Final Report for the Peer Review of the Acute and Chronic Noncarcinogenic Sections of the Arsenic Development Support Document and Follow up Conference Call

Peer Review organized by
Toxicology Excellence for Risk Assessment (*TERA*)
2300 Montana Ave, Suite 409, Cincinnati, OH 45211, 513-542-7475

October 14, 2009

PREPARED IN COOPERATION WITH THE
Texas Commission on Environmental Quality

Table of Contents

Table of Contents 2
Introduction 3
General Issues 3
Health-Based ^{Acute}ReV and ^{acute}ESL$_{noncancer}$ 4
Health-Based ^{Chronic}ReV and ^{Chronic} ESL$_{noncancer}$ 7
References 10
Appendix A - Peer Review Charge 12
Appendix B - Panel Written Comments 16
Appendix C - Follow-up Conference Call Charge 35
Appendix D - Biographical Sketches for Reviewers 39
Appendix E - List of Conference Call Observers 43
Appendix F – Comments on PBPK 45

Introduction

The Toxicology Division of the Chief Engineer's Office, 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 Effects Screening Levels (ESLs) and Reference Values (ReV) for arsenic (CAS. No 7440-38-2). The toxicity values were developed using RG-442 *Guidelines to Develop Effects Screening Levels, Reference Values, and Unit Risk Factors* (TCEQ 2006). ESLs are chemical-specific air concentrations set to protect human health and welfare. Short-term ESLs are based on data concerning acute health effects, odor potential, and vegetative effects, while long-term ESLs are generally based on data concerning chronic noncarcinogenic and/or carcinogenic health effects. ESLs are used in the evaluation of air permit applications as well as proposed rules and regulation (e.g. Permits by Rule). ReVs and URFs, used as the basis of ESLs, are used in the evaluation of air monitoring data and in the development of Protective Concentration Levels for remediation sites.

TERA is supporting the Texas Commission on Environmental Quality (TCEQ) in conducting an expert external peer review of acute and chronic noncarcinogenic sections of the *Development Support Document for Arsenic and Inorganic Arsenic Compounds, Preliminary Draft, April 2009*. The review materials, including draft document, charge to reviewers, and key references (available at http://www.tera.org/Peer/arsenic/index.html) were distributed to the panel in July 2009. Reviewers submitted written comments that addressed the charge questions in August 2009. These written comments represent the panel's review of the arsenic acute and chronic noncarcinogenic section of the DSD and are attached as Appendix A. On August 25, 2009, *TERA* facilitated a follow-up conference call between the panel and TCEQ. Conference call materials (available at http://www.tera.org/Peer/arsenic/index.html), including a focused charge and reviewer comments were distributed prior to the call; members of the public were allowed to listen to the call. The purpose of this call was to allow TCEQ to ask the panel questions regarding their written comments and to allow the panel members to discuss issues on which there were divergent opinions in the written comments. Therefore, this report of the follow-up conference call, along with the written comments submitted by the panel, comprises the complete peer review of the arsenic acute and chronic noncarcinogenic sections of the DSD.

General Issues

Are the peer reviewers aware of any information or references that would support (or refute) the position that data for arsenic trioxide should be considered relevant for arsenic pentoxide?

The panel was not aware of specific data on the comparative toxicity of arsenic trioxide and arsenic pentoxide. The panel agreed that arsenic pentoxide is converted to arsenic trioxide, and the use of arsenic trioxide data would be a health protective approach. One reviewer stated that arsenic pentoxide also has a separate mode of action, but would likely cause effects at higher doses than arsenic trioxide. So using arsenic trioxide as a surrogate may be conservative; it

would not be predictive but it would be protective. This reviewer suggested that TCEQ add a discussion of this issue to the DSD.

Does the panel have additional recommendations related to the rationale for using 10 µM as a size-selective cut point for the ESL?

One panel member asked TCEQ to clarify what they meant by "size-selective cut point". TCEQ clarified that the monitoring data was based on particles of 10 um or smaller because particles size are respirable. Therefore, TCEQ had originally intended to make the ESL applicable to just certain size particles. But based on the initial panel's written comments, TCEQ decided to use a total particulate matter approach rather than using the cut-off of 10 microns. Another panel member suggested that a short discussion addressing the possibility of systemic effects from inadvertent exposure to larger particle sizes due to mucociliary clearance from inhalation exposure be added to the report as further justification of using total particulate for the assessment.

Health-Based ^{Acute}ReV and ^{acute}ESL$_{noncancer}$

The panel felt that the Nagymajtenyi et al. (1985) study is not an adequate study to develop an ^{acute}ReV and should be replaced by the Holson et al. (1999) study as the key study. One reviewer noted an alternative explanation (Reviewer #2) of the Nagymajtenyi et al. (1985) findings and indicated that this study might be used to develop a supporting value. If such an approach were used would the following derivation be reasonable? The critical effect would be maternal toxicity, with the highest dose a LOAEL and the mid-dose of 2.9 mg/m^3 a minimal LOAEL or a NOAEL. No BMCL modeling or RDDR calculation would be included due to weaknesses in the study. Is the use of the Nagymajtenyi et al. (1985) to derive only a supporting value to supplement Holson et al. (1999) scientifically defensible?

The panel discussed the unreliability of the study design and the lack of adequate documentation for the Nagymajtenyi et al. (1985) study. One panel member stated that he did not have confidence in this study because the exposure could have been by the oral pathway in addition to inhalation. The panel member noted that animals in Nagymajtenyi et al (1985) were exposed to a "fog" that would have deposited chemical on their coat. Therefore, the animals could have had oral exposure in addition to inhalation exposure when they groomed. In contrast, the animals in the Holson study were washed after exposure, so ensuring exposure was only by inhalation. Thus, the Holson et al. (1999), study would be more reliable for developing an inhalation assessment.

TCEQ stated they felt the Nagymajtenyi et al. (1985) study was more conservative, and was more of an acute duration study. Acute studies are preferred to drive 1-hr ESLs. However, repeat-dose, multi-day studies could be used to develop short-term ESLs if an acute study was not available. TCEQ asked the panel if it would be useful to still derive a value based on the Nagymajtenyi et al. (1985) study as a co-critical or supporting study if the Holson et al. (1999) study were selected as the key study for the revised assessment. One panel member responded that if the study is not credible enough to use as the critical study, then no value should be derived from it. The panel agreed that even though the Nagymajtenyi et al. (1985) study was qualitatively similar to the Holson et al (1999) study, the Nagymajtenyi et al. (1985) study is

unreliable and should not be used for any quantitative estimates in support of the Holson et al (1999) study. Rather, the panel recommended that a paragraph be added that discusses the limitations of the study as well as the interpretations of the data and its qualitative consistency with the Holson et al. (1999) study.

One panel member discussed the issue of pretreatment of the rats in the Holson et al. (1999) study. Sprague-Dawley rats are resistant to arsenic due to the arsenic binding with the rat's hemoglobin. The rats were pretreated to ensure that all sites on the hemoglobin were occupied before assessing reproductive toxicity. The panel suggested that TCEQ should include a discussion on the pretreatment protocol to provide more confidence in the study.

The panel concluded that the Nagymajtenyi et al. (1985) study is unreliable and should not be used for any quantitative estimates or as a co-critical study in support to the Holson et al. (1999) study. The panel also recommended a paragraph be added that discusses the limitations of the study as well as the interpretations of the data and its consistency with the Holson et al. (1999) study. Additional information on the pretreatment of the rats in the Holson et al. (1999) study should also be included to provide more confidence in the study.

The critical effects observed in the Holson et al. study (1999) were respiratory effects, (i.e., point-of-entry effect) and decrease in body weight. If the Holson et al. (1999) study were used as the basis for the ^{Acute}ESL would route-to-route extrapolation to derive values based on other critical endpoints observed in acute oral studies add significantly to confidence in the ESL derived from this study?

One panel member stated that in the range finding studies, many animals died early due to suffocation that resulted from the accumulation of dust in the lungs. The panel member added that the amount of arsenic in the lungs would probably not be enough to be designated a critical effect, but it should be noted that pulmonary irritation was seen in the animals and could be noted as a portal-of-entry effect. Another panel member agreed and added that route-to-route extrapolation would only be useful as support for a weak study; however the Holson et al. (1999) study is robust, so route-to-route would not be needed.

TCEQ added that all the females died at the top three exposures due to suffocation and deaths were also seen in repeated-doses at the 25 mg/m^3 exposure. The study did not do any pathology and it was reported that the lungs did not show any pulmonary irritation. There were also signs of gastrointestinal tract lesions, which are consistent with arsenic exposure, and that decreased body weight was listed as a critical effect. One panel member replied that the study reported rales in the lungs which are a sign of lung inflammation. The panel member suggested that the assessment more clearly describe the observed effects and characterize those which are likely respiratory effects and those which may indicate systemic effects.

The panel agreed that a quantitative route-to-route extrapolation approach was not necessary since a reliable inhalation study was available and would not add to the quantitative assessment. One panel member noted that comparisons to the oral database would help alleviate any concerns about whether the Holson et al. (1999) identified the likely critical effects.

Do the data support the notion that concentration and duration both play a role in the toxicity of arsenic? If so, for the Holson et al (1999) study, rats were exposed to arsenic as a solid particulate. Should Haber's Rule as modified by ten Berge, with n = 3, be used to extrapolate the 6-hr per day exposure duration to a 1-hr exposure duration?

One panel member noted that it is important to determine which dose-metric is relevant to toxicity in order to determine if a duration adjustment is appropriate. This panel member indicated that since toxicity of arsenic is related to area under the curve, then cumulative exposure is the important dose metric. And since cumulative exposure is the important dose metric, then it is appropriate to use Haber's Law. When Haber's law is appropriate, using the ten Berge modification is more health protective. Another panel member noted that the ten Berge modification would not be as important for exposure to particulates versus exposure to a vapor or gas.

The panel agreed that the 6-hour concentration should not be used directly to represent the effects for a 1-hour exposure; rather, the 6-hr exposure from Holson et al. (1999) should be adjusted to a 1-hr equivalent using Haber's Law with the ten Berge modification since the critical effects may reflect the effects of cumulative dose and systemic effects are possible. TCEQ noted that their guidelines require the use of an "n" equal to a value of 3 as a default for the ten Berge modification unless other data suggest a better value. The panel agreed that there are no chemical-specific data to support a different choice of "n", so a default of 3 is appropriate.

If used as a supporting study, the Nagymajtenyi et al. (1985) study exposed mice to arsenic as a liquid aerosol. Should the Haber's Rule as modified by ten Berge, with n = 3, be used to extrapolate the 4-hr per day exposure duration to a 1-hr exposure duration?

The panel agreed, based on previous discussions, that the Nagymajtenyi et al. (1985) study should not be used in any qualitative analysis, so no answer to this charge question was needed.

Are the oral studies Stump et al. 1999 and Baxley et al. 1981(Reviewer #2) adequate justification for assuming that mice are not more sensitive than rats to developmental toxicity? Holson et al. (1999) states rats are more sensitive than mice based on Lewis and Sweet (1985). Would inclusion of a discussion of these studies enhance confidence in the selection of Holson et al. (1999) as the critical study?

One panel member stated that, based on the literature for arsenic, mice and rats appear to have similar sensitivity. The panel member noted that the overall database suggests that high exposure concentrations or doses could result in reduced birth weights; the other observed effects are consistent with maternal toxicity. This reviewer also stated that the available studies support the conclusion that arsenic does not cause malformations at environmentally relevant concentrations. The Lewis and Sweet (1985) reference is no longer in print and could not be reviewed.

The panel concluded that it would benefit the assessment to add a discussion of the oral studies (Stump et al. 1999 and Baxley et al. 1981) which support the overall database, including the Holson et al. (1999) study. A discussion should also be added to state that the general literature did not suggest sensitivity difference between the rat and mouse (which further supports the

selection of the Holson et al. (1999) study over the poorly documented study by Nagymajtenyi et al. (1985).

TCEQ asked if the panel would comment on whether arsenic is a developmental toxicant. The answer to this question is not simple and requires some explanation. "Developmental toxicant" is a broad term that replaced "teratogen," which refers to a substance that causes fetal malformations. Developmental toxicity encompasses multiple fetal effects that include alterations in fetal weight, death of offspring, and functional deficits in addition to malformations. The presence of any of these findings is sufficient to signal developmental toxicity. One panel member stated that arsenic does not cause malformations via typical routes of exposure because the dam cannot remain sufficiently healthy and take in enough arsenic to cause fetal malformations. However, the issue of decreased fetal weight from exposure to arsenic is more complicated. If dams receive a high enough concentration of arsenic to cause toxicity but not death, then reduced fetal weight could occur. This response could also possibly occur in humans. Consequently, based on the possibility that sufficiently high maternal arsenic exposure could result in fetal weight deficits, arsenic could be considered developmentally toxic at very high exposure concentrations. There is possible confounding in that the intoxicated pregnant animals will likely stop eating and reduced fetal weights could be an indirect effect. The panel member added that environmentally relevant exposures (low concentrations) would not cause any fetal effects, but at concentrations high enough to cause severe toxicity in the pregnant woman could lead to fetal effects (likely reduced fetal weight). Importantly, such effects would occur in association with serious maternal toxicity. Because of the presence of concomitant maternal toxicity, arsenic should ***not*** be considered a selective developmental toxicant. (A selective developmental toxicant is one that causes effects in the offspring at exposures that are not harmful to the pregnant female and is very serious). The panel did not have a discussion on uncertainty factors during the conference call, but did include discussion of uncertainty factors in their individual written comments (see Appendix B).

Health-Based ^{Chronic}ReV and ^{Chronic} ESL$_{noncancer}$

Based on the written comments, panel member's had varying opinions as to whether the limitations in the Lagerkvist and Zeturland (1994) and Lagerkvist et al. (1986) studies precluded their use to develop a ^{chronic}ReV. Nevertheless, all the panel members noted significant concerns with these studies. As an alternative approach, would an unit risk factor (URF) and ^{chronic}ESL based on excess lung cancer mortality in four different cohorts of workers exposed to arsenic be a more robust basis for the ^{chronic}ESL and would an ESL derived on this basis protect against sensitive noncancer effects?

To provide background, TCEQ indicated that their guidelines do not necessarily require them to derive separate noncancer and cancer values. Rather, they must evaluate the effects from long-term exposure. TCEQ indicated that this is the first time they are considering moving away from deriving a separate chronic noncancer value for a known carcinogen. They felt this approach would be health protective because a quantitative assessment based on the carcinogenicity usually provides a lower toxicity value. A panel member asked if TCEQ would be using a linear extrapolation or a non-linear assessment approach (point of departure based on a cancer endpoint and application of uncertainty factors) for the carcinogenic assessment. TCEQ replied that they would do the typical cancer assessment assuming a linear dose-response because the arsenic

mode of action for lung tumors is not well defined. Overall, the panel agreed that deriving only a cancer assessment would adequately characterize the long-term effects of arsenic and would likely be protective of potential noncancer effects of arsenic. However, in order to address this uncertainty, the panel suggested that TCEQ include a comparative analysis of the range of values that could be derived from the different endpoints and qualitatively discuss the results to support the selected value.

The panel discussed the most appropriate data set for developing a chronic inhalation assessment. One panel member stated that the most appropriate data to use for the chronic inhalation assessment is from the occupational epidemiology related to lung cancer. If chronic noncancer effects data were to be used, then the epidemiology literature related to cardiovascular or gastrointestinal effects should be considered as the basis for a chronic assessment. TCEQ should use the drinking water epidemiology studies to examine these endpoints. However, such an approach was not preferred because it would incorporate route-to-route extrapolation considerations and lower the confidence in the quantitative assessment. Another panel member asked if the use of the lung cancer studies would be protective of sensitive noncancer effects. Another panel member replied that most studies found gastrointestinal effects at high levels of exposure, which were higher than what you would get from a workplace inhalation exposure. However, the panel members agreed that a comparative analysis should be done to be to document this consideration.

TCEQ asked if it would be helpful to develop a reference value based on the Lagerkvist and Zeturland (1994) study and present it in the DSD to give a bound or range of potential values. A panel member replied that presenting a range is a reasonable approach that has been used by other agencies, such as the U.S. EPA. TCEQ clarified that they could develop a value for the chronic noncancer for the purpose of showing that the cancer value is protective. A panel member replied that the problem with that approach is it gives credibility to a value that is based on a single study that is not completely credible, whereas the value based on the cancer endpoint is based on an entire database and not a single study. The panel member discouraged this approach because a value derived from the Lagerkvist and Zeturland (1994) would give undue credibility to the Lagerkvist and Zeturland (1994).

The panel concluded that the chronic value should be based on the cancer inhalation database. The noncancer studies are weak and do not provide adequate data for derivation on a chronic ESL with confidence. The use of the most appropriate noncancer effects data would be based on cardiovascular and gastrointestinal effects seen in drinking water studies, although this approach also was not preferred because the route-to-route extrapolation from oral to inhalation would introduce significant uncertainty. It was suggested that a qualitative discussion of the relative sensitivity of the cancer and noncancer endpoints would provide more support for the use of the cancer inhalation data as the final basis for the chronic assessment.

Do the overall data support a causal relationship between arsenic exposure and cardiovascular effects?

One panel member stated that heart disease was seen in arsenic exposures from drinking water, specifically data from Taiwan. Cardiovascular effects can lead to heart disease, so it should be considered. Another panel member noted that studies showed confounding exposure to lead; it is

biologically plausible that the observed effects were related to arsenic exposure. However, it is not possible to rule out the effects due to co-exposure with lead. The panel concluded that based on the drinking water data, cardiovascular effects due to arsenic exposure appear to be biologically plausible, but additional examination of the overall data would be needed to provide a firmer conclusion.

As part of a sensitivity analysis to better inform the weight of evidence, the TD is considering supplementing the analysis of the chronic ESL based on the epidemiology studies by using experimental studies and toxicity values from oral routes of exposure to calculate an inhalation toxicity value: (1) USEPA's RfD (Tseng et al. 1977), (2) ATSDR's MRL (Tseng et al. 1977) and (3) Cal EPA's inhalation REL (Wasserman et al. 2004). This would involve route-to route extrapolation or the use of PBPK models to derive an inhalation ReV, for comparison purposes only. These values will be compared to the air concentration corresponding to a 1 in 100,000 excess risk for lung cancer mortality using the URF derived by the TD. Given the uncertainties associated with route-to-route extrapolation (see the ESL methodology document for the current policy of using route-to-route extrapolation), would such an approach be sufficiently robust to inform the selection or evaluation of the chronic ESL? Do any data limitations as highlighted in the ESL methodology document preclude the meaningful use of route-to-route extrapolation for arsenic with the above mentioned oral toxicity studies?

One panel member stated that route-to-route extrapolation is not necessary because of the availability of robust lung cancer mortality data, and it may cause confusion if used. Another panel member added that, generally, route-to-route extrapolation can be a useful approach, but agreed it is not needed for this assessment. The panel agreed the Wasserman et al. (2004) study needs further review and has not yet been supported with studies by other investigators. The panel member stated that TCEQ should only use the verified endpoints, such as cardiovascular and gastrointestinal effects as seen in the drinking water studies. Other potential effects, such as intellectual development, incorporate many additional variables and are difficult to effectively measure.

However, the panel did suggest that route-to-route extrapolation would be a useful tool as part of the sensitivity analysis to evaluate the protectiveness of a cancer value for noncancer effects. If done as part of the sensitivity analysis, the extrapolation should use the identified endpoints for cardiovascular or gastrointestinal effects as the primary non-cancer effects of concern.

If the TD were to develop a route-to-route extrapolation approach to supplement the ESL derived based on the occupational epidemiology what studies should be used? Three agencies have used Tseng et al. (1977) and Wasserman et al. (2004) as key studies in the development of chronic values (RfD, MRL, and chronic REL) Are these the appropriate studies to use as the key studies in the development of chronic values based on route-to-route extrapolation?

A panel member began the discussion stating that the Wasserman et al. (2004) study would not be appropriate for use due to the many uncertainties. Another panel member agreed and added that the Tseng et al. (1977) study should be used along with oral animal data to determine if the cancer value is health protective.

TCEQ clarified with the panel member that they were recommending the use of the oral animal data. TCEQ then asked if they should use the human oral data which would include less uncertainty that using animal oral data. One panel member replied that the animal data may give support for the human data. Another panel member added that the animal data is not predictive of human response, and they should review epidemiological studies to determine which studies would be the most appropriate to use in the sensitivity analysis. A panel member added that there are some epidemiological studies from Taiwan and should also review the epidemiological literature for more studies, such as the Tseng et al. (1977) study. They also recommended review articles, such as the ATSDR toxicological profile, on chronic health effects of arsenic to find useful human studies.

The panel agreed that the Wasserman et al. (2004) study had too many uncertainties and would not be appropriate to use in a quantitative route-to-route extrapolation. If oral studies were to be examined as a supplement to the inhalation studies, the epidemiology studies, rather than toxicology studies should be the focus. Before proceeding along this course the panel felt that the pool of available epidemiological studies should be critically reviewed prior to selecting any studies or endpoints for further quantitative assessment.

The preferred method for route-to-route extrapolation is the use of PBPK modeling, which provides the best estimate of a toxicant's internal and biologically effective dose as a function of exposure. The peer reviewers suggested that TD determine if available PBPK models (e.g., Mann et al. 1996) are adequate to conduct route-to route extrapolation. Has the Mann et al. (1966) model been accepted by the scientific community as a valid model that could be used for this purpose? No federal or state agency has used this model to develop inhalation toxicity values based on oral studies. In fact, in their June 2008 Draft TSD for Noncancer RELs, Cal EPA states the following "while some PBPK modeling has been applied to inorganic arsenic and its methyl metabolites, the modes of toxic action and relevant internal dosimetry are not sufficiently understood at present to use this modeling directly in REL development". Is the panel aware of other models or other route-to-route extrapolation approaches that should be considered?

The panel discussed the use PBPK modeling to enhance the route-to-route extrapolation. One reviewer stated that he had a PBPK modeler review the current status of available models for arsenic. A written summary statement was provided for inclusion into this report (see Appendix F). Panel members noted that the use of PBPK models do not involve all the key uncertainties, TCEQ asked if enough is known about the mode of action to pick the appropriate dose metric. A panel member replied that the metabolites of interest would need to be known. TCEQ added that the U.S. EPA states that the mode of action for arsenic is unknown. Another panel member agreed and recommended precluding it for use quantitatively, but indicated that it might be used qualitatively as part of the overall sensitivity analysis.

References

Baxley MN, Hood RD, Vedel GC, et al. 1981. Prenatal toxicity of orally administered sodium arsenite in mice. Bull Environ Contam Toxicol 26:749-756.

Holson, JF, DG Stump, CE Ulrich, et al. 1999. Absence of prenatal developmental toxicity from inhaled arsenic trioxide in rats. *Toxicol Sci* 51:87-97.

Nagymajtényi, L, A Selypes, G Berencsi. 1985. Chromosomal aberrations and fetotoxic effects of atmospheric arsenic exposure in mice. *J Appl Toxicol* 5(2):61-63.

Lagerkvist, BJ, B Zetterlund. 1994. Assessment of exposure to arsenic among smelter workers: A five-year follow-up. *Am J Ind Med* 25(4):477-488.

Lagerkvist, B, H Linderholm, GF Nordberg 1986. Vasospastic tendency and Raynaud's phenomenon in smelter workers exposed to arsenic. *Environ Res* 39:465-474.

Stump DG, Holson JF, Fleeman TL, et al. 1999. Comparative effects of single intraperitoneal or oral doses of sodium arsenate or arsenic trioxide during in utero development. Teratology 60:283-291.

Tseng W. 1977. Effects and dose-response relationships of skin cancer and blackfoot disease with arsenic. Environ Health Perspect 19:109-119.

Wasserman GA, X Liu, F Parvez, et al. 2004. Water arsenic exposure and children's intellectual function in Araihazar, Bangladesh. *Environ Health Perspect* 112(13):1329-1333.

Appendix A - Peer Review Charge

Peer Review of the
Acute and Chronic Noncarcinogenic Sections of the
Arsenic Development Support Document
Charge Questions

The Toxicology Division of the Chief Engineer's Office, 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 Effects Screening Levels (ESLs), Reference Values (ReV) for arsenic (CAS. No 7440-38-2). The toxicity values were developed using RG-442 *Guidelines to Develop Effects Screening Levels, Reference Values, and Unit Risk Factors* (TCEQ 2006). ESLs are chemical-specific air concentrations set to protect human health and welfare. Short-term ESLs are based on data concerning acute health effects, odor potential, and vegetative effects, while long-term ESLs are generally based on data concerning chronic noncarcinogenic and/or carcinogenic health effects. ESLs are used in the evaluation of air permit applications as well as proposed rules and regulation (e.g. Permits by Rule). ReVs and URFs, used as the basis of ESLs, are used in the evaluation of air monitoring data and in the development of Protective Concentration Levels for remediation sites.

We are asking you to provide a review of the scientific approaches used by TCEQ in developing the acute and chronic non-cancer toxicity values that are described in this draft document. 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.

There are a number of policy decisions the TCEQ has made and included in this assessment that they do not seek comment on. These risk management goals were approved by the Commissioners and Executive Director of the TCEQ and are consistent with other TCEQ programs. ESLs developed in accordance with these no significant risk levels are intended to prevent adverse effects potentially associated with cumulative and aggregate exposures as defined in Section 1.2 of RG-442 ESL Guidelines (TCEQ 2006).

Therefore, please do not spend your time commenting on the following policy decisions:

- The use of a hazard quotient (HQ) of 1 for assessing an individual chemical with a nonlinear dose-response assessment for the ReV
- In situations of cumulative and aggregate exposure, the use of an HQ of 0.3 to calculate short-term and long-term health-based ESLs for chemicals with a nonlinear (threshold) dose-response assessment (i.e., health-based ESLs = 0.3 x ReV).
- Assumption of a lifetime exposure of 70 years.

General Issues

Please consider all aspects of the arsenic DSD and evaluate strengths and weaknesses of the procedures used to develop acute and chronic toxicity values 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 or making tutorial comments. Reviewers should identify scientific uncertainties and suggest ways to reduce or eliminate those uncertainties.

- Were procedures outlined in the ESL Guidelines followed by the TCEQ to perform arsenic's toxicity assessment? If references to accepted procedures in federal, state, or other appropriate guidance documents were made in the ESL Guidelines, were those accepted procedures followed?
- Does the arsenic DSD clearly describe the approaches used by TCEQ to perform the toxicity assessment (i.e., hazard identification and dose-response assessment).

Health-Based Acute ReV and ^{acute}ESL

Chapter 3 of the arsenic DSD describes the approaches used to derive the health-based acute ReV and acute ESL. The key decision points are listed below. For each decision, please consider the scientific defensibility of the decision and any additional approaches or analyses, or additional information that could improve that decision. Please indicate if there are other issues specific to developing acute toxicity factors that have not been adequately addressed in the document.

- The choice of the critical study (Nagymajtenyi et al. 1985)
- The choice of critical effect (decrease in fetal body weight)
    - Was the most appropriate critical effect selected? If not, what would be a more appropriate critical effect?
    - Is the endpoint relevant for human risk assessment?
- Benchmark dose modeling:
    - Was the output from the most appropriate model selected?
    - Was the appropriate critical effect size selected (5% decrease in mean fetal body weight compared to control means)
    - Should the point of departure (POD) be based on the maximum likelihood estimate or the 95% lower confidence limit of the benchmark response?
- The choice of point of departure
- The choice of dosimetric adjustments
    - Was the most relevant, appropriate, and defensible dose metric selected?
    - Were the appropriate default exposure duration adjustments conducted?
    - Were the appropriate default dosimetry adjustments from animal-to-human exposure conducted? Specifically, were the appropriate estimates for conducting the regional deposition dose ratio (RDDR) chosen when the key study did not report the required parameters?
- The choice of uncertainty factors.
    - Have all of the appropriate uncertainty factors been considered and are the values assigned to the uncertainty factors clearly justified and defensible?
    - Would you make recommendations for a different approach to select uncertainty factors to calculate the acute ReV?

Health-Based Chronic ReV and ^{chronic}ESL$_{noncancer}$

Chapter 4 of the arsenic DSD describes the approaches used to derive the health-based chronic ReV and chronic ESL for noncancer endpoints. The key decision points are listed below. For each decision, please consider the scientific defensibility of the decision and any additional approaches or analyses, or additional information that could improve that decision. Please indicate if there are other issues specific to developing chronic toxicity factors that have not been adequately addressed in the document.

- The choice of the critical studies (Lagerkvist and Zetturland (1994) and Lagerkvist et al. (1986))
    - Were the relevant occupational epidemiologic studies selected for the non-cancer estimates? Were defensible reasons provided to exclude the other available studies?
- The choice of critical effect
    - Was the most appropriate critical effect selected? If not, what would be a more appropriate critical effect?
    - Is the endpoint relevant for human risk assessment?
- The choice of dosimetric adjustments
    - Was the most relevant, appropriate, and defensible dose metric selected?
    - Was the time-weighted average (TWA) exposure for arsenic (as As_2O_3) based on the Lagerkvist and Zetturland (1994) study and the Lagerkvist et al. (1986) study calculated correctly by ATSDR (2007)?
- The choice of point of departure.
- The choice of uncertainty factors.
    - Have all of the appropriate uncertainty factors been considered and are the values assigned to the uncertainty factors clearly justified and defensible?
    - Would you make recommendations for a different approach to select uncertainty factors to calculate the chronic ReV?

Welfare-Based Acute and Chronic ESLs

The TD did not find any data to allow the derivation of welfare-based acute or chronic ESLs. Please indicate if there are other issues specific to developing welfare-based ESLs that have not been adequately addressed in the document.

Appendix B - Panel Written Comments

Reviewer 1
Peer Review Comments on
Acute and Chronic Noncarcinogenic Sections of the Arsenic Development Support Document Prepared by Texas Commission on Environmental Quality

General Issues

- Were procedures outlined in the Effects Screening Level Guidance followed by the TCEQ?

The sequence of events outlined in the Guidance was followed quite well. As noted below, the implementation of the guidance is questioned in terms of the selection of the critical study for the acute values.

- Does the Arsenic Development Support Document clearly describe the approaches used by the TCEQ?

The Support Document is clear in its account of the procedures used and the reasoning behind many of the decisions (except as noted immediately below and in the following sections).

- The regions of the respiratory tract (page 19, paragraph 2, line 4) include "pulmonary" and "thoracic." How do these differ from each other?

- On page 22, section 4.1.1.1, line 1, it is not clear what the authors mean by stating that long-term exposure to arsenic causes "vascular."

- On page 23, section 4.1.1.1.2, line 2, it is not clear how blood pressure can be expressed as a percentage.

- On page 24, section 4.1.1.1.2, paragraph 2, it is not clear why the authors do not expect oral ingestion via the mucociliary escalator after inhalation exposure. They obviously are aware of the possibility as they discuss it in section 4.1.1.1.4.1 on pages 25-26.

- On page 24, section 4.1.1.1.2, last paragraph, it is not clear whether the 9 years refers to duration of exposure or time elapsed since exposure ceased.

Health-Based Acute ReV and ^{acute} ESL

- Choice of the critical study (Nagymajtenyi et al, 1985)

The Nagymajtenyi et al (1985) study is a disappointing choice for a key study. The paper is poorly documented. The Materials and Methods are inconsistent with statements made in the Results. The authors provide no rationale for their selection of some, but not all, fetuses for examination. The authors' naïve classification of commonly observed variations as malformation further erodes the confidence that can be placed in the study. Furthermore, the methodology of exposure and the means for measurement are not reported. In addition, the authors did not report any attempt to prohibit or reduce oral exposure to the aerosol due to preening after exposure had ceased. Each of these items will be discussed below.

Table 1 of the original paper provides some insight into the study's problems. The authors do not state how gestational days were calculated, making comparison of their findings

to those of others difficult. The number of litters per group is unbalanced. However, the authors report 100 fetuses examined per group. Is this the total number available? If so it is highly improbable that each group would have exactly 100 fetuses. Are the reported number of dead fetuses in addition to the 100 examined fetuses? Are all of the dead fetuses included? If so, why were all of the live fetuses not included? How were the fetal weights calculated? Is this the arithmetic mean of 100 fetuses, the arithmetic mean of al fetuses, or the litter means for each group? The authors do not declare whether the indication of variance from the mean is standard error or standard deviation. Either way, from the size of the variance, it appears that the N for the calculation was the number of fetuses (N=100) rather than the mean litter weights. This artificially decreases the variance in the study and inflates the apparent statistical significance. The authors claim a large number of fetuses were identified with retarded growth, but objective measures for this parameter are undefined.

Table 2 of the original paper lays out the "skeletal malformations" seen in half of the fetuses examined in Table 1. The authors do not specify how the fetuses were selected. Did they look at the only smallest ones? Only the largest ones? An equal number of fetuses from each litter? In addition, and most importantly, none of the "malformations" listed should be classified as such. They are all variations that are commonly seen in conjunction with the delayed growth seen in low birth weight fetuses. Thus, this Table is a reinforcement of the well-known condition that skeletal delays are seen in the presence of growth retardation. The preceding problems suggest that the authors are inexperienced in the field of developmental toxicity and are not qualified to conduct and interpret developmental toxicity studies.

Key issues in all inhalation study involve the means of generating a uniform atmosphere and the methods for measuring the atmosphere to ensure the animals are exposed to an appropriate amount of test substance. The authors make no effort to describe either of these procedures. With respect to aerosol generation, the reader is referred to a paper written in German and published as a supplement in what appears to be a minor European journal. What concentrations of arsenic trioxide did they use in the solutions that were aerosolized? How were the chambers kept uniformly filled with vapor? A major problem at higher concentrations is condensation on the walls of the chambers. How was this avoided? Was it monitored? When in the course of the exposure period was the chamber concentration monitored? Was the measurement always at the same time? Why was it done only once per exposure period? How was the sample obtained? Where in the chamber was it taken? None of these details are given in the paper. These shortcomings provide a very low sense of confidence in the value of the paper.

The authors claim in the Methods to have exposed the animals on gestational days 9, 10 and 12 but in the Results they state exposure occurred from days 9-12. Which is correct? Once the exposure period ended, how were the animals handled? Were they wiped down to remove excess aerosol that had condensed on their coats? If not, how did they account for oral exposure due to preening? Such preening-associated exposure could be larger than the inhalation exposure. This is especially true for arsenic trioxide which has a sweet taste. The consequence of not controlling for preening is that the dose obtained in the mice is a combination of inhalation plus oral exposures; the first is quantified only by the target chamber concentration and the latter (which may be dominant) is not even considered.

In summary, the Nagymajtenyi et al (1985) is a mediocre choice for discussion in the TCEQ Support Document and it is an absolutely terrible selection as the critical study. The data are not reliable and are of no value for risk assessment.

There are additional issues concerning the use of Nagymajtenyi et al for developing an acute value. If developmental endpoints were to be used in this section, the Holson et al (1999)

paper is far superior. The rationale for the extended exposure in Holson et al was to insure that the rat hemoglobin was at a steady state with the arsenic trioxide (due to some reported peculiarities with the interaction between arsenic and rat hemoglobin). The TCEQ rejected the Holson paper because the exposure regime was too long. If the TCEQ were attempting to develop information for developmental toxicity, the Holson paper should be selected. However, given the Support Document's emphasis on length of exposure, the Holson paper (although explainable due to peculiarities with the rat) may not be a suitable replacement. The discussion in the Holson paper points out some potential issues that should be considered when using inhalation data for arsenic that have been developed in rats.
The responses to the remainder of the comments are meaningless if they are believed to be applicable to the Nagymajtenyi et al study.

- Choice of critical effect (decrease in fetal body weight)

Fetal body weight tends to be the most sensitive endpoint in developmental toxicity studies. The endpoint is a continuous variable that can be adequately modeled. It is relevant to humans. The major issue is the level of decrease from the control ***litter mean*** that should have been used as marker of effect. I prefer to 10% value compared to concurrent control. In the event a 5% value is used, a comparison to historical control data is needed to ensure that a heavier than normal control value (i.e. at the upper end of normal) does not result in false positives. Similarly, a lighter than usual control value (i.e. at the lower end of normal) could result in a false negative.

- Benchmark dose modeling

The selected critical study is inappropriate as detailed above. The issues concerning a 5% decrement in fetal body weights were outlined in the previous section. To restate the issues: 1) data should be presented as litter means, *not* arithmetic means of the fetuses in each dose group; 2) if a 5% decrement is used, historical control data should be consulted as well as the concurrent control values.

- The choice of point of departure

This item cannot be addressed because the Nagymajtenyi et al study is unacceptable.

- The choice of dosimetric adjustments

This section is uneven. The authors are scholarly in their approach, but they miss some important issues. First, the Nagymajtenyi et al study exposed the animals with an aerosol. The information about mean aerodynamic diameters from papers like the Holson et al study are not relevant because the latter study exposed the animals to arsenic trioxide dust. Second, the mice likely experienced a major amount of their exposure by means of the oral route. This confounds the remainder of the dosimetric discussion. However, if the Nagymajtenyi et al study had been satisfactory, then their discussion would be quite good.

- The choice of uncertainty factors

The selection of uncertainty factors in the present document follows the guidance commonly used by other regulatory agencies such as the USEPA. As such, there is nothing to fault with the reasoning of how to apply uncertainty factors. The discussion is moot for this particular section of the TCEQ document because the Nagymajtenyi et al study is not an acceptable source of data for risk assessment.

Health-Based Chronic ReV and ^{chronic} ESL $_{noncancer}$

- The choice of the critical studies

Smelter studies are difficult studies to use for assessment of a single agent's contribution to declining health. The Ronnskar smelter environs included exposures to copper, lead, silver, gold, hydrogen sulfide and other substances. Furthermore, exposure concentrations to all of these substances were not available. The selected studies are the best of the ones available; I concur with the authors' assessment of "medium" quality.

- The choice of the critical effect

The critical effects chosen (Raynaud's phenomenon and reduced nerve conduction velocity [NCV]) are relevant to human health and quality of life. It must be noted, however, the Raynaud's disease is far more obvious and readily diagnosed than NCV, which requires special equipment and trained personnel.

- The choice of dosimetric adjustments

The dosimetric adjustments laid out in section 4.1.5 are consistent with the methods used by the EPA and other groups charged with performing risk assessments. The methods are understandable, transparent and defensible.

- The choice of the point of departure

The substitution of doses for estimated air concentrations leaves much to be desired, especially when the document says at one point that exposures for the first 16 years were "below 500 $\mu g/m^3$" (page 23 line 1) then later it is stated to be "500 $\mu g/m^3$ "(page 23, section 4.1.1.1.2, second paragraph) and then using that higher value to calculate the TWA (page 24, formula).

- The choice of uncertainty factors

As mentioned in the discussion of the acute effects, the selection of uncertainty factors in the present document follows the guidance commonly used by other regulatory agencies such as the USEPA. In this case, however, due to the conservatism in the estimation of doses mentioned under point of departure above, the UF for database uncertainty could be lowered to 7.

Reviewer 2
Technical Review Comments: Acute and Chronic Noncarcinogenic Sections of the Arsenic Development Support Document

General Issues:

The document was well written, and few suggestions regarding minor general comments and editorial suggestions are provided as strike and replace or inserted comments in the draft document. There are a few considerations of a general nature the TCEQ may want to consider to enhance the analyses.

1. Reliance on the ATSDR Profile. The document relies heavily on the recent assessment and toxicological profile developed by the ATSDR. Reliance on ATSDR review as an efficient way to summarize non-critical data may be appropriate. However, reliance (or apparent reliance) on the conclusions and interpretations of key findings should be approached cautiously, particularly when other assessments developed through robust peer-reviewed processes result in different opinions. When the TCEQ indicates support for judgments developed in other documents, the current documentation would be strengthened by adding a short description of the key data that support the TCEQ conclusion. As an example, on page 10 of the draft TCEQ document the relative potency of arsenates and arsenites is compared and the conclusion of ATSDR is presented. Such text could be enhanced by a brief summary of the key data that supports the conclusion that trivalent arsenic if more potent than pentavalent arsenic, rather than requiring the document user to trust the ATSDR finding or go back and compare relative toxicity data from the ATSDR Toxicological Profile or other sources. Providing a brief independent presentation of the data would also highlight uncertainties in such conclusions. The two primary valence states 3+ and 5+ have very different modes of action, and thus relative potency for acute toxicity may or may not be representative of relative potencies for all endpoints. I am not sure of the relative potency of arsenic metal (valence 0), and data were not presented to allow one to compare. Presenting a synopsis of the data on comparative toxicity would highlight such uncertainties. The general point here is that TCEQ should consider providing data to support general conclusions for key issues, rather than just stating agreement with the conclusion made by other agencies.

2. Scope of the Assessment – A large amount of background information on the forms of arsenic and arsenic compounds in the environment is presented. On my first read I thought the selection of compounds that were described was somewhat random in nature. It was not until page 10 of the assessment that the actual scope of the data assessment effort was presented. I would recommend that somewhere at the beginning of the document a section be added that clearly identifies the scope of the document. After reviewing the entire document the following seems to be the intent regarding chemical form:

   The document is meant to cover all inorganic compounds in the form of particulates (<10 uM), including all metals and salts with valence states 0, +3, and +5. The ESL would not cover arsine or other specialty arsenic-containing gases or organic forms of arsenic. It

would be helpful to make clear that data are shown and summarized only for selected example compounds most commonly associated with airborne environmental exposures – which presumably is the reason for not presenting data on the many soluble arsenic salts. Data for arsenic trioxide were used as the basis for the assessment, because inhalation data are more available, in the environment and body pentavalent forms are transformed into trivalent forms, and trivalent arsenic is more potent than pentavalent forms (and thus would be a health-protective surrogate). The rationale for the size selective requirement was not made clear in the document. One could see an argument for the 10 uM cutpoint from a pragmatic basis due to the statement on page 8, "anthropogenic arsenic in the atmosphere occurs as fine particles with a mass median diameter of about 1 μm…" However, since the primary effect of concern is systemic toxicity particles anywhere in the inhalable range could be of concern and a standard that include larger inhalable particulate might also be of interest. If most arsenic particulate is "fine" in nature then respirable and inhalable fraction will not differ much, and one does not lose anything by basing the standard based on the inhalable fraction.

3. Route-to-route extrapolation. Page 31 of the document notes that "Cal EPA has proposed an inhalation assessment using an oral study, route-to–route extrapolation is typically not recommended for metals, especially if inhalation studies are available, as explained in the USEPA's RfC document (EPA 1994)." I am not sure that is statement is wholly accurate. If one has robust data for the route of exposure of interest that is generally preferable, but where data by the route of interest is limited, then route-to-route extrapolation can be a preferred option, even for metals. Since the database for oral toxicity is significantly greater than inhalation for non-cancer effects (particularly for acute endpoints), the document might be enhanced by exploring this issue further. The utility of the oral data depends on what we know about absorption kinetics, and in particular, to what degree the PBPK models can be used to answer such questions. The document seemed to dismiss these options too readily, without a full exploration of how such data might inform the choices that were made.

4. Relationship of the proposed Chronic ESL to background exposures. Pages 8 and 9 of the document provide information related to ambient levels of arsenic. The cited information suggests that the proposed chronic ReV and ESL are within the background range of exposure in urban areas and in the upper range of current dietary doses (based on simplified assumptions about uptake kinetics). How one should deal with that is a policy choice, but it seems that the current document would have noted this consideration.

Health-based Acute Assessment:

Selection of the Critical Study

I agree with the conclusion of the authors that the human studies related to developmental endpoints are too limited to serve as the basis for the ESL. The acute inhalation toxicity studies for animals are also somewhat limited and the selection of a robust critical study is difficult. Overall, the document needs to do a better job of highlighting the strengths and weaknesses associated with each of the three possibilities: the mouse inhalation developmental study

(Nagymajtenyi et al. 1985); the rat inhalation developmental toxicity study (Holson et al., 1999); or extrapolation from acute oral toxicity data.

As the TCEQ authors note the Nagymajtenyi et al. (1985) study provides limited documentation, which generates significant uncertainties.

The limitations of this study have been described in detail in other reviews. Briefly, key uncertainties or criticisms include:

- Uncertainties related to the exposures: 1) lack of a description of the exposure generation system, although reference to a Non-English article is cited by the authors, 2) no characterization of the particle size distribution, 3) it was unclear if exposure are measured as As_2O_3 or As, 4) the approach for how ingested arsenic was accounted for was unclear assuming whole-body exposures were used.
- Uncertainties related to the fetal weight: 1) lack of reporting of maternal effects, 2) do the fetal weight values reported in Table 1 of the publication reflect litter averages or averages from 100 or some lesser number of fetuses, 2) are the reported errors values the SD or the SEM and what value of "N" was used in this calculation. The value of the sample size clearly has a large impact on the determination of statistical significance.
- Uncertainties related to the skeletal effects: 1) the reported findings were not "malformations" as currently defined is such studies, 2) the sampling approach (how were the 50 fetuses selected?) was not clear, and thus litter effects cannot be determined.

Differences in the willingness to use the study have become apparent. For example, Cal EPA selected this study as the basis for their acute value, while the AEGL committee chose not to use this study. The only compelling reason to use this poorly documented study would be if one thinks that mice are more sensitive than rats, and thus the robust study by Holson et al. (1999) might not represent findings in the most sensitive species. The document should explore and explain this question more thoroughly. In a quick analysis, there do not appear to be significant rat versus mouse differences in developmental toxicity based on a cursory review of the oral dosing database (as presented in the ATSDR review). In fact, the data for mice and rats are very consistent. Stump et al. (1999) reported a NOAEL of 15 mg/kg-day and a LOAEL of 23 mg/kg-day for trivalent arsenic administered once on GD 9 in rats. Baxely et al. (1981) reported a NOAEL of 11 mg/kg-day and a LOAEL of 23 mg/kg-day for trivalent arsenic administered on GD 8-15 in mice. In both species maternal effects accompanied developmental effects. Comparisons of oral LD50 values also fail to show significant differences between mice and rats, in which the ranges of LD50 values are overlapping. Thus, the oral toxicity database suggests that rats and mice should show similar developmental toxicity profiles and that maternal effects occur at or below developmental effects. An independent peer review panel convened by TERA (TERA, 1999) also evaluated this specific question of whether arsenic is a selective developmental toxicant. The peer review panel concluded that "At the experimental oral and inhalation doses tested, which generated frank maternal toxicity and lethality, no prenatal structural effects were induced in laboratory animals. Moreover, inhalation of arsenic trioxide produced no other developmental effects at concentrations that induced frank maternal toxicity. By the oral route (gavage and diet), developmental toxicity (post-implantation loss and/or decreased fetal weight) was seen only occasionally and at the highest dose level, which also induced maternal toxicity."

The reluctance to discount a potentially sensitive study is understandable, but the TCEQ document should explore other alternative explanations of the reported fact pattern. An alternative interpretation of the

Nagymajtenyi et al. (1985) study that is consistent with the available data and more congruent with the overall body of literature is available. One could view the results of Nagymajtenyi et al. (1985) as identifying a clear LOAEL at the highest dose of 28.5 mg/m^3 with a marginal LOAEL or NOAEL of 2.9 mg/m^3. This study did not report on the occurrence of maternal effects, but such effects were likely. This conclusion is based on two considerations. First, that the pattern of "developmental" effects observed – decreased fetal weight and delayed skeletal ossifications are developmental delays that are often associated with maternal stress. Second, the high dose is above the MTD reported in the Holson et al. (1999) studies that was associated with maternal lethality (and as noted above mice and rats have similar acute toxicity profiles). A high degree of maternal toxicity in the exposed mice would explain the findings in the high dose group in Nagymajtenyi et al. (1985), and the minimal effects at the lowest dose would be consistent with a statistical artifact based on the fetus versus litter-basis for analysis. One could consider the mid-dose of 2.9 mg/m^3 a minimal LOAEL or a NOAEL, again reflective of the onset of maternal stress and resulting fetal weight decreases. This interpretation of the data is not inconsistent with the findings of Holson et al. (1999) who reported a clear NOAEL of 3 mg/m^3 with a LOAEL for clinical signs of toxicity and decreased body weight of 10 mg/m^3. The finding of minimal effects at 2.9 mg/m^3 in mice, while no effects in the rats were observed at 3 mg/m^3, is not a stark difference, given the possibility of some small species differences in sensitivity, the use of liquid aerosol versus particulate exposures in the mouse study (which might enhance bioavailability), and uncertainties surrounding the exposure measurements (including) potential oral intakes in the mouse study.

Overall, if TCEQ decides to use the Nagymajtenyi et al. (1985) with its many uncertainties, then the interpretation of the study should be framed in the context of the Holson et al. (1999) study as well as extrapolation from the acute oral studies. Such an approach of weighing the uncertainties and using other data to build the case for the selected study will result in a more compelling basis for the Acute ESL.

Estimation of the Point of Departure and BMD Modeling

I have major reservations about using the BMD modeling approach given the uncertainties in the Nagymajtenyi et al. (1985) study. The rationale given that the mouse study is the better choice appears to be predicated on two arguments: 1) relative species sensitivity, 2) days of exposure. I am not sure either argument is compelling. As noted above, mice do not appear to be more sensitive than rats. The rationale for not using the Holson study based on study duration grounds does not seem compelling, given that the Nagymajtenyi et al. (1985) was also for multiple (4) days of exposure, and the observed effects (clinical signs of toxicity) were consistent with acute toxicity in the Holson study. If TCEQ considers mice to be more sensitive, then a better description of the data supporting that conclusion is needed. I think a better approach would be to consider 3 mg/m^3 as a NOAEL or perhaps a minimal LOAEL (based on the combined findings of the two inhalation studies) and forgo modeling given the significant uncertainties involved in the data set.

If TCEQ judges that the exposure measurements in the mouse study are reliable and decides to proceed with BMD modeling, then approaches based only the central tendency estimate should be used. This is because the $BMCL_{05}$ as well as the BMC_{1SD} outputs depend on the reported measure or variability. We do not know the basis for the SD or SEM values in the selected study. The BMC_{05}, based on central tendency would not be impacted by this major uncertainty, and the mean of the litter means should be the same as the mean of the individual fetus measurements (with the significant assumption that the fetuses that were weighed were selected in a fashion that represents the overall litter experiences). I did not find any technical errors in the mechanics of the model run that was presented in the current document.

Dosimetry Considerations

Given the uncertainties in the selected critical study, and the absence of a rat component to this model, not using the PBPK models to refine the POD estimate may have been appropriate, but the rationale given in the document is not compelling. Couldn't one use the model by simply reducing an uptake parameter or setting exposure via the oral pathway to zero? I think the greater concern might be the issue of tissue bioavailability – how much reaches the developing fetus – is blood AUC a good metric? On the other hand, if the effects are actually secondary to maternal stress then, blood AUC might be sufficient. Given the limitations in the inhalation studies, using this model to extrapolate from the mouse or rabbit oral studies might be useful if resources allow. This would at least help to inform the reasonableness of the proposed approach.

Regarding the proposed particle dosimetry approach, I think the authors have stretched the data too far. I would suggest the data for the Nagymajtenyi et al. (1985) study are too limited to apply the EPA RDDR methodology. If TCEQ were to obtain the cited Non-English paper that was cited as describing the exposure generation system then perhaps more data to estimate particle size characteristics would be available to estimate particle characteristics. A paradigm for developing such estimates from limited data is laid out in the RfC guidance. Alternatively one could use the Holson et al. (1999) study and apply standard dosimetry approaches. If that study were chosen, then the MPPD model could be used, since a module is available for the rat. I did rerun the RDDR software using the parameters described in the TCEQ document. I obtained the same output. Although deposition in the entire respiratory tract is of interest, the primary effect of concern is systemic toxicity and the RDDR for the extrarespiratory effects should have been used. Thus, the RDDR is 5.995 not 2.067. This changes the POD_{HEC} derived by TCEQ from the mouse study to 1.2 mg/m^3.

Choice of UF

Some of the comments above regarding the selection of the critical study and effect, would impact the ultimate selection of UF values. The comments here reflect on the UF choices based on the POD selected by the TCEQ.

The use of 3 for interspecies differences is appropriate, as a default after application of the EPA default dosimetry. If as recommended in these comments, such dosimetry modeling is not used, then the default factor of 10 would be appropriate. The document could include an enhanced discussion of how kinetic data and the PBPK models can or cannot further be used to refine the estimate of species differences in kinetics. Adequate data on human versus rat or mouse variability in toxicodynamics are not available.

The default factor of 10 for human variability in sensitivity seems appropriate. The document could provide a more complete description of the considerations related to uncertainties human variability. In some places in the document references genetic polymorphisms, but that issue as well as other pertinent sources of variability could be noted more fully in this section.

A factor of 10 was used to account for uncertainties in the database. In the context of acute limits the definition of a "complete" database is not well defined. Given that we have two acute (or at least short-term) inhalation studies in two species, developmental toxicity studies, and a supporting oral database that could be further exploited, I am not sure what additional studies are needed to address the concern that the critical adverse effect was not identified. The absence of adequate human studies is a concern, but that issue is addressed by the interspecies factor. The quality of the mouse study is low, but one could view it as providing a conservative estimate of potency. Based on these considerations a factor of 10 seems too high. A reduced factor of 3

seems more appropriate, and an argument could be made that there are no significant data gaps that are not already covered by the other factors. If there are concerns related to endpoints not covered in the existing studies (e.g., subtle neurobehavioral effects) then the rationale for data gaps should be made more clearly.

Health-based Chronic Assessment:

Selection of the Critical Study

The chronic inhalation database is limited and reliance on human studies is appropriate. The limitations in the studies have been noted to some degree in the current document. Overall, based on the TCEQ descriptions, the studies appeared to be well conducted and the observed effects are consistent with the expected toxicity profile of arsenic. It would be useful to have seen a presentation of a critical evaluation of the selected series of critical studies using a systematic evaluation approach such as the Hill Criteria or a related evaluation framework. Of particular concern are the uncertainties related to the exposure estimates and the confounding effect on concomitant lead exposures. Lead exposure itself has been associated with the effects attributed by the authors to arsenic exposure, including decreased nerve conduction velocity (Krieg EF Jr, Chrislip DW, Brightwell WS. A meta-analysis of studies investigating the effects of lead exposure on nerve conduction velocity. Arch Toxicol. 2008 Aug;82(8):531-42) and peripheral artery disease (Navas-Acien A, Guallar E, Silbergeld EK, Rothenberg SJ. Lead exposure and cardiovascular disease – a systemic review. Environ Health Perspect. 2007 Mar;115(3):472-82). Thus, reliance on these studies for dose-response is tenuous, both because the exposure levels are at best a gross estimate, and second because the effects of lead are not controlled.

Because the primary uncertainty involved is in the exposure estimates and the contribution of confounding exposures (e.g., to lead), TCEQ should consider presenting a more complete argument for use of the inhalation studies versus using alternative approaches. The key studies that serve as the basis for the U.S. EPA RfD and ATSDR Chronic oral MRL (Tseng et al., 1977) as well as the Cal EPA Chronic REL (Wasserman et al., 2004) should be discussed and the rationale for not using route-route extrapolation needs to be presented based on a thoughtful analysis of the various uncertainties, rather than a blanket statement citing EPA dosimetry guidance. Particularly since human PBPK models are available. The value in this effort is that if the derivation of chronic ESL via three different data sets yields reasonably consistent results, then the overall weight of evidence is much stronger for the approach that is ultimately selected. None of the options available are without significant uncertainty, which suggests the need to explore multiple lines of evidence.

Estimation of the Point of Departure

Based on the selection of the inhalation studies as critical, the time-weighted averaging approach adjusted for occupational versus general population minute volumes is appropriate and was done correctly. As noted above, the document would be strengthened by providing an analysis of potential POD estimates based on a route-route extrapolation approaches using the critical ingestion studies in humans or at least by presenting why such approaches are not reasonable.

Choice of UF

Some of the comments above regarding the selection of the critical study and effect, would impact the ultimate selection of UF values. The comments here reflect on the UF choices based on the POD selected by the TCEQ.

Since the NOAEL was note estimated from the critical studies, a factor for extrapolation form a LOAEL was used. The assessment relies on the ATSDR interpretation of the severity of the observed effects as a rationale or assigning different UF values for Lagerkvist and Zetturland (1994) where a factor of 3 was used and Lagerkvist et al. (1986) study where a factor of 10 was used. The approach seems reasonable, but a more fully developed rationale regarding the adversity of each observed endpoint should be presented in the document.

A full factor of 10 was used to address human variability in sensitivity. As noted above there are insufficient data to move from this default. However, a more robust presentation of the key data gaps and issues surrounding human variability in response would be useful.

The assessment also uses a default factor or 10 for database insufficiencies. It is not clear from the rationale presented that a full factor of 10 is needed. The rationale that "while several epidemiology and/or occupational studies for arsenic workers exist, none of the studies had documentation on the actual exposure concentrations. In addition, co-exposure to other pollutants and smoking were not adequately documented in the studies." This concern is largely an indictment of the current derivation approach, rather than the appropriate basis for adding an UF. The note of concurrent exposures would actually bias studies to identifying a greater concern than is real, and would not justify adding an additional factor. The limitations in the inhalation studies might be partly alleviated by reliance on the oral toxicity studies. TCEQ notes the absence of a two-generation reproductive study for arsenic, but the database for reproductive and developmental endpoints from the oral database also helps to alleviate this concern. The limited inhalation studies supplemented by the oral toxicity database might support a factor less than 10 for database considerations, if the full array of data is considered.

Welfare-based Assessment:

I am not aware of any specific data that alter the conclusions of TCEQ regarding the welfare-based assessment, but did not do an independent search of the relevant literature.

Reviewer 3
Peer Review Comments on
Acute and Chronic Noncarcinogenic Sections of the Arsenic Development Support Document Prepared by Texas Commission on Environmental Quality

General Issues

- *Were procedures outlined in the ESL Guidelines followed by the TCEQ to perform arsenic's toxicity assessment? If references to accepted procedures in federal, state, or other appropriate guidance documents were made in the ESL Guidelines, were those accepted procedures followed?*
- *Does the arsenic DSD clearly describe the approaches used by TCEQ to perform the toxicity assessment (i.e., hazard identification and dose-response assessment).*

The TEQM appears to have followed the basic template set out in the ESL Guidelines. However, the Guideline are said to be based upon the traditional 4-step risk assessment process, which includes a thorough hazard assessment. The TEQM has not performed a thorough hazard assessment and has instead focused primarily on key studies. This detracts from transparency and obscures the process by which the key studies were chosen.

The TEQM has also relied too heavily on material from secondary sources, as opposed to first-hand review. This is especially the case regarding the ATSDR document, which is relied upon heavily for both critical review and technical calculations. This reliance on ATSDR detracts from reader confidence and further obscures transparency, especially given the direct adoption of ATSDR figures for the occupational TWA, which appears to have been taken by ATSDR from a secondary source.

The TEQM has not attempted critical review of the epidemiologic literature, which suggests a lack of familiarity with the limitations of this discipline. Instead, critical assessment has usually been taken directly from the studies themselves or from the ATSDR document, with the TEQM accepting these arguments at face value. A conspicuous exception to this practice is the use of isolated findings on finger spasticity (Lagerkvist et al. 1986) as the basis for the chronic assessment.

The ATSDR has concluded that the critical non-cancer endpoints of concern to chronic assessment probably relate to respiratory and gastrointestinal effects, and that data are not sufficient to derive a chronic MRL. The US EPA has also concluded that one cannot establish a chronic reference level for As. The ATSDR has similarly concluded that the most likely acute effects involve the respiratory or gastrointestinal systems, and that the acute risk from inhalation exposure is probably low. Yet, the TEQM has derived reference levels for both acute and chronic exposure based on developmental and cardiovascular effects, respectively. This departure from precedent set by federal agencies, especially given the heavy reliance on the ATSDR document in the DSD, raises concerns about the transparency of the TEQM decision-making process.

Health-Based Acute ReV and ^{acute}ESL

- *The choice of the critical study (Nagymajtenyi et al. 1985)*

My main area of expertise is in human health and epidemiology, rather than animal toxicology. However, the choice of the Nagymajtenyi study seemed questionable to me. The ESL document indicates that factors to consider when selecting a key study include "evidence of

a doe-response relationship, reproducibility of findings, mechanism or MOA, and consistency with other studies." The results from the Nagymaijteni study suggest a dose-response relationship for fetal weight, but they do not fit the other criteria listed in the ESL. Specifically, the results have not been reproduced and are not consistent with findings from the Holson study, which the TEQM acknowledges has a "relatively better exposure design."

The main rationale for choosing a study in which confidence is only moderate appears to be that (1) Nagymajtenyi et al. deals directly with acute exposure, which fits the ESL specifications, and (2) that mice are "more sensitive" to reproductive/developmental effects from arsenic exposure than are rats. However, the latter seems to be based solely on the fact that Nagymajtenyi et al. reported positive effects at lower exposure levels than Holson, not because of any known a priori species sensitivity. Furthermore, the shorter duration of Nagymajtenyi et al. belies the fact that Holson et al. found no effect using a better design and longer and higher exposure levels. This casts doubt on the findings of Nagymajtenyi et al., given that one would expect the much heavier exposure administered by Holson et al. to have reproduced or exacerbated the effects reported by Nagymajtenyi.

The 2007 ATSDR document notes the limitations of Nagymajtenyi et al. (1985) and suggests that the effects reported in that study may have been due to maternal toxicity, which was not reported by Nagymajtenyi et al. The TEQM should provide a better rationale for why this study was chosen as key, given the noted limitations. It is not sufficient to say that this study was chosen strictly based on exposure duration, because the ESL methodology allows for exposure adjustment. The TEQM should also provide some description of the acute mouse study by Aranyi et al. (1985), which is mentioned in the ATSDR criteria document but not in the DSD.

- *The choice of critical effect (decrease in fetal body weight)*

This chosen critical effect is relevant to human risk assessment and would seem appropriate, if Nagymajtenyi et al. (1985) had provided evidence that it was not driven by maternal toxicity. However, this omission raises questions about the appropriateness of this endpoint, especially given the fact that no changes in fetal weight were noted by Holson et al. (1999) from exposure that was at least an order of magnitude higher. A variety of acute or intermediate effects have been reported in laboratory rodents, including respiratory and immunological effects, so the TEQM needs to more fully explain their choice of this particular developmental effect.

- *Benchmark dose modeling:*
    - *Was the output from the most appropriate model selected?*

The power model provided the best fit to the data, so its choice seems appropriate at first glance. However, a goodness of fit (GOF) p-value of 0.3 does not necessarily imply a good fit to the data, given that GOF models have weak power to detect lack of fit (ie, failure to reject does not assure that the model fit is good). Given this characteristic of GOF tests, a typical situation involves choosing the best fitting model from among a group of models with adequate fit (ie, p-value >0.05). In the current situation, the fit of all other models was significantly bad. The TEQM needs to suggest rationale for why other, simpler (eg, linear) models provided such poor fit to the data. It would also be appropriate to provide regression diagnostics for all models tested.

    - *Was the appropriate critical effect size selected (5% decrease in mean fetal body weight compared to control means)*

The 5% change used to define an adverse effect on fetal body weight seems appropriate given the citations to Kavlock et al. (1995) and Allen et al. 1996.

- *Should the point of departure (POD) be based on the maximum likelihood estimate or the 95% lower confidence limit of the benchmark response?*

The ESL methodology suggests that a point of departure can be based on either a lower limit or MLE. However, the MLE represents the best estimate of the true value. The use of a lower limit adds conservatism, but this does not seem appropriate given UFs that already combine to 300-fold.

- *The choice of point of departure*

Use of BMC modeling to derive the POD seems appropriate given the continuous data available on fetal body weight.

- *The choice of dosimetric adjustments*
  - *Was the most relevant, appropriate, and defensible dose metric selected?*
  - *Were the appropriate default exposure duration adjustments conducted?*

The dose metric based on 4-hour average airborne exposure seems appropriate given lack of data on blood levels or area under the curve.

- *Were the appropriate default dosimetry adjustments from animal-to-human exposure conducted? Specifically, were the appropriate estimates for conducting the regional deposition dose ratio (RDDR) chosen when the key study did not report the required parameters?*

The TEQM used parameter estimates from Holson et al. (1999), ATSDR, and EPA as inputs to the RDDR. This would seem to be appropriate for parameters such as average mouse weight and respiratory area, but not for geometric standard deviation or MMAD. It is not sufficient to say that atmospheric arsenic emissions are generally in the respirable range, because the key question relates to the size of the particulates generated specifically by Nagymajtenyi et al. Similarly, one has no assurance that the GSD reported by Holson et al. will be similar to that within Nagymajtenyi et al. The TEQM should formally state that these are simplifying assumptions that are not based on data, rather than suggesting that they are reasonable read-across estimates.

- *The choice of uncertainty factors.*

The TEQM appear to have considered all appropriate uncertainty factors, and the overall 300-fold uncertainty is not inconsistent with extrapolation from animal data. However, the 10-fold uncertainty attached to data limitations may be excessive, given that the better-designed study by Holson et al. (1999) found no effect from exposure more than 10-fold above that reported by Nagymajtenyi et al. Therefore, confidence in the database would seem to be more heavily weighted toward uncertainty about whether the effect reported by Nagymajtenyi is real, rather than whether it might be more severe than reported. For this reason, a 3-fold UF may have been appropriate (for a cumulative UF of 100). The critical effect (decreased average fetal body weight) is not considered severe, which is also consistent with an UF of 3.

Health-Based Chronic ReV and $^{chronic}ESL_{noncancer}$

- *The choice of the critical studies (Lagerkvist and Zetturland (1994) and Lagerkvist et al.*

(1986))

- *Were the relevant occupational epidemiologic studies selected for the non-cancer estimates? Were defensible reasons provided to exclude the other available studies?*

The TEQM bases their assessment of chronic risk on epidemiologic evidence, which provides a direct relationship to human health without extrapolation from animal studies. However, epidemiologic evidence is observational rather than experimental in nature, so that findings can be (and often are) driven by bias. Therefore, one must carefully review the weight and limitations of epidemiologic evidence before concluding that an association is both real and accurately described. Unfortunately, the TEQM does not appear to have performed a comprehensive, independent review of the epidemiologic literature. Rather, they have focused their chronic assessment upon a literature base consisting of essentially one study with a weak epidemiologic design.

Lagerkvist and Zetturland (1994) and Lagerkvist et al. (1986) are different iterations of the same study, which is a continuation of Blom et al. (1985). This study has a cross-sectional design, which is an approach that provides only a snapshot in time. The 5 years of "followup" provided by Lagerkvist and Zetturland (1994) does not alter this situation, because the base population is not a well-defined cohort and because no longitudinal (eg, person-year) analysis has been attempted.

Unlike a cohort study, which seeks to include all individuals with a given exposure and to follow them over time, a cross-sectional study provides only information on those who are currently working. One has no knowledge of who left or entered the base population, or what happened to them over time. Such an approach allows no strict temporal relationship between exposure and illness. Furthermore, such a design has a greater potential for selection bias compared to a cohort approach. In the current study, this potential for biased selection is exacerbated by the fact that the referent population is not selected randomly or comprehensively. In fact, there is evidence of differential participation, in that participation appears to be greater among those with exposure (85%) than among the referents (76%). Furthermore, only 50 of the 198 self-selected referents were actually included, and selection of this 25% subset is not well described, except for saying that those chosen formed the best matches on smoking and use of vibrating hand tools (Blom et al. 1985). Therefore, Blom et al. (1985) represents a cross-sectional design in which 85% of exposed workers are compared to only 19% of potential referents. There is a considerable potential for differential selection in such an approach, which is to say that there is substantial potential for selecting referents that are different from the exposed group on risk factors other than exposure.

There is also a substantial potential for confounding in this Scandinavian study. The exposed workers had been coexpopsed to sulfur dioxide and heavy metals, substance to which the referents had not been exposed. Some risk factors for Reynaud's phenomenon, such as stress and caffeine consumption, were apparently ignored. Other unexplored factors that may have been important include genetics (ie, familial predisposition) and work practices. Uncontrolled or residual confounding becomes especially important when investigating subtle changes, such as subclinical differences in peripheral circulation, rather than gross effects (eg, incidence of heart disease).

The two groups were said to be similar on smoking status, but the exposed workers actually had a greater prevalence of current smoking (51%) compared with the referents (37%) (Lagerkvist and Zetturland, 1994). No information was provided on smoking intensity (eg, cigarettes per day) among the two groups. Alcohol consumption is said to be similar between

the two groups, but again, no data distributions are provided. These omissions are important, because nonsignificant differences between groups can still result in important confounding, especially for subtle effects such as changes in blood pressure or nerve conduction. For example, Lagerkvist et al. (1986) reported that exposure to vibration was not significantly associated ($p>0.1$) with white fingers, suggesting that vibration would not be an important confounder. However, examination of their table 2 reveals that 16% (9 out of 56) of those with vibration exposure had white fingers, compared with only 7-8% (3 out of 39) of those without exposure. Although not statistically significant, this 2-fold difference would seem important.

Confounding effects can be alleviated through well-conducted statistical modeling. However, the model covariates explored in the key studies appear to have been limited to age, smoking, and use of vibrating hand tools, and it is not clear if these were entered simultaneously or sequentially. The authors of the key studies provide no assurance of extensive model-building exercises (eg, forward or backward elimination) that would have simultaneously addressed all potential covariates. These authors also provided no regression diagnostics to assure that the models fit the data well, and were not influenced by multicollinarity or influential observations.

Symptoms and risk factors such as past exposures, trauma, or medical conditions were identified via questionnaire, which raises the issue of recall bias. That is to say, those who are concerned about exposure to As might remember past exposures differently, or might pay greater attention to their symptoms or health conditions, than would auto workers without this recall stimulus.

The potential limitations raised above do not mean that the findings from the key studies are necessarily incorrect, but they do detract from confidence in these findings. Validation from other, better-designed, studies is needed to confirm these associations, especially if one is going to base regulatory guidance upon them.

The TEQM provided little critical examination of the issues raised above. This omission suggests a toxicologic/experimental perspective, with less understanding of the limitations of observational research. This is a major limitation for any risk assessment based on epidemiologic data.

The TEQM appears to have performed only a cursory review of the occupational epidemiology literature on As exposure. More specifically, they did not provide a complete review of the cohort studies that have been performed among workers with As exposure. Cohort designs are generally considered to be superior to a cross-sectional approach, and such studies could provide supportive evidence for or against inhaled As as a cause of cardiovascular disease. A more substantive review of this literature was provided by ATSDR (2007), which concluded that:

> "none of these studies provided conclusive evidence that the observed increase in risk was due to arsenic exposure. The studies in the ASARCO and Anaconda copper smelter workers failed to find a clear dose-response relationship with arsenic (Enterline et al. 1995; Welch et al. 1982), while a follow-up study of the Ronnskar smelter workers not only found lack of a dose-response, but also that the risk of cardiovascular disease was no longer elevated in the cohort (Järup et al. 1989). The studies in orchard workers and tin miners were limited by confounding exposures to copper, lead, and radon, respectively (Qiao et al. 1997; Tollestrup et al. 1995). The risk of cardiovascular disease mortality in the tin miners not only showed no dose-response relationship with arsenic exposure, but was positively associated with radon exposure, suggesting that radon may have been responsible for the increased cardiovascular risk in this cohort (Xuan et al. 1993)."

This assessment by ATSDR does not suggest substantive support for inhaled arsenic as a cause of cardiovascular disease. Lack of supporting evidence from the Ronnskar study (Järup et al. 1989) is particularly pertinent, as this is the same Swedish facility studied by Blom et al (1985), Lagerkvist and Zetturland (1994,) and Lagerkvist et al. (1986).

- *The choice of critical effect*
    - *Was the most appropriate critical effect selected? If not, what would be a more appropriate critical effect?*

The TEQM chose changes in NCV and prevalence of Reynaud's phenomenon as critical effects, but only results from Lagerkvist et al. (1986) were used in the chronic ReV. This means that prevalence of Reynaud's phenomena and cold-induced changes in digital spasticity represent the critical effect, which is said to be related to cardiovascular disease.

As stated earlier, this critical effect is taken from one, essentially unsupported, epidemiologic study with substantial potential for bias. This raises concerns about the adequacy of the underlying data used in the chronic health assessment. That is to say, it is not clear that the changes reported by Lagerkvist et al. (1986) are directly due to As exposure, rather than to other factors peculiar to the As workers included in this cross-sectional study. It is also not clear that the changes reported in Lagerkvist et al. (1986) are indicative of increased risk for cardiovascular disease, given that no measures of cardiovascular disease were explored and that exposed workers were generally not clinically abnormal. Reynaud's phenomenon is actually quite common in Nordic countries, and most cases represent nuisance conditions that do not lead to more serious illness.

The occupational and drinking-water literatures identify numerous effects that have been strongly linked to arsenic, including dermal, hematopoietic, neurological, gastrointestinal, and respiratory outcomes. These are summarized over several pages of tables in the ATSDR (2007) document. Dermal and gastrointestinal effects would be most influenced by dermal or oral exposure. However, neurologic, hematopoietic, and (especially) respiratory effects would be more directly influenced by inhalation exposure. The TEQM would be on much safer ground if they chose one of these latter endpoints as the critical effect.

Cardiovascular effects could have been used as the critical effect, but this would have to be based on drinking water studies, which would have required route-to-route extrapolation. A relatively recent review of the literature pertaining to the epidemiologic association between heart disease and As exposure was conducted by Navas-Acien et al. (*Am J Epidemiol* 2005;162(11):1037-1049)). These authors concluded that it is "impossible to establish firm conclusions regarding the effects of arsenic on cardiovascular outcomes at concentrations lower than those observed [for drinking water] in Taiwan" and that "the cardiovascular effects from chronic, low-level exposure to arsenic are unknown." It should be noted that this review did not include the Lagerkvist and Zetturland (1994) or Lagerkvist et al. (1986) studies, although it did include the Jarup et al. occupational cohort.

    - *Is the endpoint relevant for human risk assessment?*

Secondary Reynaud's phenomenon (if real) can be a potentially serious condition, and is therefore relevant to human risk assessment.

- *The choice of dosimetric adjustments*
    - *Was the most relevant, appropriate, and defensible dose metric selected?*

The ATSDR appears to have used a TWA approach to determine long-term average As exposure, which was greater before 1975 (500 ug/m3) and lower after 1975 (50 ug/m3). There are two main problems with this approach. First, the 500 and 50 ug/m3 figures appear to be crude estimates that were not validated by substantial industrial hygiene measurements. Secondly, TWA exposure may not be pertinent for a non-cancer effect for which a non-linear dose-response is anticipated. The effects measured in any epidemiologic study (if real) may have been driven by past averages that were above the TWA estimate, or even by peak exposures that were much higher. It should be noted that neither Lagerkvist and Zetturland (1994) nor Lagerkvist et al. (1986) found a correlation between current urinary As levels and the critical effects, suggesting that such effects would have been driven by past exposure. It is also important to remember that most workers had a history of arsenic-induced dermatitis, and that approximately 25% had nasal-septal changes indicative of high-level exposure.

- *Was the time-weighted average (TWA) exposure for arsenic (as As_2O_3) based on the Lagerkvist and Zetturland (1994) study and the Lagerkvist et al. (1986) study calculated correctly by ATSDR (2007)?*

A search of the ATSDR document did not uncover this calculation, which appears to have been taken from another source.

- *The choice of point of departure.*

The points of departure were based on the LOAELs from the Lagerkvist and Zetturland (1994) and Lagerkvist et al. (1986) studies. This is appropriate if one assumes that these studies, effects, and TWA estimates are appropriate (see above).

- *The choice of uncertainty factors.*
  - *Have all of the appropriate uncertainty factors been considered and are the values assigned to the uncertainty factors clearly justified and defensible?*

The combined uncertainty of 1000-fold seems excessive given the use of human data. Specifically, the database UF of 10 seems high given that there is an adequate database for As. There is considerable uncertainty surrounding the findings from the Lagerkvist et al. (1986) study, but this has to do with the low-level effects reported in that study, rather than the database as a whole. Other studies have not consistently reported cardiovascular risk at such low levels of inhaled exposure. The key studies have also not demonstrated a direct relationship between prevalence of Reynaud's and subsequent cardiovascular disease. Therefore, an UF of 3 for database uncertainty would seem more appropriate (combined uncertainty of 300 fold).

- *Would you make recommendations for a different approach to select uncertainty factors to calculate the chronic ReV?*

NA (see above discussion)

Welfare-Based Acute and Chronic ESLs

NA

Appendix C - Follow-up Conference Call Charge

Follow-up Conference Call on the Peer Review of the Acute and Chronic Noncarcinogenic Sections of the Arsenic Development Support Document

Charge Questions

The purpose of this conference call is provide TCEQ an opportunity to ask clarifying questions regarding the comments submitted by the peer review panel and to obtain feedback on additional options for developing the ESL values that TCEQ is considering in response to the individual panel members' written comments.

As an aid to the panel, clarifications on methodology have been identified by the Toxicology Division (TD) of the TCEQ that respond to some of the panel comments. First, TD relies on toxicity assessments conducted by other federal, state, and international agencies that have undergone a peer-review process as a starting point in their toxicity assessments, because of time- and resource constraints. However, the TD obtains copies of key studies and supporting studies and critically reviews these studies. TD does not routinely adopt toxicity values developed by other organizations. The toxicity assessments conducted by others are critically reviewed. These toxicity values may be adopted if procedures outlined in the TCEQ ESL Guidelines document are followed. Second, the legislature dictates that the TD develop an acute (usually 1-hour averaging time) and chronic values to evaluate all chemicals. These values are used to evaluate emissions from facilities during the air permit review process and to evaluate ambient air monitoring data. Other federal or state agencies may decide not to develop values, but the TD must have procedures in place to evaluate air emissions from all chemicals. Since the statutes require that ESL values be developed, databases that are lacking present a complex challenge for the TD.

GENERAL QUESTIONS:

1. Are the peer reviewers aware of any information or references that would support (or refute) the position that data for arsenic trioxide should be considered relevant for arsenic pentoxide?

2. Does the panel have additional recommendations related to the rationale for using 10 µM as a size-selective cut point for the ESL?

ACUTE ASSESSMENT:

1. The panel felt that the Nagymajtenyi et al. (1985) study is not an adequate study to develop an acute ReV and should be be replaced by the Holson et al. (1999) study as the key study. One reviewer noted an alternative explanation (Reviewer #2) of the Nagymajtenyi et al. (1985) findings and indicated that this study might be used to develop a supporting value. If such an approach were used would the following derivation be reasonable? The critical effect would be maternal toxicity, with the highest dose a LOAEL and the mid-dose of 2.9 mg/m^3 a minimal LOAEL or a NOAEL. No

BMCL modeling or RDDR calculation would be included due to weaknesses in the study. Is the use of the Nagymajtenyi et al. (1985) to derive only a supporting value to supplement Holson et al. (1999) scientifically defensible?

2. The critical effects observed in the Holson et al. study (1999) were respiratory effects, (i.e., point-of-entry effect) and decrease in body weight. If the Holson et al study were used as the basis for the Acute ESL would route-to-route extrapolation to derive values based on other critical endpoints observed in acute oral studies add significantly to confidence in the ESL derived from the Holson et al. (1999) study?

3. Does the data support the notion that concentration and duration both play a role in the toxicity of arsenic?

    a. If so:
        i. For the Holson et al (1999) study, rats were exposed to arsenic as dust particulate. Should the Haber's Rule as modified by ten Berge, with n = 3, be used to extrapolate the 6-hr per day exposure duration to a 1-hr exposure duration?

        ii. If used as a supporting study, the Nagymajtenyi et al. (1985) study exposed (mice to arsenic as an aerosol). Should the Haber's Rule as modified by ten Berge, with n = 3, be used to extrapolate the 4-hr per day exposure duration to a 1-hr exposure duration?

4. Are the oral studies Stump et al. 1999 and Baxley et al. 1981(Reviewer #2) adequate justification for assuming that mice are not more sensitive than rats to developmental toxicity? Holson et al. (1999) states rats are more sensitive than mice based on Lewis and Sweet (1985). Would inclusion of a discussion of these studies enhance confidence in the selection of Holson et al. (1999) as the critical study?

CHRONIC ASSESSMENT

1. Based on the written comments, panel member's suggest that the Lagerkvist and Zeturland (1994) and Lagerkvist et al. (1986) studies are not adequate studies to develop a chronic ReV. As an alternative approach, would an unit risk factor (URF) and chronic ESL based on excess lung cancer mortality in four different cohorts of workers exposed to arsenic be a more robust basis for the chronic ESL and would an ESL derived on this basis protect against sensitive noncancer effects? Do the overall data support a causal relationship between arsenic exposure and cardiovascular effects?

2. As part of a sensitivity analysis to better inform the weight of evidence, the TD is considering supplementing the analysis of the chronic ESL based on the epidemiology studies by using experimental studies and toxicity values from oral routes of exposure to calculate an inhalation toxicity value: (1) USEPA's RfD (Tseng et al. 1977), (2) ATSDR's MRL (Tseng et al. 1977) and (3) Cal EPA's inhalation REL (Wasserman et al. 2004). This would involve route-to route extrapolation or the use of PBPK models to derive an inhalation ReV, for comparison purposes only. These values will be compared

to the air concentration corresponding to a 1 in 100,000 excess risk for lung cancer mortality using the URF derived by the TD. Given the uncertainties associated with route-to-route extrapolation (see the ESL methodology document for the current policy of using route-to-route extrapolation), would such an approach be sufficiently robust to inform the selection or evaluation of the chronic ESL? Do any data limitations as highlighted in the ESL methodology document preclude the meaningful use of route-to-route extrapolation for arsenic with the above mentioned oral toxicity studies?

3. If the TD were to develop a route-to-route extrapolation approach to supplement the ESL derived based on the occupational epidemiology what studies should be used? Three agencies have used Tseng et al. (1977) and Wasserman et al. (2004) as key studies in the developed of chronic values (RfD, MRL, and chronic REL) Are these the appropriate studies to use as the key studies in the development of chronic values based on route-to-route extrapolation?

4. The preferred method for route-to-route extrapolation is the use of PBPK modeling, which provides the best estimate of a toxicant's internal and biologically effective dose as a function of exposure. The peer reviewers suggested that TD determine if PBPK models (Mann et al. (1996) are available to conduct route-to route extrapolation. Has the Mann et al. (1966) model been accepted by the scientific community as a valid model that could be used for this purpose? No federal or state agency has used this model to develop inhalation toxicity values based on oral studies. In fact, in their June 2008 Draft TSD for Noncancer RELs, Cal EPA states the following "while some PBPK modeling has been applied to inorganic arsenic and its methyl metabolites, the modes of toxic action and relevant internal dosimetry are not sufficiently understood at present to use this modeling directly in REL development". Is the panel aware of other models or other route-to-route extrapolation approaches that should be considered?

Appendix D - Biographical Sketches for Reviewers

Dr. Andy Maier. Dr. Maier is serves as both a senior toxicologist and is the Director of *TERA*. Dr. Maier has 14 years of professional work experience in the areas of environmental health, occupational hygiene, and toxicology. He provides overall science and strategic direction for the organization and provides oversight for risk and safety assessment projects. In his capacity as a toxicologist and risk assessor, he has led numerous projects, and has co-authored technical reports, human health risk assessment documents, or toxicity summaries covering more than 100 individual substances for government and private industry sponsors, including comprehensive documents for derivation of RfDs, RfCs, cancer risk assessments, and acute and occupational exposure limits. Completing such risk documents required detailed understanding of various dose metrics (including biomarkers of exposure) as well as the critical evaluation of the level of severity and potential biological significance of observed effects for identification of adverse effect levels. Dr. Maier has demonstrated expertise in inhalation, oral and dermal toxicology, acute and chronic hazard identification and dose-response for cancer and non-cancer risk assessment and regulatory toxicology. He has served in various capacities in support of peer review meetings, including serving as chairperson, facilitator, expert panel member, and report developer. He completed his PhD in Toxicology conducting research on the mechanisms of toxicity of mixtures of metals (including arsenic) and polycyclic aromatic hydrocarbons. His current research interests include developing approaches for using biomarker data to support dose-response assessments and improved methods for developing occupational exposure limits.

Dr. Maier received his M.S. in Industrial Health from the University of Michigan and his Ph.D. in Toxicology from the University of Cincinnati. He is a Diplomate of the American Board of Toxicology, as well as a Certified Industrial Hygienist. He has served on several volunteer risk assessment committees and holds a position as an Adjunct Associate Professor with the University of Cincinnati.

Dr. John Desesso. Dr. DeSesso is a Senior Managing Scientist for Exponent. He specializes in the areas of developmental and reproductive toxicology, general toxicology, risk assessment, and human health effects of environmental agents and pharmaceuticals. His research interests include normal and abnormal development, with emphasis on the mechanisms by which chemical and physical agents influence developing organisms.

Dr. DeSesso holds adjunct professorships at Georgetown University School of Medicine, Rosalind Franklin University of Medicine and Science (Chicago Medical School), the Graduate School of Public Health at San Diego State University, and University of North Texas Health Sciences Center. DeSesso has been invited frequently to serve as the chairman of scientific sessions at national and international scientific meetings, especially those involving mechanisms or amelioration of developmental toxicity and human health risk assessment. He has published more than 260 papers, chapters, and reports, and has presented at 200+ conferences and symposia.

Dr. DeSesso's broad background in health science has been recognized in diverse scientific positions. Prior to joining Exponent, he held various positions at Noblis/Mitretek Systems, where he directed commercially and governmentally funded public interest research efforts on the potential adverse effects of environmental agents. Projects include assessing the potential for inorganic arsenic contamination to cause human birth defects, investigating the possible

reproductive toxicity of a commonly used ingredient of consumer products, and assessing the possible teratogenicity of a novel ingredient used in veterinary medicine. He provided key expertise in human health risk assessment, including negotiation of risk assessment protocols on a site-specific basis. He wrote the Air Force approach to ecological risk, and is a co-author of multiple in-depth hazard evaluations of the widely encountered environmental contaminant trichloroethylene that addressed its potential human carcinogenicity and its potential to cause adverse effects in developing babies. The published findings of these studies contribute to its ongoing regulatory re-evaluation.

During 15 years at The MITRE Corporation, he wrote health hazard assessments for the EPA Office of Toxic Substances in support of its Pre-Manufacturing Notification Program. For EPA's Office of Pesticide Programs, Dr. DeSesso analyzed toxicity data submitted for registration of pesticides. For the EPA Office of Health and Environmental Assessment, he wrote assessment documents concerning the carcinogenicity, toxicology, and reproductive effects of existing chemical substances. For the FDA, Dr. DeSesso led a team that assessed the relevance to human health of direct bladder exposure studies conducted in rodents. He was named to the EPA/ILSI National Work Group on Rodent Bladder Cancer, where he helped to establish rational and scientific approaches for applying rodent bladder cancer study results to human health risk assessment. He also served on a team that studied the health effects of refined menhaden oil during FDA's consideration of its generally recognized as safe (GRAS) status.

Dr. DeSesso is active in 14 scientific societies and was elected President of the Teratology Society (1994–1995) and of the Mid Atlantic Reproduction and Teratology Association (MARTA; 2001–2002). He serves as an Executive Board member of the Federation of American Societies of Experimental Biology (FASEB; 1999–2004; 2006–present). He is a frequent contributor to continuing education courses that deal with toxicology and risk assessment. He has provided public commentary regarding scientific issues before federal agencies, state legislatures, and congressional committees.

Dr. John Bukowski. Dr. Bukowski is a senior associate at WordsWorld Consulting, a public health and medical-communications consultancy located in Dayton, Ohio. He provides research assistance on epidemiology and public/occupational health, as well as general assistance on issues relating to clinical medicine. Dr. Bukowski has 20 years of experience in epidemiology and public health, which includes service within government, academia, and private industry. Prior to joining WordsWorld, he was a senior scientist and epidemiologist for ExxonMobil Biomedical Sciences, focusing on such varied topics as children's health, reproductive health, neurological health, solvent exposure, risk assessment, and emerging health issues. He has also served as a research scientist within both the New Jersey Department of Environmental Protection (NJDEP) and the U.S. EPA. Dr. Bukowski has a broad range of clinical experience, including 7 years as a practicing veterinarian and service as the founding Director of the Clinical Research Centre at the University of Prince Edward Island, Canada. At UPEI, he oversaw all clinical and environmental research for the CRC, including a series of case-control studies on the associations between clinical birth outcomes and agricultural contamination of PEI ground water. He also authored several major reports for provincial organizations, including a report to the PEI Cancer Research Council on the carcinogenic potential of agricultural pesticides applied on the Island.

Dr. Bukowski holds a Ph.D. in epidemiology from the University of Medicine and Dentistry of New Jersey. He also holds a Masters in Public Health from the University of Michigan, and a doctorate in veterinary medicine from Michigan State University.

Appendix E - List of Conference Call Observers

Reviewers

Dr. John Bukowski
WordsWorld Consulting

Dr. John Desesso
Exponent

Dr. Andy Maier (Facilitator)
Toxicology Excellence for Risk Assessment (*TERA*)

Texas Commission on Environmental Quality (TCEQ)

Mr. Mike Aplin

Ms. Angela Curry

Dr. Neeraja Erraguntla

Mrs. Shannon Ethridge

Dr. Roberta Grant

Mr. Joseph (Kip) Haney

Dr. Michael Honeycutt (Toxicology Division Director)

Mrs. Allison Jenkins

Mrs. Lindsey Jones

Dr. Carla Kinslow

Dr. Jong-Song Lee

Mr. Darrell McCant

Dr. Tracie Phillips

Mr. Manuel Reyna

Attendees

Dr. Latrice Babin
Harris County Public Health & Environmental Services

Dr. Lisa JN Bradley
AECOM

Mr. Larry Carlson
Tenaska

Dr. Andrew Chiu
Ontario Ministry of the Environment

Dr. Elena Craft
Environmental Defense Fund

Ms. Maria Hegstad
Risk Policy Report

Ms. Patricia Nance
Toxicology Excellence for Risk Assessment (*TERA*)

Mr. James Gilmore
Ontario Ministry of the Environment

Dr. Grazyna Kalabis
Ontario Ministry of the Environment

Mr. Alejandro Nava-Ocampo
Ontario Ministry of the Environment

Ms. Tania Onica
Ontario Ministry of the Environment

Dr. Jacqueline Smith
Harris County Public Health & Environmental Services

Ms. Joan Strawson
Toxicology Excellence for Risk Assessment (*TERA*)

Mr. Timothy C. Wolfson
Babst Calland Clements & Zomnir, PC

Appendix F – Comments on PBPK

Screening-level review of arsenic PBPK models for application in risk assessment
Prepared by Lisa M. Sweeney, Ph.D., DABT
Toxicology Excellence for Risk Assessment
August 25, 2009

Summary
Route-to-route extrapolation (oral to inhalation) of existing human studies pertaining to arsenic using physiologically based pharmacokinetic modeling is possible, and is an option that could be further explored. Interspecies extrapolation from mice, rabbits, or hamsters, to humans is possible, but is not currently an option for rats.

Methods
The literature review was limited to articles available in the reviewer's personal reference collection (Menzel et al., 1994; Mann et al., 1994, 1996a and b; Yu, 1993 and 1999), abstracts available through PubMed, and articles retrieved for this review (Gentry et al., 2004 and El-Masri and Kenyon, 2008).

Results
Models in Experimental Animal Species
No detailed, published information on PBPK models of arsenic for rats was identified. Menzel et al. (1994) provide minimal description of development of a rat PBPK model for arsenic ingestion. Yu (1993), in a doctoral thesis, describes a PBPK model of ingested arsenic in rats, developed with minimal calibration data. If PBPK modeling of rat studies were to be considered, it would be advisable to use the more recent mouse model of Gentry et al. (2004) (discussed in greater detail below) as a starting point.

Mann et al. (1996a, b) presented results of arsenic PBPK model development for hamsters, rabbits, and humans. Additional model development for marmoset has been reported, but no details were provided (Mann et al., 1994). Limitations in the model documentation in Mann et al. (1996a, b) have been at least partially remedied by the publication of the mouse PBPK model by Gentry et al. (2004). For example, Mann et al. (1996a, b) provided descriptions of how to permeability constants were calculated, but in some cases, the units don't work out correctly (e.g, equation A5 of Mann et al., 1996a). Gentry et al. (2004) used the same permeability coefficients used by Mann et al. (2006a, b), and provided the calculated values for the various tissues and the four forms of arsenic considered in the model (AsV, AsIII, MMA, and DMA). Gentry et al. (2004) have developed the most robust of the mammalian PBPK models, using mouse data for arsenic, MMA, and DMA administered by iv or the oral route.

Models in Humans
Detailed descriptions of arsenic PBPK models have been provided by Mann et al. (1996b), Yu (1993 and 1999), and El-Masri and Kenyon (2008). Only Mann et al. (1996b) provide a description of disposition by the inhalation route. There are, however, problems with the documentation of the Mann et al. (1996b) human model. For example, the paper lists a human skin volume of ~6.2 kg, but a more commonly used value is 2.6 kg (El-Masri and Kenyon, 2008). It is unclear if this value is a typographical error, or was actually used in the model. If an erroneous value was used, the impact on the model predictions is uncertain. In general, one can

conclude that uptake by inhalation was adequately described by Mann et al. (1996b), as their model acceptably described the kinetics of both inhalation and oral studies of arsenic. Mann et al. (1996b) caution that the individual parameters in the model should not be used outside the context of the model, as their estimation is mutually interdependent. This reviewer agrees with this caution; some of the parameters (including some estimated from the hamster and rabbit models, such as partition coefficients) differ dramatically from those estimated for mice using more extensive data sets. If consideration is to be given to route-to-route extrapolation using PBPK models, it is recommended that the more recent human model of El-Masri and Kenyon (2008) be reviewed in depth, and that the inhalation modeling of Mann et al. (1996b) be reviewed in light of recent developments in particulate inhalation modeling.

References

El-Masri HA, Kenyon EM. Development of a human physiologically based pharmacokinetic (PBPK) model for inorganic arsenic and its mono- and di-methylated metabolites. J Pharmacokinet Pharmacodyn. 2008 Feb;35(1):31-68.

Gentry PR, Covington TR, Mann S, Shipp AM, Yager JW, Clewell HJ 3rd. Physiologically based pharmacokinetic modeling of arsenic in the mouse. J Toxicol Environ Health A. 2004 Jan 9;67(1):43-71.

Mann S, Droz PO, Vahter M. A physiologically based pharmacokinetic model for arsenic exposure. II. Validation and application in humans. Toxicol Appl Pharmacol. 1996 Oct;140(2):471-86.

Mann S, Droz PO, Vahter M. A physiologically based pharmacokinetic model for arsenic exposure. I. Development in hamsters and rabbits. Toxicol Appl Pharmacol. 1996 Mar;137(1):8-22.

Mann SB, Droz PO, and Vahter M. A physiologically based pharmacokinetic model for the four major arsenic species in mammals. In: Arsenic Exposure and Health, Chappell WR, Abernthy CO, and Cothern, CR, eds., 1994, pp 219-231.

Menzel DB, Ross M, Oddo SV, Bergstrom P, Greene H, and Roth RN. A physiologically based pharmacokinetic model for ingested arsenic. In: Arsenic Exposure and Health, Chappell WR, Abernthy CO, and Cothern, CR, eds., 1994, pp 209-218.

Yu D. Two problems in risk. Ph.D. dissertation thesis, nuclear engineering, University of California, Los Angeles (1993).

Yu D. A physiologically based pharmacokinetic model of inorganic arsenic. Regul Toxicol Pharmacol. 1999 Apr;29(2 Pt 1):128-41.