentrations, irrelevant to this point as that may be).

There are more problematic clarity issues with the material in Appendices A and B; these will be described and discussed below (question 6e).

2.1.2 Reviewer 2

In general the methods are described clearly. However, as noted below, it would be useful to include a more extensive nonmathematical description in the main document of the analyses described in Appendix A, as well as the conclusions and implications. In addition, it was not clear what the author considered to be the weight of evidence regarding the mode of action (MOA) for carcinogenicity. In some places it seemed that the author leaned towards a mutagenic MOA but in other places it seems that the author considered the weight more towards a non-mutagenic MOA. (Note that this is a biological consideration, separate from the conclusion to use a linear extrapolation approach as a default based on insufficiency of the data for a nonmutagenic MOA.)

2.1.3 Reviewer 3

Yes, the draft DSD is very well organized, well written, and detailed in describing the steps used to estimate the unit risk factor (URF). The formulas used as well as

assumptions and defaults are well described, and the extensive appendices are quite helpful in understanding formulas as well as concepts.

2.1.4 Reviewer 4

The approaches used by TCEQ to develop the proposed URF for carcinogenicity resulting from inhalation exposure to isoprene are carefully and clearly described. The available carcinogenicity data in mice and rats are presented and/or referenced and the rationale for choosing hepatocellular carcinomas in male mice as the dataset for calculating the URF (the most sensitive adverse health effect relevant to humans, as required by TCEQ RG-442) is explained. The arithmetic involved is presented in detail.

2.2 Were procedures outlined in RG-442 *TCEQ Guidelines to Develop Toxicity Factors* (TCEQ 2012) followed by the TCEQ in this assessment? (Charge Question 2)

2.2.1 Reviewer 1

The procedures outlined in RG-442 *TCEQ Guidelines to Develop Toxicity Factors* (TCEQ, 2012) appear to have been followed by TCEQ in this assessment.

2.2.2 Reviewer 2

The methods were followed to a large degree, but there were important exceptions or areas of incomplete documentation. Specifically, the guidelines do not seem to have been followed in the choice of endpoints to be considered for the point of departure (POD). In addition, the documentation of the methods used for the dosimetric adjustment and calculation of human equivalent concentration (HEC) is less clear and transparent than I would expect based on the guidelines, and does not seem to apply all of the considerations expected for the guidelines, although I agree with the final results for the dosimetric adjustments.

2.2.3 Reviewer 3

The only step where the draft seems to disagree with the TCEQ Guidelines is in the use of EC_{10} instead of LEC_{10} for calculating the URF. The rationale given in the draft is strong and reasonable: that the EC_{10} is sufficiently conservative because of other steps, including using a nonthreshold model when some evidence of a threshold concentration exists, so that the lower bound on EC_{10} need not be used. I strongly recommend stating that the decision to use a lower confidence limit on the EC_{10} is not a science issue but is based on policy or regulatory concerns. There is extensive statistical literature on the properties of the confidence limit and its numerical stability compared with the median or mean value, but that is not sufficient justification. All the LEC tells us is the data spread for that model and data set used, and so is a highly misleading indicator of the

total numerical uncertainty in the estimate of the URF. Indeed, a poor data set (that might be used because it is still the best one) would have a large confidence interval that only reflects the spread of that data set and would not provide assurance of scientific credibility. If improvements to the uncertainty evaluation are needed, then one might consider: recommending the EC_{10} but also showing the confidence intervals to indicate quality, and expanding the analysis to include not just the chosen EC_{10} but also the results of model averaging or the range of model results across all candidate data sets (including more endpoints).

2.2.4 Reviewer 4

The basic procedures outlined in TCEQ RG-442 (which I consulted) were followed, but were modified to involve participation of a biostatistical consulting firm to work up the dose-response data. This firm introduced a metric based on theoretical work by Armitage & Doll (*Brit J Cancer* (1954) 8:1-12; cited as a "classic paper reprint" in the DSD references as Armitage & Doll, 2004) that introduces the concept of "stages" of carcinogenesis and the concept that many human cancers, notably carcinomas, arise through a series of 7 discrete steps, or stages. This metric is introduced into a series of equations on which the final URF calculation is based. This paper is not cited and this approach is not included in TCEQ RG-442. (See further comments on this in my response to questions 6d and 12, below).

2.3 Please identify any relevant studies or data that have not been cited and would affect an important part of the assessment and explain how they would impact the assessment specifically. (Charge Question 3)

2.3.1 Reviewer 1

I know of no other studies or data relevant to this assessment.

2.3.2 Reviewer 2

I am not aware of any uncited relevant studies.

2.3.3 Reviewer 3

None.

2.3.4 Reviewer 4

All relevant rodent bioassays have been cited. The literature survey seems complete.

3. Peer Reviewer Responses to Charge Questions on Cancer Assessment and Unit Risk Factor (URF)

3.1 Section 4.2.3 briefly presents carcinogenic weight of evidence classification information and conclusions of authoritative bodies and TCEQ's weight of evidence conclusion. Is TCEQ's weight of evidence conclusion appropriate? (Charge Question 4)

3.1.1 Reviewer 1

Yes, given the information presented, both the recitation of the carcinogenic effects observed in experimental species and the conclusions from IARC and NTP, TCEQ's conclusion concerning the carcinogenic weight of evidence appears to be appropriate.

3.1.2 Reviewer 2

Yes, the conclusion is supported by multiple animal studies and is consistent with the conclusions of other authoritative bodies.

3.1.3 Reviewer 3

Yes, based on the animal evidence and lack of human evidence.

3.1.4 Reviewer 4

The RG-442 guidelines state that TCEQ will perform a carcinogenic dose-response assessment for chemicals considered "likely to be carcinogenic to humans" or "carcinogenic to humans." The draft DSD accurately cites the conclusion of the U.S. National Toxicology Program (NTP), in the 12th Report on Carcinogens (RoC), that isoprene is "reasonably anticipated to be a human carcinogen" based on the criterion that there is clear evidence of carcinogenicity from studies in experimental animals. The bioassays indicate there is an increased incidence of malignant and/or a combination of malignant and benign tumors in multiple species (in this case, rats and mice) at multiple tissue sites. The NTP RoC listing is exactly parallel to the classification of isoprene by the International Agency for Research on Cancer (IARC), also cited in the draft DSD, as "possibly carcinogenic to humans" on the basis of the same published carcinogenicity bioassay data, which the international IARC working group concluded provided sufficient evidence of carcinogenicity of isoprene in experimental animals. "Sufficient evidence" for IARC is provided when a causal relationship has been established between the agent and an increased incidence of malignant neoplasms or of an appropriate combination of benign and malignant neoplasms in two or more species of animals, i.e., the same criterion used by the RoC. Accordingly, TCEQ's weight of evidence conclusion that isoprene is likely

to be carcinogenic to humans is based on the concordant conclusions of two authoritative bodies, and is entirely appropriate. It should be noted that both the RoC and the IARC Monographs on the Evaluation of Carcinogenic Risk to Humans present carcinogenic hazard identifications, and do not undertake quantitative risk assessments.

3.2 Section 4.2.4 discusses isoprene's carcinogenic mode of action (MOA). Have the authors clearly and accurately summarized the available data and hypotheses for isoprene's mode of action? (Charge Question 5)

3.2.1 Reviewer 1

While I agree with TCEQ's ultimate decision to treat isoprene as if there were no threshold, I believe the discussion in Section 4.2.4 (and elsewhere when MOA is discussed) is incredibly inconsistent. The final paragraph of Section 4.2.4 starts with the statement that "[s]cientific evidence suggests that carcinogenic effects observed from isoprene exposure are mediated by its genotoxic metabolite …." The tenor of this sentence and the preceding paragraphs is such that it appears as if the genotoxic MOA is fairly well (perhaps very well) established. However, the final sentence of Section 4.2.4 states that the "isoprene MOA is not well understood." That seems like a weak rationale when a much stronger case could be made. I have additional comments about the inconsistencies within this document with respect to MOA, presented in response to subsequent questions.

Moreover, as will be discussed below, the analysis in Appendix A[1] does *not* support the idea that exposure intensity has a greater impact than exposure duration, despite the claim in line 19, p. 9. This entire paragraph (lines 11-25, p. 9) is a rather unsophisticated mish-mash of unsupported statement and confusing information. There appears to be confusion about non-linear and threshold behavior; you can have the first without having the second. Despite what Placke et al., (1996) may claim, there is no reasonable way to conclude from their partial-lifetime-exposure study that a threshold is evident; very reasonable, non-threshold models describe their data. In fact, the presentation by Sielken et al. (Appendix A) made a point of the fact that the multistage models fit to the data (after adjustments of dose and numbers at risk) were almost all linear. There are indeed features of the Placke et al., (1996) data and analysis results that need explaining, but reasonable explanations can be provided (see discussion below) that do not necessitate invocation of a threshold.

In conclusion, the non-threshold approach used by TCEQ is completely justified. It is just that the document has done a relatively poor (and weaker-than-needed) job of supporting it.

[1] The reference to Appendix B on line 18 of p. 9 is misleading and should be removed; Appendix B is just the mathematical details supporting the calculations of Appendix A and has no bearing on the issue under discussion, i.e., the relative impact of exposure intensity vs. exposure duration.

3.2.2 Reviewer 2

The general conclusion is appropriate, but the discussion is conflating several concepts. The choice of the appropriate dose metric is related to, but only one component of, determining the MOA and appropriate extrapolation method. Determining the appropriate dose metric can aid in understanding the shape of empirical dose-response curves, and can linearize curves that initially appear nonlinear. For example, if metabolism is saturated, a curve may appear to have a high-dose plateau, while use of the appropriate internal dose metric could result in a linear dose-response. In the case of isoprene, the statement about exposure intensity having a greater impact than exposure duration on response frequency (a statement that should be reconsidered, based on Reviewer 1's comments) speaks to the relevant choice of dose metric, but does not address the issue of MOA, particularly whether the chemical acts via interaction with DNA or another MOA. Although the text ultimately concludes that the MOA is not well understood and so linear dose-response assessment should be done, the interweaving of the issue of dose metric confuses the issue. Consideration of dose metric is important, but should be conducted as an initial step in the MOA evaluation. For example, if a different conclusion had been reached regarding the dose metric, it might have been appropriate to conduct linear extrapolation, but using a different dose metric.

It is also unclear what the author considers to be the weight of evidence regarding the MOA. While I understand that a complete MOA analysis is not done if the data are not considered strong enough to support a nonmutagenic MOA, it would aid in transparency to communicate where the preponderance of the data is considered to lie. In some places the author states that cancer is due to a genotoxic metabolite, but much of the presented genotoxicity data are negative, and no weight of evidence evaluation of the genotoxicity data is provided.

The text twice notes that the methyl group on the diepoxide suppresses the cross-linking activity of the isoprene metabolite relative to that of the corresponding butadiene metabolite. It would be useful to briefly expand on this idea, explaining the relevance of the cross-linking activity to the butadiene MOA and how that informs the potential MOA for isoprene. Similarly, the NTP stop-exposure study may provide some useful insight regarding MOA, but the data are not presented in such a way as to facilitate such an evaluation.

3.2.3 Reviewer 3

Yes, the summary is clear and accurate. The most important aspect seems to be that because of insufficient information on MOA, the dose-response evaluation resorts to a default use of a nonthreshold model. The MOA discussion (section 4.2.4) could be improved by removing the concentration vs. duration issue, or by justifying how that issue informs the understanding of MOA.

3.2.4 Reviewer 4

The authors effectively present the existing evidence that supports, but does not prove, a mutagenic MOA for isoprene carcinogenicity involving the isoprene dioxirane metabolite. No alternative MOA is presented or discussed. The TCEQ RG442 guidelines (Section 5.7.5.1.2) make it clear that mutagenic activity *in the target tissue* early in the carcinogenic process must be demonstrated to support a mutagenic MOA for carcinogenicity. Such information is not available for mouse liver or lung, or for lymphoid tissue. Accordingly, the authors correctly state that the MOA for isoprene carcinogenicity is not well understood, and cannot be assumed to be mutagenic.

3.3 Were the most appropriate studies (Melnick et al., 1994a; Melnick et al., 1999; and Placke et al., 1996) selected for the dose-response assessment and was their selection sufficiently described and justified? (Charge Question 6.a)

3.3.1 Reviewer 1

There is relatively little to say about the study selection. I agree with the decision not to use the epidemiological studies. In that case, the three rodent bioassays appear to provide the best information about isoprene carcinogenicity.

3.3.2 Reviewer 2

Yes, these studies were well-conducted cancer bioassays in rats and mice. However, the Melnick studies are usually cited as NTP, with citation by the first author only when the study is published in a peer-reviewed journal. Citation to NTP also provides some immediate context not available just by saying at the beginning of the sentence that this was an NTP study.

Given the complexity of the data, it would be very useful for the authors to provide tables of exposure scenarios, concentrations and observed tumors, to aid in assimilating the results. Figure 4-1 helps with that, but this should be done for all the tumor targets before narrowing down to the ones being focused on. Tables 1 and 2 of Placke et al., (1996) and some of the summary tables in the appendix would help.

3.3.3 Reviewer 3

I am not sufficiently qualified in toxicology to judge that. The justification seemed well described.

3.3.4 Reviewer 4

These 3 studies contain all the available isoprene carcinogenicity bioassay data in rats and mice, and so it is clearly appropriate to consider all of them for the dose-response assessment. Melnick et al., (1999) is based on a final NTP technical report, which should be referenced as it is the primary data source for rats:
NTP (National Toxicology Program). 1997 draft. NTP Technical Report on the Toxicology and Carcinogenesis Studies of Isoprene (CAS No. 78- 79-5) in F344/N Rats (Inhalation Studies). Technical Report No. 486. NIH Publication No. 97-3976.

However, Placke et al., (1996) clearly provides the most extensive dataset for carcinogenicity in mice, and appropriately was chosen for the dose-response. This selection was adequately described and justified.

3.4 Adjustments were made to the data to account for differences between the exposure durations and times of response observation, continuous exposure duration, and number of study animals, (Section 4.2.6.1). Are these adjustments biologically appropriate? Were the correct approaches used to adjust the data for each? (Charge Question 6.b)

3.4.1 Reviewer 1

TCEQ is to be commended for finding alternative means of dealing with the less-than-lifetime exposure scenarios and with the varying durations of the selected studies. The bottom line is that I think the adjustments were performed correctly, despite some issues and potential confusion of the supporting material in Appendices A and B.

3.4.2 Reviewer 2

Although these adjustments are biologically appropriate in general, I found the relationship (or apparent lack thereof) to the analysis of the appropriate dose metric confusing. In particular, the analysis in Appendix A points to concentration being the primary determinant of response, with ppm x weeks and ppm x hr x weeks being less predictive, was not clear. In that case, it was not clear why it was appropriate to move forward with the Armitage-Doll modeling where duration of exposure is a key determinant.

I will also note that the one metric that seemed most predictive – ppm and hours/day – did not seem to be part of the modeling, to the degree I was able to follow. In Appendix A, Table 8, compare groups 3 and 4; also compare groups 9 and 10 vs. 11 and 12. In particular, the results for the latter four groups, where response is constant for a given ppm and hrs/day (regardless of weeks of exposure), but varies with hrs/day, suggests that hrs/day of exposure is an

important determinant of response, at least at this high concentration. This suggests that short-term repair may relate to hours/day of exposure, and may be important in determining the ultimate response. That said, the details of the methods used for these adjustments are beyond my area of expertise. It would be useful to include in the main text a less mathematical description of the analysis by Sielken et al. of the various dose metrics (i.e., concentration alone vs. concentration and duration), the results and their implications for the modeling. Based on the data of Placke et al., (1996), please address the appropriateness of conducting the stated adjustment.

3.4.3 Reviewer 3

The adjustments mostly derive from the use of the Armitage-Doll multistage model and prior history of such adjustments by the US EPA and other regulatory agencies, and so, as applied in this draft they are appropriate and mathematically consistent. The biological justification in general for such adjustments is not strong, being mostly supported by empirical studies, not by biological processes that have a duration component.

3.4.4 Reviewer 4

Adjustments are generally made to bioassay data when a significant number of test animals do not survive to the end of the study, and various approaches have been used by different authors/agencies to accomplish this. The adjustments made in Section 4.2.6.1 are carefully described, and are certainly acceptable as one rational approach to data adjustment. I'm not sure there is a single "correct" approach.

3.5 Hepatocellular carcinoma, alveolar/bronchiolar carcinoma, and histiocytic sarcoma were selected as human-relevant cancer endpoints for the dose-response assessment. Was the selection of these endpoints clearly explained and justified? Do you agree with what was chosen? (Charge Question 6.c)

3.5.1 Reviewer 1

I have concerns about the endpoints selected for deriving the POD. These concerns are primarily based on a comparison of the EC_{10} values presented in Section 4.2.6.2 (Table 4-2) relative to EC_{10} values shown in Appendix A. For example, Section 4.2.6.2 states that the cancer must be malignant (why?) and have a statistically significant (at what confidence level?) dose effect. However, there are tumors analyzed in Appendix A that satisfy those criteria and were not included. One example is hemangiosarcoma in the spleen in female mice (Placke et al., 1996), for which the EC_{10} was 16.51 (for m=1), less (more "sensitive") than any of those shown in Table 4-2. Moreover, the reliance on Placke et al., (1996) alone, as slipped into the discussion on line 20, p. 14, has not been discussed or

justified. There are other malignant, statistically significant endpoints from the NTP studies that could be considered, and which would be more health protective in the sense of estimating lesser values for the EC_{10}.

It is also the case that the Sielken et al. analysis (Appendix A) provided EC_{10} estimates for some of the endpoints selected in Section 4.2.6.2, but for NTP and Placke et al., (1996) combined. For example, liver carcinoma was a selected endpoint. It was also statistically significant in the NTP study. Why was that study not selected for that particular endpoint or, more importantly, why was no discussion provided for or against the combination of that endpoint across the two studies? Such a combination would have provided a lesser value for the EC_{10}, compared to the one presented in Table 4-2.

In this respect, it appears that TCEQ has not made selections that are consistent with adequate health protection.

3.5.2 Reviewer 2

While the authors explained their rationale for choosing the three specified tumor endpoints, based on their malignancy, this choice does not seem to be consistent with the TCEQ guidelines. I could not find any place where the TCEQ guidelines specifically address the issue of benign vs. malignant tumors, but the TCEQ guidelines for that section heavily reference the EPA (2005) guidelines, which are very clear that the determination of whether to include benign tumors should be made on a case-by-case basis, and benign tumors should be included in certain cases. As stated in the EPA guidelines (p. A-5), "*The default is to include benign tumors observed in animal studies in the assessment of animal tumor incidence, if such tumors have the capacity to progress to the malignancies with which they are associated.*"[2] This means that benign and malignant tumors of the same cell type should have been combined, based on the evidence that the benign tumors progressed to malignant ones. Thus, the dose-response assessment should include alveolar/bronchiolar carcinoma *and* adenoma, and hepatocellular carcinoma *and* adenoma. Additional analysis of the adversity of the benign tumors in other tissues would be needed prior to excluding them. Furthermore, harderian gland tumors cannot be excluded merely because the site does not occur in humans; any such exclusion needs to be conducted on a *mode of action* basis. Harderian gland tumors are commonly seen with chemicals acting via a mutagenic MOA, which is consistent with the observations for isoprene, based on the (conservative) MOA conclusion of the assessment.

Although the inclusion of multiple additional endpoints would generally entail additional work, visual inspection of the data would indicate that the response for

[2] The EPA guidelines further state (p. 2-2), "Observation of only benign neoplasia may or may not have significance for evaluation under these cancer guidelines. Benign tumors that are not observed to progress to malignancy are assessed on a case-by-case basis."

some of the endpoints is sufficiently low that it may not be necessary to model those endpoints. And of course, Sielken et al. modeled "all" endpoints. Which raises the question of why the most labor-intensive task, of doing the detailed modeling using a range of different approaches, was conducted prior to the choice of endpoints. One usually completes the hazard characterization first, determining which endpoints are relevant to humans based on biology, prior to conducting the dose-response modeling. Doing the modeling first and then choosing the endpoints for consideration of the point of departure raises the specter of "cherry-picking" the data.

3.5.3 Reviewer 3

I am not sufficiently qualified in toxicology to judge that. The justification seemed well described.

3.5.4 Reviewer 4

Hepatocellular carcinoma (HCC), alveolar/bronchiolar carcinoma, and histiocytic sarcoma *in male mice* were chosen as human-relevant malignant neoplastic endpoints. It is important to note that none of these three malignant tumor entities occurred in excess in rats, which casts doubt on the idea that their occurrence can be extrapolated to humans or any other species. It is explained that these three were chosen because they are all malignant and of types that can occur in humans, and because malignancy rather than any lesser category of abnormal growth was considered to be of most relevance to humans. The problem with this argument is that both the lung tumors and the hepatocellular tumors are types that clearly evolve through several morphologically identifiable steps ("stages?" see questions 6d and 12, below), while no non-malignant precursor lesion is known for histiocytic sarcoma. The lung tumors evolve in the sequence alveolar/bronchiolar hyperplasia → adenoma → adenocarcinoma, and the liver tumors in the sequence foci of alteration → hepatocellular adenoma → HCC. Because this progressive sequence is well recognized (neither alveolar/bronchiolar adenoma nor hepatocellular adenoma is an end-stage lesion), and the histological criteria for differentiating adenoma from carcinoma are in some cases difficult to apply, the adenomas and carcinomas in lung and in liver are commonly grouped together in the tabulation of bioassay results by most authoritative bodies; *viz*. alveolar/bronchiolar adenomas *and* carcinomas; hepatocellular adenomas *and* HCCs. In my opinion, this "lumping" approach should have been adopted as a second alternative to the "malignancies only" method, and the full dose-response calculations should be carried out in parallel using both datasets. The 3-dimensional graphs in Appendix A include this approach, but the authors did not carry it further.

3.6 Benchmark dose modeling was conducted on the adjusted data for the endpoints identified, with the EC_{10} calculated for each cancer stage (m = 1, 2, 3). Was it appropriate to base the final URF on the number of stages with the lowest EC_{10}? Do

you agree with the selection of the best estimate, EC_{10} , (e.g., rather than the lower bound of the estimate, the LEC_{10}) as the point of departure (POD), and did TCEQ authors provide sufficient justification for this selection? (Charge Question 6.d)

3.6.1 Reviewer 1

It is my firm belief that the selection of the EC_{10} from the model with m=1, because it produced the lowest EC_{10}, is exactly the wrong decision. The reasons for that belief are outlined here.

First, it should be noted that the Sielken et al. (Appendix A) analysis ignored, without any justification, some modeling choices or options. All of their analyses (whether using m=1, 2, or 3) assumed that it was only the first transition that was dose dependent. Although I believe there is some support for that decision (see next paragraph), it is not provided; options in which other transitions were the dose-dependent ones were never explored. Moreover, there are options that would consider more than one transition to be dose-dependent; they too were not considered, nor was the decision not to consider them even acknowledged. All of that is to say that selection of the m=1 option because it gave the lowest EC_{10} is based on consideration of a small subset of model choices, so I have no confidence that, even if that were a worthwhile basis for selection, the true minimum EC_{10} was identified.

Second, and more importantly, there are reasons to believe that the m=1 selection runs counter to the evidence from the Placke et al., (1996) data. To understand this, I must discuss and critique the analysis in Appendix A, Section 4, which purports to show that exposure intensity has greater impact than exposure duration. The gist of the calculations and analysis are in Table 9 of Appendix A.

The claim is that because the EC_{10}s for the ppm metric are similar and the EC_{10}s for the ppm-x-weeks or ppm-x-weeks-x-hours are different, across data subsets, then this shows that intensity is more important. That is simply not the case. Once the EC_{10}s for the ppm metric are calculated for the three subsets of data, the results for the other two metrics are entirely predictable and determined, *regardless* of the importance of intensity vs. duration. That is because the various dose metrics are simple multiples of one another. For example, for the Duration = 80 weeks subset, the doses expressed as ppm-x-weeks are exactly 80 times the doses in ppm. And, the EC_{10} for that subset across the two metrics also differs by exactly 80 (ignoring round-off). The same is true for the other subsets, but with a multiple equal to the number of weeks of exposure. The ppm-x-weeks-x-hours metric is always 8 times greater than the corresponding ppm-x-weeks metric, and the EC_{10} values for the former are exactly 8 times greater than the EC_{10} values for the latter. This is just a simple fact of the modeling/mathematics and does not inform us about relative importance.

I do agree that the consistency of the estimated EC_{10}s from the ppm metric, across the three subsets that differ by duration of exposure, is important and needing of an explanation (or at least a meaningful discussion).[3] That is, on the face of it, it is unexpected that an exposure intensity experienced over a duration of 20 weeks should produce the same risk as when experienced over 40 or 80 weeks. However, an explanation is relatively straight-forward within the context of the Armitage-Doll multistage model that is under consideration here. If the dose-dependent transition between stages is the first one, and if there are many subsequent transitions or there are some subsequent transitions that have a relatively slow rate of occurrence, then these results are understandable. That scenario would postulate that after 20 weeks, any subsequent transitions from normal to first-stage cells would have little or no chance of progressing to full tumors, because the time needed to progress through the subsequent transitions is insufficient (because there are many of them or because there is one or more "rate-limiting" later transition). Thus, any dosing from week 21 to week 80 (or longer) would be "wasted" in the sense that the first-stage transitions occurring then would have no impact on tumor incidence, because the subsequent transitions do not have enough time to be completed.[4]

Therefore, in relation to the question of whether the m=1 model(s) should be chosen because they give the lowest EC_{10}, the answer must be no. Such models do not have any subsequent transitions beyond the dose-dependent one, so they could not satisfy either of the postulated conditions that would give approximately equal risk regardless of duration (i.e., many such transitions or slow transitions after the first one). Incidentally, this is not a threshold argument; all of the above can be true with no threshold behavior of any kind. Moreover, this is why I can accept that Appendix A only considered models for which only the first transition is dose-dependent (as alluded to above). However, Appendix A and the DSD are silent about that rationale; no explanation for the limitation to such models is provided.

With respect to the question about using EC rather than LEC values, I do not think that using EC values is the best policy. The consistent use of LEC values properly accounts for, within and across assessments, differences in experimental design. This is the one uncertainty, i.e., that associated with finite sample sizes, for which well-developed and rigorous numerical procedures exist. In other words, it is an uncertainty that can easily be addressed and used to improve consistency without resorting to hand-waving and other more nebulous comparisons of competing features that would tend to increase or decrease URFs.

[3] I have concerns about comparing the modeling results across subsets of the data using only one point from the estimated dose-response curves, i.e., the dose that gives a 10% extra risk. Rather, I would prefer to compare and contrast the entire dose-response curves. But, the important points of my critique can be made by reference to the summary value, the EC_{10}.

[4] These considerations do not include any effects of "latency" or the time needed for a fully cancerous cell to proliferate to become an observable tumor. But any such latency would have the same effect of making a 20 week exposure similar to an 80 week exposure if the latency was long enough.

Why sacrifice that rigor for some qualitative notion that using EC values is less conservative so as to compensate for the fact that a non-threshold approach was used (a reason cited in the DSD)? That sort of comparison is fraught with problems and poorly defined trade-offs. Since, as argued elsewhere in my review, there may be no compelling reason to think that there is a threshold; this trade-off may not even be applicable. So, here again, for this choice I think TCEQ could be faulted for not choosing a suitably health-protective option.

3.6.2 Reviewer 2

If the assessment is based solely on carcinomas (introducing a substantial nonconservative element), I cannot support the use of the EC_{10} based on the current analysis, which seems to be based primarily on balancing uncertainties and conservatisms. While such balancing is an important part of the assessment (see discussion of uncertainties), the number of earlier issues that are not fully considered precludes me from commenting on the specifics of these choices related to the POD, and raises the potential that the current assessment may not be adequately health protective.

3.6.3 Reviewer 3

I agree with the use of EC_{10} instead of LEC_{10} (see my comments previously about science vs. policy) and with the selection of the lowest value. The justification for selecting the lowest value seems mostly based on lack of knowledge (e.g., insufficient MOA information to help decide how many stages were operating) and so is consistent with approaches used by other regulatory agencies. This seemed very well described, especially the formula derivations and proofs as well as worked out numerical examples in the Appendix.

3.6.4 Reviewer 4

The use of the concept of cancer stage is very confusing in this analysis, and at least to me it appears ill-defined and in some respects arbitrary. See comments on this in my answer to Question #12 below. But, that said: I agree with selection of the best estimate, $EC_{10,}$ calculated for "cancer stage m = 1" and I agree that the authors provided proper justification for this selection. The URF corresponding to cancer stage m=1 was the largest value (URF = 8.1E-04 per ppm or 8.1E-07 per ppb), which resulted in the most conservative (i.e., lowest) calculated 10^{-5} risk air concentration, and it is the only value of m (m being defined as "the number of stages in the multistage carcinogenic process") that makes sense, ***as data for only <u>one</u> biological stage, full-blown malignancy, were included in the dose-response analysis.***

3.7 Were the analyses in the Appendix on the data for the critical effects correctly performed and were the conclusions adequately justified? (Charge Question 6.e)

3.7.1 Reviewer 1

I have already discussed the erroneous analysis or interpretation of Section 4 of Appendix A. Although I believe the calculations to be done correctly, there are other concerns about the material in the Appendix, as follows.

Many of those concerns relate to clarity and consistency of the presentation. In the equation in the middle of Appendix p. 28, for example, the subscripts on the "T" and "t" upper ends for the integrations do not appear to be subscripts at all; they appear to be parts of the integrands. An alternate notation (e.g., "t(i)" rather than "t_i") could clarify this. In general the equations and some parts of the text should be better proofed to see where erroneous subscripting etc. makes the intent of the calculations harder to determine (e.g., in most of the equations for adjusted number at risk, where a key subtraction symbol is subscripted so that it is hard to understand that subtraction is going on).

Then on p. 29, there are other hindrances to understanding. The references to Appendix A there should be to Appendix B. At the bottom of that page, the general description and specification of the equivalent dose, D, uses the index j to identify the dose-dependent transition ("j=1"), but the earlier, second equation on p.29 uses the index i. While in general indexing is not affected by the labeling, it is confusing here because j is not defined anywhere and it is not immediately apparent what it is referring to. Moreover, I believe that the proper definition would set j (or i) to 0, since the first transition is actually indexed with 0 (as shown in the definitions of $H(T_e)$ on pages 28 and 29). I think many of the table headings (Tables 10-15 at least) would need to be changed to "j=0" to be consistent with this proposed change. These are obstacles to coherent communication of the main idea.

Most important, for the sake of clarity and consistency, is the fact that Appendix A seems to confuse the parameters T_e and T_{end}. T_e is the time such that, if one were using the stepwise function d(t), the risk at T_e would equal the risk at T (normal lifetime) using D. It has nothing to do with the duration of the experimental observation period (which is T_{end}). This error was propagated to p. 13, line 9 of the main document. That may be an understandable confusion since the Appendix itself seems to erroneously refer to T_e as the duration of observation, in the following instances (there may be others I have missed):

a. In the last example on Appendix page 30, where it says that $T_e = 105$ (the duration of observation for some of the Placke et al., (1996) dose groups). Incidentally, I believe the calculated values of D shown in this example (and the others in Table 12) are correct, in that they used $T_e = 104$.
b. In the second full paragraph on Appendix p. 50, where the correct reference should be to T_{end} (T_e has no bearing on adjusted numbers at risk)

c. Repeatedly in Appendix B, T_e is referred to as the necropsy time (pp. 97, 99, 101, 103, 104, 107, 108, etc.).

For all intents and purposes, T_e should be set equal to T for all calculations of adjusted dose, since the desire is to find the lifetime (T) equivalent (D) for an intermittent exposure (d(t)) that occurs over the same length of time (T_e). I must confess that the "concept" of T_e is a difficult one to present and explain, as evidenced by my own rather convoluted definitions presented here.

Some additional issues/comments related to Appendix A are as follows.

In Table 12, it should be noted that the m=1 adjusted doses are the same for the two rows having ppm=2200, but either weeks=80 and hr/day = 4, or weeks =40 and hr/day = 8 (first and second from the last lines of the table). This fact highlights an assumption that is being made but that is not mentioned anywhere. That assumption is that it is okay to "average out" hrs/day (and days/week) in a standard/simple way, but that longer term intermittency needs the more complete (and involved) adjustment associated with intermittent exposure (i.e., for "on-off" patterns occurring over weeks or months as opposed to within weeks or within days). Why? In reality, the animals are being exposed to a much more complicated pattern of starting and stopping exposure which, in theory, may make some difference in the relevant adjustment of dose. To return to the example from Table 12 cited at the start of this paragraph, why would one necessarily think that 80 weeks of 4 hr/day exposure to 2200 ppm would give the same equivalent dose as 40 weeks of 8 hr/day 2200 ppm exposures? At the very least, the document should acknowledge that that assumption is being made and perhaps indicate likely consequences of it.

Table 16 presents results of statistical tests, but those tests did not use the adjusted doses nor the adjusted numbers at risk. Why not? A big part of the effort in this document was to get those adjusted numbers and it seems just as appropriate to use those in the statistical testing as in the dose-response modeling. Likelihood-based techniques could be used in place of Cochran-Armitage or Fisher's exact tests to accomplish that.

The characterization of the BMDS method for calculating the LEC (fourth paragraph on p. 50) is incorrect. That procedure does not mediate the identification of the LEC by finding the "largest slope that is not statistically detectable as a bad fit." The optimization method used in BMDS works directly with the potential LEC values, finding the smallest such value for which maximization of the likelihood under the constraint that the EC is equal to that value is "not statistically detectable as a bad fit." The incorrect description of the LEC calculation method is repeated on p.85.

I like the idea implicit in Tables 19 and 20 that endpoints can be combined across studies. I would even go so far as to suggest that even if the response of interest

was not statistically significant in one study, but was in another study, then those data could and should be combined (or at least examined to determine if such combination is appropriate). As stated earlier, it appears that TCEQ ignored the results in Appendix A related to such combinations, and I think that is a shame.

On the other hand, I think that Sielken et al. have missed an opportunity to do some additional and more rigorous exploration of the merits of various model options (e.g., choice of the value of m). Given my earlier suggestion that m=1 is the worst choice, the question remains as to what value of m should be used. One way to address that is to determine for what values of m[5] exhibit the greatest "consistency" across studies (and across the dose-group subsets within the Placke et al., (1996) study, as defined in Table 8). This could be done, for each choice of m by fitting the entire combined data set with a single multistage model; then fitting separate multistage models to each study (or each subset of Placke et al., (1996)). Comparing the maximum log-likelihoods (or associated BIC values) would determine for which values of m the combined fit was acceptably close to the separate-model fit. The value of m that gave the largest log-likelihood, for example, could be selected. Or, if several values for m were "close enough," TCEQ could then select the one with the smallest EC_{10} as the health-protective choice.

For Figures 1 – 18, the combinations that were not run (for which there was no URF or URF_UB calculated) should be distinguished from those that were run. My suggestion is just to have those "columns" (if one can call a square with 0 height a column) be greyed out. Then it will be much clearer what we should be focusing on in those figures and not just artifacts of the graphical display. For Figures 7-12 (and to some extent 13-18) there are a lot of combinations that were not run and that all congregate in one side of the plots. Why not just take those out, especially since these kinds of plots are not well suited for a quantitative appraisal but rather for a qualitative sense of relative differences? Also in reference to these figures (p. 61) it is not clear what is meant by "including severity" on line 4. Is that a reference to benign vs. malignant?

What is supposed to be Appendix C (but which is still labeled Appendix B) appears to be material irrelevant to the DSD. I recommend removing it from the document.

3.7.2 Reviewer 2

This is beyond my area of expertise.

[5] I would extend the possible values above m=3; no reason is ever given for restricting m to be less than or equal to 3.

3.7.3 Reviewer 3

There are way too many calculations presented in the Appendix to check. Their descriptions indicate the values are correctly calculated and my spot checking of some numbers for plausibility support that conclusion. The sections that use T_e and T_{end} need to be proofed, because there seem to be some inconsistencies. I first assumed T_e was the end of the exposure period, and T_{end} was the end of the study (thus end of observation period). Those imply two different concepts- the adjustment because of dosing that does not occur for the full lifetime, vs. the adjustment because the observation period is not lifetime (regardless of the dosing regimen). If these two concepts are to be part of the analyses, then the summary of each key study should include that information. Thus, section 4.2.5.3 on the Placke et al., (1996) study should include the observation period, even if just to say it was the same as the dosing period (i.e., 20, 40 or 80 weeks).

3.7.4 Reviewer 4

These analyses appear to have been carefully carried out. The conclusions, which in some respects are at variance with those of the NTP, are carefully described and appear to me to be adequately documented and justified.

3.8 Did the dosimetric adjustments and conversion into human equivalent concentrations follow TCEQ guidance (Section 4.2.6.3)? (Charge Question 7)

3.8.1 Reviewer 1

I am not intimately familiar with the TCEQ guidance, but I had raised a question about the rationale for treating isoprene as a Category 3 gas when it was initially identified as a Category 2 gas. Why change to Category 3 rather than Category 1?

3.8.2 Reviewer 2

The authors are correct in recognizing that isoprene is a "Category 2" gas, causing both systemic and respiratory tract effects. However, the guidance to use the Category 1 or Category 3 guidelines, "whichever is more relevant" is intended to mean that the dosimetric adjustments should be done on an endpoint-by-endpoint basis, using the adjustment approach that is most relevant *for that endpoint.* The intent is not to choose at this point (prior to the dosimetric adjustment) which endpoint is most relevant. This approach is further described in EPA's most recent documentation on the topic (U.S. EPA, 2012)[6]. Although this report was published after TCEQ's guidance, it is consistent with the then-current thinking in practice at the time of the writing of TCEQ's guidance.

[6] US EPA (2012). Advances in Inhalation Gas Dosimetry for Derivation of a Reference Concentration (RfC) and Use in Risk Assessment. EPA/600/R-12/044. Available at http://cfpub.epa.gov/si/si_public_record_report.cfm?dirEntryId=244650

This means that one needs to separately consider the systemic effects and respiratory tract effects of isoprene. Additional mechanistic consideration is needed in this case for the respiratory tract effects of isoprene. The "Category 1" – type DAF calculation is based on direct reaction with the respiratory tract tissue and is used for respiratory tract effects of chemicals that are rapidly reactive or highly water soluble. As described in the chloroprene IRIS assessment, chloroprene is neither rapid reactive nor highly water soluble, and so the respiratory tract effects from this chemical are likely due to the absorption and systemic distribution of the chemical, rather than direct reactivity. Consideration of the chemical properties of isoprene, specifically its log K_{ow} and reactivity, indicate that the same is likely to be true for isoprene – that the respiratory tract effects are likely due to absorption and systemic distribution of the chemical.

This means that the final conclusions of the analysis of the DAF for isoprene based on systemic effects and applying the Category 3 approach is correct. *However*, TCEQ has not adequately presented the line of reasoning supporting this conclusion. The information presented in the previous paragraphs is a critical part of the rationale for why the (more conservative) Category 1 adjustment was not applied for the lung tumors.

Finally, as noted by the assessment, there are substantial interspecies differences in metabolism that could lead to substantial interspecies differences in sensitivity. A key assumption of the default Category 3 dosimetry is that the parent chemical is the toxic form. For isoprene it seems more likely that a metabolite is the toxic form, which would lead to marked differences in dosimetry between humans and mice. I would encourage the authors to address the potential for developing a chemical-specific adjustment factor (CSAF) based on some measure of the relevant isoprene metabolite (or range of CSAFs based on different metabolites) in blood. It appears that the data are insufficient for derivation of a CSAF, but this is not really addressed until the discussion on uncertainties; it should be addressed early in the document. (I will also note that differences in maximal metabolic velocity do not necessarily translate into differences in amount metabolized, depending on whether enzyme capacity or blood flow to the organ is rate limiting. However, in this case it appears that differences in toxicity are consistent with differences in metabolic capacity.)

3.8.3 Reviewer 3

Yes, both the calculations and the explanation are well done. One area that could use revision concerns the decision not to include duration in the dose-response model. First, if that decision is only based on the studies shown in Table 9, then that constraint should be justified. The text suggests that similar results have been shown elsewhere but those other results are too briefly summarized to understand what they actually evaluated. Table 9 does show that concentration is the important aspect, not duration, but that might not be obvious to many readers. The

accompanying text should explain how Table 9 shows that, such as: If ppm x weeks were the more appropriate exposure metric (a la Haber's principle, Gaylor 2000), then the EC_{10} estimates for that metric should be roughly equal and those for ppm should be varied. Instead, the consistent results only appeared for exposure given merely as ppm alone.

[reference: Gaylor, DW (2000). The use of Haber's Law in standard setting and risk assessment. Toxicology 149(1):17-19.]

3.8.4 Reviewer 4

Isoprene on inhalation produces both respiratory and remote effects and is therefore classified as a Category 2 gas. According to the TCEQ Guidelines (TCEQ 2012), dosimetry for Category 2 gases is under review by USEPA, and therefore dosimetry equations for Category 3 gases (tumors at remote sites) were applied. This is consistent with TCEQ guidance. Appendix A (p. 81) states clearly that there are no data from which a Dose Adjustment Factor (DAF) can be calculated for isoprene, and that the factor used was simply "borrowed" from the structurally related substances 1,3-butadiene and chloroprene. This ought to have been brought forward into the main document instead of just being buried in the Appendix; it is appropriate as a working method, but clearly adds to the uncertainties in this exercise (see question 11).

3.9 The final URF was derived using a non-threshold approach using the best estimate excess cancer risk resulting from continuous exposure to isoprene at 1 ppb in air for each cancer stage and then selecting the most conservative EC_{10} (cancer stage m=1) for use in deriving the ESL. Was this appropriate and does it result in the most appropriate URF and $^{chronic}ESL_{nonthreshold(c)}$? (Charge Question 8)

3.9.1 Reviewer 1

I have stated concerns above related to the following issues asked in this question.

a. I do not think using the best estimate, EC_{10} (as opposed to the lower bound estimate, LEC_{10}) is most appropriate.
b. I do agree with the non-threshold approach, not because it is a default to resort to when there is lack of information. In the case of isoprene, there appears to be sufficient evidence arguing actively *against* a threshold hypothesis (at least for the data cited up to this point in the DSD – see additional comments below), and none of the statements from Placke et al., (1996) nor from Appendix A carry any weight in that argument.
c. I have argued against using the m=1 model for reasons related to the apparent lack of dependence of the risk on duration (at least for the durations that have been studied) and the fact that that suggests to me, in the context of the multistage model, that there must be transitions subsequent to the initial, dose-dependent transition.

As a consequence of the fact that I disagree with two of the three components mentioned here for URF and ESL derivation, I must conclude that the overall result is that the DSD has not identified the most appropriate URF and $^{chronic}ESL_{nonthreshold(c)}$. Add to that the fact that I also questioned the choice of the endpoints to consider, and I am afraid I have little confidence in the derived $^{chronic}ESL_{nonthreshold(c)}$.

3.9.2 Reviewer 2

As noted above, I have concerns about the choices of the types of tumors used for the modeling, and the decision to focus only on malignant tumors, excluding benign tumors of the same cell type. While I do not have the expertise to comment on the choice of the cancer stage and use of EC_{10} in combination, I would recommend that TCEQ support the choice of approach with other precedents in which the multistage Armitage-Doll model was used as the basis for the assessment.

3.9.3 Reviewer 3

Yes. The uncertainties were extremely well described and show clearly how using any other approach would be misleading about its accuracy and biological justification. As I stated previously, there should be a clear distinction between the science and policy issues. While the science might lead to the chosen EC_{10} the question of "appropriateness" might be more of a policy issue, e.g., where a balance might be sought between health protection and reasonable cost. One such example is the US EPA Office of Water approach where the maximum contaminant level goal (MCLG) is first determined and then the enforceable MCL is set, the latter also considering available technology and cost. Both the MCLG and MCL are published.

3.9.4 Reviewer 4

In my opinion, given that TCEQ guidance mandates a non-threshold approach for isoprene as a result of incomplete data to define an MOA, the above approach was appropriate. However, I am reluctant to endorse a chain of calculations that lead to an ESL that may be lower than the isoprene level present in human breath. Inhalation ESLs are chemical-specific air concentrations set to protect human health and welfare; $^{chronic}ESL_{nonthreshold(c)}$ calculation is mentioned in passing on p. 18 of the DSD but the actual figure is not given.

3.10 Did the document provide sufficient justification for the decision that isoprene has not been demonstrated to have a mutagenic MOA for liver carcinogenicity? (Charge Question 9)

3.10.1 Reviewer 1

If anything, I think the document has given too much weight to the possibility that there might not be a mutagenic component. As stated earlier, the document needs to make a decision about that, and I think there is enough of a basis for doing so. Once decided, the document need not try to "balance out" non-health-protective choices by saying that the assessment is already conservative enough because a non-threshold, mutagenic MOA has driven choice of POD and extrapolation approach, etc.

3.10.2 Reviewer 2

The conclusion appears to be justified, but the somewhat incomplete presentation of the material makes it hard to evaluate the support for the conclusion. The evaluation of mutagenic MOA should first address the data for isoprene's mutagenicity, addressing each of the three types of genotoxicity studies: point mutation, clastogenesis, and DNA damage. Data gaps in the standard testing battery should be noted. For example, it appears from the writeup that there are no gene mutation assays *in vitro* in mammalian cells, but this is not explicitly stated. Similarly, it appears that there are no *in vivo* mutation assays in mammalian systems (which is different from the finding of mutations in tumors). The inconsistency between the chromosome aberration assay and micronucleus tests is noted, but details of the micronucleus test are not provided in the text. In this case, the difference may reflect interspecies differences, particularly in metabolism, since the micronucleus test was in the sensitive species (mouse), while the chromosome aberration test was in CHO cells, from hamsters. DNA damage assays such as SCEs have the least weight with regard to evaluating mutagenic potential, since they only indicate that the chemical (or metabolites) has the potential to interact with DNA, not whether such interaction results in mutation.

3.10.3 Reviewer 3

I am not sufficiently qualified in mutagenicity to judge that. However, the justification seemed well described and consistent with TCEQ guidance. The uncertainty from conflicting evidence about mutagenicity supports that decision.

3.10.4 Reviewer 4

The document accurately cites the very stringent criteria that TCEQ requires to assign a mutagenic MOA for carcinogenesis in any specific organ/cell type, namely, that a mutagenic event related to the carcinogenic process must be

demonstrated *in the target tissue* prior to the beginning of the carcinogenic process (TCEQ RG442 section 5.7.5.1.2). Such cell/tissue mutation data are not available for mouse liver; hence, despite the strong evidence summarized in section 4.2.4 of the draft isoprene document that supports a mutagenic/genotoxic MOA for isoprene in general, the authors are forced to conclude that there is no conclusive evidence of such an MOA in mouse liver.

3.11 Was the decision not to apply age-dependent adjustment factors (ADAFs) (Section 4.2.7) to the URF, to account for potential increased sensitivity of children, justified and properly considered given TCEQ guidance on evaluating the carcinogenic MOA (see Section 5.7.5 of TCEQ 2012)? (Charge Question 10)

3.11.1 Reviewer 1

Here is another instance where the MOA discussion in the document is schizophrenic. In previous sections, the argument was that evidence suggested mutagenicity, so a non-threshold approach was selected. In Section 4.2.7, the final paragraph puts the emphasis on how isoprene has *not* been demonstrated to have a mutagenic MOA so no ADAFs should apply. As stated in response to the previous question, a decision needs to be made, however tough it might be to do so. But once made, follow through with that decision in all respects. This section reads like the complete obverse of the previous sections and introduces new references and considerations that were not cited when earlier MOA was under discussion. So, if the earlier (and newly cited) data suggest mutagenic MOA, then there needs to be follow through on ADAF decisions as well as modeling decisions. It does not appear that the isoprene assessment has been consistent in that regard.

3.11.2 Reviewer 2

It appears to be justified, but an improved evaluation of the chemical's mutagenic potential is needed for a clear conclusion.

3.11.3 Reviewer 3

Yes.

3.11.4 Reviewer 4

Section 5.7.5 of TCEQ RG442 (2012) clearly states that ADAFs will be applied only for carcinogens for which early-life exposure data are available, or (by default) those that operate by a mutagenic MOA. There are no early-life exposure data for isoprene, and no conclusive evidence that a mutagenic MOA operates in isoprene-induced mouse liver carcinogenicity, which is the basis of the URF

calculation. In my opinion, TCEQ was wise, given all the uncertainties surrounding isoprene, not to attempt age-dependent adjustment of the URF.

3.12 Section 4.2.8 presents an uncertainty analysis. Have all the key uncertainties been identified? Are the conclusions regarding these uncertainty issues and their impact on the URF correct and sufficiently discussed? (Charge Question 11)

3.12.1 Reviewer 1

It is generous to call Section 4.2.8 an "analysis." It is a recitation of the typical uncertainty considerations, all of which appear to be listed.

Specific issues that I identified are the following. With respect to site concordance (4.2.8.2), I think it is at least worth calling out the fact that in the two experimental species, there was absolutely no overlap of tumor sites. This is important because one of the criteria for endpoint selection was related to picking sites relevant to humans. But if there is no overlap even among closely-related animal species (rodents) then why would we expect site concordance between rodents and humans? I realize that "site not relevant to humans" as used in this document refers to organs that are not present in humans (forestomach, Harderian gland, etc.). but I have always wondered why such sites should be so routinely ignored, unless there is something unique about the metabolism, cellular interactions, DNA damage, etc. that occurs in those organs and would not occur in other (human) organs. This is default decision (not to consider sites "not relevant to humans") that warrants greater consideration here, because of the lack of site concordance among the rodents.

With respect to the dose-response assessment uncertainty (4.2.8.3), the discussion is about model uncertainty. That should relate to the differences in risk estimates produced by different models. In fact, since only the multistage model was used (albeit with different dose metrics), there is no basis for assessing model uncertainty in this analysis. To cite various statements from EPA about what *might* happen, across models, at the lower end of the observable range, is to throw up red flags that may or may not have any bearing on the isoprene assessment. Moreover, the last sentence in Section 4.2.8.3 about the closeness of the EC and LEC values is a complete non sequitur; considering the fit of one model has nothing to do with model uncertainty.[7]

Finally, in the interest of thoroughly beating that dead horse, Section 4.2.8.4 again raises the issue of threshold and non-threshold behavior. Once again some new considerations are introduced here at the end of the assessment. ***All*** of the data and observations relevant to the determination of thresholds or lack thereof should have been presented in one place, the first time some decision was needed about

[7] I would also argue strongly against using the similarity of the values of the EC and LEC to make determinations of "a tight data fit."

that issue. In this particular instance, the document has flopped back to the "yes, there is evidence of a threshold" camp. The document suffers badly from having several separate, schizophrenic and disjointed considerations of that issue, ones that present no unified or consistent conclusions or emphasis on the likelihood of such a threshold.[8]

3.12.2 Reviewer 2

Rather than simply focusing on two aspects of the dose-response assessment (choice of m=1 and use of the EC_{10}), I would encourage the authors to array all key decision points in the assessment, starting with the hazard characterization, with a qualitative statement, and perhaps even a semi-quantitative one, regarding the impact of the decision made on the final assessment. This approach can provide a much more complete picture of the results and implications of the balancing of uncertainties. For example, brief inspection of the modeling results indicates that the decision to exclude adenomas of the same cell type as carcinomas was nonconservative, by a factor of 5-fold or more. Conversely, the choice of the most sensitive animal species may introduce conservatism of a factor of at least a few.

Perhaps the biggest gap in the uncertainty discussion was the consideration of the *implications* of endogenously-produced isoprene. A key statement in this regard is the concentration of isoprene in the breath of nonsmokers. Since NTP (1999) calculated the concentration in air from the estimated amount exhaled per day, it would be useful to go back to the original studies, to evaluate reliability of the data and ensure consistency of the calculations. Given that Patty's Industrial Hygiene (Clayton and Clayton, as cited by HSDB) is the basis for the key numbers, I would recommend going back to the *original* studies, since I have seen unit conversion errors compounded when there are multiple citations of secondary sources.

If the data are indeed correct, that the lower end of the range of the concentration of isoprene in breath in nonsmokers is 50 $\mu g/m^3$ that may support the idea that the calculated concentration corresponding to a 1E-5 risk (33 $\mu g/m^3$) is adequately health-protective and likely over-protective. However, as noted in the text, most of the endogenously produced isoprene is metabolized, and the carcinogenicity of isoprene is likely due to its metabolites. Further consideration of the implications of the endogenous exposure is needed. Although it is possible that endogenous levels contribute to the overall disease burden, it is also appropriate to consider the endogenous exposure to a chemical as part of the risk management approach.

[8] The phrase "as well as exhaled" on line 12 of p. 21 makes no sense or is horrible grammar.

3.12.3 Reviewer 3

Yes. While there are other more detailed approaches for uncertainty analysis of dose-response modeling, such as model averaging and use of information criteria for model selection (e.g., the Akaike IC used by EPA), the discussion makes a strong argument that the other uncertainties and data weaknesses would not justify the extra determinations. The evaluation might be improved if the various decisions could be arranged in sequence. If the MOA is to be analyzed first, e.g., for evidence of mutagenic ability, then subsequent statistical modeling should not affect that MOA evaluation. The same goes for adjusting the dose metric to reflect duration (or not to): if the dose metric is based on biological grounds, then that decision precedes the dose-response modeling. If many decision points cannot be strongly supported because of lack of information, it would be useful to have a summary table identifying all the decisions that were default approaches because of inadequate data or understanding.

3.12.4 Reviewer 4

This section succinctly discusses interspecies differences, including metabolic differences and the lack of site concordance for tumor development between mice and rats. It would have been appropriate to note that in a large series of NTP bioassays of a wide variety of chemicals, tumor site concordance between mice and rats was very limited for most organ sites. One should not expect that tumor development at a given site or tissue in any one rodent species will accurately predict tumor sites in any other species, including humans (Haseman JK, Lockhard AM: *Environ. Health Persp.* 101, 50-54, 1993). The section goes on to discuss uncertainties involved in the dose-response analysis and linear low-dose extrapolation. These four issues are certainly the key uncertainties that impact the use of rodent bioassay data and their use to calculate the URF for isoprene, and they are well discussed. I concur strongly with the cautious note sounded on p. 21 (Section 4.2.8.4) on the very serious doubt that linear low-dose extrapolation is appropriate at all for isoprene in view of its endogenous production and metabolism in humans, and the very incomplete evidence regarding its carcinogenic MOA in mice and rats.

3.13 Please identify any other relevant issues or questions that are important for the review of this assessment. (Charge Question 12)

3.13.1 Reviewer 1

There are no additional considerations, issues, or questions that I need to raise at this point.

3.13.2 Reviewer 2

Minor/clarifying comments and questions:

P. 9: "Based on the analyses conducted by Sielken et al. (2012) (Appendix A and B) on behalf of TCEQ, they conclude that…" Does "they" refer to Sielken et al.? (in which case the sentence should be restructured to Sielken et al. concluded…) Or does it refer to TCEQ?

P. 15, Figure 4-1: Please add text explaining the vertical lines in the figure.

Also, as a process note, I understand the utility of focusing reviewers on key sections of the DSD. However, the text/tables on physical/chemical properties should always be provided, since this information provides critical context to issues such as determination of the appropriate dosimetric adjustment approach. I had to refer to other documents for key information such as the isoprene K_{ow}.

3.13.3 Reviewer 3

One suggestion to improve reader understanding would require significant revision, namely, to replace the 3D bar charts with standard 2D bar charts. All modern texts and journal articles on effective visualization and interpretation of quantitative information recommend against use of 3D bar charts for two main reasons. First, is that bars in front can hide shorter bars behind them. Second, the commonly chosen perspective view distorts the perception of the bar heights.

It also might be helpful to add a paragraph to the main text to explain what is contained in Appendix A: Dose-Response Modeling and Inhalation Toxicity Factors for Isoprene Report. In particular, explain why all the tables and graphs are included if only the Placke et al., (1996) study is deemed suitable for use in estimating the URF.

There are two minor revisions that should be made to improve the presentation; neither one impacts the procedures nor final calculations.

[p 17, line 26] The first relates to the decision to round off the URF and use that rounded value for calculating the ESL. I performed some example calculations to show that the early rounding can give different answers from those obtained by waiting to round only the final ESL value. E.g., consider URF=7.4091. Using the rounded value of URF=7.4 gives ESL=14 (rounded from 13.51), whereas using the more accurate value of URF=7.4091 gives ESL=13 (rounded from 13.4969) showing a slight difference between early rounding vs. calculating using the full value and then rounding, respectively. While such differences are small and rarely occur, some note should be added so the reader understands that TCEQ is aware of this issue.

[Appendix A, Dose-Response Modeling and Inhalation Toxicity Factors for 2 Isoprene Report, p 26]. The second relates to the description of the results from evaluating whether concentration or duration was more important in estimating the EC value. Table 9 shows EC_{10} values for 3 metrics (ppm, ppm*week, ppm*week*hr), all for liver adenoma/carcinoma, all for 8h/d exposure. The three data sets varied by the number of weeks (20, 40, 80). Thus the data are not varied by hours/day so that aspect cannot be checked. The approach and results are then valid, but redundant. The evaluation actually only compares (ppm) with (ppm*week). The resulting EC_{10} values for (ppm*week*hr) are proportional to those for (ppm*week) so do not add anything.

One editorial suggestion: Retitle Appendix A (the report on modeling) because it has its own Appendix A (Equivalent Doses), which can be confusing.

3.13.4 Reviewer 4

Isoprene is one of the very few chemical carcinogens (other than steroid hormones) that are produced endogenously in the human body. While Section 11 on uncertainties presents a succinct summary of the issues attending endogenous formation of isoprene, in my view an effort should be made to discuss the 10^{-5} excess cancer risk concentration in air calculated from the URF (12 ppb, or 33 $\mu g/m^3$ based on HCC in male mice from Placke et al., (1996) in comparison with the significantly higher levels of isoprene in the breath of non-smokers (50-400 $\mu g/m^3$; NTP Background Document for Isoprene, p. 5). How do you regulate meaningfully to protect public health when your "acceptable concentration" is lower than what normal people routinely exhale?

Another important relevant issue is the use of the "cancer stage" terminology without any clarification of what is meant. Armitage & Doll (1954, reprinted 2004), which is quoted repeatedly, was a trail-blazing effort to estimate, on the basis solely of epidemiological data, the number of successive events, "stages," through which **human** malignancies evolve, starting from a normal target cell and proceeding through successively different steps/lesions, and they arrived at the estimate of seven for at least some human carcinomas. That is very close to how, for example, the Vogelstein model of human polyp-derived colon carcinoma develops. Today, the multistage view of the evolution of many (not all!) cancers involves either a series of successive morphological changes that are histologically demonstrable, or successive genetic changes at the gene or chromosome level, or both together (as with human colon cancer). A completely separate but conceptually related use of the term "stage" to describe individual cases of human cancers is used in diagnostic pathology to describe the extent of neoplastic disease: Stage 1, confined to the organ/tissue of origin; stage 2, extending beyond the organ of origin but not invasive into neighboring tissues; stage 3, local invasion; stage 4, distant metastasis. These stages are clinically very important (get diagnosed at stage 1, find a good surgeon and live; get diagnosed at stage 4, fight like mad but prepare to die). Clinical stages for

established malignancies may also involve successive genetic changes beyond those involved in the original transition to malignancy, but the term "stage" is used very differently in the clinical sense than it is in the tumor biology sense, and it is unfortunate that it is not more clear just what is meant by this term as used in the DSD. In this DSD and its appendices "stage" is simply a word and an integer that can be plugged into certain mathematical formulae (although Appendix A does present the basic concept of progressive evolution of a cancer stepwise from a single normal cell: p. 31, Section 5.3). The DSD document would be immensely improved, and much more understandable, if an effort were made to reconcile the tangible, observable concept of "stage" with its use in the equations that have their origin in Armitage & Doll.\

4. References

Armitage P, Doll R (2004) The age distribution of cancer and a multi-stage theory of carcinogenesis. British journal of cancer 91(12):1983-9 doi:10.1038/sj.bjc.6602297Armitage & Doll (1954) The age distribution of cancer and a multi-stage theory of carcinogenesis. Brit J Cancer (1954) 8:1-12.

Clayton, G.D., F.E. Clayton (eds.) Patty's Industrial Hygiene and Toxicology. Volumes 2A, 2B, 2C, 2D, 2E, 2F: Toxicology. 4th ed. New York, NY: John Wiley & Sons Inc., 1993-1994.

Gaylor, DW (2000). The use of Haber's Law in standard setting and risk assessment. Toxicology 149(1):17-19.

Haseman JK, Lockhard AM. (1993) Correlations between chemically realted site-specific carcinogenic effects in long-term studies in rats and mice. *Environ. Health Persp.* 101, 50-54, 1993

Melnick RL, Bridge DA, Bucher JR, et al. (1994a) NTP technical report on toxicity sutdies of isoprene (CAS No. 78-79-5) administered by inhalation to F344/N rats and B6C3F1 mice. In: Program NT (ed). National Institute of Environmental Health Sciences

Melnick RL, Bucher JR, Chhabra RS, et al. (1999) NTP technical report on the toxicology and carcinogenesis studies of isoprene (CAS No. 78-79-5) in F344/N rats (inhalation studies). In: NTP (ed). National Institute of Environmental Health Sciences

NTP (National Toxicology Program). 1997 draft. NTP Technical Report on the Toxicology and Carcinogenesis Studies of Isoprene (CAS No. 78- 79-5) in F344/N Rats (Inhalation Studies). Technical Report No. 486. NIH Publication No. 97-3976

NTP (1999) NTP Report on Carcinogens Background Document for Isoprene. National Toxicology Program, Research Triangle Park, NC

Placke ME, Griffis L, Bird M, Bus J, Persing RL, Cox LA, Jr. (1996) Chronic inhalation oncogenicity study of isoprene in B6C3F1 mice. Toxicology 113(1-3):253-62

Sielken RL, Valadez-Flores C, Bretzlaff RS (2012) Dose-Response Modeling and Inhalation 23 Toxicity Factors for Isoprene. Sielken & Associates Consulting Inc.

US EPA (2012). Advances in Inhalation Gas Dosimetry for Derivation of a Reference Concentration (RfC) and Use in Risk Assessment. EPA/600/R-12/044. Available at http://cfpub.epa.gov/si/si_public_record_report.cfm?dirEntryId=244650

Appendix A - Charge Questions and Instructions for Peer Reviewers

This page left intentionally blank.

Experts Selected by TERA to Peer Review TCEQ Isoprene Section 4.2 Carcinogenic Potential - Developmental Support Document, April 2013

TERA independently selected the following four experts to provide independent peer review of the TCEQ document. Each has been screened for conflict of interest. None of the selected experts has a conflict of interest with the review of this document.

Bruce Allen, M.S.

Mr. Bruce Allen received a Master of Biomathematics with a minor in statistics from North Carolina State University. He has 30 years of experience related to human and environmental health and safety assessment. He has expertise as a biomathematician involved in risk assessment, modeling, statistical analysis, and clinical trials, having worked for a variety of government and private clients. Mr. Allen has extensive experience with the quantitative aspects of risk analysis such as dose-response analysis; statistical appraisal of data, models, and modeling results; and developing rigorous approaches to decision making in risk assessment contexts. His expertise in doseresponse analysis extends to modeling, including biologically motivated modeling of cancer, noncancer, and genotoxic endpoints as well as genomics data. Mr. Allen's statistical expertise includes computer-intensive approaches such as Monte Carlo simulation, bootstrap analysis, and Markov chain Monte Carlo approaches for Bayesian analyses. In addition he is an expert in other techniques for uncertainty analyses, data quality objectives, quality control/assurance, statistical analyses for site risk assessments, and analysis of clinical trials data. Mr. Allen has provided expert testimony, is a frequent peer reviewer of risk assessment documents, and has served as manager for numerous projects including multi-disciplinary, multi-year efforts. Mr. Allen has authored or coauthored more than 50 journal articles and has been given over 30 invited presentations at scientific meetings on bio-statistics, modeling and a wide range of quantitative risk
assessment issues.

Lynne Haber, Ph.D., DABT

Dr. Lynne Haber received a Ph.D. in biology from the Massachusetts Institute of Technology. She has extensive experience in the developing and reviewing documents conducting hazard characterization and dose-response assessments for chemicals, and in risk assessment methods development. She has led the development of numerous assessment documents, and has been a coauthor or technical reviewer of 100's more. She currently serves as the Associate Director of TERA (Toxicology Excellence for Risk Assessment). She has served as a panel chairperson or panel member for scientific peer reviews organized by TERA, and U.S. and international government agencies, and

has served on two panels for the US National Academy of Sciences. Dr. Haber is active in communicating her findings to the broader scientific community through participation in professional societies, routine publication of her work, authoring book chapters (including lead author of the chapter on noncancer risk assessment for Patty's Toxicology, 2001, 2011), service as an editorial reviewer for scientific journals, and through presentation of invited lectures. She has done research into issues such as methods for improving the scientific basis for uncertainty factors by addressing genetic polymorphisms and risk to children; consideration of mode of action in cancer risk assessment; and use of biomarker data in risk assessment. She has served in several leadership positions the Society for Risk Analysis (SRA) Dose-Response specialty group and as an officer of the Society of Toxicology (SOT) Risk Assessment Specialty Section (RASS), and is a Diplomate of the American Board of Toxicology (ABT). She is one of the lead teachers for TERA's Dose-Response Assessment Boot Camp, developed a course on issues related to children's risk assessment, and presents specialized courses to diverse groups of risk assessors and at professional society meetings.

Rick Hertzberg, Ph.D.

Dr. Richard Hertzberg received a Ph.D. in biomathematics from the University of Washington. He is an adjunct professor in the Department of Environmental Health at Emory University where he teaches graduate courses in risk assessment in environmental health. Dr. Hertzberg has extensive experience with mathematical modeling for quantitative risk assessment, specializing in biomathematical modeling and health risk assessment of chemical mixture exposures. Prior to Emory University he led the research program on mixture risk assessment at the US EPA, National Center for Environmental Assessment (NCEA) and was instrumental in writing the EPA mixture risk guidelines. He initiated the use of categorical regression for dose-severity modeling, and the interaction-based hazard index for mixture risk assessment. In addition to his position at Emory, he is a private consultant, specializing in biomathematical methods for environmental health risk assessment of exposures to chemical mixtures including categorization by method and application and development of quantitative methods for cumulative risk assessment of chemical and nonchemical stressors for cumulative risk assessment, including interactions. During his career Dr. Hertzberg received two silver medals and four bronze medals for mixture risk assessment methods and cumulative risk assessment guidance. He also received the Distinguished Achievement Medal in environmental statistics from the American Statistical Association.

Jerry Rice, Ph.D.

Dr. Jerry Rice received a Ph.D. from Harvard University. He is an internationally recognized expert in experimental carcinogenesis and in carcinogenic hazard identification. Dr. Rice is a Distinguished Professor of Oncology and a Member of the Cancer Prevention and Control Program at the Lombardi Comprehensive Cancer Center, Georgetown University Medical Center. Prior to Georgetown University, Dr. Rice managed the IARC Monographs on the Evaluation of Carcinogenic Risks to Humans program at the World Health Organization, and is now an ad-hoc advisor for the program. Dr. Rice also serves as science expert on the Joint FAO/WHO Expert Committee on Food Additives. He has published more than 250 research and review papers on tumor promotion, identification of chemical carcinogens, transplacental carcinogenesis and the molecular pathology of human and experimental tumors during his 30-year career at the NCI and has received countless invitations to present his work. He is a member of the Society of Toxicology, Toxicology Forum, European Association for Cancer Research, and the International Society for Differentiation. Dr. Rice has received numerous PHS Medals for outstanding contributions to the field of cancer research.

Appendix B - Charge Questions and Instructions for Peer Reviewers

This page left intentionally blank.

Scientific Peer-Review of the Carcinogenic Section (Section 4.2) of the Isoprene Development Support Document Review Guidelines

Introduction and Instructions

The peer reviewers are asked to provide their opinions and comments on specific and general questions. For each response (including the Yes/No questions), please explain your reasoning and considerations, discuss scientific support for your comments and opinions, and identify the sources you consulted to construct your response. Please address each charge question by adding your answers to this Word document; and reference the TCEQ document page, paragraph, and line number, where appropriate.

Your written review should be returned to nance@tera.org by email no later than May 24, 2013.

Background

The Toxicology Division of the Texas Commission on Environmental Quality (TCEQ) has prepared a draft Development Support Document (DSD) that outlines the hazard assessment and dose-response processes used to derive health-protective Effects Screening Levels (ESLs) and Reference Values (ReV) for Isoprene. The draft DSD includes Section 4.2, which documents the derivation of an inhalation unit risk factor (URF) based on liver carcinoma incidence in mice and air concentrations corresponding to the policy-based 1 in 100,000 excess risk level. These toxicity values are used in the evaluation of air permit applications and ambient air data and were developed using RG-442 TCEQ Guidelines to Develop Toxicity Factors (TCEQ 2012).

We are asking you to provide a review of the scientific approaches used by TCEQ in developing the URF for isoprene as described in the Carcinogenic Potential (Section 4.2) of the draft DSD. Information on the metabolism of isoprene is provided in Section 3.1.3 as additional background information for your review, but you are not being asked to review this section. The DSD is a summary document and does not provide a detailed description of every aspect of the toxicity assessment for a chemical. References to appropriate papers or documents are provided if more detailed information is needed. Please contact Ann Parker (parker@tera.org) if you wish to see a copy of any of the cited references.

There are a number of policy decisions the TCEQ has made and included in this assessment that they do not seek comment on. For example, risk management goals were approved by the Commissioners and Executive Director of the TCEQ and are consistent with other TCEQ programs. Therefore, please do not spend your time commenting on the policy-based excess risk level (1E-05) and default lifetime exposure assumption of 70 years.

General

Please evaluate strengths and weaknesses of the procedures used to develop the URF based on the specific questions described below. Where possible, try to put the strengths and weaknesses in perspective by indicating their relative magnitude. Please try to avoid emphasizing minor technical details. Reviewers should identify scientific uncertainties and suggest ways to reduce or eliminate those uncertainties.

1. **Does the draft DSD clearly describe the approaches used by TCEQ to develop the URF?**

2. **Were procedures outlined in RG-442 *TCEQ Guidelines to Develop Toxicity Factors* (TCEQ 2012) followed by the TCEQ in this assessment?**

3. **Please identify any relevant studies or data that have not been cited and would affect an important part of the assessment and explain how they would impact the assessment specifically.**

Cancer Assessment and Unit Risk Factor (URF)

The draft Isoprene DSD describes the approaches used to evaluate carcinogenicity and derive the URF and the chronic ESL (at the 1E-05 excess risk level) for cancer in Section 4.2. Please review the key decisions made by TCEQ in deriving these values.

In formulating your response to each question, please consider and comment on the consistency of the assessment with TCEQ's RG-442 guidelines, the scientific appropriateness of the decision or conclusion, and any additional approaches or additional information that would improve that decision/conclusion.

4. **Section 4.2.3 briefly presents carcinogenic weight of evidence classification information and conclusions of authoritative bodies and TCEQ's weight of evidence conclusion. Is TCEQ's weight of evidence conclusion appropriate?**

5. **Section 4.2.4 discusses isoprene's carcinogenic mode of action (MOA). Have the authors clearly and accurately summarized the available data and hypotheses for isoprene's mode of action? [Please keep in mind that the purpose of the DSD is to document the derivation of the URF and ESL as opposed to being a comprehensive weight of evidence paper on the MOA. Ultimately, if data on the MOA are not sufficient to justify an alternate approach to linear low-dose extrapolation, the DSD only needs to generally summarize the primary proposed MOAs, MOA issues, and justify use of the default extrapolation method.]**

6. **Please comment on the following key decisions in the TCEQ assessment. For each, please discuss if the conclusions and choices are supported by the available data and discuss any additional information, data, or analyses that could improve the decision.**

    a. **Were the most appropriate studies (Melnick et al., 1994a; Melnick et al., 1999; and Placke et al., 1996) selected for the dose-response assessment and was their selection sufficiently described and justified?**

    b. **Adjustments were made to the data to account for differences between the exposure durations and times of response observation, continuous exposure duration, and number of study animals, (Section 4.2.6.1). Are these adjustments biologically appropriate? Were the correct approaches used to adjust the data for each?**

    c. **Hepatocellular carcinoma, alveolar/bronchiolar carcinoma, and histiocytic sarcoma were selected as human-relevant cancer endpoints for the dose-response assessment. Was the selection of these endpoints clearly explained and justified? Do you agree with what was chosen?**

    d. **Benchmark dose modeling was conducted on the adjusted data for the endpoints identified, with the EC_{10} calculated for each cancer stage (m = 1, 2, 3). Was it appropriate to base the final URF on the number of stages with the lowest EC_{10}? Do you agree with the selection of the best estimate, EC_{10}, (e.g., rather than the lower bound of the estimate, the LEC_{10}) as the point of departure (POD), and did TCEQ authors provide sufficient justification for this selection?**

    e. **Were the analyses in the Appendix on the data for the critical effects correctly performed and were the conclusions adequately justified?**

7. **Did the dosimetric adjustments and conversion into human equivalent concentrations follow TCEQ guidance (Section 4.2.6.3)?**

8. **The final URF was derived using a non-threshold approach using the best estimate excess cancer risk resulting from continuous exposure to isoprene at 1 ppb in air for each cancer stage and then selecting the most conservative EC_{10} (cancer stage m=1) for use in deriving the ESL. Was this appropriate and does it result in the most appropriate URF and $^{chronic}ESL_{nonthreshold(c)}$?**

9. **Did the document provide sufficient justification for the decision that isoprene has not been demonstrated to have a mutagenic MOA for liver carcinogenicity?**

10. **Was the decision not to apply age-dependent adjustment factors (ADAFs) (Section 4.2.7) to the URF, to account for potential increased sensitivity of children, justified and properly considered given TCEQ guidance on evaluating the carcinogenic MOA (see Section 5.7.5 of TCEQ 2012)?**

11. **Section 4.2.8 presents an uncertainty analysis. Have all the key uncertainties been identified? Are the conclusions regarding these uncertainty issues and their impact on the URF correct and sufficiently discussed?**

12. **Please identify any other relevant issues or questions that are important for the review of this assessment.**

Appendix C – Public Comments

This page left intentionally blank.

TERA received one set of public comments on the Isoprene document. The comments are provided below:

Public Comment:

RE: Peer Review of Section 4.2, Carcinogenic Potential, of the *Development Support Document for Isoprene, Draft April 2013*

Thank you for the opportunity to provide comments on the Texas Commission on Environmental Quality's draft Development Support Document (DSD) outlining the hazard assessment and dose-response processes used to derive health-protective Effects Screening Levels and Unit Risk Factors for isoprene. These comments are being made on behalf of the members of the International Institute of Synthetic Rubber Producers, Inc. (IISRP), a non-profit trade association representing the interest of the synthetic rubber industry. Our members, who produce and/or consume isoprene, have a keen interest on this matter as we have conducted significant scientific research on this substance and have provided the findings to governments and/or regulatory bodies.

Our scientists have reviewed the Draft of Section 4.2 of the DSD and have determined it to be representative of the available data and support the analysis of the data. We therefore have no additional comments.

This review has been based upon the scientific data only and not upon any policy decisions made by the TECQ. As noted the industry has been involved in health research of isoprene over the years and we would be pleased to assist in addressing any issues that might arise from the review.

Sincerely,
Sent via email

James L. McGraw
Managing Director & CEO
IISRP
2077 S. Gesner Road # 133

Houston, TX 77063
Phone: 713.783.7511 ext 222
jlmcgraw@iisrp.com