Prepared for
Superfund Settlement Project (SSP)
RCRA Corrective Action Project (RCAP)

Document type
Report

Date
January 2024

Comments on the BMD Model Code and Modeling Results for the Draft IRIS Toxicological Review Inorganic Arsenic

RAMBØLL Bright ideas. Sustainable change.

Contents

1. Introduction and Overview

This document provides comments on the Benchmark Dose (BMD) model code and BMD modeling results used in the development of the United States Environmental Protection Agency (USEPA) Integrated Risk Information System's (IRIS) External Review Draft of the Toxicological Review of Inorganic Arsenic (Toxicological Review; USEPA 2023a) and the associated Supplemental Information (USEPA 2023b). USEPA released the BMD model code and BMD modeling results provided for the draft IRIS Toxicological Review of Inorganic Arsenic on December 29, 2023 for public comment, with the comment period ending January 16, 2024.

While the comments set forth in the following pages identify numerous deficiencies in the BMD model code and BMD modeling results, it does not represent a comprehensive review of all of the modeling conducted for all of the endpoints considered for dose-response modeling in the Toxicological Review. A 15-day comment period for a highly technical analysis supported by voluminous documentation in the absence of instructions on the step-by-step processes for running the models is not adequate; therefore, this review provides specific comments on those cancer and noncancer endpoints the reviewers were able to evaluate within the comment period. Overall, we found the modeling poorly documented, with inconsistencies between information in the modeling files and the documentation provided in the Toxicological Review (USEPA 2023a) and the associated Supplemental Information (USEPA 2023b).

Most troubling, the USEPA appears to have made the decision not to include a number of figures in the Toxicological Review (2023a), which they had already generated, that would have provided a much more informative presentation of the fit of the modeling to the individual studies for each endpoint. A critical deficiency that we previously noted in the Toxicological Review (2023a) was that the comparisons of model predictions for extra risk with study results in Figures 4-7 through 4-13 only used "summarized" study data, which the model appears to fit exactly. However, we were able to find R code in the BMD modeling materials that generates plots showing the comparison of model prediction of relative risk with individual study results for every endpoint (see example in section 3.2, below). These comparisons should be included in the Toxicological Review (2023a) to provide a more transparent assessment.

The complexity of the meta-regression analysis precludes any independent quantitative evaluation within the timeframe of the SAB review, but these newly discovered plots demonstrate that there is significant uncertainty in risks estimated by the agency's dose-response modeling that is not transparently presented in the Toxicological Review. These uncertainties are also reflected in the disparity between the risk estimates obtained by USEPA (2023) with estimates obtained under different modeling assumptions (Shao et al. 2021), as was discussed in our previous comments.

Unfortunately, the necessary files to investigate the model sensitivity analyses conducted by the USEPA (2023a) are still not available (see next section). In particular, it is not possible to determine whether the USEPA considered any modeling options where studies of extremely high exposures, which would be uninformative regarding the carcinogenicity of inorganic arsenic at the much lower concentrations of concern for the IRIS assessment, were excluded. Given the strong evidence that the dose-response for the carcinogenicity of arsenic is nonlinear, this would be a reasonable option for the cancer endpoints.

2. General Comments

- The documentation and methods used to identify studies and data for inclusion in Bayesian meta-regression dose-response analyses are lacking in transparency. While the methods for study and data selection are presented in the Systematic Review Protocol

(USEPA 2023c), it is impossible to follow the decision process through the Systematic Review (USEPA 2023a), Systematic Review Supplemental Data (USEPA 2023b), and the BMD data files.

- Methods used to convert data for use in the regression are complex and convoluted. It takes multiple spreadsheets to understand what was used for each set of doses and the parameters used to perform the Monte Carlo conversions. This makes it impossible to review all the endpoints in the files given in the time allotted for public comment.
- Data are missing, for example the spreadsheet for calculating effective counts mentioned on page C-17 of the Supplemental document (USEPA 2023b) is not provided. It appears that calculations for effective counts are conducted in code in R, but the code is difficult to follow and not well documented.
- Organization of the files provided is given without a mapping to show how they interconnect and where information is located. In addition, the number of significant digits past the decimal differs from one file to another, and also in the tables presented in the supplemental material. This adds difficulty in duplicating the final values from the modeling.
- Other general items that make it difficult to quickly evaluate the modeling conducted and verify the results include:
    - The studies are not in alphabetic order in the R input files.
    - Variation in reporting standard deviations directly or as 3×SD.
    - Most of the references to YASALW.xla are to specific sites on the USEPA drives and must be changed, as the reviewer of the files would be unable to access USEPA drives; however, no instructions were provided to achieve this in the absence of accessing the USEPA drives.
- There appears to be little attention to detail with results incorrectly reported for some of the studies in the tables in the supplemental file. For example, in the values reported in Table C-38 which include the beta values (Mean, Standard error of the mean, mean standard deviation and 2.50%, 25%, 50%,75%, and 97.5%) reported for Pan et al. 2013 (study #4) in the output file are a repeat of the Grau Perez (2007) beta and do not match the values reported for study 4 in the file **Full_output Diabetes mle.csv,** which appears to contain the results that were reported in Table C-37.
- In the Supplemental section C2.2 Neurocognitive Effects Exposure-Response Modeling Results and specifically on supplemental page C-188 it is stated that "… USEPA conducted dose response analyses for only two studies (Wasserman et al., 2014; Wasserman et al., 2004) for which the authors provided raw data on exposures, outcomes, and covariates". However, no files were provided in the modeling files provided by the USEPA on the modeling of the neurocognitive effects.
- Regarding the dose-response modeling approach, the Toxicological Review (2023a) indicates that "a sensitivity analysis using a more complex double Hill model that allows for negative response estimates was conducted to verify the reasonableness of this assumption (see Appendix C, Section C.1.1 Sensitivity Analysis of Possible Non-monotonic Dose-Response Relationships)." In the supplemental section, Sensitivity Analysis of Possible Non-monotonic Dose-Response Relationships, the use of fractional polynomial models and "double Hill" models are discussed, and tables of results are provided. In addition, the supplemental file states "each subset of the fractional polynomial model group has its own pair of R and Stan programs to implement it." However, these files were not provided by the USEPA. Without these additional files, it is difficult to ascertain exactly what data was used in evaluating these alternate models or how complete the analysis of alternate models was.

3. Comments by Topic

3.1 Study Selection

- Page 4-5 of USEPA (2023a) states that, "To promote consistency and to document the meta-regression data set selection, at least 2 reviewers evaluated studies according to the above considerations and provided qualitative ratings for each element and provided an overall rating (good/criteria met, fair, poor/criteria not met) for each criterion, with a brief description of the basis for the choice. Reviewers then discussed the ratings and resolved differences or refined concerns. To help focus the discussions, the number of study limitations for each study were tabulated, and studies with the most markdowns were further examined for suitability."
    - The decisions of the reviewers' evaluations of each study should be presented in the supplemental data and data files. A review of the file iAs_study_selection_for-dose_response_Jan2021.xlsx indicates some information that appears to correspond to the evaluation of each of the studies for inclusion in qualitative and quantitative analysis. However, to follow the decision process in this file, more details are needed to understand and transparently document the reviewer's decision making process and the overall results in regard to studies that were selected and studies that were eliminated from consideration for qualitative and quantitative analyses.
- According to the Systematic Review Protocol (USEPA 2023c), epidemiologic studies containing exposure- or dose-response data were subject to risk-of-bias (RoB) evaluations to assess aspects of internal validity of study findings based on study design and conduct for hazard identification. According to USEPA (2023a), medium or high confidence studies with exposure or dose response data were considered for dose response. The studies evaluated for risk of bias for diabetes are presented in Figures 3-22 and 3-23. However, not all of the studies that are presented in Figures 3-22 and 3-23 considered medium- or high-confidence studies are presented in the file iAs_study_selection_for-dose_response_Jan2021.xlsx. Therefore, it seems that not all studies that underwent risk of bias evaluation and received a medium- or high-confidence rating were considered for qualitative analysis.
    - Overall, a transparent method for mapping each study's progress through risk of bias and confidence evaluations to consideration and decisions regarding inclusion in the qualitative evaluations is needed.

3.2 BMD Code Comments

R Script Comments (File: MR hier all_10-5-2022-AL.R):

1. A few discrepancies were noted between the input data (mle doses-Baris Unlagged-mg_FitDrop.csv) and what is in the supplement:
    a. Chen et al. (2010) dose is 21.6 in data from the code, but 18.2 is reported in Table C19
    b. Highest dose for Steinmaus et al. (2013) is missing in the data provided in the code
    c. Highest dose for Lin et al. (2018) is missing in the data provided in the code
2. Lines 213-258: Setting up and running Stan model
    a. What is PY_ref? –All of the data variables and parameters should have at least a brief description/definition in the Stan code if not in the R code as well

b. Where does b_sigma_scale (=5) come from?

c. Note that max_treedepth and adapt_delta have been set to be higher than normal. This isn't necessarily an issue, but could indicate problems with model identifiability, etc. This should be discussed in the supplement since setting to default max_treedepth and adapt_delta results in low ESS/n_eff .

d. Consider including some visual diagnostics in the supplement to support the Stan run: https://mc-stan.org/bayesplot/articles/visual-mcmc-diagnostics.html

3. Lines 414-446: Plot across studies

   a. Defining values

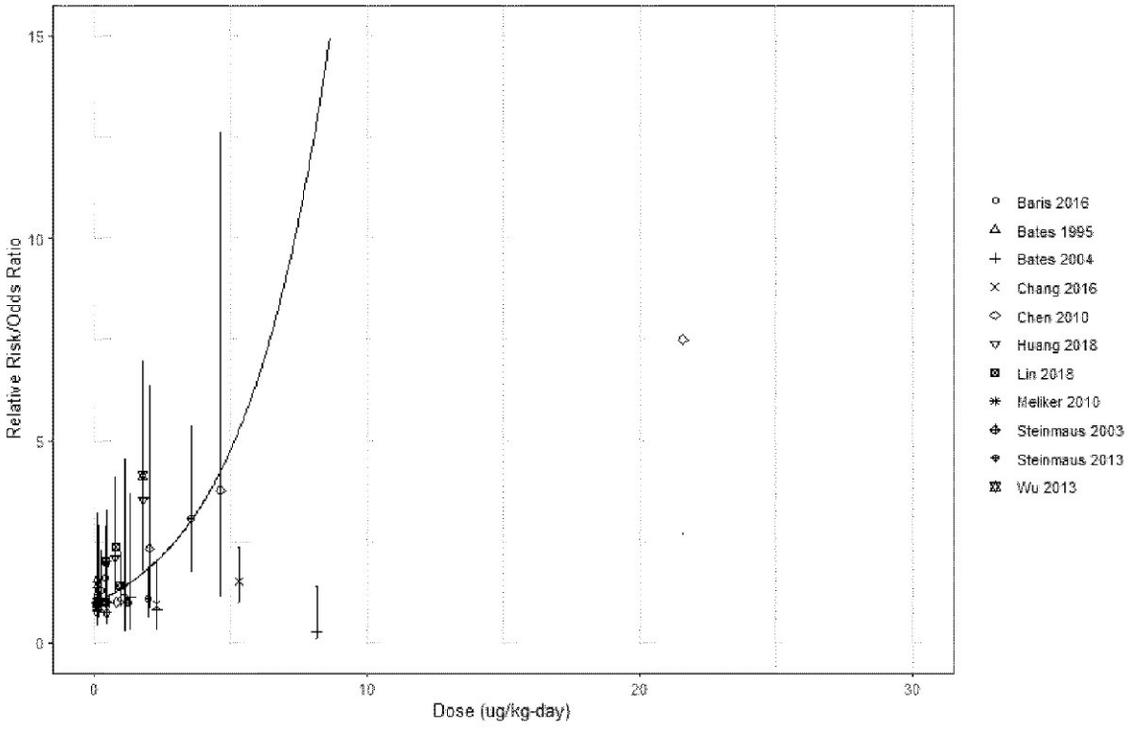
      i. Ref_dose_overall: This seems to be the background exposure, but the value is 0.071429 (ug/kg-day?) which is the background from Allen et al. but *not* the dose suggested in the EPA report of 0.0365 ug/kg-day (e.g. Subtext for Table ES-1 page 16). Please explain this discrepancy,

      ii. astar.all: Please describe this quantity. What is it, and why was a value of -0.5 chosen? And why does that same value apply for all endpoints?

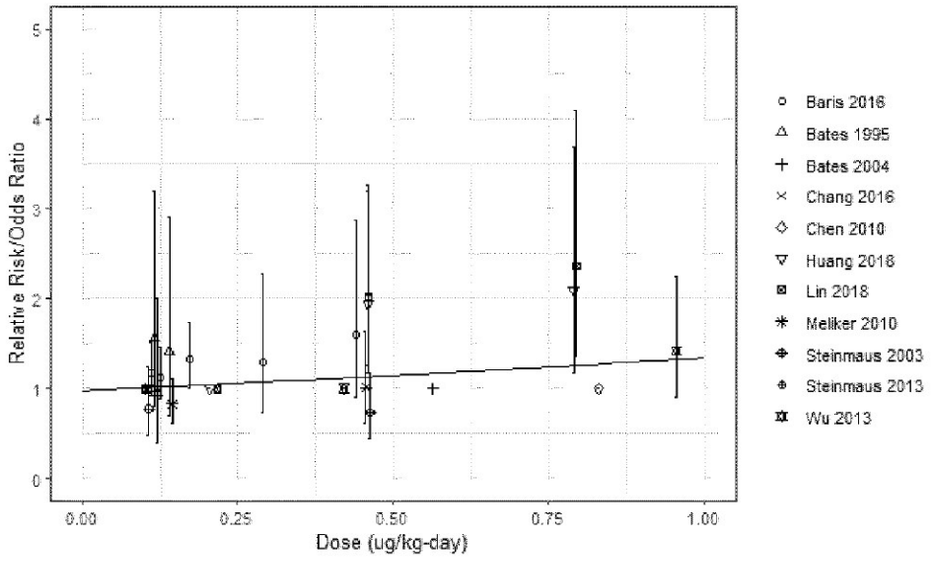
   b. Plots

      This summary plot of RR/OR vs. Dose (first figure below, generated in the R code) does not appear in the document or supplement, and it is unclear why it would be omitted. It demonstrates the variability of the studies around the overall model fit and seems particularly important at doses ≤ 1 µg/kg-day (second figure below) which the authors focus on in other plots. How do the values for these plots then get translated into the extra risk values (e.g. for Figure 4-7)? Where is the code to do so? Why are those plots not generated as part of the R code?

Bladder Cancer (Data\Bladder Cancer\MLE Meta-Regressions-Post Interagency Review\mle-fit drop\ DR_combined_BC FitDrop - new bsigma.png)



Lung Cancer (Data\Lung\MLE Meta-Regressions-Post Interagency Review\Doses-MLE\DR_combined_BC FitDrop - new bsigma.png)



4. Overall comment: There is really no reason for the analyses to be as disjointed as they are, and the lack of clear file structure, documentation, or analysis flow (there should at minimum be a diagram explaining this) indicates poor computational analysis practice which would not meet the bare minimum for publication much less for a regulatory guidance document.

3.3 BMD Data Files and Modeling

3.3.1 Bladder Cancer Files

The file BladderCancer\MLE Meta-Regressions-Post Interagency Review\mle-fit drop\ Full_output.txt seems to contain the pooled results given in Table C-20 (USPEA 2023b) Summary of bladder cancer Bayesian analysis output using MLE dose estimates. It is unclear how all of the output maps to Table C-20 (USEPA 2023b). Below are comparisons made with the Intake Uncertainties files for Bladder Cancer and the Conversion Factor Validation Spreadsheet_v4.xlsx as well as the input files for the MLE Meta Regression using the mle-fit drop input and output file.

Baris et al. 2016

- Age mean and SD used in workbook but not given in Table C-18, but used in simulation calculations.
- Units for LE in Table C-18 are µg/L but in the Baris et al. workbook and in the Conversion Factor Validation Spreadsheet_v4.xlsx on the Baris et al. 2016 worksheet units are reported as µg/day.
- Conversion Factor Validation Spreadsheet_v4.xlsx indicates on the Baris et al. 2016 worksheet that Solver was used for average age, but it appears that a weighted average (both cases and total weighted averages give a value of 65 when rounded).
- Why are the cases and controls by dose group not listed in the workbook? They can only be found in the input file for R
- Cases and controls in input files do not match Table C-19 – neither raw counts nor effective counts – but the dose values for analysis are the same. Control counts listed on Main!C22-C27 do not match the R input files.

Bates et al. 1995

- Age mean and SD used in workbook are not given in Table C-18, but used in simulation calculations.
- Units for LE in Table C-18 are µg/L but in the Baris et al. workbook and in the Conversion Factor Validation Spreadsheet_v4.xlsx on the Bates et al. 1995 worksheet units are reported as µg/day.
- Why are the controls by dose group given in the Bates workbook but the cases are not listed?

Bates et al. 2004

- Age is not used directly in Monte Carlo calculations of Bates et al. 2004, it is used only to determine factor f, so the equation given in Table C-18 is wrong.
- Why are the controls by dose group given in the Bates workbook but the cases are not listed?

Chang et al. 2016

- No explanation or justification for determining percent males versus females used in estimating the gender specific Urinary creatinine/kg-day is provided.
- Table C-19 provides effective counts, but no raw counts. However, the R input file includes raw values and the Effective counts output from R show the matching effective counts.

Chen et al. 2010

- Conversion Factor Validation Spreadsheet_v4.xlsx Chen 2010 worksheet reports Dietary intake (ug/kg-day) as 0.65 ± 0.33, which is what is used in the Chen et al. workbook, Table C-18 reports it as 65 ± 3.33.
- Equation given in Table C-18 does not match what was included in the Chen et al. 2010 intake uncertainty workbook, which does not use age or BW in the calculations directly but does use RDWE - Reported avg duration of well exposure (yrs)
- Average age at diagnosis – points to a table at right in workbook for avg start and end ages. Start age is calculated based on the mid-point of 10 year ranges in columns L to M rows 5 to 9 (age range and count) but this table is not present in any form in Chen et al. 2010 intake uncertainty workbook. The end age is the average age calculated plus the average years of follow up. Since the average years of follow up is provided in the Chen et al. 2010 intake uncertainty workbook with a standard deviation (3.6), why is a single point used for the age at diagnosis, which is used to calculate the most likely value for the BetaPert distribution of f. This makes the calculation of f less certain.
- Conversion Factor Validation Spreadsheet_v4.xlsx on the Chen et al. 2010 worksheet refers to a Hsieh 2008 for the duration of exposure. Not sure which of two Hsieh (2008a or b) references are being referred to that are provided in the Toxicological Review (USEPA 2023a).
- There are multiple citations of the Taiwanese Exposure Handbook (or just Taiwanese Handbook) in the Conversion Factor Validation Spreadsheet_v4.xlsx, with – no reference provided for this document in either the main (USEPA 2023a) or supplemental file (USEPA 2023b).
- The highlighted values in the copy of Table C-3 below are as given in Chen_2010_NE_Taiwan_bladder-08-10-22.xlsx. However, it should be noted that the reported MLE is larger than the 95th percentile. Also, the yellow highlighted value has an incorrect placement of comma.
- Cases and effective counts are only supplied in the input and output R files.

Table C-3. MLE, low and high MCMC dose estimates for three different WCR and RDWE distribution assumptions[a]

Exposure ranges (µg/L)	Most Likely (MLE; µg/kg-d)			Low (5th percentile µg/kg-d)			High (95th percentile µg/kg-d)		
	Log-normal	Normal	Uniform	Log-normal	Normal	Uniform	Log-normal	Normal	Uniform
Meliker et al. 2010									
0–10	0.103	0.109	0.106	0.101	0.100	0.099	0.110	0.116	0.112
10–100	0.145	0.145	0.152	0.136	0.124	0.142	0.154	0.154	0.162
100–1,000	0.455	0.450	0.457	0.333	0.266	0.336	0.723	0.729	0.706
Chen et al. 2010b									
0–400	0.830	0.835	0.771	0.810	0.804	0.753	0.851	0.897	0.788
400–1,000	1.106	1.275	0.928	1.078	1.013	0.908	1.136	1.325	0.949
1,000–500	2.042	2.463	1.460	1.956	1.800	1.418	2.120	2.753	1.503
5,000–10,000	4.646	13.561	2.942	4.402	3.551	2.820	4.912	6.973	3.072
10,000–100,0000	21.595	22.264	12.685	18.196	15.074	10.634	26.152	37.053	15.284

[a] Dose estimates obtained from "NORMAL", "UNIFORM" and "LOGNORMAL" results, Main tab, Meliker2010_CE5- Ln_ugperday-08-08-22.xlsx and Chen_2010_NE_Taiwan_bladder-08-10-22.xlsx, Supplemental Material, bladder cancer "Intake Uncertainty…" folder, EPA HERO database).

Huang et al 2018

- No explanation or justification for determining percent males versus females used in estimating the gender specific Urinary creatinine/kg-day.
- Table C-19 provides effective counts but no raw counts. However, the R input file includes raw values and the Effective counts output from R show the matching effective counts.

Lin et al. 2018

- No explanation or justification is provided for determining percent males versus females used in estimating the gender specific Urinary creatinine/kg-day.
- Table C-19 provides effective counts but no raw counts. However, the R input file has raw values and the Effective counts output from R show the matching effective counts.

Meliker et al 2010 – no specific comments

Steinmaus et al. 2003 – no specific comments

Steinmaus et al. 2013

- The Conversion Factor Validation Spreadsheet_v4,xlsx WCR (Water consumption rate (L/day)) includes a value of 1.7 ± 0.9 which matches the Steinmaus et al. 2013 workbook. However, Table C-18 reports a value of 1.7 ± 0.09.

Wu et al. 2013

- No explanation or justification is provided for determining percent males versus females used in estimating the gender specific Urinary creatinine/kg-day

3.3.2 Diabetes

Using examples from what appears to be the MLE modeling for Diabetes in directory Diabetes\MLE MR-Post IAR-No Rangel-Moreno2022\mle, below are comments on issues with the uncertainty workbooks, input files, and correlation between output files and summary tables in the supplemental document (USEPA 2023b).

Coronado Gonzalez 2007

- No explanation or justification for determining percent males versus females used in estimating the gender specific Urinary creatinine/kg-day.

Pan et al. 2013

- Intake Uncertainty worksheets defines **DI - Dietary intake (ug/kg-day) (see below)** as having a mean of 1.44 and 3×SD of 1, but there is no reference or code to indicate the source for these values.

Grau Perez 2007

- No explanation or justification is provided for determining percent males versus females used in estimating the gender specific Urinary creatinine/kg-day.
- Control or cohort numbers are not mentioned in Table C-37 but are supplied in the R input file. These values should be supplied in the documentation of the Toxicological Review, not buried within modeling files.

James et al. 2013

- Control or cohort numbers are not mentioned in Table C-37 but are supplied in the R input file.

3.3.3 Lung Cancer

Using examples from the MLE modeling for Lungs in directory Lung\MLE Meta-Regressions-Post Interagency Review\Doses-MLE, below are comments on issues with the uncertainty workbooks, input files, and correlation between output files and summary tables in the supplemental document.

Argos et al. (2014)

- No explanation or justification was provided for determining percent males versus females used in estimating the gender specific Urinary creatinine/kg-day.

García-Esquinas et al. (2013)

- Results for beta in Table C-30 do not match the beta values in the Final output.txt.

Chen et al. (2010a)

- The mean value for DI is calculated in the workbook but no justification for SD is provided?

- The Taiwanese handbook again is mentioned as a source in this workbook, but no citation is provided in the IRIS Assessment documentation USEPA (2023a, 2023b).

Dauphiné et al. (2013) – no specific comments

D'Ippoliti et al. (2015) Males & D'Ippoliti et al. (2015) Females –

- Only one uncertainty workbook and one spreadsheet is provided in the Conversion Factor Validation Spreadsheet_v4.xlsx. In the R input file, dosing information is the same for both genders; however, the cases and number of controls differ.

Ferreccio et al. (2000) – no specific comments

Mostafa et al. (2008) Smokers & Mostafa et al. (2008) Non-Smokers

- Two different uncertainty workbooks are provided, but only one spreadsheet is included in the Conversion Factor Validation Spreadsheet_v4.xlsx worksheet which is mislabeled as Mostafa 2009. The difference in two workbooks is that the Smoker worksheet (Mostafa_2009_DW_Smoke-02-10-21.xlsx) has a 3×SD value of 3.3 for LE - Low (e.g. tap water) exposure (ug/L) and non-smoker workbook (Mostafa_2009_DW_NSmoke-02-10-21.xlsx) has a value of 9.9 for the 3×SD for LE. The Conversion Factor Validation Spreadsheet_v4.xlsx worksheet for Mostafa 2009 lists a mean and std for LE as 2.5 ± 3.3. Table C-28 in the supplemental document (USEPA 2023b) lists LE for both smokers and non-smokers as 2.5 ± 3.3. This difference makes some minor difference in the doses – which are reported with enough significant digits to see the differences in the R input file.

Steinmaus et al. (2013) – no specific comments

4. References

USEPA. 2023a. IRIS Toxicological Review of Inorganic Arsenic [CASRN 7440-38-2]. United States Environmental Protection Agency, Integrated Risk Information System, Center for Public Health and Environmental Assessment, Office of Research and Development. Washington, DC. EPA/635/R-23/166a.

USEPA. 2023b. IRIS Toxicological Review of Inorganic Arsenic Supplemental Information [CASRN 78-30-44]. United States Environmental Protection Agency, Integrated Risk Information System, Center for Public Health and Environmental Assessment, Office of Research and Development. Washington, DC. EPA/635/R-23/166b.

USEPA. 2023c. Updated Problem Formulation and Protocol for the Inorganic Arsenic IRIS Assessment [CASRN 7440-38-2]. Supplemental Information – Appendix A. United States Environmental Protection Agency, Integrated Risk Information System, Center for Public Health and Environmental Assessment, Office of Research and Development. EPA/635/R-23/166b.

All remaining references cited are provided in USEPA (2023a, 2023b), with the exception of those that are noted to be missing.