



Technical Memorandum

Comments on the Science Advisory Board (SAB) Review of the U.S. EPA Draft Toxicological Review of Inorganic Arsenic for the Integrated Risk Information System (IRIS)

Prepared for

Wood Preservative Science Council

Prepared by

Exponent
15375 SE 30th Place, Suite 250
Bellevue, WA 98007

with

Pamela Mink
Emory University
Atlanta, GA

November 15, 2010

Comments on SAB Review of the U.S. EPA Draft Toxicological Review of Inorganic Arsenic

This technical memorandum provides comments in response to the draft report (dated October 2010) of the Science Advisory Board (SAB) Workgroup that conducted a focused review of the U.S. EPA draft Toxicological Review of Inorganic Arsenic. Specifically, these comments address key scientific issues concerning the treatment of evidence from epidemiological studies and the dose-response assessment of EPA's draft cancer risk assessment of inorganic arsenic. Some of these comments are noted by the SAB Workgroup review, whereas others are not sufficiently addressed, in part because of their limited charge. These comments are in addition to those previously submitted to the record, although key comments that are still pertinent are summarized here with additional information.

Our comments primarily relate to EPA Charge Questions 1 and 2 to the 2010 SAB, and in particular to certain aspects concerning epidemiological studies and the dose-response relationship at low doses, respectively.

1.0 Summary of Comments

Overall, the dose-response assessment presented in the draft Toxicological Review (U.S. EPA 2010b) is based primarily on the work of Morales et al. (2000) and NRC (1999, 2001) and is largely uninformed by the past 10 years of additional epidemiological studies and mechanistic research on the toxicology and mode of action of arsenic. The proposed assessment used studies on the arseniasis endemic area of Southwest Taiwan (Wu 1989; Chen et al. 1988, 1992), that examined bladder and lung cancer. The general approach for calculating the proposed cancer slope factor is a linear extrapolation from cancer risks at high doses for the Southwest Taiwanese population to a low-dose comparison population, which largely ignores the shape of the dose-response relationship at lower doses (e.g., <150 $\mu\text{g/L}$, ppb of arsenic in drinking water) even within this population. This approach also ignores the wealth of other low-dose arsenic exposure studies that indicate a sublinear dose-response or lack of significant risk at low doses, as well as consistent evidence for such a dose-response from mode-of-action studies.

Although an approach that overestimates risk will err in the direction of protection of public health, in the case of arsenic, which occurs naturally, greater accuracy using the best scientific data is particularly important. The cancer slope factor predicted by the draft Toxicological Review will result in cancer risk estimates for arsenic that are of "significant public health concern" as noted by the 2010 SAB Workgroup. Such risks even for background exposures to inorganic arsenic in soil, water, and food (Tsuji et al. 2007) complicate risk communications and risk management decisions for even ordinary exposures. A full presentation of possible cancer slope factors based on the available weight of evidence using both linear and non-linear models without a comparison population would provide more information to risk managers on the underlying science, and greater transparency on the effects of modeling assumptions.

1.1 Evaluation of Epidemiological Studies

- As noted by the SAB Workgroup, the evaluation and selection of epidemiological studies was inconsistent and was not conducted using clearly defined and transparent *a priori* criteria. What constituted a significant weakness or strength was not presented. It is also unclear why each individual study is evaluated in isolation and compared to the Southwest Taiwan data, rather than evaluating the weight of evidence from multiple studies through the use of more sophisticated statistical techniques such as meta-analysis. Justification for the exclusion of each study within the context of the criteria-based review was not provided.
- Exposure misclassification and small study size or low power appear to be the primary reasons that all studies other than those of Southwest Taiwan were excluded from consideration, or were considered to be consistent with the risk assessment derived by the draft Toxicological Review. However, the effect of exposure misclassification has been incorrectly interpreted, and the evaluation of power was inconsistent and did not consider the direction or precision of the relative risk estimates (particularly in low exposure categories), or weight and consistency of evidence at low doses from multiple studies, including Southwest Taiwan. Limited statistical power does not necessarily preclude the detection of statistically significant associations. Consistency and precision of estimated relative risks from the “low exposure” studies, when analyzed individually or as a meta-analysis, are more informative than a simple *post-hoc* power calculation.
- With regard to misclassification, confounding, and other biases, the SAB Workgroup noted that “...the key issue is whether the quantitative consequences of bias are of sufficient magnitude to be of concern. Methods are available for this purpose (see, for example: Lash, Fox, and Fink: *Applying Quantitative Bias Analysis to Epidemiological Data*, Springer, 2009). The SAB suggests that the IRIS assessment include a simple table that identifies potential biases (misclassification of exposure, misclassification of disease, omitting confounders, etc.) and the potential magnitude and direction of bias in inferences that are drawn from the study data. A simple summary could then relate these sources of bias to their impact in the data and methods used in the IRIS assessment.”
- The SAB Workgroup noted that power calculations are not the only means by which the relative power of studies should be evaluated and that “Power calculations are useful in planning a study, but after the study is completed, the most informative presentation of epidemiologic findings that combines both the observed results and reflects the power of the study is the relative risk point estimates for a specific exposure comparison and the associated confidence interval.”

- The dose-response evaluation presented is inconsistent with the substantial and growing database from other studies. Epidemiological studies since the 2007 cutoff for literature included in the draft Toxicological Review show consistency with earlier findings of a lack of increased risk at low doses. EPA National Center for Environmental Assessment (NCEA) staff have published the results of an assessment of low-level arsenic studies since 2007 (Gibb et al. 2010). However, this assessment uses the same *post-hoc* power calculations as a means of eliminating the newer studies and contains a number of other technical inaccuracies in evaluating whether evidence from other low-level epidemiological studies is consistent with EPA's estimates of risk based on Southwest Taiwan. The draft Toxicological Review and Gibb et al. (2010) also fail to acknowledge the effect of study design in evaluating power. Many of the low-dose arsenic exposure studies are case-control studies, a more efficient study design that does not need as large a study population to have the same power as a cohort or ecological study.
- The proper approach for assessing statistical consistency of risk estimates from Southwest Taiwan with the weight of evidence from other studies is through use of meta-analysis and comparison of relative risk (RR) estimates and associated confidence limits. Meta-analysis of previous and new studies on the association of low-dose arsenic exposure with bladder cancer indicates consistent, statistically stable, and robust estimates of relative risks that are slightly above 1.0 for smokers and non-smokers combined, and below 1.0 for non-smokers, although neither estimate significantly differs from 1.0 (i.e., no significant increase or decrease in risk from arsenic exposure). The upper confidence limits¹ for both of these relative risk estimates are below the estimated relative risk at 50 $\mu\text{g}/\text{L}$ based on EPA's cancer risk assessment using the Southwest Taiwanese data. Thus, EPA's proposed cancer slope factor results in risks that are higher and statistically inconsistent with risks observed in low-dose epidemiological studies.
- The summary relative risk estimate of below 1.0 from the meta-analysis for non-smokers is in the direction of decreasing bladder cancer risk with increasing arsenic exposure (Summary RR estimate = 0.83, 95% CI of 0.65 to 1.06). As a result, EPA's presumed bias toward the null as a result of non-differential exposure misclassification (assuming a dichotomous comparison, which is incorrect) and low statistical power do not explain the consistent (although generally not significant) *inverse association* (RR<1.0) observed in low-dose studies of non-smokers. This observation clearly demonstrates that exposure misclassification cannot explain the lack of consistency of the epidemiological data with the linear dose-response relationship assumed by EPA.

¹ For collapsed exposure categories above and below 50 $\mu\text{g}/\text{L}$.

- The epidemiological data at low doses in Southwest Taiwan and in other studies consistently reflect a sublinear dose-response relationship or threshold for significant risk below an arsenic concentration in water around 100 to 200 $\mu\text{g}/\text{L}$. Exposure misclassification resulting from variation from multiple well water concentrations within villages does not appear to explain the lack of positive dose-response relationship at low doses in Southwest Taiwan.

1.2 Dose-Response Assessment

- The linear dose-response relationship derived by the draft Toxicological Review essentially arises from the use of a comparison or reference population to anchor the lower end of the dose-response relationship. The portion of the Southwest Taiwan database used is largely at high doses. The estimates are statistically stable because of the large sample size at high doses with elevated risk combined with the Poisson linear model and a huge comparison population anchoring the low end of the dose-response relationship. Such statistical stability should not be confused with accuracy or representativeness in depicting the actual relationship at low doses even for Southwest Taiwan.
- The comparison population is not an appropriate reference or control group because of the differences in socioeconomics, lifestyle, and other factors between the arseniasis endemic area and the greater Southwest Taiwan region included in the comparison population.
- The sensitivity analysis conducted by the draft Toxicological Review to evaluate the effect of the comparison population or type of model was incomplete, and never showed the effect of using a non-linear model with no comparison population or examining only the low-dose range without a comparison population. Thus, the changes presented without the comparison population are constrained and would be even greater than the 88 percent decrease presented for female bladder cancer if, for example, the effect were examined for only the low-dose data (e.g., <150 ppb) or if non-linear modeling were used.
- In addition to the epidemiological studies, the growing weight of evidence on the mode of action for carcinogenesis of inorganic arsenic continues to indicate mechanisms that would be associated with thresholds for significant risk at low doses. The EPA 2005 cancer risk guidelines allow for the consideration of non-linear dose-response relationships for chemicals that are not direct mutagens, and permit reliance on the toxicological weight of evidence even when some uncertainty is present. Detailed comments on mode of action are described in other comment submissions (e.g., by Dr. Samuel Cohen).

- Given the considerable weight of scientific evidence supporting a non-linear dose-response relationship without the comparison population, EPA should present the cancer risk assessment for arsenic using this approach as well. Derivation of the cancer slope factor for ingested inorganic arsenic using a linear no threshold approach and the non-linear approach would provide full transparency of possible values based on the underlying scientific evidence. Such an assessment would provide a more complete evaluation of the science for risk managers.

2.0 Charge Question 1: Evaluation and Selection of Epidemiological Studies

In Charge 1, EPA asked the 2010 SAB Workgroup to comment on the Agency's response to the 2007 SAB's recommendations regarding the evaluation of the epidemiological literature. EPA states that they have performed "an extensive review and evaluation of all available human studies for iAs [inorganic arsenic] using the criteria suggested by the SAB" (U.S. EPA 2010a). Furthermore, EPA states that they agree with the 2007 SAB conclusion "that the Taiwanese data were the best available for determining the carcinogenic risk due to exposure to iAs" and that "there were no other additional epidemiological studies that had comparable utility to the Taiwanese dataset (Wu 1989; Chen et al. 1988, 1992)" (U.S. EPA 2010a).

The specific recommendations from the SAB (2007) included the following:

*The Panel also suggests that published epidemiology studies of US and other populations chronically exposed from 0.5 to 160 µg/L inorganic arsenic in drinking water be **critically evaluated, using a uniform set of criteria and that the results from these evaluations be transparently documented in EPA's assessment documents.** If, after this evaluation, one or more of these studies are shown to be of potential utility, the low-level studies and Taiwan data may be compared for concordance. Comparative analyses could lead to further insights into the possible influence of these differences on population responses to arsenic in drinking water.* (Executive Summary, p. 7, emphasis added)

*All of these studies, including those from Taiwan, Chile, Argentina and the U.S. as described above, should be judged by the same set of criteria, with the comparative assessment of those criteria across studies clearly laid out in a tabular format. Some of the criteria have been listed in the previous paragraph. **The relative strengths and weaknesses of each study need to be described in relation to each criterion.** The caveats and assumptions used should be presented so that they are apparent to anyone who uses these data. Included in the risk assessment background document should be a complete and transparent treatment of variability within and among studies and how it affects risk estimates. **The present lack of transparency in the application of the criteria in the process of study selection was pointed out by several panel members.*** (p. 39, emphasis added)

2.1 Lack of Transparency of Selection Criteria and Study Selection

The SAB Workgroup comments note that the draft Toxicological Review (U.S. EPA 2010b) does not fulfill the objective of a transparent evaluation of other epidemiologic studies of low-level arsenic exposure using a uniform set of criteria. The draft Toxicological Review states that, “Each publication was evaluated using a uniform set of criteria, including the study type, the size of the study population and control population, and the relative strengths and weaknesses of the study” (U.S. EPA 2010b). **Nevertheless, the application of the criteria in the process of study selection, and particularly in the process of study elimination, was not transparent.** Furthermore, it was not stated *a priori* what features would be required for a study to be “of potential utility.” Our previous comments from March 2010 provided examples of the lack of explanation and transparency along with inconsistencies in the draft Toxicological Review. Previous comments from members of the 2007 SAB, also submitted to the record in March 2010, were in agreement that the evaluation of the epidemiological studies in the current draft Toxicological Review is insufficient.

2.2 Misinterpretation of Epidemiological Concepts: Exposure Misclassification and Study Power

In addition to the above issues, there are several instances where epidemiologic concepts are discussed incompletely or inaccurately. Two such instances (exposure misclassification and sample size or study power) appear to constitute the major rationale for discounting the many low-dose studies in the United States and other countries and selecting Southwest Taiwan as the best study on which to base estimates of cancer risk at lower doses examined by risk assessments of arsenic. The draft Toxicological Review states that exposure misclassification may result in an apparent lack of statistical significance of a positive dose-response relationship at low doses, particularly for populations in the United States who are more mobile than in Taiwan. The document also notes that many studies lack the sample size and hence study power of the Southwest Taiwanese database. These types of assertions incorrectly interpret or apply epidemiological concepts and hence are inaccurate.

2.2.1 Exposure Misclassification

Nondifferential misclassification does not always produce bias in the direction of the null (i.e., relative risk of 1.0). There are certain conditions under which the direction of bias is predictable, but these conditions are not met in the epidemiologic studies of health effects of arsenic in drinking water. This concept is explained in relatively simple terms in Rothman’s textbook, *Epidemiology: An Introduction* (2002):

*Nondifferential misclassification of a dichotomous exposure will always bias an effect, if there is one, toward the null value. **If the exposure is not dichotomous, there may be bias toward the null value; but there may also be bias away from the null value,** depending on the categories to which individuals are misclassified.* (p.101, emphasis added)

The more advanced textbook, *Modern Epidemiology* (Third Edition), by Rothman, Greenland, and Lash (2008) notes further that the condition of nondifferentiality (that is, classification of exposure has the identical sensitivity and specificity among cases and noncases, and neither disease nor uncontrolled risk factors result in different accuracy for cases compared to noncases) “may seldom be met exactly” (p. 355). Thus, the following statement in the draft Toxicological Review (U.S. EPA 2010b) **is incorrect and misleading**:

Therefore, studies with low levels of exposure that are ecological in nature (no individual exposure) are more prone to misclassification, which means they are biased toward the null hypothesis (pp. 95 and 147).

Specifically, it is not true that being prone to misclassification means being biased toward the null hypothesis. It should be noted that the illustration of bias toward the null in the paper cited in the draft Toxicological Review, by Cantor and Lubin (2007), is based on a **binary** (i.e., dichotomous) exposure. The epidemiologic studies cited in the draft Toxicological Review measure arsenic in drinking water on a continuous scale and then, for the vast majority of the studies, create multi-level categorical variables to characterize arsenic exposure. For further information on this issue, the reader is referred to Rothman et al. (2008), Dosemeci et al. (1990), Kristensen (1992), Flegal et al. (1991), Wacholder et al. (1991).

The point here is not to downplay the importance of accurate classification of exposure, but rather to emphasize that the direction of bias associated with misclassification is never certain, and is not always even predictable.

Even if exposure misclassification resulted in regression to the null for epidemiological studies of arsenic as alleged in the draft Toxicological Review and by Cantor and Lubin (2007) and Gibb et al. (2010), the consistent relative risk estimates in the direction of <1.0 for non-smokers in low-dose studies argues against such an effect masking significant positive risks for arsenic exposure (see Section 3.0 below). Exposure misclassification resulting from well water variation within villages also does not appear to explain the lack of a positive dose-response for low-dose arsenic exposure in the Southwest Taiwan data set (see Section 4.2 below).

2.2.2 Statistical Power

The draft Toxicological Review praised or criticized epidemiologic studies for statistical power or study size, but there is limited discussion of these issues or of the arguably more important issue of precision. We noted examples of this in our previous comments from March 2010. The EPA NCEA staff paper (Gibb et al. 2010) also eliminated each of the recent low-dose epidemiological studies because of insufficient power or study size compared to the Southwest Taiwan data set. Nevertheless, the SAB Workgroup noted that power calculations are not the only means by which the relative power of studies should be evaluated and that:

Power calculations are useful in planning a study, but after the study is completed, the most informative presentation of epidemiologic findings that combines both the observed results and reflects the power of the study is the

relative risk point estimates for a specific exposure comparison and the associated confidence interval.

Criteria for consideration of sample size also need to take into account study design (e.g., case-control versus weaker study designs such as ecological). The necessary sample size calculations presented in the power evaluation of studies by Gibb et al. (2010, Table 3) are based on a cohort study and do not even approximate sample size requirements for sufficient power of a case-control study. The authors acknowledge that sample size calculations would be different for different parameters, including different study designs, but imply that the estimates they provide are conservative. In fact, the sample size requirements for a case-control study would be considerably lower.

Gibb et al. (2010) also note that the predicted high risks would not be detectable in studies of populations in the United States because of higher background cancer incidence than in Taiwan. However, their evaluation does not consider the results for never smokers for which the background incidence rates for lung and bladder cancer are much lower. Moreover, the background cancer rates used by Gibb et al. (2010) to support their argument are based on rates in Southeastern Asia and North America rather than rates in Taiwan and the United States. The differences between the rates reported by Gibb et al. (2010) are much greater than the differences between either Taiwan and the United States or the differences between Eastern Asia (where Taiwan is located, <http://globocan.iarc.fr/population.htm>) and North America. If the actual reported age-adjusted incidence rates for Taiwan and the United States are used, Taiwan has a 2.0 times lower bladder cancer rate for males and females combined (not four times lower as reported by Gibb et al.) and a 29% lower lung cancer rate for males (rather than 40%) and a 47% lower lung cancer rate for females (rather than a 3 times lower rate or 67% lower rate [IARC 2008]).

Finally, despite criticisms about small sample size, statistically significant results were observed in some of the analyses of never smokers who had lived at least 50% of their lifetime (until diagnosis) in study towns in the Bates et al. (1995) study (see Table 3, Bates et al. 1995). Thus, limited statistical power does not always preclude the detection of statistically significant associations. **Consistency and precision of estimated relative risks from the “low exposure” studies, when analyzed individually or as a meta-analysis, are more informative than a simple *post-hoc* power calculation.**

2.3 Overall Inconsistent Presentation of Study Evaluation and Selection

The draft Toxicological Review is inconsistent in its listing of strengths and/or weaknesses across epidemiologic studies with similar characteristics, nor are the strengths and weaknesses described in the context of criteria, as requested by the SAB (2007). Furthermore, some of the weaknesses listed are based on misinterpretation of epidemiologic concepts.

The draft Toxicological Review states,

The Taiwanese database is still the most appropriate source for estimating bladder and lung cancer risk among humans (specifics provided in Section 5)

because of: (1) the size and statistical stability of the database relative to other studies; (2) the reliability of the population and mortality counts; (3) the stability of residential patterns; and (4) the inclusion of long-term exposures.

Nevertheless, the SAB (2007) Panel listed several limitations of the Southwest Taiwan database, including the following:

...its ecologic character, lack of smoking information, limited precision of exposure estimates, especially among villages with multiple wells, and the possible issue of compromised nutrition among segments of the exposed population.

These relative strengths and limitations are not reconciled in the present draft Toxicological Review.

In conclusion, as also noted by the 2010 SAB Workgroup, EPA has provided only a partial response to the 2007 SAB's recommendation to perform a critical evaluation of the relevant epidemiologic literature, "using a uniform set of criteria and that the results from these evaluations be transparently documented in EPA's assessment documents" (SAB 2007). The review and evaluation of the epidemiologic literature was conducted without describing a methodological approach, the relative strengths and weaknesses of each study were not described in relation to *a priori* criteria, justification for the final decision regarding each study was not provided, and the process of the evaluation was neither transparent nor reproducible.

2.4 Recommendations

In order to achieve what the 2007 SAB requested, it is recommended that EPA return to their charge and address their suggestions systematically and completely. Previous comments submitted by one of the 2007 SAB members and by EPRI included a draft format for summary tables with specific guidelines recommended by that SAB. Furthermore, as noted by the current SAB Workgroup, it would be informative to conduct sensitivity analyses and/or uncertainty analyses to estimate the magnitude and direction of the biases identified in the review of the epidemiological studies. The SAB Workgroup pointed to a specific text, namely, Lash, Fox, and Fink: *Applying Quantitative Bias Analysis to Epidemiological Data*, Springer, 2009, but other approaches are also available. EPA appears to be taking the view that biases inherent in the Southwest Taiwan studies are less substantial and/or influential than the biases from the low-dose epidemiological studies. Therefore a quantitative assessment of bias in all of the studies would be helpful.

3.0 Weight of Evidence from Low-Dose Epidemiological Studies

The current draft Toxicological Review (U.S. EPA 2010b) lacks references to the epidemiologic literature published during the past 3 years. There are several important primary articles (Mostaf et al. 2008; Heck et al. 2009; Chen et al. 2009, 2010; Meliker et al. 2010) and two

meta-analyses (Chu and Crawford-Brown 2006, 2007; Mink et al. 2008) that were not included in the draft Toxicological Review and should be added. Our previous comments from March 2010 presented a table summarizing the main features and results of these studies. An EPA NCEA staff publication (Gibb et al. 2010) attempts to address these studies individually using an approach based on study power similar to that used by the draft Toxicological Review to evaluate previous studies.

However, rather than evaluate each low-dose study individually for power *post hoc*, the appropriate method for evaluating whether the cancer risk assessment based on the Southwest Taiwan data is consistent with the available evidence from low-dose studies is to conduct a meta-analysis and to evaluate relative risk point estimates and associated confidence intervals as indicated by the 2010 SAB Workgroup. Two such meta-analyses (Chu and Crawford-Brown 2006, 2007, and Mink et al. 2008) have been published that have not been acknowledged by the draft Toxicological Review or the EPA NCEA staff publication (Gibb et al. 2010). Limitations of the Chu and Crawford-Brown meta-analysis have been summarized (Brown 2007b; Crawford-Brown 2007; Mink et al. 2008). **The purpose of our meta-analysis was to examine the potential association between low-level exposure to arsenic in drinking water and bladder cancer, using meta-analysis to improve precision and increase statistical power (Mink et al. 2008).** We updated our meta-analysis to include the two recent studies of bladder cancer (Chen et al. 2010; Meliker et al. 2010). Because Chen et al. (2010) did not stratify on smoking status, their study is included only in the updated analyses restricted to ever smokers and never smokers combined. Chen et al. (2010) is an update to Chiou et al. (2001), and therefore Chiou et al. was removed from the updated analyses. In our previous comments in March 2010, we included all exposure groups of Chen et al. (2010) even though a considerable number of study participants had drinking water arsenic exposure concentrations in excess of 100 to 300 ppb, which is much higher than the vast majority of data from the low-dose studies. We have therefore updated the Mink et al. (2008) calculations with exclusion of exposure categories of >300 ppb or >100 ppb in Chen et al. (2010). We also excluded the “unknown” exposure category in Chen et al. (2010) because high exposures may have also occurred in this category.

The overall pattern of results (Table 1) is similar to those of Mink et al. (2008). Summary relative risk estimates (SRRE) decreased slightly for “ever” and “never” smokers combined and for “ever” smokers, and increased slightly for “never” smokers; the 95% confidence interval corresponding to the SRREs became narrower, indicating an improvement in precision. The SRRE for “ever” smokers was somewhat attenuated, but there was still evidence of heterogeneity. In both the previous and the updated meta-analysis, the results for “never” smokers are the most robust. Indeed, for the purpose of dose-response assessment, these data from “never” smokers should provide statistically stable estimates. The background incidence for both bladder and lung cancer for “never” smokers is also lower than the general population incidence. These findings in combination with the consistency of the findings for “never” smokers in the individual studies indicate that low level exposure to arsenic in drinking water alone is unlikely to contribute to a significant increase in bladder cancer incidence.

Table 1. Updated meta-analysis of Mink et al. (2008)

Analysis	SRRE	95% CI	P-value for Heterogeneity
Never smokers	0.83	0.65–1.06	0.894
Ever smokers	1.19	0.97–1.45	0.038
All subjects	1.08 ^a	0.95–1.24	0.223
	1.06 ^b	0.94–1.20	0.429

^a Excluding >300 ppb and unknown exposure categories of Chen et al. (2010)

^b Excluding >100 ppb and unknown exposure categories of Chen et al. (2010)

These meta-analysis results can be used to evaluate whether cancer risk estimates predicted by NRC (2001) and the draft Toxicological Review are consistent with the results of low-dose epidemiological studies. NRC (2001) predicted a range of relative risk estimates of 1.22 to 2.57 for bladder cancer from drinking water with arsenic at 50 ppb, depending on the model used. EPA NCEA staff (Gibb et al. 2010) present a relative risk of 1.64 for bladder cancer at 50 ppb based on NRC (2001) which is the approach used in the draft Toxicological Review. The updated meta-analysis SRREs using collapsed categories above and below 50 ppb for all subjects² and “never” smokers³ are well below these predicted relative risks based on high dose exposure in Southwest Taiwan. The upper confidence limits for combined smoking categories slightly overlap the low end of the NRC (2001) relative risk ranges but are below the 1.64 relative risk presented by Gibb et al. (2010). **Thus, the risk estimates from NRC (2001) and thereby the draft Toxicological Review are overestimated at low doses and are not statistically consistent with the results of low-dose exposure studies.**

SRREs for never smokers are consistently below 1.0, although not significant. Following the logic presented in the draft Toxicological Review regarding exposure misclassification and lack of power at low doses causing bias to the null and decreased statistical significance, **it is quite possible that with increased statistical power and precision, the relative risk estimates for the lowest exposure groups would indeed achieve statistical significance, and would be statistically significantly below 1.0.**

The database from low-exposure populations is growing and many of the issues that are raised in the draft Toxicological Review (U.S. EPA 2010b) have been addressed in more recent, updated studies (e.g., Chen et al. 2009, 2010). **Furthermore, the draft Toxicological Review has indicated that the data from the Taiwan studies, while imperfect, still have utility. It has not been demonstrated why data from other imperfect studies could not also be used in a weight of evidence or quantitative evaluation.**

² SRRE=1.06, 95% CI of 0.83 to 1.35, excluding >300 ppb and unknown groups of Chen et al. (2010); SRRE=1.02, 95% CI of 0.82 to 1.26, excluding >100 ppb and unknown groups of Chen et al. (2010).

³ SRRE=0.82, 95% CI of 0.61 to 1.09.

4.0 Charge Question 2: Dose-Response Modeling

The dose-response assessment presented in the draft Toxicological Review uses a low-dose linear (Poisson) model in which the lower end of the curve originates at the data point for a comparison population representing all of Southwest Taiwan. **The combined effect of this approach, however, largely ignores the shape of the dose-response relationship at low doses based on this database, the consistent weight of evidence from other epidemiological studies at low doses, and mechanistic data on the mode of action of arsenic carcinogenicity indicating threshold or sublinear dose-response relationships.** Using a non-linear approach based on the weight of evidence is consistent with the 2005 EPA cancer risk guidelines, even when uncertainty is present regarding the exact mode of action (U.S. EPA 2005). Although the 2007 SAB did not acknowledge this part of the 2005 cancer risk guidelines, they did recommend that the effect of non-linear models be considered as well.

4.1 Effect of Comparison Population and Linear Dose-Response Modeling

The dose-response model presented in the draft Toxicological Review uses a low-dose linear model and anchors the lower end of the dose-response curve at the data point for the Southwest Taiwan regional comparison population, assuming this region has zero arsenic in drinking water (p. 120; U.S. EPA 2010b). Anchoring the lower end of the dose-response curve at the data point for the comparison population essentially ignores the shape of the data at low doses (Morales et al. 2000; Brown 2007a). The draft Toxicological Review presents a limited sensitivity analysis and reports that use of the comparison population does not make much of a difference (e.g., at most an 88 percent decrease in risk for female bladder cancer without the comparison population; page 139). However, this analysis uses the full range of the well water concentration data and the low-dose linear model. Whether or not a comparison population is used has less of an effect if a linear relationship is forced through the entire range of data.

SAB (2007) recommended that EPA test the sensitivity of the model to the choice of the reference population (Southwest Taiwan) and to the assumption of linearity by also using an alternative hazard model with a dose contribution that is multiplicative and non-linear in form. However, the sensitivity analysis presented in the draft Toxicological Review (pp. 139–140) is limited and constrained, and consequently does not show the full effect of the various assumptions in combination. **For transparency in showing the full effect of these assumptions, the draft Toxicological Review should present a more comprehensive sensitivity analysis that shows the effect of using non-linear models with and without the comparison population.**

4.2 Evidence Indicating that the Comparison Population is Inappropriate or Needs Adjustment for the Arseniasis Endemic Area

NRC (2001) recognized that the potential differences other than arsenic exposure between the study population and comparison population could affect the results and that the comparison population appeared to have much lower bladder and lung cancer rates even at the same low

doses as in the study villages, suggesting that these populations differed in ways other than arsenic exposure:

A potential disadvantage, however, of using an external comparison group is that the analysis can be biased if the study population differs from the comparison population in important ways. (p. 190; NRC 2001; emphasis added)

In general, estimated ED_{01s} tend to be lower for models that included an external comparison population, primarily because the lung and bladder cancer rates seen in the comparison populations were much lower than rates seen even in the study villages with low exposures. (p. 191, NRC 2001; emphasis added).

The draft Toxicological Review accepts the NRC (2001) recommendation to use the Southwest Taiwan comparison population without evaluating whether the rationale is sound. NRC (2001) recommended using a comparison population largely based on 1) the results of Tsai et al. (1999), which showed that cancer risks were similar whether the study population was compared to a Southwest Taiwan comparison population or an all of Taiwan comparison population, 2) uncertainty in mechanistic data at that time, and 3) the possibility of exposure misclassification affecting the dose-response relationship for the study area. Nevertheless, the Southwest Taiwan comparison population area includes the counties of Chiaya and Tainan, which also have large urban areas that have long been settled. Tainan City was the first capital of Taiwan established by the Dutch in the early 1600s.

The assumption of equivalence in socioeconomic and lifestyle between the study area townships and the comparison population region should be examined. The greater Southwest Taiwan region is not as impoverished with the severely undernourished conditions of the arseniasis endemic villages.⁴ Moreover, Tsai et al. (1999) found that the arsenic-exposed villages, when compared to these other reference populations, had significantly increased cancer mortality from all causes and many other cancers (e.g., brain, bone, nasal cavity, colon, intestine, stomach) that were not consistently related to arsenic in the wealth of other epidemiological studies. By contrast, the two reference populations, Southwest Taiwan region and all of Taiwan, were more similar in cancer rates. **Significantly increased risk of cancer from all causes as well as non-arsenic related cancers likely reflects poorer health conditions or factors other than arsenic that might be affecting cancer risk for the arsenic exposed areas compared to the other two reference populations.**

Since NRC (2001), considerably more information is available on the carcinogenic mode of action for inorganic arsenic and its metabolites. The overall weight of evidence from this information supports a non-linear mode of action at low doses (see comments submitted by Dr. Samuel Cohen and EPRI).

NRC (2001) and the draft Toxicological Review suggest that the comparison population is needed because exposure misclassification has caused a lack of dose-response relationship at

⁴ Such nutritional deficiencies have been shown to increase the toxicity of arsenic in this population and others (e.g., Chen et al. 2001; Milton et al. 2004; Mitra et al. 2004; Gamble et al. 2005, 2007; Spallholtz et al. 2004; Yang et al. 2002; Miyazaki et al. 2005).

low doses in the arsenic exposed villages. Exposure misclassification is a serious problem for the Southwest Taiwan study population because the median well water concentration was used to represent exposure for a village despite large variation in well water concentrations among wells within many study villages. However, as noted above, exposure misclassification does not always bias associations toward the null (i.e., toward a relative risk of 1.0) nor would it always produce a lack of a dose-response relationship particularly for continuous data such as arsenic in drinking water.

Brown (2007a) examined the effect of variation in arsenic well water concentration within a village by conducting separate dose-response assessments for villages with low variation ($<25 \mu\text{g/L}$) and those with high variation ($>24 \mu\text{g/L}$) in arsenic well water concentrations within a village. If exposure misclassification resulting from well water variation is masking a positive dose-response, then the “low variation” villages should show more of a positive dose-response trend for cancer risk than the “high variation” villages. Instead, the opposite was observed. Brown (2007a) found that if only the 23 villages with low variation in well water concentration were considered, no dose-response relationship was apparent, although considerable variation in cancer risk among villages was apparent, unrelated to well water concentration. By contrast, the 19 “high variation” villages, which should be more prone to exposure misclassification, indicated more of a positive dose-response relationship (particularly for females), with a gradual increase in risk at low exposures and a steeper increase in risk at higher median well water concentrations (e.g., $>200\text{-}300 \mu\text{g/L}$; Brown 2007a). **Therefore, the available data do not indicate that exposure misclassification from multiple well water concentrations within villages is the cause of the lack of dose-response relationship at low doses. These results accordingly do not support the use of a comparison population as recommended by NRC (2001).**

4.3 Recommendations for Dose-Response Assessment

1. Conduct and report a full sensitivity analysis of model shape (e.g., linear, non-linear), with and without the comparison population, for the whole dose-response range as well as the lower dose region (e.g., <100 to $200 \mu\text{g/L}$).
2. Evaluate and report the effect of using an area term in the analysis that accounts for the underlying difference in risk for the comparison population versus study area villages, which appears to be unrelated to arsenic exposure.
3. Incorporate elements of a weight of evidence evaluation in the dose-response assessment including the wealth of epidemiological studies and mode of action information on arsenic carcinogenicity, as allowed by the EPA Cancer Risk Guidance (U.S. EPA 2005) and report results of the evaluation as a range of CSFs, such as linear and non-linear modeling without the comparison population.

References

Bates, M.N., Smith, A.H., Cantor, K.P., 1995. Case-control study of bladder cancer and arsenic in drinking water. *Am. J. Epidemiol.* 141(6), 523–530.

Brown, K.G., 2007a. How creditable are cancer risk estimates from the S.W. Taiwan database for arsenic in drinking water? *Hum. Ecol. Risk Assess.* 13, 190.

Brown KG. 2007b. Letter to the editor: Inorganic arsenic in drinking water and bladder cancer: A meta-analysis for dose-response assessment. *Int. J. Environ. Res. Public Health* 4(2):193–194.

Cantor, K.P., Lubin, J.H., 2007. Arsenic, internal cancers, and issues in inference from studies of low-level exposures in human populations. *Toxicol. Appl. Pharmacol.* 222(3):252–257.

Chen C-L, Chiou H-Y, Hsu L-I, Hsueh Y-M, Wu M-M, Chen C-J. 2009. Ingested arsenic, characteristics of well water consumption and risk of different histological types of lung cancer in northeastern Taiwan. *Environ. Res.* DOI: 10.1016/j.envres.2009.08.010.

Chen C-L, Chiou H-Y, Hsu L-I, Hsueh Y-M, Wu M-M, Wang Y-H, Chen C-J. 2010. Arsenic in drinking water and risk of urinary tract cancer: A follow-up study from northeastern Taiwan. *Am. Assoc. for Cancer Res.* DOI: 10.1158/1055-9965.EPI-09-0333. Available at: www.aacrjournals.org.

Chen CJ, Chen CW, Wu MM, Kuo TL. 1992. Cancer potential in liver, lung, bladder and kidney due to ingested inorganic arsenic in drinking water. *Br. J. Cancer* 66(5), 888–892.

Chen CJ, Kuo TL, Wu MM. 1988. Arsenic and cancers. *Lancet*: letter to the editor. February 20, 1988.

Chen, C-J., Y-M. Hsueh, M-P. Tseng, Y-C. Lin, L-I. Hsu, W-L. Chou, H-Y. Chiou, I-H. Wang, Y-L. Chou, C-H. Tseng, and S-H. Liou. 2001. Individual susceptibility to arseniasis. pp. 135–143. In: *Proc. of the Fourth Annual International Conference on Arsenic Exposure and Health Effects*. June 18–22, 2000. W.R. Chappell, C.O. Abernathy, and R.L. Calderon (eds). Elsevier Science Ltd.

Chiou H-Y, Chiou S-T, Hsu Y-H, Chou Y-L, Tseng C-H, Wei M-L, Chen C-J. 2001. Incidence of transitional cell carcinoma and arsenic in drinking water: A follow-up study of 8,102 residents in an arseniasis-endemic area in northeastern Taiwan. *Am. J. Epidemiol.* 153(5), 411–418.

Chu H-A, Crawford-Brown D. 2007. Letter to the editor: Inorganic arsenic in drinking water and bladder cancer: A meta-analysis for dose-response assessment. *Int. J. Environ. Res. Public Health* 4(4):340–341.

- Chu HA, Crawford-Brown DJ. 2006. Inorganic arsenic in drinking water and bladder cancer: A meta-analysis for dose-response assessment. *Int. J. Environ. Res. Public Health* 3(4):316–322.
- Crawford-Brown D. 2007. Letter to the editor: Response to letter from K.G. Brown. *Int. J. Environ. Res. Public Health* 4(2):194.
- Dosemeci M, Wacholder S, Lubin JH. 1990. Does nondifferential misclassification of exposure always bias a true effect toward the null value? *Am. J. Epidemiol.* 132(4):746–748.
- Flegal KM, Keyl PM, Nieto FJ. 1991. Differential misclassification arising from nondifferential errors in exposure measurement. *Am. J. Epidemiol.* 134(10):1233–1244.
- Gamble, M.V., X. Liu, H. Ahsan, J.R. Pilsner, V. Llievski, V. Slavkovich, F. Parvez, D. Levy, P. Factor-Litvak, and J.H. Graziano. 2005. Folate, homocysteine, and arsenic metabolism in arsenic-exposed individuals in Bangladesh. *Environ. Health Perspect.* 113(12):1683–1688.
- Gamble, M.V., X. Liu, V. Slavkovich, J.R. Pilsner., V. Ilievski, P. Factor-Litvak, D. Levy, S. Alam, M. Islam, F. Parvez, H. Ahsan, and J.H. Graziano. 2007. Folic acid supplementation lowers blood arsenic. *Am J. Clin. Nutr.* 86:1202–1209.
- Gibb, H, Haver C, Gaylor D, Ramasamy S, Lee JS, Lobdell D, Wade T, Chen C, White P, and Sams R. 2010. Utility of recent studies to assess the National Research Council 2001 estimates of cancer risk from ingested arsenic. *Environ. Health Perspect.* doi: 10.1289/ehp.1002427 (available at <http://dx.doi.org/>) Online 28 October 2010.
- Heck JE, Andrew AS, Onega T, Rigas JR, Jackson BP, Karagas MR, Duell EJ. 2009. Lung cancer in a U.S. population with low to moderate arsenic exposure. *Environ. Health Perspect.* 117(11):1718–1723.
- IARC. 2008. Globocan 2008 Country Fast Stat. Available at: <http://globocan.iarc.fr/>. Accessed November 10, 2010.
- Kristensen P. 1992. Bias from nondifferential but dependent misclassification of exposure and outcome. *Epidemiology.* 3(3):210–215.
- Meliker JR, Slotnick MJ, AvRuskin GA, Schottenfeld D, Jacquez GM, Wilson ML, Goovaerts P, Franzblau A, Nriagu JO. 2010. Lifetime exposure to arsenic in drinking water and bladder cancer: A population-based case-control study in Michigan, USA. *Cancer Causes Control.* DOI: 10.1007/s10552-010-9503-z.
- Milton, A.H., Z. Hasan, S.M. Shahidullah, S. Sharmin, M.D. Jakariya, M. Rahman, D. Dear, and W. Smith. 2004. Association between nutritional status and arsenicosis due to chronic arsenic exposure in Bangladesh. *Int. J. Environ. Health Res.* 14(2):99–108.

- Mink, P.J., D.D. Alexander, L.M. Barraj, M.A. Kelsh, and J.S. Tsuji. 2008. Low-level arsenic exposure in drinking water and bladder cancer: A review and meta-analysis. *Regul. Toxicol. Pharmacol.* 52:299–310.
- Mitra, S.R., D.N. Guha Mazumder, A. Basu, G. Block, R. Haque, S. Samanta, N. Ghosh, M.M. Hira Smith, O.S. von Ehrenstein, and A.H. Smith. 2004. Nutritional factors and susceptibility to arsenic-caused skin lesions in West Bengal, India. *Environ. Health Perspect.* 112(10):1104–1109.
- Miyazaki, K., C. Watanabe, K. Mori, K. Yoshida, and R. Ohtsuka. 2005. The effects of gestational arsenic exposure and dietary selenium deficiency on selenium and selenoenzymes in maternal and fetal tissues in mice. *Toxicology* 208(3):357–365.
- Morales KH, Ryan L, Kuo TL, Wu M-M, Chen C-J. 2000. Risk of internal cancers from arsenic in drinking water. *Environ. Health Perspect.* 108(7):655–661.
- Mostafa MG, McDonald JC, Cherry NM. 2008. Lung cancer and exposure to arsenic in rural Bangladesh. *Occup. Environ. Med.* 65:765–768. DOI: 10.1136/oem.2007.03785.
- NRC. 1999. Arsenic in drinking water. National Research Council, Washington, DC.
- NRC. 2001. Arsenic in drinking water: 2001 update. National Research Council, Washington, DC.
- Rothman KJ, Greenland S, Lash TL. 2008. *Modern Epidemiology* 3rd Ed. Philadelphia: Lippincott, Williams and Wilkins.
- Rothman KJ. 2002. *Epidemiology: An Introduction*. New York: Oxford University Press.
- SAB. 2007. Advisory on EPA's assessments of carcinogenic effects of organic and inorganic arsenic: A report of the U.S. EPA Science Advisory Board. EPA-SAB-07-008. U.S. Environmental Protection Agency, Science Advisory Board, Washington, DC.
- Spallholtz, J.E., L.M. Boylan, and M.M. Rhaman. 2004. Environmental hypothesis: Is poor dietary selenium intake an underlying factor for arsenicosis and cancer in Bangladesh and West Bengal, India? *Sci. Total Environ.* 323:21–32.
- Tsai S-M, Wang T-N, Ko Y-C. 1999. Mortality for certain diseases in areas with high levels of arsenic in drinking water. *Arch. Environ. Health* 54(3):186–193.
- Tsuji, J.S., L.J. Yost, L.M. Barraj, C.G. Scrafford, and P.J. Mink. 2007. Use of background inorganic arsenic exposures to provide perspective on risk assessment results. *Regul. Toxicol. Pharmacol.* 48:59–68.
- U.S. EPA. 2005. Guidelines for carcinogen risk assessment. EPA/630/P-03/001F. U.S. Environmental Protection Agency, Risk Assessment Forum, Washington, DC.

U.S. EPA. 2010b. Toxicological review of inorganic arsenic (CAS No. 7440-38-2). EPA/635/R-10/001. Draft. U.S. Environmental Protection Agency, Washington, DC. Available at: www.epa.gov/iris.

U.S. EPA. 2010a. NCEA's Proposed Charge to External Reviewers for the IRIS Toxicological Review of Inorganic Arsenic (cancer). U.S. Environmental Protection Agency, Washington, DC.

Wacholder S, Dosemeci M, Lubin JH. 1991. Blind assignment of exposure does not always prevent differential misclassification. *Am. J. Epidemiol.* 134(4):433-437.

Wu M-M, Kuo T-L, Hwang Y-H, Chen C-J. 1989. Dose-response relation between arsenic concentration in well water and mortality from cancers and vascular diseases. *Am. J. Epidemiol.* 130(6):1123-1132.

Yang, L., W. Wang, S. Hou, P.J. Peterson, and W.P. Williams. 2002. Effects of selenium supplementation on arsenism: An intervention trial in Inner Mongolia. *Environ. Geochem. Health* 24:359-374.