SCIENCE ADVISORY BOARD
A Federal Advisory Committee to the U.S. Environmental Protection Agency

November 19, 2024

EPA-SAB-25-004

The Honorable Michael S. Regan
U.S. Environmental Protection Agency
1200 Pennsylvania Avenue, N.W.
Washington, D.C. 20460

Subject: Transmittal of the DRAFT Science Advisory Board Report titled "Review of EPA's draft IRIS Toxicological Review of Inorganic Arsenic," dated October 2023.

Dear Administrator Regan,

The EPA's Office of Research and Development, Integrated Risk Assessment Program requested that the Science Advisory Board (SAB) review the IRIS Toxicological Review of Inorganic Arsenic. In response to the EPA's request, the SAB assembled the Inorganic Arsenic Review Panel with subject matter experts to conduct the review. Please find enclosed the final report from the SAB.

The SAB Inorganic Arsenic Review Panel met virtually on January 5, 2024, to listen to a presentation by EPA staff, and then in-person on January 24-26, 2024 to deliberate on the Agency's charge questions. Additional virtual meetings were held on July 8 and 16, 2024, to discuss the panel's draft report. Extensive oral and written public comments were considered throughout the advisory process. This report, reviewed and approved during the SAB October 15-16, 2024, public meeting, conveys the consensus advice of the SAB. In addition, Appendix A contains recommended changes that are primarily editorial in nature and Appendix B highlights important Tier 2 recommendations that can be readily addressed.

The SAB wishes to commend the EPA IRIS Program for the thoroughness of the report and for the briefings they provided at our January meetings. The SAB agreed with many of the conclusions presented in the draft IRIS document. The SAB had several concerns regarding the presentation of the meta-regression analyses and identified areas that would benefit from further clarification to enhance transparency and increase the utility of the IRIS document.

We note that the IRIS document is highly responsive to previous recommendations from the National Academies of Sciences, Engineering and Medicine, focusing on epidemiological studies, appropriate databases, and thorough description of the population, exposures, comparators, and outcomes. The confidence conclusions are justified and strengthened by the wealth of human data in the conclusions of the hazard assessment for the cancer and noncancer disease endpoints. The systematic review approach and the methods used by EPA for including/excluding studies in the draft IRIS review of inorganic arsenic are described in detail and are generally presented in a consistent and logical manner across the four health outcomes. The SAB concurs that the epidemiological data are of a sufficiently broad range to make low dose extrapolation needed only occasionally to bridge across studies. The focus on chronic exposure and the inclusion of both studies with environmental samples and biomarkers of arsenic exposure is appropriate.

While the SAB includes many recommendations within this report, we would like to highlight the following. The assessment should reconsider whether setting a reference dose (RfD) for inorganic arsenic is scientifically supportable, and a detailed presentation of this decision should be provided in the IRIS document. This discussion should address the following issues: 1) Whether an RfD is appropriate for a contaminant with serious non-cancer effects that exhibit a low-dose linear (non-threshold) dose-response; 2) The need for a science-policy decision on a risk level that does not represent appreciable risk if an RfD for inorganic arsenic is developed; 3) Consideration of an alternative approach for assessment of non-cancer effects of inorganic arsenic based on a data driven model relating risk and dose, with recognition of the need to select a target risk level for non-cancer effects if this approach is used.

If a decision is made to develop an RfD for inorganic arsenic, then the SAB recommends the following steps: 1) Predicted risk at the RfD should be presented for each non-cancer effect for which no threshold is identified and indicated the risk level is not considered to represent an appreciable risk; 2) Ischemic heart disease (IHD) rather than cardiovascular disease is recommended as the key endpoint for diseases of the cardiovascular system; 3) EPA should evaluate whether fatal occurrences of IHD or incidence of IHD (including both fatal and non-fatal occurrences) is most appropriate as the basis for the RfD; 4) Feasibility of performing dose-response evaluation and identifying a point of departure (POD) for neurodevelopmental effects should be evaluated. If determined that development of a POD for neurodevelopmental effects is not feasible, then application of a database uncertainty factor (UF) > 1 to account for potentially more sensitive neurodevelopmental effects should be considered; and 5) A benchmark response (BMR) of 1% instead of 5% (i.e., $BMDL_{01}$ instead of $BMDL_{05}$) should be considered in light of the severity (substantial incidence of mortality) of IHD and Type 2 diabetes.

The SAB found the cancer slope factor to be scientifically justified but recommended that the IRIS document provide the “background” risk estimates from background dietary and drinking water inorganic arsenic exposure and state that these risks (1 in 1000; 2 in 1000) are much

greater than the cancer risk range usually considered by EPA in developing health-based criteria (i.e., typically 1 in 10,000 to 1 in 1,000,000 risk).

With regard to other health outcomes, the SAB recommends that the EPA prioritize the analysis on Type 2 over Type 1 diabetes. The evidence is robust for a causal relationship between inorganic arsenic and Type 2 diabetes but there are relatively few studies supporting the inclusion of Type 1 diabetes. It is also recommended that EPA include more cohort studies with prenatal and early childhood exposure assessments of inorganic arsenic to provide further evidence on neurobehavioral and birth outcomes, particularly those employing biomarkers of exposure. The EPA should provide a rationale for the selection of one study (Kile et al., 2016) from the 68 high and medium confidence studies evaluated for adverse pregnancy outcomes and specifically address the adequacy of this single study for determining a POD that is protective for this outcome. Further, additional dose-response evaluation is recommended to assess whether neurodevelopmental toxicity is the most sensitive effect.

The report includes multiple suggestions for a more practical and accessible explanation of the meta-regression modelling and suggestions for additional sensitivity analyses. EPA should apply a fit statistic that penalizes for model complexity to determine if the simple linear logistic model is sufficient, provide convergence diagnostics for all assessed outcomes, and examine goodness-of-fit using posterior predictive plots. The IRIS document should clarify that PBPK models provided the scientific basis for the dose conversions but were not directly used to estimate daily dose. More details should be provided on how the lifetable calculations were performed, including a worked example and an explanation of required modifications when age-specific incidence rates are not available. Risk levels for non-cancer and cancer effects at the same dose of iAs should be presented and compared, with recognition that both types of health effects from low-dose exposure to iAs are of concern.

As the EPA finalizes its draft IRIS Toxicological Review of Inorganic Arsenic, the SAB encourages the Agency to address the noted concerns and consider the provided advice and recommendations. The SAB recognizes that there are a large number of Tier 1 and Tier 2 recommendations but notes that many can be promptly addressed to enhance transparency and clarity. The SAB appreciates this opportunity to review the IRIS Toxicological Review of Inorganic Arsenic and looks forward to the EPA's response to these recommendations.

Sincerely,

/s/

Kimberly Jones, Ph.D.
Chair
EPA Science Advisory Board

/s/

Peter S. Thorne, M.S., Ph.D.
Chair
EPA Inorganic Arsenic Review Panel

NOTICE

This report has been written as part of the activities of the EPA Science Advisory Board, a public advisory committee providing extramural scientific information and advice to the Administrator and other officials of the Environmental Protection Agency. The Board is structured to provide balanced, expert assessment of scientific matters related to problems facing the Agency. This report has not been reviewed for approval by the Agency and, hence, the contents of this report do not represent the views and policies of the Environmental Protection Agency, nor of other agencies in the Executive Branch of the Federal government, nor does mention of trade names or commercial products constitute a recommendation for use. Reports of the EPA Science Advisory Board are posted on the EPA website at https://sab.epa.gov.

The SAB is a chartered federal advisory committee, operating under the Federal Advisory Committee Act (FACA; 5 U.S. Code 10). The committee provides advice to the Administrator of the U.S. Environmental Protection Agency on the scientific and technical underpinnings of the EPA's decisions. The findings and recommendations of the Committee do not represent the views of the Agency, and this document does not represent information approved or disseminated by EPA.

U.S. Environmental Protection Agency
Science Advisory Board Inorganic Arsenic Panel

CHAIR

Dr. Peter S. Thorne, University of Iowa Distinguished Chair and Professor, Department of Occupational & Environmental Health, College of Public Health, Director of Human Toxicology Program, University of Iowa, Iowa City, IA

MEMBERS

Dr. Aaron Barchowsky, Associate Professor, Department of Environmental and Occupational Health, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA

Dr. Hugh Barton, Independent consultant, Mystic, CT

Dr. Mark Borsuk, Professor of Civil and Environmental Engineering, Pratt School of Engineering, Duke University, Durham, NC

Dr. Sylvie M. Brouder, Professor and Wickersham Chair of Excellence in Agricultural Research, Department of Agronomy, Purdue University, West Lafayette, IN

Dr. Aimin Chen, Professor of Epidemiology, Department of Biostatistics, Epidemiology and Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA

Dr. Wei-Chun Chou, Assistant Professor, Department of Environmental Sciences, University of California, Riverside, CA

Dr. Matthew Gribble, Associate Professor, School of Medicine, Department of Medicine, Division of Occupational, Environmental and Climate Medicine, University of California, San Francisco, CA

Dr. Margaret Karagas, Professor and Chair, Department of Epidemiology, Geisel School of Medicine at Dartmouth, Hanover, NH

Dr. Ana Navas-Acien, Leon Hess Professor and Chair, Department of Environmental Health Sciences, Mailman School of Public Health, Columbia University, New York, NY

Dr. Michael Pennell, Professor, Division of Biostatistics, College of Public Health, Ohio State University, Columbus, OH

Dr Isaac Pessah, Distinguished Professor Emeritus, Molecular Biosciences, School of Veterinary Medicine, University of California, Davis, Davis, CA

Dr. Gloria Post, Research Scientist, Division of Science and Research, New Jersey Department of Environmental Protection, Trenton, NJ

Dr. Christopher States, Professor, Department of Pharmacology and Toxicology, University of Louisville School of Medicine, Louisville, KY

Dr. Daniele Wikoff, Chief Scientific Officer, Principal, Practice Director, Health Sciences, ToxStrategies, Inc, Ashville, NC

Dr. Carol Wood, Oak Ridge National Laboratory (retired), Harriman, TN

SCIENCE ADVISORY BOARD STAFF

Dr. Diana Wong, Designated Federal Officer, U.S. Environmental Protection Agency, Science Advisory Board Staff Office, Washington, DC

U.S. Environmental Protection Agency
Science Advisory Board

CHAIR

Dr. Kimberly L. Jones, Associate Provost and Professor of Civil and Environmental Engineering, Department of Civil and Environmental Engineering, Howard University, Washington, DC

MEMBERS

Dr. C. Marjorie Aelion, Dean Emerita and Professor Emerita, Dean Emerita, School of Public Health and Health Sciences and Professor Emerita, Department of Environmental Health Sciences, University of Massachusetts Amherst, Amherst, MA

Dr. David T. Allen, Gertz Regents Professor of Chemical Engineering and Director of the Center for Energy and Environmental Resources, Department of Chemical Engineering, The University of Texas, Austin, TX

Dr. Susan Anenberg, Professor and Chair, Department of Environmental and Occupational Health, Milken Institute School of Public Health, George Washington University, Washington, DC

Dr. Florence Anoruo, Assistant Professor of Plant and Environmental Science and Associate Research Scientist, Department of Biological and Physical Sciences, South Carolina State University, Orangeburg, SC

Dr. Joseph Arvai, Director and Dana and David Dornsife Professor, Wrigley Institute for Environment and Sustainability, University of Southern California, Los Angeles, CA

Dr. Maxmilian Auffhammer, Professor, Department of Agricultural and Resource Economics, University of California, Berkeley, Berkeley, CA

Dr. Roland Benke, Director, Renaissance Code Development, LLC, Austin, TX

Dr. Veronica J. Berrocal, Professor, Department of Statistics, University of California Irvine (UCI), Irvine, CA

Dr. Tami Bond, Walter Scott, Jr. Presidential Chair in Energy, Environment and Health, Department of Mechanical Engineering, Colorado State University, Fort Collins, CO

Dr. Elizabeth W. Boyer, Professor of Environmental Science, Department of Ecosystem Science and Management, Penn State University, University Park, PA

Dr. Sylvie M. Brouder, Professor, Wickersham Chair of Excellence in Agricultural Research and Director of Global Food Security, Department of Agronomy, Purdue University, West Lafayette, IN

Dr. Jayajit Chakraborty, Professor, Mellichamp Chair in Racial Environmental Justice, Bren School of Environmental Science & Management, University of California, Santa Barbara, Santa Barbara, CA, CA

Dr. Aimin Chen, Professor, University of Pennsylvania, Philadelphia, PA

Dr. John DiGiovanni, Professor and Coulter R. Sublett Chair in Pharmacy, Division of Pharmacology and Toxicology and Department of Nutritional Sciences, Dell Pediatric Research Institute, The University of Texas at Austin, Austin, TX

Mr. Earl W. Fordham, Deputy Director, Office of Radiation Protection, Division of Environmental Public Health, Washington Department of Health, Richland, WA

Dr. Rebecca Fry, Professor, Department of Environmental Science and Engineering, School of Global Public Health, University of North Carolina, Chapel Hill, Chapel Hill, NC

Dr. Joshua Graff Zivin, Pacific Economic Cooperation Chair in International Economic Relations, Department of Economics, School of Global Policy and Strategy, University of California, San Diego, La Jolla, CA

Dr. John Groopman, Edyth H. Schoenrich Professor of Preventive Medicine, Department of Environmental Health and Engineering Associate Director for Population Sciences Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins Bloomberg School of Public Health Johns Hopkins School of Medicine, Johns Hopkins University, Baltimore, MD

Dr. Steven P. Hamburg, Chief Scientist, Environmental Defense Fund, Providence, RI

Dr. Wendy J. Heiger-Bernays, Emeritus Clinical Professor, School of Public Health, Boston University, Boston, MA

Dr. Selene Hernandez-Ruiz, Director, Laboratory and Analytical Services Division, Water Resources Mission Area, U.S. Geological Survey, Lakewood, CO

Dr. Lea Hildebrandt Ruiz, Associate Professor, McKetta Department of Chemical Engineering, University of Texas at Austin, Austin, TX

Dr. David Keiser, Professor, University of Massachusetts Amherst, Amherst, MA

Dr. Mark W. LeChevallier, Principal, Dr. Water Consulting, LLC, Morrison, CO

Dr. Angela M. Leung, Associate Professor of Medicine, David Geffen School of Medicine; VA Greater Los Angeles Healthcare System, University of California Los Angeles, Los Angeles, CA

Dr. Hui Li, Professor of Environmental Soil Chemistry, Department of Plant, Soil and Microbial Sciences, Michigan State University, East Lansing, MI

Director Lisa Lone Fight, Geospatial/Environmental Scientist, Science, Technology, and Research Department, MHA Nation, Mandan, Hidatsa and Arikara Nation (Three Affiliated Tribes of the Fort Berthold Indian Reservation), New Town, ND

Dr. Lala Ma, Carl F. Pollard Associate Professor of Health Economics, Department of Economics, Gatton College of Business and Economics, University of Kentucky, Lexington, KY

Dr. Jade Mitchell, Professor and Associate Chair, Biosystems and Agricultural Engineering, Michigan State University, East Lansing, MI

Dr. Enid Neptune, Associate Professor of Medicine, Department of Medicine, Division of Pulmonary and Critical Care Medicine, Johns Hopkins University, Baltimore, MD

Dr. Carla Ng, Associate Professor, Department of Civil and Environmental Engineering, University of Pittsburgh, Pittsburgh, PA

Dr. Sheila Olmstead, Professor and Cornell Atkinson Scholar, Jeb E. Brooks School of Public Policy, Cornell University, Ithaca, NY

Dr. Austin Omer, Sustainable Systems Agronomist, Crop Science Commercial, Bayer U.S., Morton, IL

Dr. Gloria Post, Research Scientist, New Jersey Department of Environmental Protection (Retired), Langhorne, PA

Dr. Amanda D. Rodewald, Garvin Professor and Senior Director of Center for Avian Population Studies, Department of Natural Resources and the Environment, Cornell Lab of Ornithology, Cornell University, Ithaca, NY

Dr. Jonathan M. Samet, Professor, Departments of Epidemiology and Environmental and Occupational Health, Colorado School of Public Health, Aurora, CO

Dr. Jeremy Sarnat, Associate Professor, Gangarosa Department of Environmental Health, Rollins School of Public Health, Emory University, Atlanta, GA

Dr. Drew Shindell, Distinguished Professor, Nicholas School of the Environment, Duke University, Durham, NC

Dr. Genee Smith, Assistant Professor, Department of Environmental Health and Engineering, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD

Dr. Daniel O. Stram, Professor, Department of Population and Public Health Sciences, Keck School of Medicine, University of Southern California, Los Angeles, CA

Dr. Peter S. Thorne, University of Iowa Distinguished Chair & Professor, Department of Occupational & Environmental Health, College of Public Health, University of Iowa, Iowa City, IA

Dr. Godfrey Arinze Uzochukwu, Senior Professor and Director, Waste Management Institute, North Carolina Agricultural and Technical State University, Greensboro, NC

Dr. Wei-Hsung Wang, Professor, Center for Energy Studies and Director of the Radiation Safety Office, Louisiana State University, Baton Rouge, LA

Dr. June Weintraub, Deputy Director for Environmental Health, Center for Environmental Health, California Department of Public Health, Sacramento, CA

Dr. Sacoby Wilson, Professor and Director of the Center for Community Engagement, Environmental Justice, and Health (CEEJH), Maryland Institute for Applied Environmental Health, School of Public Health, University of Maryland-College Park, College Park, MD

Dr. Douglas Wolf, Independent Consultant, Independent Consultant, Pittsboro, NC

Dr. Yiliang Zhu, Professor, Division of Epidemiology, Biostatistics, and Preventive Medicine Department of Internal Medicine, School of Medicine, University of New Mexico, Albuquerque, NM

SCIENCE ADVISORY BOARD STAFF

Dr. Shaunta Hill-Hammond, Designated Federal Officer, U.S. Environmental Protection Agency, Science Advisory Board Staff Office, Washington, DC

TABLE OF CONTENTS

ACRONYMS AND ABBREVIATIONS

ADME	absorption, distribution, metabolism, and excretion
AIC	Akaike's information criterion
AOP	adverse outcome pathway
BIC	Bayesian information criterion
BMD	benchmark dose
BMDL	benchmark dose lower confidence limit
BMR	benchmark response
CAAC	Chemical Assessment Advisory Committee
CI	confidence interval
cIMT	carotid intima-media thickness
CL	confidence limit
CSF	cancer slope factor
CVD	cardiovascular disease
DCS	diseases of the cardiovascular system
DMA	dimethylarsinic acid
EPA	Environmental Protection Agency
HAWC	Health Assessment Workplace Collaborative
iAs	inorganic arsenic
IHD	ischemic heart disease
IRIS	Integrated Risk Information System
LOAEL	lowest observed adverse effect level
MMA	monomethylarsonic acid
MOA	mode of action
NASEM	National Academies of Science, Engineering, and Medicine
NOAEL	no observed adverse effect level
NRC	National Research Council
ORD	Office of Research and Development
PBPK	physiologically based pharmacokinetic
PECO	populations, exposures, comparators, and outcomes
POD	point of departure
RfD	oral reference dose
RRB	relative risk over the background exposure
SAB	Science Advisory Board
UF	uncertainty factor
UF_A	animal-to-human uncertainty factor
UF_H	human variation uncertainty factor
UF_L	LOAEL-to-NOAEL uncertainty factor
UF_S	subchronic-to-chronic uncertainty factor
UF_D	database deficiencies uncertainty factor
U.S.	United States

1. INTRODUCTION

The U.S. Environmental Protection Agency (EPA) has developed an assessment titled the "draft IRIS Toxicological Review of Inorganic Arsenic" in support of the Agency's online database, the Integrated Risk Information System (IRIS). IRIS is produced and maintained by EPA's Center for Public Health and Environmental Assessment within the Office of Research and Development (ORD). Draft IRIS assessments contain information about chemicals that encompass hazard identification and a dose-response assessment, two of the four steps in the human health risk assessment process. When used by risk managers in combination with information on human exposure and other considerations, draft IRIS assessments support the Agency's regulatory activities and decisions to protect public health. The IRIS Program is developing this assessment of inorganic arsenic (iAs) at the request of multiple EPA national and regional programs. The methods used in the assessment are summarized in the iAs protocol (link provided in Appendix A of the Draft IRIS document) and have been reviewed by the National Academies of Sciences, Engineering, and Medicine (NASEM; formerly the National Research Council) [NRC, 2013]. Methods and problem formulation decisions were heavily informed by prior NASEM input (NRC, 2014)**;** (NASEM, 2019).

The assessment under review updated a prior IRIS assessment of iAs (US EPA, 1995) that included an oral reference dose (RfD) and a determination of carcinogenic potential, and cancer slope factor (CSF) for carcinogenic effects. The "draft IRIS Toxicological Review of Inorganic Arsenic" includes a review of the available scientific literature on the noncancer and cancer health effects in humans exposed to iAs. The systematic review protocol for iAs and appendices for toxicokinetic information, dose-response modeling, and other supporting materials were provided as Supplemental Information—Appendix A: Updated Problem Formulation and Protocol for the Inorganic Arsenic IRIS Assessment—Appendices B, C, D, E, F and G to the draft Toxicological Review. The EPA's Office of Research and Development requested that the Science Advisory Board (SAB) conduct a scientific peer review of EPA's draft IRIS assessment. In response to the EPA's request, the SAB identified subject matter experts to augment the SAB Chemical Assessment Advisory Committee (CAAC) and assembled the SAB Inorganic Arsenic Review Panel to conduct the peer review. The SAB Inorganic Arsenic Review Panel met virtually on January 5, 2024 to hear a presentation by EPA staff, and then at an in-person meeting on January 24-26, 2024 to deliberate on the Agency's charge questions. Another virtual meeting was held on July 8 and 16, 2024 to discuss the panel's draft report. Oral and written public comments were encouraged and considered throughout the advisory process.

This report is organized by the charge questions raised by the Agency (in italics) and are followed by the consensus response and recommendations. The SAB identified numerous instances in which the analyses and conclusions in EPA's draft IRIS assessment could be revised to be more thorough and transparent. Additional information is presented in Appendices at the end of the report. The SAB includes key recommendations that are necessary to improve the critical scientific concepts, issues, and/or narrative within the EPA's draft IRIS assessment. Tier 1 recommendations are essential for improving the transparency of the EPA's conclusions and to bolster the supporting evidence for them. Tier 2 recommendations are included for EPA to consider as they revise their assessment, and Tier 3 recommendations represent suggestions to inform future reviews or research efforts.

A list of acronyms and abbreviations can be found at the front of this report to assist in orienting the reader to the terminology used throughout the SAB's responses to the charge questions. Comments that are primarily editorial in nature are presented in Appendix A. Appendix B lists Tier 2 recommendations that are important for clarity and which can be readily addressed. All materials and comments related to this report are available at: https://sab.epa.gov/ords/sab/r/sab_apex/sab/advisoryactivitydetail?p18_id=2631&clear=18&session=21631571003591

2. RESPONSE TO CHARGE QUESTIONS

2.1. Charge Question 1. Systematic Review Methods and Documentation

The IRIS Toxicological Review of Inorganic Arsenic describes and applies a systematic review protocol for identifying and screening pertinent studies. The protocol is described in brief detail in Section 1.5.1 (Literature Searching and Screening) and in full detail in Appendix A (Updated Problem Formulation and Protocol for Inorganic Arsenic IRIS Assessment). If applicable, please identify additional peer-reviewed studies of inorganic arsenic that the assessment should incorporate[1].

The IRIS Toxicological Review of Inorganic Arsenic has been highly responsive to previous recommendations (e.g., NRC, 2013; NASEM, 2019), focusing on epidemiological studies, databases that are appropriate, and good description of the population, exposures, comparators, and outcomes (PECO). The focus on the chronic exposure and the inclusion of both studies with environmental samples and biomarkers of arsenic exposure is appropriate.

Given the magnitude of the iAs literature, the focus on priority outcomes of bladder cancer, lung cancer, cardiovascular outcomes, diabetes, pregnancy and birth outcomes, and neurodevelopmental effects seems reasonable although some outcomes were not pursued regarding the 2013 and 2019 recommendations (e.g., skin lesions, skin cancer, renal and liver cancers, nonmalignant respiratory disease). Comments on the selection of outcomes are provided under charge question #4. A suggestion to update the search could be to look for the 'cited by' references for each screened study to see if any further articles are identified that were not retrieved by systematically querying the health literature databases.

As indicated in charge question #2, for cancers, it was assumed that causal relationships exist, and the focus was to "update the quantitative risk estimates" (pages ES-1, line 27). This seems reasonable to the committee and no changes are recommended. The focus on medium and high-confidence studies is methodologically appropriate to ensure the most valid pooled-estimates for population inference.

A common thread to articles that did not come up in the searches is that they often evaluated iAs in the context of an exposure mixture. While mixture methods may have less statistical power than traditional, single-exposure models, they offer insights into the health impacts associated with a person's exposure experience. Specific studies that may have been missed are listed with the outcomes discussed under charge question #3.

The inclusion of articles describing PBPK models is appropriate although the terminology used is confusing as probably it is unnecessary to indicate whether the models meet the PECO criteria. It

[1]Newly identified studies (i.e., studies identified by EPA or the public that meet PECO criteria but were not addressed in the external review draft, for example due to recent publication) will be characterized by EPA and presented to the peer review panel. This characterization will focus on EPA's judgment of whether the studies would have a material impact on the conclusions (i.e., identified hazards or toxicity values) in the external review draft. The peer review panel is asked to review EPA's characterization and provide tiered recommendations to EPA regarding which studies, if any, to incorporate into the assessment before finalizing as well as their expected impact.

would be preferable to indicate that this was an additional specific search category for PBPK studies, and that these were not required to have outcomes and some of the other PECO requirements.

The inclusion of the supplemental material is very useful for the current IRIS as well as for future efforts. The effort that the EPA has expended to summarize such a large body of literature has to be commended. This effort provides an important compilation of iAs and health effect literature. It is possible that for some outcomes for which the amount of evidence is currently insufficient and could not be incorporated into the meta-analysis, that the information will become available in the future.

The process of the systematic review, including the use of two reviewers, is reported in a thoughtful manner. SAB members had different viewpoints on the need to update the search to include papers published since 2019. Most agreed with describing the updated systematic review conducted in 2022, and changing the information provided if the level of evidence was only moderate or new papers published were influential. Other members indicated that for transparency and completeness, the EPA should consider incorporating the newer studies if they are suitable for dose-response modeling, since the influence of a study may not be fully understood until the modeling is done.

An impressive number of relevant studies have been identified. For ischemic heart disease outcomes some members recommended the study by Monrad et al. (2017) as a potentially relevant study to be included in the dose-response meta-analysis, in particular the sub-study conducted in the region of Aarhus, Denmark, where there was more variability in iAs levels in drinking water. Other members disagreed with this inclusion due to substantial measurement error in the assignment of those water estimates to the study participants and the limited variability as the exposures are quite low. Potentially a sensitivity analysis could be recommended as a compromise.

The Table in Appendix A is detailed and complete and straightforward to review.

Additional studies of interest are highlighted throughout the review document.

Tier 3:

- Literature searching could be augmented by looking for the 'cited by' references for each screened study to see if any further articles are identified that were not retrieved by systematically querying the health literature databases.
- Consider conducting a sensitivity analysis including the Aarhus, Denmark sub-study data.

2.2. Charge Question 2. Systematic Review Documentation

As recommended in the 2019 NASEM review of the Inorganic Arsenic protocol, bladder cancer and lung cancer were accepted as hazards and only considered for the ability to update dose-response analyses. Similarly, the following health outcomes were included for evaluation of both hazard and, as appropriate, dose-response analyses: diseases of the circulatory system, diabetes, pregnancy outcomes, and developmental neurotoxicity. For these latter health effects, the Toxicological Review provides an overview of individual study evaluations, and the results of those evaluations are made available in the Health Assessment Workplace Collaborative (HAWC) (linked here HAWC). Note that a

"HAWC FAQ for assessment readers" document, linked here (scroll to the bottom of the page, and the document is available for download under "attachments"), is intended to help the reviewer navigate this on-line resource.

a. *Please comment on whether the study confidence conclusions for the Inorganic Arsenic studies are scientifically justified and clearly described, considering the important methodological features of the assessed outcomes. Please indicate any study confidence conclusions that are not justified and explain any alternative study evaluation decisions.*

EPA has produced an excellent IRIS review of iAs toxicity by providing confidence that the studies are scientifically justified and presented clearly. The confidence conclusions are justified and strengthened by using the wealth of human epidemiological data in the conclusions of the hazard assessment for the cancer and noncancer disease endpoints. The epidemiological database includes studies with a broad range of exposures supporting hazard and dose-response analyses. Thus, mode of action (MOA) considerations are limited to establishing the biological plausibility for the health outcomes of interest. The systematic review approach and the methods used by EPA for including/excluding studies in the draft IRIS for iAs are described in detail and presented in a consistent and logical manner across the four health outcomes listed in the charge question. Overall, the systematic approach used by EPA for including/excluding studies for analyses had several strengths compared to those used in past IRIS documents (e.g., Fig 1-1) and adhered to more recent NASEM recommendations (NASEM, 2019). The approach relied on keyword and objective delineators of study design and their characteristics when searching the HAWC, PubMed, and Web of Science databases. The HAWC system provides transparency on the selected articles and categories to which they were assigned. The individual study evaluation ratings and the rationale for the rating can be easily viewed for the studies that met the criteria for medium or high confidence. The criteria used to assign risk of bias evaluations and confidence ratings are presented clearly. Automated machine-learning algorithms were used to prioritize relevant study characteristics and quality assessed based on PECO criteria. There is sufficient information about the processes and decision-making points used by EPA to assign study confidence ratings for each health outcome. The quality and selection criteria were consistent across the health outcomes and within a health outcome. Overall, the confidence conclusions for the iAs studies are scientifically justified and presented clearly.

While the report has many strengths, there was concern that the search strategies are not intuitive, and it was difficult finding the low confidence or uninformative studies which impacted the SAB's ability to review rationale for placement of the studies in these categories. There is an incredible wealth of studies in HAWC and there was an insufficient amount time to review many of the individual studies to appreciate whether the EPA approach and protocol was appropriate. The SAB found it difficult to review the numerous papers that were excluded. There is a suggestion to simplify the search process and improve the accessibility of the studies.

While the study confidence rankings are justified and reasonably described overall, the determinations for risk of bias, which in turn could impact the confidence ratings, need to be more carefully assessed. There was concern that best practices for sample and data collection and controlled analyses as a cause for variation was not critically evaluated when comparing across studies. EPA should clarify that individual study evaluations were not outcome-specific and were not always sufficiently refined. As one example of how this is important, a study may not have adjusted for a covariate or considered it as

a potential confounder because it is in the causal pathway. Another example is that studies of urinary arsenic should not be downgraded for exposure assessment characterization if they did not adjust for seafood intake but did speciate arsenic and exclude arsenobetaine or if they were conducted in a population where seafood consumption is rare (in this case total urinary arsenic in the absence of speciation is an adequate biomarker).

EPA should better describe how a study can have low confidence in the source of the modifiers and covariates known to have impact on the outcome, or - in several cases - not assess covariates known to be associated with the outcome - and still receive high and medium confidence, particularly as it relates to utilizing these studies in the dose-response analyses. The EPA could consider using Sohel et al. (2009) as an example, as this study was used for dose-response modeling of the cardiovascular endpoint but did not assess for smoking (or other known risk factors for cardiovascular disease). It may be the case that the EPA has already considered these aspects, but it is not clearly reported in the assessment (or HAWC) and may provide an opportunity to increase transparency by giving an example of how residual and/or uncontrolled bias has been addressed.

An important issue raised in evaluating the evidence is whether the EPA could include discussion of MOA and mechanistic studies to improve confidence of the plausibility of the hazard identification and identification of points of departure at low levels of iAs exposure. The exclusive use of human data to establish hazard and dose-response relationships is sufficient to support the conclusions and is desired to reduce uncertainty in the outcome of the analyses. *In vitro* and *in vivo* animal and human mechanistic data could be used to demonstrate the MOA and pathogenic mechanisms for iAs-promoted cardiovascular disease. However, the animal data should not be combined with the human data in the dose-response analyses as it would unnecessarily raise uncertainty in the RfD due to requiring the factoring of cross-species uncertainty.

Recommendations:

Tier 1:

- In the HAWC, the individual study evaluation ratings and the rationale for the rating can be easily viewed for the studies that met the criteria for medium or high confidence. The committee recommends having similar tables for the studies that were considered but did not meet the criteria, i.e., low confidence or uninformative, for assessing the study confidence conclusions (i.e., in HAWC and as supplemental materials/appendices). It is also recommended that the literature trees be expanded such that navigating to the low and uninformative studies is parallel to that of the medium and high studies.
- While the justification of study confidence rankings is reasonably described overall, the determinations for risk of bias, which in turn could impact the confidence ratings, need to be carefully assessed and transparently and consistently reported. It is recommended that the EPA clarify that individual study evaluations (appraisals) were not outcome-specific and were not sufficiently refined to the topic. It is also recommended that EPA describe how a study can have low confidence in the source of the modifiers and covariates and still receive high and medium confidence for the dose-response analyses. P. 1-9 suggests that some of the new data did not undergo study evaluation but was only considered for dose-response criteria. The EPA should

clarify these statements, the workflow, and the results. Further comments on evaluation ratings are provided under charge question #3 for specific outcomes.

- Data from mice and cell-based studies show sensitivity to the pathogenic effects of iAs on the vasculature and cardiac tissues within the lower range of human exposures. The panel recommends that the EPA include mechanistic studies to increase evidence for the plausibility of the outcomes for the human dose-response relationship and RfD determination for cardiovascular disease.

  The statement that animal models, especially rodent models, are less sensitive to adverse effects is not true for adverse cardiovascular effects. The SAB suggests a review of the literature where mouse models respond with the preclinical signs of human disease within the low range of human exposure. Examples should include dose-response studies in the mouse model of atherosclerosis that demonstrate a threshold for chronic dosing near 10 µg/L (Makhani, 2018). Vascular remodeling and endothelial dysfunction are observed with thresholds between 1 and 10 µg/L (Soucy, 2003; Straub, 2008). The likely mechanism is reactive oxygen species generation as a result of iAs chronically activating endothelial and cardiac cell NADPH oxidases and decreasing nitric oxide bioavailability (Straub 2008). Generation of NADPH derived superoxide and hydrogen peroxide deprives the vasculature of nitric oxide needed to dilate vessels, suppress inflammation, and reduce the inflammatory cell infiltration that is the mechanism for atherogenesis. At human relevant chronic dosing, iAs also alters arginine metabolism to reduce nitric oxide causing a pathogenic plasma asymmetric dimethylarginine (ADMA) profile (Dheer et al., 2015), a known risk factor for cardiac disease in humans (Osorio-Yanez et al., 2013). Other subchronic and chronic rodent studies with exposures less than 100 µg/L find pathogenic structural cardiac remodeling and gene expression (Soucy 2005, Hays 2008). Thus, as suggested in NRC 2013, animal data could be included in the hazard and dose-response assessment to provide enhanced evidence for plausibility and MOA information that would strengthen the certainty of the RfD for cardiovascular disease. However, the SAB agrees that animal data should not be used in determining human dose-response relationships given the vast amount of relevant epidemiologic data.

Tier 2:

- ES.2 line 27-29: EPA should clarify whether or why regression analysis did not include dose-response data above 100 µg/L when well water in the U.S. ranges to more than 1000 µg/L.
- The EPA should eliminate all references to discredited dogma that animal and cell-based models are insensitive to pathogenesis and disease promotion. This is not true for iAs promoted cardiovascular disease and likely cancer promotion in studies that adhere to human relevant dosing.
- EPA should improve the quality and increase the resolution of the Figures to improve legibility and clarity.

b. *Results from individual inorganic arsenic studies are presented and synthesized in the health outcomes sections. Please comment on whether the presentation and analysis of study results are clear, appropriate, and effective to allow for scientifically supported syntheses of the findings across sets of studies*.

The SAB agrees that the overall presentation of the results, including the figures and tables, is laid out reasonably. However, it is not easy to follow the sequence of results from individual studies from one section to another and the corresponding results in the figures and text.

There are ways they could be more effectively organized and described.

Specifically, the EPA could present the logic and sequence of individual studies in the text, tables, and figures. For example, there are studies described in the text that do not appear in the evaluation tables; but the reasons for this are not always evident. These could be studies deemed low confidence or uninformative but there is not an easy way to determine this. Further, there is a lack of concordance between studies presented in the evaluation tables and those appearing in the forest plots. A statement (and footnotes) about those with medium or high confidence that were considered but not included in the forest plot should be made explicit and vice versa.

In general, the titling and labeling of tables and figures needs to be clarified and elaborated in several places. On the forest plot Figure 3-5 and similar figures, it is not always clear whether the outcome is incidence or mortality. This needs to be specified, and outcomes should be grouped or provided in separate forest plots (one for incident IHD and another for fatal IHD). Further, Figure 4-1 is designed to show the number of studies or datasets for each of the outcomes. This figure should also show the categories of CVD as defined above. Additionally, there is no justification provided for the suggestion of a cut point of $\leq$25 or >25 datasets. The SAB recommends removing this and placing the number of available studies under the outcome on the x-axis (i.e., N=25, for outcomes with 25 datasets). The SAB also recommends denoting the studies that were selected for dose-response analyses. The lifetable analyses tables and figures were the most confusing. We recommend the table headings (e.g., for Table 3-4) to include what the values for the µg/kg-day categories represent, e.g., 0.12 for background diet, that 1.12 relates to 100 µg /L. The CSF equation on footnote 26 on page 4-21 should be more prominent and accessible in the report.

Additional detail is needed on specific aspects of results. For instance, for studies using urinary arsenic, it is important to indicate whether the investigators speciated arsenic and excluded arsenobetaine, and if they did not do so, whether the study was conducted in non-seafood eating populations and how seafood intake was controlled in the analyses. Studies of pregnancy and early childhood are appropriately covered in text under a separate section (e.g., on page 3-30); however, these do not consistently appear in the forest plots. The SAB recommends that they be included where appropriate. Also, the EPA should review the assignment of the study design for cross-sectional studies. It is not consistently clear from the figures why they are cross-sectional. For instance, do they all look at prevalence rather than mortality or incidence? This needs to be articulated.

The report needs to clearly distinguish between studies from highly exposed populations and lower exposures. This distinction is inconsistent in the report. For instance, studies from highly exposed populations are covered in a separate section sometimes, while other times they are not. The SAB recommends that the data and results be presented separately, where possible, although this is not necessary for the dose-response analysis. Life stages can be viewed as susceptible windows of exposure (i.e., as effect modifiers) as indicated in the text. They can also be classified with respect to their timing of the occurrence of the outcome as during pregnancy (e.g., gestational hypertension or diabetes), or childhood (e.g., childhood blood pressure or other CVD-related intermediary outcomes

measured during childhood) (e.g., page 3-41 lines 24-27). These two distinct aspects of life stage need to be encompassed in the presentation of the results.

The sections on modes or mechanisms of action should be discussed in more detail. Mechanisms that contribute to observed differences are only superficially addressed and more likely reflect biomarkers of later disease processes once they have already been diagnosed rather than the underlying causes of disease initiation and progression.

Although there was clarity and consistency for studies meeting high or medium confidence criteria, it was difficult to delve into the numerous papers that were excluded given the time allotted. EPA should provide more details about the specific metrics used to discount studies from consideration.

Biomarkers of effect are potentially consequences of the disease process rather than causal biomolecules. On the other hand, urine biomarkers of exposure reflect the exposure assessment derived from measurements of water concentration, and other sources.

As indicated under Question 2b, EPA should address possible differences among studies in quality control measures used at collection sources, and non-standardization of sample collection pipelines, including processing and storage conditions and storage times prior to analytical assessment of iAS. For example, changes in sample volumes during different storage conditions could appreciably affect iAs determinations in the low to mid ppb range. Contamination of samples with iAs is generally not a major concern compared to other metals (e.g., lead or zinc).

The four non-cancer health outcomes listed in the charge question are considered for risk and dose-response modeling independently of one another. EPA should comment on the interplay across individual health outcomes as they relate to risk modification. Additionally, diabetes may mediate the relationship between iAs exposure and DCS. EPA should comment whether participants enrolled in iAs studies were assessed for more than one health outcome and if this would have any bearing on the interpretations and conclusions reached.

Neurodevelopmental impairments with low-level iAs exposure not only impairs early cognitive function in mice but also significantly shortens the trajectory for progressing neurodegenerative disorders in a genetically susceptible model relevant to diseases of aging shown in recent animal studies (e.g., Nino et al 2019) and in humans (e.g. Jiang et al 2023). Assessment of iAs exposures and neurodegenerative health outcomes is missing from the IRIS assessment.

Recommendations:

Tier 1:

- The logic and sequence of inclusion of individual studies in the text, tables, and figures should be included.
- The report needs to clearly distinguish between studies from highly exposed populations and lower exposures.
- Figure 4-1 shows the number of datasets for each of the outcomes. Instead of the apparent use of a cut point of <25 or >25 datasets, the SAB recommends placing the number of available

studies under the outcome on the x-axis. The SAB also recommends denoting the studies that were selected for dose-response analyses.
- The report needs clarification on the key confounders, covariates, and risk modifiers and how these were incorporated into the assessment and synthesis.

Tier 2:
- Include assessment of iAs exposures and neurodegenerative health outcomes in the IRIS assessment.
- Additional detail is needed on specific aspects of results. For studies using urinary arsenic, EPA should indicate whether or not arsenic was speciated. If arsenic was not speciated or does not include measurement of arsenobetaine, the use of total urinary arsenic is an adequate biomarker of iAs exposure if seafood intake is rare in the population.
- Studies of pregnancy and early childhood should be included in the forest plots.
- The EPA should review the assignment of the study design for cross-sectional studies.
- Although there was clarity and consistency for studies meeting high or medium confidence criteria, EPA should provide more details about the specific metrics used to discount studies from consideration.
- For the four non-cancer health outcomes, EPA should comment on the interplay across individual health outcome as it relates to risk modification.

2.3. Charge Question 3. Noncancer Hazard Identification

For each health effect prioritized for hazard identification in the assessment based on the protocol for inorganic arsenic and outlined below, please comment on whether the available epidemiological data (the primary focus of these analyses based on recommendations from the NASEM) have been clearly and appropriately synthesized to describe the strengths and limitations. Please also comment on whether the weight-of-evidence decisions for hazard identification are scientifically justified and clearly described, and appropriately consider health effects in susceptible subpopulations or lifestages (e.g., children) to the extent possible, given the available data.

a. *For diseases of the circulatory system, the Toxicological Review concludes the currently available **evidence demonstrates** that inorganic arsenic causes cardiovascular effects in humans given sufficient exposure conditions. This conclusion is based on studies of humans that assessed exposure levels of <10 µg/L to 930 µg/L showing increased ischemic heart disease and hypertension, as well as related cardiovascular disease endpoints of atherosclerosis and repolarization abnormalities (e.g., QT prolongation).*

A major effort has been expended to synthesize the available epidemiological evidence for iAs and diseases of the circulatory system. Overall, the effort is comprehensive and detailed, allowing one to obtain a complete picture that includes both clinical and subclinical outcomes. The process of how the studies were identified, evaluated, and how they were considered depending on the level of confidence is clearly explained. Given the number of studies identified for the hazard identification phase already in the 2019 search, it makes sense that the additional studies identified in the 2022 search were only included for the dose-response meta-analysis and not additionally included for the hazard identification phase, since the inclusion of additional studies was not going to be influential.

Overall, the weight-of-evidence decisions for hazard identification are scientifically justified, clearly described, and appropriately consider health effects in susceptible subpopulations (e.g., effect modification by sex for QT interval prolongation and ischemic heart disease; by methylation capacity for atherosclerosis).

The most significant concern regarding the evidence for the cardiovascular outcomes, is that while in the screening and descriptive process it is appropriate to include the different clinical and subclinical outcomes, for the meta-analysis the SAB recommends that IRIS focuses on ischemic heart disease as the key endpoint. There are four reasons for this: 1) IHD (also called coronary heart disease in some studies) is a more specific cardiovascular outcome, characterized by atherosclerosis as the main underlying pathophysiological process. 2) There are substantial epidemiological and experimental animal models providing support that iAs exposure enhances atherosclerosis development – this evidence could also be summarized as it would strengthen the presentation. 3) CVD is an umbrella term that includes many different outcomes, such as stroke and other cardiovascular health endpoints beyond IHD (e.g., myocarditis, pericarditis, heart failure, etc.). Although IHD probably represents the largest type of CVD endpoint in most populations, the specificity is diminished and thus CVD is an inferior endpoint compared to IHD. The 2013 and 2019 NASEM committees recommended that the EPA focuses on IHD as the key cardiovascular endpoint. 4) In addition to the comments above, for CVD incidence, there are only two studies meeting the criteria for inclusion in the meta-regression (dose-response meta-analysis). For these reasons, the use of studies that evaluate the association between iAs exposure and IHD is strongly recommended.

There are several types of studies looking at the association of iAs exposure and IHD. Some studies have reported both fatal and non-fatal events together (incident IHD). Other studies have also or instead reported the association with IHD mortality alone. Including studies that report both fatal and non-fatal events (incident IHD) has the advantage of including aspects of the disease that result in a health burden and impact to society that is larger compared to mortality alone. However, focusing only on fatal events has the advantage of selecting an outcome that is severe and more easily quantified in the population than clinical non-fatal events, and for which research studies are easier to conduct as the identification of non-fatal events requires more complex research surveillance system. Selecting either IHD mortality (fatal events only) or IHD incidence (both fatal and non-fatal events) would be appropriate for the dose-response analysis.

Figure 3-3 is clear and helpful, although the category of IHD is not clearly identified. It is not clear if it is included within the term “atherosclerosis”, which is a broader term and can include IHD but also other outcomes such as peripheral artery disease, carotid plaque, carotid intima media thickness, or coronary artery calcification. This might be related to the “additional sub-tagging” but because CHD/IHD is an important outcome, it deserves its own circle in the figure.

For pages 3 through 6, the SAB recommends combining the terms “ischemic heart disease” and “coronary heart disease” under the same category and use only one of the terms. Both terms refer to the same condition, but different authors and journals generally use one term or another.

Figure 3-5, panels a, b, c is a helpful figure with multiple outcomes and useful information. Several enhancements could further increase the utility of this figure including specifying the range of the X values and better organization of the studies. Currently it is not clear if it is by outcome, levels of

exposure, year of publication, alphabetical by author, or something else. For Figure b, the effect estimate is not clear as it says it is a "regression coefficient-drinking water-continuous exposure" but given the values, it seems more likely that this is an odds ratio, relative risk, or hazard ratio. Please clarify all the effect estimates.

Several comments on inclusion of other streams of evidence are warranted. The cross-sectional and ecological studies can be considered of less value in general, although for completeness of the synthesis of the evidence it is appropriate to have them. The inclusion of the natural experiment is relevant. As iAs levels decline in more communities, including the US, research leveraging these natural experiments will provide very important evidence to evaluate the impact of the interventions and provide further evidence for the risk assessment of iAs. The inclusion of the intermediate endpoints and risk factors provides additional support for the conclusions for the clinical outcomes. On page 3-18, the sentence about carotid intima-media thickness (cIMT) that reads, "Coronary atherosclerosis is typically clinically assessed using ultrasonography to measure cIMT where …" is not correct as this likely refers to carotid atherosclerosis. It is hard to assess atherosclerosis of the coronary arteries with ultrasound - normally CT scans and the assessment of calcification is needed.

For Figure 3-13-a, the SAB recommends removing the study by Li et al. for hypertension for iAs metabolism (iAs%, MMA% and DMA%). The SAB notes that the interpretation of iAs metabolism is very different from iAs exposure. iAs metabolism seems to be a phenotype distinct from exposure and related to genetic, nutritional, and other factors that are not necessarily linear with exposure. The SAB liked the consideration of methylation capacity as described in section 3-40, however did not believe it should be included in Figure 3-13-a and as part of the association with hypertension.

The conclusion that the evidence demonstrates iAs causes cardiovascular effects in humans is robust based on the large body of evidence synthesized as part of the IRIS assessment.

Summary table 3-2 for cardiovascular / ischemic heart disease incident and mortality and cerebrovascular disease, hypertension and stroke is excellent.

Recommendations:

Tier 1:

- The assessment should focus on ischemic heart disease as the key endpoint. EPA should conduct analyses of both IHD mortality (fatal events only) and IHD incidence (both fatal and non-fatal events) for the dose-response analysis.

Tier 2:

- Remove the study by Li et al. from Figure 3-13-a.

b. *For diabetes, the Toxicological Review concludes the currently available **evidence demonstrates** that inorganic arsenic causes diabetes in humans given sufficient exposure conditions. This conclusion is based on studies of humans that assessed exposure levels of <150 µg/L to >150 µg/L showing increased incidence of diabetes mellitus (Type 1 and Type 2 diabetes).*

The available epidemiological data have been clearly and appropriately synthesized to describe the strengths of the data available for the association between iAs exposure and diabetes. Overall, the weight-of-evidence decision indicating that the *evidence demonstrates* that iAs causes diabetes is well supported by the epidemiological and animal literature, although some studies showed a null association. The weight-of-evidence decisions for hazard identification are scientifically justified and clearly described and appropriately consider health effects in susceptible subpopulations (e.g., by *AS3MT* genotype and vitamin B intake for HOMA2-IR; and folate intake as a modifier of iAs metabolism associations with type 1 diabetes).

Some specific comments follow:

In Figure 3-24, the EPA should ensure that all the relevant studies are included. For instance, the paper by Kuo et al. (PMID: 25583752) on urinary arsenic (sum of inorganic and methylated arsenic species) and incident diabetes should be included. The reason for exclusion of this study is unclear as the design is very similar to Grau-Perez 2017. The association in the study by Kuo et al. is largely null. From a reporting perspective, in HAWC, Kuo et al. is tagged both as a medium/high confidence study and as a case-control/cohort study (the subject of the figure), and thus the SAB was unable to determine the reason for its exclusion. For instance, on page 3-59 it is indicated "as well as lower-exposed areas, including the Northern Plains (Kuo et al., 2015) in the U.S., with null results observed in Utah, U.S. (Lewis et al., 1999)." However, Kuo et al. reported findings overall for the Northern and Southern Plains and also the Southwest and the association with the sum of inorganic and methylated arsenic species, a biomarker of iAs exposure was null in that study. This should be corrected, and the study added to the figure. The SAB also notes a 2024 publication that studied the association of iAs in drinking water with incident diabetes in two U.S. populations (Spaur et al. Diabetes Care 2024).

Further, there were 24 studies tagged to this category in HAWC, but fewer than 20 studies are shown in Figure 3-24. Additional narrative may be helpful in describing the correspondence between HAWC and assessment reporting (both here and for other outcomes). Per previous comments, the literature trees could also be expanded using existing HAWC tags to reflect each step in the process, and the connection to figures/tables throughout the assessment. The SAB recommends careful description of the selection and reporting of the studies on iAs and diabetes.

Overall, the evidence is robust for a causal relationship between iAs and Type 2 diabetes, albeit with some null associations. For Type 1 diabetes the evidence is primarily for iAs metabolism, not for iAs exposure and there are very few studies. It seems that there is no mention of Type 1 diabetes in the conclusions for this outcome, which is appropriate. It is not clear why Type 1 diabetes was included in the charge question. Type 1 diabetes should be excluded as the evidence is very limited. Section 3.2.2 should emphasize Type 2 diabetes results and note where possible the age of the study population and/or age at diagnosis. This is important since Type 2 diabetes is chosen for meta-regression analysis, although the report does not state this until p. 4-13. It would increase transparency to address the differentiations between Type 1 and Type 2 diabetes as part of a more granular synthesis in the evidence judgement sections to better align with the methodologies from GRADE and Bradford as described in the protocol (see previous comment on improving the transparency of reporting around

causality). That is, to expand the evidence stream summary table (table 3-4) to include all aspects that contribute to the evidence determination that iAs "causes" Type 2 diabetes.

It is also recommended that the Agency improve the transparency of reporting as related to low and uninformative studies (see recommendations for the previous charge question).

Recommendations:

Tier 1:

- Prioritize the analysis on Type 2 over Type 1 diabetes.

Tier 2:

- Include all the relevant studies in Figure 3-24.
- Be more descriptive in the selection and reporting of the studies on iAs and diabetes.
- Expand the evidence stream summary table (Table 3-4) to include all aspects that contribute to the evidence determination that iAs "causes" Type 2 diabetes.

c. *For pregnancy and birth outcomes, the Toxicological Review concludes the currently available **evidence indicates** that inorganic arsenic likely causes pregnancy and birth effects in humans given sufficient exposure conditions. This moderate epidemiology evidence generally supports a weaker hazard judgment, although the specific judgment reached is more heavily influenced by other lines of evidence than when there is robust epidemiological evidence. Although there is notable uncertainty in this judgment without reviewing the other lines of evidence (out of scope for this assessment), it is reasonable to judge that the available evidence indicates that pregnancy and birth effects are likely caused by iAs exposure, given sufficient exposure conditions. This conclusion is based on studies of humans that assessed exposure levels of <100 µg/L to >100 µg/L showing decreased fetal and post-natal growth or length of gestation.*

EPA has clearly and appropriately reviewed the iAs exposure and pregnancy and birth outcomes following the protocol of identifying relevant studies and updated the literature to 2022. EPA examined the quality of the epidemiologic studies by assessing exposure and outcome assessments, biases, and interpretation. The weight-of-evidence is justified, and the interpretation is generally reasonable.

The literature on pregnancy and birth outcomes continues to emerge, especially from studies in the US. Therefore, in evaluating whether iAs likely causes these outcomes, it makes sense to turn to the literature on more highly exposed populations along with that of exposures more typical of the US. This is especially the case for outcomes that are rare in the US, i.e., neonatal deaths. While the conclusions seem reasonable that associations for fetal losses, stillbirth, and neonatal deaths were stronger in more highly exposed populations, it would be helpful to clearly denote these studies in the figures and text. The distinction between highly exposed populations and other populations may help clarify heterogeneity of results for other outcomes as well. The iAs exposure range may be a critical value in determining the specific adverse health effects observed during pregnancy and infancy. The EPA may need to provide the evidence of associations by stratification of median exposure levels, e.g., >200, 150-199, 100-149, 50-99, 10-49, <10 µg/dL in drinking water.

This section focuses on fetal and infant mortality, birthweight, fetal growth, prematurity, and postnatal growth. It does not cover maternal pregnancy outcomes such as gestational diabetes or hypertension. Thus, EPA may consider renaming the section to reflect the content, i.e., fetal, newborn, and infant health outcomes. The EPA should consider placing studies on preterm birth before birthweight since preterm birth is also a cause of infant mortality and the issue of gestational length would more logically be addressed before the discussion of birth weight. This would place the findings on fetal growth next to those of postnatal growth which would flow logically. In the "Evidence Judgement" summaries, the individual measures can be discussed separately, e.g., head circumference, birth length, etc. The outcomes in this section can be better defined and the methods used to assess outcomes are needed. For instance, the definition of fetal growth should be provided in the first paragraph of this section and the approach used to determine these outcomes (e.g., self-report vs medical records) should be specified (page 3-81 beginning line 19). It also is necessary to indicate the timing of the event, e.g., the week of gestation used to define fetal losses and prematurity. Gestational length (or gestational age) is a determinant of birth weight, i.e., is on the causal pathway. This is mentioned in the context of Kile et al. (2016) (page 3-77, lines 8-9). As adjustment could influence whether an association was observed, some discussion on whether gestational length or gestational age was adjusted for in analyses of low birth weight would be appropriate.

Some studies were incorrectly classified as cross-sectional. For example, on page 3-80, lines 24-28, the study by Luo et al. (2017), measured iAs in whole blood collected in the first trimester and evaluated the association between exposure concentrations and birth weight. In addition to explaining any discordance between studies that appear in the evaluation figures and forest plots, it is not clear why some studies appear for one birth outcome but not another. For instance, Howe et al. (2020) which was deemed of medium or high confidence for gestational age, also analyzed birth weight but does not appear in the figures describing birth weight.

On page 3-81, lines 25-26, using total urinary arsenic might result in uncertainty in interpretation for studies that did not speciate arsenic, particularly in populations eating seafood. Otherwise, the use of urinary arsenic would tend to reduce exposure misclassification. This can be corrected. To strengthen the rationale to use drinking water iAs levels as the exposure indicator during pregnancy, it is better to provide validation data of correlation between drinking water and maternal urinary iAs concentrations. When iAs levels were measured in whole blood samples, it is better to provide the validity of using whole blood as biospecimen, for example, including the intraclass correlation coefficient during pregnancy if available.

On page 3-75, lines 4-7, for the summary of iAs and fetal loss and infant death, while "there is general consistency" is acceptable, there is a need to mention some studies did not find statistically significant positive association between iAs and fetal loss. On page 3-97, lines 14-15, the "direct toxic mechanisms" were mentioned for infant death. EPA needs to add more details to the potential mechanisms leading to infant death from iAs exposure. Diarrhea and lower respiratory tract infections were mentioned in the sentence below for the causes of infant death in developing countries. This part can be further strengthened to understand the role of iAs exposure on fetal loss and infant mortality.

The POD analysis uses low birth weight (fetal growth) and is supported by the data. Data appear to be mixed for support of preterm birth. A recent publication from the ECHO consortium did not identify

inverse associations between area-level iAs exposure and birth weight (Lewis et al., 2023). EPA may wish to cite this large U.S.-based study into the inference on iAs exposure and birth outcomes while pointing out that the study is based on water utility data and may have exposure misclassification. On page 3-81, lines 15-18, for the summary of birth weight findings from iAs exposure, it is better for EPA to mention some studies did not find the association as the evidence was not synthesized using meta-analysis or pooled analysis. The exposure dose range is probably the reason for some of the discrepancies between studies.

Additional articles that should be considered mostly include recent studies on metalloid/metal mixtures:

- Hoover JH, Coker ES, Erdei E, Luo L, Begay D, MacKenzie D; NBCS Study Team; Lewis J. Preterm Birth and Metal Mixture Exposure among Pregnant Women from the Navajo Birth Cohort Study. Environmental Health Perspectives 2023 Dec;131(12):127014. doi: 10.1289/EHP10361. Epub 2023 Dec 18. PMID: 38109118; PMCID: PMC10727039.
- Huang H, Wei L, Chen X, Zhang R, Su L, Rahman M, Golam Mostofa M, Qamruzzaman Q, Zhao Y, Yu H, Wei Y, Christiani DC, Chen F. Cord serum elementomics profiling of 56 elements depicts risk of preterm birth: Evidence from a prospective birth cohort in rural Bangladesh. Environ Int. 2021 Nov;156:106731. doi: 10.1016/j.envint.2021.106731. Epub 2021 Jun 28. PMID: 34197971.
- Howe CG, Claus Henn B, Farzan SF, Habre R, Eckel SP, Grubbs BH, Chavez TA, Faham D, Al-Marayati L, Lerner D, Quimby A, Twogood S, Richards MJ, Meeker JD, Bastain TM, Breton CV. Prenatal metal mixtures and fetal size in mid-pregnancy in the MADRES study. Environ Res. 2021 May;196:110388. doi: 10.1016/j.envres.2020.110388. Epub 2020 Oct 28. PMID: 33129852; PMCID: PMC8079562.

In the discussion of risk modifiers (pages 3-97 and 3-98), the focus is on sex. There are other reported results on risk modifiers that deserve mention. For example, Gilbert-Diamond et al. (2016) reported on pre-pregnancy BMI as another potential modifier. Other potential modifiers reported in the literature include undernutrition.

Recommendations:

Tier 1:

- The outcome definition and assessment method need to be provided for fetal, newborn, and infant health outcomes. The section can be renamed to fetal, newborn, and infant health outcomes to indicate the focus on fetal and infant, but not maternal, health outcomes. The discussion of preterm birth outcomes can be placed before birth weight.
- The context of high or low exposure scenarios of iAs for the fetal, newborn, and infant health outcome needs to be clearly provided. The EPA should stratify the exposures by categories to indicate high, moderate, and low exposure scenarios for various perinatal health outcomes.

Tier 2:

- Consider gestational length or gestational age as on the causal pathway to birth weight in the evaluation of studies of birth weight.
- Distinguish cross-sectional from prospective studies consistently in the figures and text.
- Correct discrepancies in the inclusion of studies in the forest plots across outcomes.

- Provide a balanced statement for the use of urinary iAs as the exposure biomarker, as the urinary biomarker may reduce misclassification if speciation of arsenic was used and arsenobetaine was excluded from the data analysis.
- Acknowledge some studies that did not find significant associations with adverse fetal, newborn, and infant health outcomes when summarizing the studies.
- Figure 3-31(b) on page 3-78 shows a relatively large regression coefficient and confidence intervals for the inverse association for the Kile 2016 study that are distinct from other studies. EPA should provide context for this finding.
- Include risk (effect) modifier variables, particularly pre-pregnancy BMI and nutritional status, in the reporting of the associations between iAs exposure and infant health outcomes.

d. *For neurodevelopmental effects, the Toxicological Review concludes the currently available **evidence indicates** that inorganic arsenic likely causes neurodevelopmental effects in humans given sufficient exposure conditions. This moderate epidemiology evidence generally supports a weaker hazard judgment, although the specific judgment reached is more heavily influenced by other lines of evidence than when there is robust epidemiological evidence. Although there is notable uncertainty in this judgment without reviewing the other lines of evidence (out of scope for this assessment), it is reasonable to judge that the available evidence indicates that neurodevelopmental effects are likely caused by iAs exposure, given sufficient exposure conditions. This conclusion is based on studies of humans that assessed exposure levels of <100 µg/L showing cognitive and behavioral deficits in children and adolescents.*

This section concludes that evidence indicates that iAs likely causes neurodevelopmental effects in humans. The available epidemiological data have been synthesized to describe the strengths of the data available for the association between iAs exposure and these effects. The EPA excluded several important prospective cohort studies on neurodevelopmental outcomes with prenatal and early childhood exposure assessments of iAs.

Of 72 assessed studies, 52 were deemed medium-high confidence (17 were excluded and 3 studies identified in the 2022 search update were not considered). Assessments were categorized into cognitive (short- and long-term memory, verbal comprehension, perceptual reasoning, processing speed, executive function, visuospatial function), social/behavioral/emotional, and motor effects. Of the 52 studies, one study of IQ and drinking water iAs in U.S. children was subjected to dose-response analysis (Wasserman et al., 2014) and EPA identified a LOAEL of 0.11 µg/kg/day based on the central tendency estimate of the total dose from diet and drinking water associated with exposure to 5 µg/L in drinking water, from the Wasserman et al. (2014) data (p. C-199, lines 9-12). Because the dose-response was non-monotonic at higher exposures (>20 µg/L), EPA concluded that it was "not advisable" to use the results of this modeling to estimate exposure "across the full range of exposures." EPA concluded "the currently available **evidence indicates** that iAs likely causes neurodevelopmental effects in humans given sufficient exposure conditions. This moderate epidemiology evidence generally supports a weaker hazard judgment, although the specific judgment reached is more heavily influenced by other lines of evidence than when there is robust epidemiological evidence." EPA acknowledged that "there is notable uncertainty in this judgment without reviewing other lines of evidence," which were considered "out of scope for this assessment".

EPA concluded that "it is reasonable to judge that the available evidence indicates that neurodevelopmental effects are likely caused by iAs exposure, given sufficient exposure conditions. This conclusion is based on studies of humans that assessed exposure levels of <100 µg/L showing cognitive and behavioral deficits in children and adolescents."

Recommendations:

Tier 1:

- Overall, the synthesis of the evidence on neurodevelopmental outcomes is well organized and reported. However, while cross-sectional studies of concurrent exposure measurements may provide a reasonable proxy for exposure during an etiologically relevant period, the weight of evidence is not the same as cohort studies to determine the role of prenatal and early postnatal exposure. It is recommended that EPA include more cohort studies with prenatal and early childhood exposure assessments of iAs. Recent prospective studies on neurodevelopmental outcomes include Patti et al. 2022, Wasserman et al. 2022, Dai et al. 2023, Notaro-Barandian et al. 2023, and Butler et al. 2023. The cohort study findings were listed after the cross-sectional studies and should be presented in more detail emphasizing distinct windows of exposure.

Tier 2:

- Evidence Judgement (Page 3-122): While it is true that exposure levels of <100 µg/L primarily showed cognitive effects, the statement did not provide much information about the lower limit of exposure that the association showed. EPA should indicate that the inverse association was observed at a level much lower than 100 µg/L.
- Under the premise that water concentrations are a reasonable estimate of past and current exposure in populations with long-term exposures, it should be noted that in such studies one cannot separately evaluate windows of exposures that may be relevant to neurodevelopmental studies (i.e., pre- vs. post-natal exposure). More specifically, EPA states (P3-103 L3-8) "*Many of the cross-sectional studies evaluated populations that had experienced chronic or lifelong exposure to arsenic, and thus the concurrent exposure measurements are expected to be a reasonable proxy for exposure during an etiologically relevant period."* The reasoning does not accurately reflect critical internal concentrations in the blood and developing brain due to differential distribution and metabolism during the perinatal period. Differential absorption, distribution, metabolism, and excretion and critical windows for neurodevelopmental susceptibility in the perinatal period should be highlighted in more detail.
- It is unclear why estimates for ADHD presented confidence intervals (Figure 3-41, panel c), whereas forest plots appear without confidence intervals for Bayley outcomes of cognitive, language, motor, social-emotional, and adaptive skills (Panel b). If they are not visible or not available this should be indicated by a footnote.
- There are additional studies that could be included as part of the evidence synthesis on neurodevelopmental outcomes. Many of these are newer studies and focus on iAs in the context of other metals, i.e., metal mixtures (Stein 2022; Valeri 2017; de Water 2022; Doherty 2020) and one study examined how the infant gut microbiome alters neurodevelopmental susceptibility to iAs exposure (Laue et al., 2022). Additionally, Soler-Brasco (2023) found that creatinine-corrected urinary iAs and metabolites monomethylarsonic acid (MMA) and dimethylarsinic acid (DMA) collected during pregnancy corresponded to decrements in several

neuropsychological scores at the age of 4−5 years and were associated with polymorphisms contributing to susceptibility to iAs. It is recommended that these newer studies be considered for updating the IRIS, especially if they contribute to dose-response analysis of neurodevelopmental outcomes. More generally, neurodevelopmental effects should be discussed in the context of birth weight.

Tier 3:

- Mechanistic Observations (P3-120 L1): This section does not portray the current literature on iAs MOA. For example, the highlighted citation Herrera et al. (2013) used relatively blunt methodology at extremely high exposure levels (50 mg/L GD1-LD21) to achieve a measurable response. Additional studies are cited that also use very high exposure levels to implicate changes of glucocorticoid receptor and draw parallels to endocrine effects reported by Wasserman (2007). EPA should consider more recent MOA/AOP studies performed at doses and concentration much more relevant to human exposures in the U.S. (e.g., Niño 2019).

2.4. Charge Question 4. Meta-regression Analyses

EPA performed dose-response meta-analyses, herein referred to as meta-regression (MR) analyses on bladder cancer, lung cancer, diseases of the circulatory system (DCS), and diabetes and presents the results of these analyses in Section 4.3.

a. Please comment on whether the application of a MR analysis and methods used to select studies for the MRs are clearly described and scientifically justified. If there are additional publicly available studies that warrant consideration as the basis of these analyses, please identify those studies, and outline the rationale for including them in the assessment.

The application of a meta-analysis approach is valuable for pooling results across studies to estimate a dose-response function, has well-established precedent in epidemiology as the gold standard for evidence synthesis, and is well-justified. The meta-analysis approach is especially valuable and appropriate for synthesizing evidence to estimate human dose-response estimates for a toxicant like iAs that has a large epidemiological literature base, many simultaneous toxicity mechanisms and targets making the toxicology literature much more complex to evaluate, and where observed human responses appear to differ from simplified experimental systems for many outcomes, further making the toxicological literature challenging to interpret.

The SAB agrees with EPA's approach as being scientifically justified and appropriate to rely on human evidence, synthesized through meta-analysis, as the basis for the dose-response estimation.

The priority of outcomes for dose-response modeling was based on the number of studies with RRB values below 10. An exception was made for pregnancy and birth outcomes as well as neurodevelopmental outcomes as stated on page 4-3, lines 18-22, "Higher-level dose-response analyses were also performed for pregnancy and birth outcomes (see Section 4.4) and neurodevelopmental effects (see Section 4.5), due primarily to their inclusion of potentially susceptible lifestages and their importance for EPA Program and Regional Office consideration in cost-benefit analyses." The inclusion of outcomes relevant to vulnerable populations (e.g. pregnancy, birth, and

childhood for neurodevelopmental outcomes) is justified. However, the cut point (≥25 datasets) does not represent the number of studies deemed appropriate for dose-response modeling. Rather than selecting outcomes based on these initial screening criteria, the EPA should consider other outcomes which include individual studies that meet the criteria (e.g., RRB below 10) for dose-response modeling. This particularly pertains to cancer outcomes for which quantitative risk estimation is the primary goal. The SAB recommends that EPA consider other outcomes as recommended in the NRC/NASEM reports (e.g., respiratory outcomes) or justify their exclusion.

It appears that not all pertinent results from studies identified as relevant via EPA's systematic review were included in the quantitative meta-analyses. It was unclear why cited studies were excluded from the meta-analyses. For example, in Bulka et al. 2022, Supplemental Table 3 presented expected differences in gestational age and birth weight outcomes [regression coefficients] across drinking water concentrations, comparing 5-10 µg /L and >10µg /L iAs in drinking water exposure levels vs. a reference category of 5 µg /L iAs in drinking water, which seems relevant for the meta-analyses on drinking water iAs exposures and these health outcomes. It may be worth reviewing studies identified as relevant and included in hazard identification, but not meta-analysis to confirm no usable associations are presented (in Supplementary Materials) for those papers.

Studies of gestational urinary arsenic were not considered for dose-response modeling as the metabolism of arsenic changes during pregnancy, and models to convert gestational urinary arsenic to harmonizable arsenic intake do not exist for pregnant people. There are publications relating water to urinary arsenic during pregnancy with and without creatinine adjustment (e.g., Gilbert-Diamond et al., 2011, *PNAS* 108(51):20656-20660). Thus, it may be possible to include these studies, and they would add prospective study data. Likewise, studies on children were not mentioned. Studies of water vs. urinary arsenic among children also report correlations (e.g., Wasserman *et al*., 2007, *Environmental Health Perspectives* 115(2):285-289; Wang *et al*., 2007, *Environmental Health Perspectives* 115(4):643-647). Further, data relating water arsenic concentrations to toenail concentrations exist (reviewed in Signes-Pastor *et al*. 2020, *Environmental Research* 195:110286) and have been translated back to water concentrations (Karagas, Stukel and Tosteson, 2002, *Int. J. Hyg. Env. Health* 205(1-2):85-94). With the inclusion of these data, the EPA could increase the number of studies in the dose-response modeling and provide a more precise estimation of the dose-response relationship between iAs intake and the modeled outcomes. Biomarkers represent all sources of exposure. Therefore, reliable biomarkers can characterize the health impacts of iAs at the lower end of the dose-response curve, whereas reliance on drinking water concentrations and self-reported or imputed consumption could lead to appreciable misclassification of exposure. On the other hand, studies of biomarkers can be affected by reverse causality while studies of water iAs cannot be affected by reverse causality as metabolic and pathophysiological processes cannot affect the levels of iAs in the water (i.e., drinking water iAs is an exogenous variable). As a result, combining studies that are affected by different types of bias including studies of arsenic biomarkers and studies of water arsenic is a strength as it allows one to assess the robustness of the findings across different metrics of exposure, affected by different types of bias.

Additional discussion of the methodological strengths and weaknesses of water exposure data vs. biomarker data for epidemiological studies can be found in:

- Glassmeyer, S.T., *et al.* (2023). Water, Water, Everywhere, but Every Drop Unique: Challenges in the Science to Understand the Role of Contaminants of Emerging Concern in the Management of Drinking Water Supplies. *GeoHealth*, 7(12): e2022GH000716.
- Avanasi, R., Shin, H. M., Vieira, V. M., & Bartell, S. M. (2016a). Impacts of geocoding uncertainty on reconstructed PFOA exposures and their epidemiological association with preeclampsia. *Environmental Research*, 151, 505–512. 10.1016/j.envres.2016.08.019
- Avanasi, R., Shin, H. M., Vieira, V. M., & Bartell, S. M. (2016b). Variability and epistemic uncertainty in water ingestion rates and pharmacokinetic parameters, and impact on the association between perfluorooctanoate and preeclampsia in the C8 Health Project population. *Environmental Research*, 146, 299–307. 10.1016/j.envres.2016.01.011
- Weisskopf, M. G., & Webster, T. F. (2017). Trade-offs of personal vs. more proxy exposure measures in environmental epidemiology. *Epidemiology*, 28(5), 635–643. 10.1097/ede.0000000000000686

The overall description of the methods used for selecting studies for dose-response analyses is clear. The committee recommends two ways for this section to be more transparent: 1) include the citations of the studies excluded next to the reason for their exclusion in the flow chart (e.g., Figure 4-2) or include the link or reference to the supplemental table/figure providing the reasons for exclusions and 2) provide the link or reference to the supplemental information that furnishes a description of the studies (e.g., Table C-16 for bladder cancer). A few of the reasons are not immediately obvious and need to be clarified, e.g., exclusion because the study observed a non-monotonic relation. In principle a non-monotonic relationship could be real, and thus studies with these types of relations should be included. Another criterion that needs to be clarified is the minimum number of studies included to perform a meta-analysis. In principle, it should be required that there are at least three studies to be included. If for a certain outcome only two studies are available, this is an insufficient number of studies for a meta-analysis.

Responses to previous charge questions provide additional details on the question of study/data set inclusion and exclusions. For instance, for cardiovascular disease outcomes, the committee recommended that the EPA focus on IHD incidence (incident and fatal IHD). The reasons for this recommendation were provided in response to charge question 3a (see Section 2.3).

Recommendations:

Tier 1:

- Consider inclusion of other outcomes from studies that meet the criteria (e.g., RRB below 10) for dose-response modeling. This particularly pertains to cancer outcomes for which quantitative risk estimation is the primary goal (e.g., skin cancers, skin lesions). Provide justification for any that are excluded.
- Consider inclusion of studies of gestational urinary arsenic and studies on children.

Tier 2:

- Review studies identified as relevant (included in hazard identification, but not meta-analysis) to confirm the absence of usable associations presented in Supplementary Materials for those papers.

b. Please comment on whether the modeling approaches for the MR analyses, including calculation of effective counts, estimation of iAs intake values that account for background oral and dietary exposures, the choice of logistic regression for modeling response probabilities, and hierarchical Bayesian methods to estimate pooled slopes of the relationship between extra risk and dose, are scientifically justified and clearly described.

General Comment

As noted in the response to charge question 2.2, iAs intake was calculated from both drinking water studies and urinary biomarker studies in the assessment. In the former, arsenic in drinking water was assumed to be inorganic arsenic (i.e., arsenite and arsenate). In drinking water studies, background dietary intake (DI) of iAs was added to the estimation of iAs intake (see equations 4 and 5 below). The median U.S. background iAs dose of 0.02 µg/kg-day from diet was assumed (Xue et al. 2010). In some studies, total arsenic in urine (adjusted for dilution using urinary creatine) was used for calculation of iAs intake. PBPK modeling has demonstrated that inorganic arsenic is eliminated almost exclusively in urine. So total arsenic in urine provides a valid estimate of iAs intake if other forms of ingested arsenic that may be excreted in urine, primarily arsenobetaine, are accounted for in the assessment. For studies using urinary arsenic for exposure assessment, it is important to indicate whether the investigators speciated arsenic and excluded arsenobetaine, and if they did not do so, whether the study was conducted in non-seafood eating populations and how seafood intake was assessed in the analyses. Below is the SAB's response to each modeling component mentioned in the charge question.

- **Effective Counts:**

The EPA used an "effective count" approach to account for confounding bias in the studies included in the MR analysis. The method used the adjusted association measures reported in each study (odds ratio or relative risk) to produce pseudo-observations that can be directly modeled using standard dose-response models, thereby providing causal slope estimates within and across studies. The SAB agreed that the method was scientifically justified, though some equations and definitions require clarification. In particular, inconsistency in the definition of expected effective counts for cohort studies needs to be resolved. In the footnote on p. C-15, the EPA defined the expected effective count as the expected number of cases in a group given the referent exposure but with its own confounder profile. However, since the adjusted RR was used to calculate it, the confounder profile should be the same as in the referent group. Thus, it should be the expected number of cases given the same exposure and confounder profile as in the referent group. The EPA should also clarify that adjusted RRs are being used to calculate the effective counts, not the unadjusted values. For instance, the EPA provides the following expressions for the standard error (SE) of the log(RR) and 95% confidence interval for the RR:

$$SE\big(logRR(i)\big) = \sqrt{\frac{1}{Cases(0)} + \frac{1}{Cases(i)}} \qquad \text{Eq. 6}$$

$$95\%\ upper\ confidence\ limit = e^{(\log(RR)+1.96\times SE(\log(RR))}, \text{ Eq. 7}$$

where *Cases*(*i*) refer to the number of cases in exposure group *i*, with *i* = 0 denoting the reference group. The EPA should clarify that the RR in each of these expressions are adjusted RRs.

Recommendations:

Tier 2:

- Revise the definition of expected effective counts in cohort studies to be consistent with the equation provided.
- Where appropriate (e.g., equations 6 and 7 above), replace RR with adjusted RR to clarify how effective count calculations were performed in cohort studies.

- **Estimation of iAs intake values that account for background oral and dietary exposures:**

The SAB agrees that conversion of dose to a common metric in µg/kg-day was a scientifically justified approach to support the meta-analysis, which requires a common metric to combine the epidemiological studies. Different conversions were applied for studies estimating exposure based upon drinking water intake and for urinary biomarker studies. Furthermore, drinking water studies used several metrics such as well water concentration, daily dose and cumulative exposure necessitating variant equations as discussed in the text of the Supplementary Materials (pages C-5 and C-6) and shown in Tables C-18, C-28, C-36, and C-44. Excel files provide the values used for the parameters and justification for their selection for each study.

The equations in the Tables need to be reviewed and numerous editorial corrections made (see detailed comments in Appendix A). The equations need to be presented consistently in the text, tables, and supporting spreadsheets for studies with the same metric and all terms in the equation defined.

Throughout the document (e.g., p. C-5, lines 22-23), it is stated that the dose metric is "average daily dose (µg/kg)," and the units for various factors used in the conversion equations do not include "per day" (e.g., p. C-5, lines 26-27 - water consumption rate [WCR] is in L/kg rather than L/kg-day). The standard convention in exposure calculations such as those presented in the IRIS iAs document is to include "per day" in the units where appropriate. As such, it is suggested that "per day" be added to the units in the exposure calculations throughout the main document and supplementary information where appropriate. Relevant to this point, in Appendix A (draft Systematic Review Protocol), it is stated on p. 5-11, line 1 that the dose metric is µg/kg-day, but the units used in the accompanying equation (p. 5-11, lines 7-13) do not include "per day."

The dose conversion equation for cumulative exposure in units of mg/L*years to mg/kg (p C-5) is shown below. The excerpt is copied from pg. C-5:

"The conversion from cumulative exposures to average daily dose (µg/kg) was carried out as follows:

$$dose = DI + f \times (WCR \times WE) + (1 - f) \times (WCR \times LE) \qquad \text{Eq.4}$$

where the terms in that expression are DI = dietary intake (average daily μg/kg); f = fraction of time (over lifetime up through the study) spent consuming well water (unitless); WCR = water consumption rate (L/kg); WE = well water concentration (μg/L); and LE = low-end water concentration (μg/L). The variable f was calculated as the ratio of the assumed average duration of well exposure (ADWE), generally the reported duration of drinking well water (RDWE; yr), to the average age at diagnosis (AAD; yr). It was assumed that when drinking non-well water, the subjects consumed water with the low-end water concentration. The parameter WE was derived separately for each group by dividing the reported cumulative exposure (μg/L-yr) for that group by the RDWE: $WE = \frac{CE}{RDWE}$. The values used in Eq 4, above, ideally would come from study-specific data reported in the study of interest but could also be drawn from other suitable sources (e.g., from other studies reporting on the same study population or from national authoritative sources)."

On pages C-6 and C-7, the equation is given when a study reports a daily exposure (excerpt copied below):

"For example, consider a study that used daily iAs exposure (in units of μg/day) as the dose metric rather than cumulative exposure in units of (μg iAs/L drinking water) years. Thus, the conversion to average daily μg/kg (dose) was carried out as follows:

$$dose = DI + f \times \left(\frac{WE}{BW}\right) + (1 - f) \times (WCR \times LE) \qquad \text{Eq.5}$$

where the terms in that expression are as above with the addition of BW = body weight (kg). The variable f was estimated as described above, but in the case of a daily exposure study, the parameter WE was derived separately for each group by dividing the reported daily exposure (μg/day) for that group by a BW value."

The use of WE in equations 4 and 5 with different definitions and units is confusing. In Table C-18, the Exposure or Dose Metric column indicates μg/d, DD for Baris et al 2016, Meliker et al 2010 and Steinmaus et al 2013. Using DD (μg/d) as the parameter name in equation 5 would be clearer and the equation would read Dose = DI + f * DD/BW + (1-f)*LE*WCR with units for WCR of L/(kg-day). Equation 5 with text and Tables (e.g., C-18) would need to be changed Since the equation combines iAs exposure from diet (DI), more highly contaminated water (DD), and water with lower levels (LE), text should clearly indicate that DD refers to the more highly contaminated water exposure only.

Additionally, the following sentence is incorrect: "The variable f was estimated as described above, but in the case of a daily exposure study, the parameter WE was derived separately for each group by dividing the reported daily exposure (μg/day) for that group by a BW value (p. C-7, lines 1-4)." It should be revised to say that "the daily intake from more highly contaminated water in units of μg/kg-day was derived separately for each group by dividing the reported daily exposure (DD, in units of μg/day) for that group by a BW value."

The equations for converting drinking water metrics used in the epidemiological studies to the common metric for the meta-analysis include an estimate of dietary exposure in addition to the exposure in water as shown above. Since the urinary biomarkers would reflect iAs intake from food and water, estimating dietary intake for studies in which exposure is based on iAs levels in drinking water that did not already do this provides a common basis for the analysis. El-Masri et al. (2018a, b)

demonstrated that including diet as well as drinking water exposures substantially improved the predictions of the physiologically based pharmacokinetic model. The effect was greatest at lower drinking water levels. This further supports the scientific value of estimating both water and dietary exposures for meta-analysis. Estimates of dietary exposures used data that reflected the study population as closely as possible and are documented for each study in Conversion Factor Validation Spreadsheet_v4.

It is stated on p. 1-11, Lines 14-15 that, "Studies with creatinine corrected urinary intake biomarker data were preferred..." and it appears that urinary arsenic data from studies that did not adjust for creatinine were not considered. This text should be edited for clarity to reflect the Agency's final decisions about using creatinine and specific gravity adjustments in light of the comments below.

The equation for conversion of urinary concentrations should be added to the text Appendix C as it is currently only given in the tables just noted. The equation in the Tables (e.g. Table C-18) specified with "creat." for creatinine, "BW" for body weight is:

> dose = (µg total As/g creat. × g creat./d)/BW

Regarding this equation, the document states:

> According to EPA's PBPK model (El-Masri et al., 2018a,b), iAs is eliminated almost exclusively in urine. Thus, total µg/kg-day arsenic in urine is a good approximation of µg iAs/kg-day intake, assuming arsenic intake is substantially in the form of iAs. Urinary creatinine/kg-day is estimated as = $(266.16 – 47.17*sex - 2.33*BMI + 0.66*age + 0.17*age^2)*113.12/10^6$, where sex is 0 for male and 1 for female and BMI is estimated as $BW/(Height/100)^2$.

The conversion for urinary biomarker studies was supported by a physiologically based pharmacokinetic model, published as peer-reviewed journal articles (El-Masri et al 2018 a, b; El-Masri and Kenyon 2008), that builds on several previously developed PBPK models as described in the IRIS assessment documentation. However, text should be reviewed and edited as needed to clarify that the PBPK model was not used in the conversion of urinary arsenic data to a common dose metric (µg/kg-day). In the dose-response sensitivity analyses described in the Supplementary Materials, the impact of including the urine biomarker studies is assessed for each of the endpoints in the Bayesian analyses. Across the endpoints excluding the urine biomarker studies leads to modest changes and the changes are in both the positive and negative direction. This further supports that the approach to estimating intake from urine biomarker measurement is not introducing a large uncertainty into the analyses but is facilitating the use of more of the scientific database.

However, for urinary biomarker studies that reported exposure or dose metric as µg total As/g creatinine to account for dilution, if specific gravity was available in the study, it would be relevant to conduct an additional sensitivity analysis that uses arsenic corrected for specific gravity instead of urinary creatinine, given there was some concerns that urinary creatinine could introduce some bias because of the connection between creatine and arsenic metabolism (Abuawad et al., 2023). However, data has shown that urinary arsenic corrected by urinary creatinine is more closely correlated with blood arsenic and water arsenic, compared to urinary arsenic corrected by specific gravity (Abuawad et al., 2022), probably because specific gravity does not have sufficient precision. Based on these data,

urinary arsenic corrected by urinary creatinine is an adequate method to correct spot urine samples for urine dilution.

Additionally, as mentioned on pg. 1-11, lines 5-6, non-toxic organic arsenic compounds found in seafood such as arsenobetaine can contribute to total urinary arsenic. It is recommended that additional information be provided about how seafood consumption and the presence of organic arsenic compounds may have impacted the exposure estimates based on urinary arsenic.

Recommendations:

Tier 1:

- Augment the analysis to include those studies correcting biomonitoring data for dilution with specific gravity, although it may be less precise than creatinine adjustment.

Tier 2:

- Revise the IRIS document to clarify that PBPK models provided the scientific basis for the dose conversions but were not directly used to estimate daily dose.
- The equation for calculation of dose from urinary biomarker studies should be provided and discussed in the text, and not just in the table.
- Provide additional information about how seafood consumption and the presence of organic arsenic compounds may have impacted the exposure estimates based on urinary arsenic.
- Add "per day" to units for factors used in dose calculations where appropriate (e.g., mg/kg-day; mg/L-day).
- The dose conversion equations in the text, tables, and elsewhere (e.g., EXCEL files) need to be reviewed and corrected. The text currently describes the equations for cumulative drinking water exposure in mg/L*years and daily dose in mg/day but could be expanded to specify equations for the other metrics used including water concentration in mg/L and cumulative exposure in mg.
- All the calculation spreadsheets should be reviewed as an additional quality assurance and quality control (QA/QC) step to ensure they are correct, and the terminology is consistent between the documents and the spreadsheets in light of the numerous editorial comments made on the document.

Choice of logistic regression for modeling response probabilities:

The EPA used a logistic regression model for dose-response modeling because it can integrate case-control and cohort data into the same analysis. The SAB believes this justification is reasonable but are concerned about the restrictions imposed by the logistic model, particularly the monotonicity imposed as well as the lack of a lower threshold. The EPA performed extensive sensitivity analyses of the monotonicity assumption comparing their simple linear-in-the-logit model to fractional polynomial and double hill models. The SAB commends the EPA for those explorations but believes the justification for selecting their simpler model is lacking particularly when the alternative models provided superior fit. For instance, in the sensitivity analysis for bladder cancer, two fractional polynomial models had maximum and mean posterior likelihoods that were considerably larger than the logistic model and produced consistently higher lifetime risk estimates. To support the decision to use a simpler model,

the EPA should examine fit statistics such as the Akaike's Information Criterion (AIC) and Bayesian Information Criterion (BIC) which incorporate penalties for model complexity. The EPA should expand their sensitivity analyses to include a model which either explicitly includes a lower threshold effect or which allows for a lower threshold if supported by the data.

Recommendations:

Tier 1:

- Use a fit statistic that penalizes for model complexity (e.g., AIC or BIC) to determine if the simple linear logistic model is sufficient for an individual outcome, or if a more flexible alternative model is needed.

Tier 2:

- Expand sensitivity analyses to include models with a lower threshold effect for iAs exposure.

Hierarchical Bayesian methods to estimate pooled slopes of the relationship between extra risk and dose:

Overall, the SAB agrees that the EPA's Bayesian meta-regression is a scientifically justified approach for combining slope estimates from multiple studies and is well described. However, revisions are still needed to ensure that the method is accessible to a broader audience. In particular, the hierarchical structure and meaning of model parameters (such as the overall slope and its variance) may be confusing to those not familiar with meta-analyses. A diagram describing the model structure and how it accounts for intra- and inter-study variability could make this model more understandable to a general audience.

The likelihood function used to perform Bayesian inference on the EPA meta-regression model is a hybrid of binomial (for case-control studies) and Poisson (for cohort studies) likelihood contributions for study-specific exposure groups. Since the EPA's hierarchical model is fit to non-integer outcome data (the effective counts), continuous approximations of the binomial and Poisson were utilized. Although the approach is theoretically sound, we expect that most readers (including statisticians) will not be familiar with these approximations. Thus, the EPA should provide citations to peer-reviewed publications supporting their likelihood function. If such citations do not exist, the EPA should provide the derivation of these approximations in their supplementary materials.

The SAB also raised a concern about the two-stage approach used to fit their models. In the first stage, the outcome (effective count) and the exposure was estimated. Then, in the second stage, the EPA treated the outcome and exposure as observed (not estimated) data and fit their meta-regression model. Although this is a practical approach, two stage approaches like this can produce biased coefficient estimates in the outcome model (see, for example, Tsiatis and Davidian, 2004) and may under-estimate variability and uncertainty, as noted in the Allen et al. (2020b) article which proposed the meta-regression approach used by the EPA. Allen et al. suggested a future simulation study to examine these issues. The EPA is strongly encouraged to perform these simulation studies to ensure that their models are providing accurate BMDs and BMDLs.

The SAB believes that the sensitivity analyses performed by the EPA provides a rigorous assessment of the influence of individual studies on the meta-regression. However, there was general agreement within the SAB that additional model diagnostics are needed. Assessments of convergence and model fit should be expanded. In the draft document, convergence diagnostics are provided for CVD and IHD, but are missing for the other outcomes. Goodness-of-fit evaluation is limited to assessments of study-specific curves. Since primary interest is in the pooled slope across studies, the EPA should also examine posterior predictive plots, i.e., they should use their model results to generate data for a future study and see how well it agrees with the observed data. The SAB also believes that the EPA should examine differences in participant characteristics (e.g., distribution of age, sex) between the studies included in the meta-regression and their target population for the assessment. If differences exist, the EPA should perform an analysis examining the influence of these characteristics on the study specific slopes. For instance, the slopes could be regressed on study-level summaries of demographics, or the demographic variables could be included as covariates in the meta-regression.

Finally, the SAB has concerns about the appropriateness of certain data for the meta-regression. For instance, the bladder cancer model included multiple studies from the same population measured at different times. The inclusion of these studies violates the independence assumptions of their model and results in one population having a large influence on the results. The SAB is also concerned about the use of the meta-regression when there are only two studies, as in their analysis of the incident CVD data. When the sample size is this small, the results will almost certainly be dominated by the priors.

Recommendations:

Tier 1:

- Provide a more practical and accessible explanation of the meta-regression model including a diagram of the hierarchical structure.
- Provide theoretical support for the continuous approximations to the binomial and Poisson distributions used to construct the likelihood, either through citations to peer reviewed publications or by adding detailed derivations to the supplementary materials.
- Provide convergence diagnostics for all outcomes included in the assessment.
- Examine goodness-of-fit using posterior predictive plots.

Tier 2:

- Examine differences between participant characteristics in the studies included in the meta-regression and the target population for the assessment. If differences exist, perform an analysis examining the impact of demographic factors on study-specific slopes.
- When multiple studies from the same population are included in the same meta-regression model, examine the impact of the population on the overall slope by re-running the analysis with these studies excluded.
- Clearly describe the minimal data standards required for the meta-regression, including the minimum number of studies needed.

Tier 3:

- Perform an extensive simulation study examining the impact of ignoring the estimation error of the expected counts and exposure on BMD and BMDL estimation.

c. *In applying the hierarchical Bayesian model, EPA selected priors for the pooled MR slope such that the pooled slope could not be negative, reflecting the causal determinations for bladder cancer, lung cancer, diseases of the circulatory system, and diabetes. Please comment on whether this decision is scientifically justified and scientifically justified and clearly described.*

The SAB agrees that the use of a Gamma prior on the overall mean is supported by a substantial evidence base that iAs exposure should increase risk of the outcomes analyzed. However, this assumption can place undesirable restrictions on the analysis that are not supported by the data from the individual studies. For example, in the bladder cancer analysis, about half of the posteriors for the study specific slopes give considerable weight to negative values and the posterior for the overall mean slope gives high weight toward values close to zero. In light of these findings, the EPA should perform a sensitivity analysis using a prior for the overall mean which allows negative values, but which favors positive values to be consistent with a priori knowledge.

The SAB appreciates the sensitivity analysis on the parameters of the gamma distribution, but this does not address the concern of the positivity restriction. In addition, some clarification is needed regarding the parameter values used in the sensitivity analysis. For instance, Table C-23 provides a range of values, however, it is not clear what those values correspond to.

Recommendations:

Tier 2:

- Perform sensitivity analyses removing the positivity restriction on the prior for the overall mean slope.
- Clearly list the values of parameters a and b used in the sensitivity analyses for the gamma prior.

2.5. Charge Question 5. Lifetable Analyses

EPA applied lifetable analyses to extrapolate estimates of MR pooled slopes to the desired target population (i.e., general United State population). EPA provides the results of the lifetable analysis as extra risk values calculated for a set of discrete iAs intake values, as well as polynomial trend lines with equations for the extra risk curves so that users can calculate extra risk values for each outcome at any dose they require. For each outcome below, please comment on whether the lifetable methods have been scientifically justified and clearly described.

General Comments for Question 5a-5d:

The general approach for life table estimation is scientifically justified because it follows the standard approach (Crump and Allen, 2011) used in other risk assessments, such as the IRIS Toxicological Review of Libby Amphibole Asbestos. However, in the draft document it was unclear how the general methodology was applied in this assessment including how the meta-regression slopes were incorporated into the life tables. During the in-person public meeting, the EPA provided a more detailed explanation on how the calculations were performed including screenshots of the Excel

spreadsheets used in the calculations. It was then understood that the following methodology was used:

i. Population data were used to obtain age-specific incidence rates at an assumed background dose. The U.S. median background dose of iAs was assumed to be 0.0365 µg/kg-day.
ii. Slope estimates from the Bayesian meta-regression were used to convert the incidence rate at background to incident rate at zero dose, which provides an intercept for a logistic regression model for age-specific incidence.
iii. The corrected intercept and slope were then used to calculate age-specific incidence and age-specific probability of disease at a given dose. This calculation assumed a constant exposure to iAs over a lifetime, which is a reasonable simplifying assumption since drinking water iAs often has more geographical variability than temporal variability.
iv. The age-specific probabilities were then summed to obtain the lifetime risk of disease.

The EPA needs to provide this level of transparency in the description of the Lifetable Analysis method to estimate lifetime extra risk for the target population in the supplementary materials for the iAs assessment. It is recommended that the EPA provide a worked example in the supplementary materials which translates the general equations (in page C-27 and C-28, Appendix C of the Supplemental Materials document) to quantities from the meta-regression and includes a table which clearly shows the steps of the calculation. For cases where the calculations are considerably different (e.g., when only lifetime risk is available), clear explanation should be provided on how the calculations differ from the worked example provided earlier. The SAB also recommends inclusion of more practical definitions of the components of the life table analysis (e.g., what a hazard function is) that can be understood by a more diverse scientific audience. In addition, clearer citations for the population risk values used in the lifetable calculations should be provided in the supplemental materials document.

The EPA provided extra risk curves that can be used to estimate extra risk without performing the lifetable calculations. However, SAB members did not find these plots helpful and would have preferred that the equations for each curve be presented in a table. The EPA should explain the purpose of these plots if they are to remain in the document.

The SAB also recommends that, in future assessments, the EPA use the results from the Bayesian meta-regression to generate a posterior distribution for extra risk at selected dose values. This approach will not change the overall conclusions of the assessment, including cancer slope factors and reference doses, but will provide a more complete understanding of the uncertainty of these estimates. EPA should also consider, when possible, relaxing their assumption of a constant exposure over a lifetime.

a. *The Toxicological Review estimates extra risks at various iAs doses for multiple DCS endpoints: cardiovascular disease (CVD) incidence, fatal CVD, ischemic heart disease (IHD) incidence, and fatal IHD. Age stratified mortality values were available for fatal CVD and fatal IHD and were used in the lifetable analysis. Age-stratified DCS morbidity values were not available, and a single lifetime background risk value was used in the analyses for CVD and IHD incidence. At 0.13 µg/kg-day, a lifetime extra risk for CVD incidence of 2.1×10^{-2} was estimated.*

The two available studies (Chen et al. 2013b, and Moon et al 2013) for CVD incidence were from different ranges of exposure: Chen et al. at high-dose, and Moon et al. at low-dose. The limited number of studies and the variation in the β_mean (0.04 in the Chen et al. study and 0.54 in the Moon et al. study) may suggest large uncertainty in the estimation of lifetime extra risk. Thus, due to the limited data, the EPA should reconsider the decision to include this endpoint in RfD determination in the assessment. The SAB recommends the assessment focus on ischemic heart disease as the key endpoint for cardiovascular outcome (see response to Charge Question 3a). This recommendation was also revisited in the SAB's response to Charge Question 7b.

The target population for the risk estimates is unclear. In particular, the assumption of 70% lifetime incidence of CVD requires justification as the probability changes by the starting age for lifetable calculations and differs by males and females. It is not clear whether using different starting age or stratifying by sex changes the extra risk estimation. The lifetime incidence of CVD also differs by country and race/ethnicity groups. The Leening et al., 2014 paper showed the overall lifetime risks of CVD were 67% for men and 66% for women at age 55. The study sample was from Netherlands. EPA needs to remark on the possible difference by country and race/ethnicity if a relevant U.S. reference for lifetime risk cannot be identified.

b. *The Toxicological Review estimates extra risks at various iAs doses for type II diabetes mellitus. Age-stratified diabetes mortality and morbidity values were not available, and a single lifetime background risk value was used in the analysis. At 0.13 µg/kg-day, a lifetime extra risk for diabetes of 1.8×10^{-2} was estimated.*

As in the CVD calculations, the EPA needs to provide more justification for the assumed 40% lifetime incidence of diabetes in the U.S. since the probability changes by the starting age and differs by race/ethnicity. It is not clear whether using different starting age or stratifying by race/ethnicity changes the extra risk estimation. The SAB also identified two potential typos in Table 4-11: (i) the posterior mean for the pooled slope (β_mean) is higher than the posterior mean slope in each study and (ii) the posterior mean slope in the Pan et al. study equals the 5^{th} percentile.

c. *The Toxicological Review estimates extra risks at various iAs doses for developing bladder cancer. Age stratified mortality and morbidity values were available for bladder cancer and were used in the lifetable analysis. At 0.13 µg/kg-day, a lifetime extra risk for bladder cancer of 7.9×10^{-4} was estimated.*

The SAB did not have additional comments for the bladder cancer lifetable calculations. All of our concerns are provided in our general comments for Charge Question 5.

d. *The Toxicological Review estimates extra risks at various iAs doses for developing lung cancer. Age stratified mortality and morbidity values were available for lung cancer and were used in the lifetable analysis. At 0.13 µg/kg-day, a lifetime extra risk for lung cancer 2.4×10^{-3} was estimated.*

The SAB did not have additional comments for the lung cancer lifetable calculations. All of our concerns are provided in our general comments for Charge Question 5.

Recommendations:

Tier 2:

- Provide more details in the supplementary materials on how the lifetable calculations were performed, including a worked example and an explanation of required modifications when age-specific incidence rates are not available. The explanation should be comprehensible to a diverse scientific audience.
- Provide clearer citations for the population risk values used in the lifetable calculations.
- Discuss, or if possible, examine the impact of using different starting ages and stratification by race/ethnicity and sex in the lifetable calculations for CVD incidence and diabetes.
- Remark on the possible differences by country if a relevant U.S. reference for lifetime risk cannot be identified.
- Explain the purpose of the extra risk curves and include a table containing the equations for the fitted curves.

Tier 3:

- If data are available, perform sensitivity analyses to account for proportions of study populations that are not exposed to a constant level of iAs over their lifetime.
- In future assessments, the EPA should use the results from the Bayesian meta-regression to generate a posterior distribution for extra risk at selected dose values.

2.6. Charge Question 6. Cancer Slope Factor

Based on the lifetable analyses for lung cancer and bladder cancer, linear trend lines were used to estimate a cancer slope factor (CSF). These CSFs were estimated using only risks derived in the low-dose region given non-linearity at higher doses. Please comment on whether the selected CSF values are scientifically justified and clearly described.

The CSFs for lung cancer, bladder cancer, and combined cancer risk are based on data indicating a linear dose-response relationship in the low-dose region. The 95% upper bound extra risk estimates were used to calculate the linear slopes, which are analogous to EPA CSFs. These CSFs appear to be scientifically justified for estimation of risks in the low-dose range (<0.22 µg/kg-day). However, the draft document states (Table ES-1; Figures 4-7 and 4-9; p. 4-67, lines 12-15) that the dose-response is non-linear at doses above 0.22 µg/kg-day and that "risk estimates should not be obtained using the CSF" above that dose (p. ES-3, lines 5-7).

The CSFs appear to be scientifically justified. However, several aspects of the explanations of the cancer risk estimates need to be revised to ensure that the information is clear to the reader, as discussed below:

- The paragraph that starts on p. ES-2, line 30, and ends on p. ES-3, line 9, discusses both cancer risk values based on mean extra risk values (i.e., the lifetime extra risk estimates of 7.9 for bladder cancer and 2.4 for lung cancer per 10,000 for a hypothetical U.S. cohort) and the cancer slope factor, which is based on the 95th percentile upper confidence bound on the lifetime extra

risk. For clarity, it should be stated the lifetime extra risks for the hypothetical U.S. cohort are based on the mean extra risk values (in contrast to the cancer slope factor which is based on the 95th percentile upper confidence bound on the extra risk values). This recommendation also applies where extra risk estimates are discussed elsewhere in the document.

- The CSFs are explained as "estimates of the 95% upper bound on the lifetime extra risk associated with a daily 1 µg/kg dose" (p. ES-2, line 33 – p. ES-3, line 1), and "an estimate of the 95% upper bound lifetime extra risk (above an estimate of the U.S. risk at zero iAs dose) associated with a 1 µg iAs/kg-day dose" (p. 4-67, lines 15-17). This explanation is not factually incorrect, however, expressing the CSFs in terms of a dose of 1 µg/kg/day may be confusing to readers because the CSFs are stated <u>not</u> to apply to a dose as high as 1 µg/kg/day. It is suggested that this information be rephrased to avoid potential confusion, such as "a factor that represents the slope of the dose-response relationship between the estimated 95% upper bound on the lifetime extra risk and dose. For iAs, the CSF is expressed in units of $(\mu g/kg\text{-}day)^{-1}$", and "the slope of the dose-response relationship for the estimated 95% upper bound lifetime extra risk (above an estimate of the U.S. risk at zero iAs dose) and dose, is expressed in units of $(\mu g/kg\text{-}day)^{-1}$."

- The cancer risk information was presented in terms of drinking water concentration (µg/L) or daily dose (µg/kg-day) in different places throughout the document. It would be clearer and more informative to provide this information in terms of both units where relevant (i.e., for risk estimates based on drinking water studies). For example, the Executive Summary states: "Lifetime extra risks of 7.9 and 24 were estimated for bladder cancer and lung cancer respectively, for a hypothetical U.S. cohort of 10,000 individuals exposed for a lifetime at the U.S. drinking water standard of 10 µg/L," but the daily dose of 0.13 µg/kg-day that is associated with 10 µg/L is not provided. The same paragraph then goes on to discuss the linearity below a daily dose of 0.22 µg/kg-day, but it does not state the drinking water concentration associated with this dose.

- The combined risk of bladder cancer and lung cancer with no drinking water exposure can be estimated as 0.001 (1 in 1000) by multiplying the dose from background dietary exposure (0.02 µg/kg-day) by the draft CSF (5.3E-2 per µg/kg-day). Similarly, the combined risk of bladder cancer and lung cancer from background dietary and drinking water exposure can be estimated as 0.002 (2 in 1000) by multiplying the background dose (0.0365 µg/kg-day) by the draft CSF (5.3E-2 per µg/kg-day). It is recommended that the IRIS document provide these "background" risk estimates and state that these risks (1 in 1000; 2 in 1000) are much higher than the cancer risk range usually considered by USEPA in developing health-based criteria (e.g., 1 in 10,000 to 1 in 1,000,000).

The protocol for the draft IRIS assessment, which was reviewed and approved by NASEM (2019), includes analyses of the uncertainties associated with several parameters used in CSF development. These analyses include inferred probability distributions for each conversion factor used in the estimation of iAs intake, Monte Carlo analysis to obtain the MLE, low-end, and high-end values for iAs intake, use of the effective count approach for odds ratio or relative risk to account for confounding

bias in the studies included in the meta-regression analysis, and Bayesian meta-regression analysis in the dose-response analysis.

However, the draft IRIS document does not appear to include a section that describes uncertainty characterization as it is discussed in the IRIS Handbook (USEPA, 2022). Section 8.4 of the IRIS Handbook, "Characterizing uncertainty and confidence in toxicity values," recommends that an assessment should – at least qualitatively – address principal sources of uncertainty (USEPA, 2022). The Handbook describes several factors that commonly affect the level of uncertainty, including consistency in the overall database (specifically noting that the variability among candidate values for the same outcome be evaluated, taking into account explanations for differences), dose metrics used for dose-response modeling, model uncertainty, and statistical uncertainty.

In addition to the characterizations of uncertainty already included in the draft document, it is recommended that a stand-alone discussion of the qualitative and quantitative uncertainties associated with the CSF in the context of the recommendations in the IRIS Handbook (USEPA, 2022) be added. As an example, that may help with these clarifications, the Handbook discusses uncertainty around the strength of evidence associating a dose metric with the critical effects. While the draft iAs assessment mentions use of effective count approaches to account for confounding bias in the modeling, it is not clear how this helped to characterize uncertainty (particularly if the studies modeled were rated as having low risk of bias for confounding). Thus, for this example, while the assessment of uncertainty due to confounding is discussed in Section C.1.1, clarification is needed as to how uncertainties related to factors such as systematic bias, residual bias, and uncontrolled confounding were considered for each outcome. Additionally, as discussed in the response to Charge Question 5, EPA should consider estimation of risks in the lifetable analysis using the individual iterates of the Bayesian analysis to provide a more complete picture of the uncertainty in the CSF.

The Handbook also states: "Whether it is quantitative or qualitative, characterization of uncertainty is communicated clearly and transparently to facilitate decision-making." Given the advanced methods used to qualitatively evaluate uncertainty in the CSF in draft IRIS assessment, clarifying and enhancing the reporting around the characterization of uncertainty in the CSF would strengthen the assessment and would be useful to risk managers, particularly regarding cost-benefit analysis.

<u>Recommendations</u>:

Tier 1:

- When cancer risk values are discussed, it should be made clear whether they represent the mean or the 95% upper bound estimate.
- The explanation of cancer risk that is associated with a 1 µg/kg-day dose (p. ES-2, line 33 – p. ES-3, line 1; p. 4-67, lines 15-17) be rephrased to say that the CSFs, in units of $(\mu g/kg\text{-}day)^{-1}$, are the slopes of the relationship between risk (unitless) and dose (µg/kg-day), and that that they only apply to doses up to 0.22 µg/kg-day.
- Consistent with the IRIS handbook, an uncertainty section should be added to the document summarizing uncertainties in the cancer assessment and resulting slope factor.

Tier 2:

- Cancer risks should be presented in terms of both µg/kg-day and µg/L (assuming standard consumption rate) when both units are relevant.
- The approach used to account for background exposures in the lifetable analyses should be clearly presented, and the document should discuss that the estimated "background" cancer risks (from dietary exposure, with no additional exposure from drinking water) are much higher than the cancer risk range of 1 in 10,000 to 1 in 1,000,000 usually used by EPA.
- Consider adding discussion of how uncertainties related to systematic bias, residual bias, and uncontrolled confounding were considered for each outcome.

Tier 3:

- Estimation of risks in the lifetable analysis using the individual iterates of the Bayesian analysis should be considered.

2.7. Charge Question 7. Non-Cancer RfD Candidate Values

EPA calculated a non-cancer RfD based on candidate values for each individual noncancer health outcome considered for dose-response analyses and presents the results of these analyses in Sections 4.6.

a. *EPA determined that data from the (Wasserman, 2014) study on developmental neurocognitive effects were not appropriate for candidate value derivation given the strong nonlinearity observed in the relationship between iAs exposure and IQ scores. Please comment on whether this approach is scientifically justified and clearly described.*

Summary of EPA conclusions:

EPA concluded that "the currently available evidence indicates that iAs exposure likely causes neurodevelopmental effects in humans given sufficient exposure conditions" and stated that this conclusion was primarily based on human studies showing cognitive effects at drinking water concentrations of <100 µg/L.

Wasserman et al. (2014), a study of IQ and drinking water iAs in U.S. children, was the only study subjected to dose-response analysis for cognitive and behavioral outcomes. Several other human studies of this endpoint showed monotonic effects (increased effect with increasing dose) but were not selected for dose-response analysis. For example, a cross-sectional study performed in a highly exposed population in Bangladesh (Wasserman et al., 2004) was considered but was not chosen to proceed for dose-response analysis because of concerns about the generalizability of IQ tests standardized to U.S. populations to a Bangladeshi population.

The analyses presented in Wasserman et al. (2014) were based on iAs drinking water exposure categories (<5 µg/L; 5-10 µg/L, etc.). Using data provided by the authors, EPA conducted an additional analysis with iAs drinking water concentration as a continuous variable (Tables 4-15 andC-80 and Figure C-53). Although there was a negative relationship between drinking water iAs concentration and full-scale IQ and other IQ measures, the relationship was not statistically significant overall at the $p < 0.05$ level. Specifically, the negative relationship of drinking water iAs with IQ was clear at low iAs levels

(up to approximately 20 µg/L), which included most of the subjects, but was irregular at higher exposure levels (approximately 20 µg/L to 115 µg/L) which included fewer subjects.

Importantly, Wasserman et al. (2014) state that their data indicate that a drinking water concentration of 5 µg/L "may represent an important threshold." Consistent with this conclusion, EPA identified a LOAEL of 0.11 µg/kg-day, which is the central tendency estimate of the total dose from diet and drinking water associated with exposure to 5 µg/L in drinking water, from the Wasserman et al. (2014) data (p. C-199, lines 9-12).

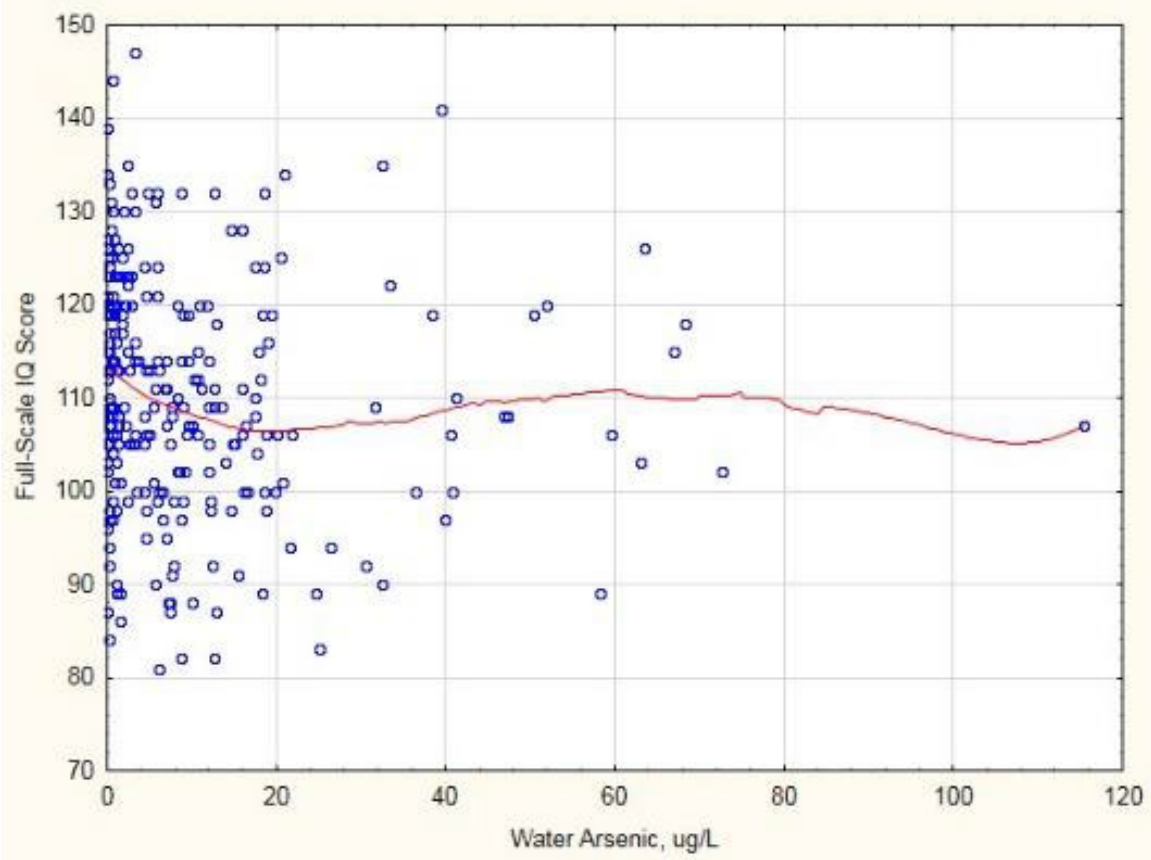

Figure 2-1. Kernel smoothed fit of full-scale IQ to water arsenic concentration.

However, because the dose-response was non-monotonic at higher exposures (>20 µg/L), EPA concluded that it was "not advisable" to use the results of this modeling to estimate exposure "across the full range of exposures." Additional limitations of Wasserman et al. (2014) mentioned by EPA include:

- Ethnic and socioeconomic differences between the study population and children in the U.S. general population.
- IAs exposures in the study were "relatively low," possibly limiting the ability to detect effects at higher doses.
- Uncertainties about dietary arsenic intake in the study population, although this was acknowledged to be less of an issue than in studies from other countries (e.g., Bangladesh).

- Lead exposures, which are associated with decreased IQ in children's IQ, were not measured or controlled for.

Commentary

Limitations and exclusion criteria for studies considered for dose response

The draft IRIS document stated that the dose-response data from Wasserman et al. (2014) were not appropriate for RfD derivation because of the "nonlinearity" of the dose-response relationship (p. 4-56, lines 17 and 30; Figure C-53), and the charge question states that there is "strong nonlinearity." However, it is unclear that this is a scientifically supportable rationale for exclusion of this study, since a linear relationship is not required for data used in dose-response evaluation, and many datasets used as the basis of PODs and RfDs are not linear. Additionally, the draft document states that the number of subjects in the higher exposure group was relatively small, and that, while the overall dose-response over the entire exposure range was not statistically significant (Table C-80), there was a strong monotonic dose-response relationship in the lower exposure range (<20 μg/L). This lower dose range (<20 μg/L) included most of the study's subjects and is also most relevant to environmental exposures in the U.S. population. Furthermore, the SAB concluded that non-monotonicity at the higher exposures (shown in Figure C-53) may not actually be as clearcut as suggested in the draft IRIS document.

The SAB also considered other limitations of Wasserman et al. (2004) and Wasserman et al. (2014) mentioned by EPA, as follows:

- EPA stated that Wasserman et al. (2004) and, to a lesser extent, Wasserman et al. (2014) are not appropriate for RfD development because of ethnic/socioeconomic differences between the study population and children in the U.S. general population (Wasserman et al., 2014) and because the IQ tests that were used may not be culturally relevant in Bangladesh (Wasserman et al., 2004).

  However, the results of these two studies were expressed as changes compared to the lowest exposed subject group, and ethnic/socioeconomic factors would not necessarily impact the magnitude of the dose-related effects of iAs when based on comparison to this referent group. Additionally, uncertainties related to socioeconomic and ethnic differences may be accounted for by application of a UF_H for interindividual variability.

  Relevant to this issue, it was noted that the RfD for pregnancy and birth outcomes is based on a single Bangladeshi study and that U.S. data were used to interpret adversity from this study in the draft document. It is not clear why this approach is not also applicable to neurodevelopmental outcomes.

  Clarification is needed as to why lack of information on lead exposure was not considered in evaluating bias in Wasserman et al. (2014), which was assigned "low" for accounting for confounding variables, "probably low" for co-exposure, citing "no evidence of co-exposure" and "definitely low risk of bias" for exposure characterization. However, consideration of lead exposures is a major concern only if iAs and lead exposure are positively correlated, and that it

is not clear that there is a scientifically supportable reason to believe that this is the case in Wasserman et al. (2014). Specifically, the source of iAs in drinking water in the study area is likely natural occurrence in geological formations while the lead in drinking water arises from corrosion of plumbing. The SAB is not aware of a reason that iAs would be expected to be correlated with lead in drinking water here.

- EPA stated that the "relatively low" exposures in the study potentially limited the ability to detect effects at higher doses. However, this statement needs to be clarified as it appears to be inconsistent with the exclusion of other studies because the exposure range is above the range relevant to the U.S. population.

The lack of consideration of birth weight in Wasserman et al. (2014) may also be a potential concern, although it is not clear that adjustment for birth weight is necessary for the age range (8-10 years) of the subjects in this study. Clarification is needed as to whether EPA believes that lack of information on birth weight is a relevant consideration in evaluating bias for this study, which was assigned "low" for accounting for confounding variables.

Study selection for dose-response evaluation of neurodevelopmental effects

Additional information is needed on the selection of studies for dose-response modeling of neurodevelopmental/behavioral impairments, including additional explanation of the rationale for exclusion of all but two of the 52 studies that were considered (Wasserman et al., 2004; Wasserman et al., 2014). Specific documentation of criteria not met for each study that was excluded should be provided. It would be helpful to know if the two studies that were ultimately selected for dose-response analyses were also the most sensitive and/or where the magnitude of the effects associated with iAs exposure in those two studies rank relative to the overall evidence base.

Additionally, EPA stated that Wasserman et al. (2014) was excluded from further dose-response analysis based on lack of generalizability. However, generalizability was not mentioned as an exclusion criterion in the systematic review or dose-response criteria. Therefore, additional rationale and/or acknowledgment of use of generalizability as a criterion should be provided, particularly because this same approach does not seem to have been applied to other studies.

Consideration of neurodevelopmental effects in dose-response evaluation and RfD development

It would be appropriate and informative for EPA to identify the dose levels (e.g., LOAEL and/or NOAEL) at which neurodevelopmental cognitive effects occur and to also, if possible, develop a candidate RfD for these effects. This evaluation could potentially consider both the U.S. study (Wasserman et al., 2014) and the Bangladeshi study (Wasserman et al., 2004), for which EPA obtained individual exposure and response data and covariates from the authors, as well as any other studies that EPA may identify as appropriate after consideration of the SAB's recommendations.

An RfD for neurodevelopmental effects would be useful for comparison with the RfDs for other endpoints (see response to charge question 9 below), even though it might not be considered for use as the final RfD due to greater uncertainty than for the other candidate RfDs. It would provide

information about whether the final RfD based on another endpoint (e.g., DCS) is protective of all neurodevelopmental outcomes, which would inform the need for application of a database uncertainty factor (UF_D) to protect for potentially more sensitive effects, as discussed further in the response to Charge Question 8a. A UF_D is applied when there are gaps in the toxicological database or when available information indicates that additional studies might indicate that a lower RfD is appropriate (USEPA. 2022. ORD Staff Handbook for Developing IRIS Assessments, p. 8-13).

As discussed earlier in this response, it is unclear that Wasserman et al. (2014) should be excluded from consideration for RfD development, and, specifically, that the shape of the dose-response relationship did not appear to be a supportable reason for exclusion. Therefore, it is recommended that a dose-response evaluation of only the lower portion of the exposure range (up to about 20 µg/L) in this study be conducted. This exposure range includes most of the study's subjects, and it is the range most relevant to U.S. environmental exposures. The results of this evaluation will provide information on whether associations in the lower part of the exposure range (<20 µg/L) are statistically significant and potentially useful for development of a POD and associated RfD.

Additionally, Wasserman et al. (2014) state that a drinking water iAs concentration of 5 µg/L "may represent an important threshold." Consistent with this conclusion, EPA has identified a LOAEL of 0.11 µg/kg/day from the Wasserman et al. (2014) data (p. C-199, lines 9-12), which is the central tendency estimate of the total dose from diet and drinking water associated with exposure to 5 µg/L in drinking water. If a BMD/BMDL cannot be developed from the Wasserman et al. (2014) data, EPA should consider development of an RfD based on this LOAEL, or, at a minimum, comparison of this LOAEL to the PODs for other non-cancer effects to determine whether neurocognitive effects are more sensitive. Relevant to this point, EPA (2012) BMD modeling guidance (https://www.epa.gov/sites/default/files/2015-01/documents/benchmark_dose_guidance.pdf) and the EPA (2022) ORD Staff Handbook **(**https://cfpub.epa.gov/ncea/iris_drafts/recordisplay.cfm?deid=356370) say that a LOAEL may be used as the POD for RfD development if the data do not support BMD modeling.

Recommendations:

Tier 1:

- EPA should reconsider whether uncertainties related to socioeconomic and ethnic differences preclude the use of Wasserman et al. (2004) and Wasserman (2014) for RfD development and consider whether these factors may be accounted for by application of a UF_H for interindividual variability.
- EPA should reanalyze the dose-response relationship in Wasserman et al. (2014) and consider whether this relationship, particularly in the lower dose range (<20 µg/L), can be used as the basis for POD development.
- EPA should consider whether the approach for the proposed RfD for birth outcomes, which uses U.S. data to interpret adversity from a study in a non-U.S. population, is also applicable to neurodevelopmental effects. If it is concluded that this approach is not applicable to neurodevelopmental effects, a rationale should be provided as to why it is appropriate for birth outcomes.

- Additional information should be provided about the rationale for exclusion of the 52 studies of neurodevelopmental behavioral effects considered (except for Wasserman et al. (2004) and Wasserman et al. (2014)). This discussion should include which criteria were not met for each excluded study. It should also discuss whether the two studies selected for dose-response analyses were the most sensitive and/or where the magnitude of the effects reported in those two studies fell relative to the overall evidence base.

Tier 2:

- Regarding potential lead exposure in Wasserman et al. (2014), EPA should discuss whether there is reason to expect that exposure to lead and iAs would be positively correlated in this study and that EPA clarify why lack of information on lead exposure was not considered in evaluating bias in this study.
- EPA should clarify why "relatively low" exposures in Wasserman et al. (2014) potentially limited the ability to detect effects at higher doses, since this statement appears to be inconsistent with the exclusion of other studies in which exposure was above the range relevant to the U.S. population.
- EPA should clarify whether lack of information on birth weight is a relevant concern in evaluating study bias for Wasserman et al. (2014).
- EPA should discuss whether generalizability was used as a criterion for study selection. If so, the rationale and/or acknowledgment of use of generalizability as a criterion should be provided.
- EPA should identify the dose levels (e.g., LOAEL and/or NOAEL) at which neurodevelopmental cognitive effects were observed and develop a candidate RfD for these effects if possible. Wasserman et al. (2014), Wasserman et al. (2004), and other studies that EPA may identify as appropriate should be considered for RfD development.
- If a BMD/BMDL cannot be developed from the Wasserman (2014) data, EPA should consider development of an RfD based on this LOAEL, or, at a minimum, comparison of this LOAEL to the PODs for other non-cancer effects to determine whether neurocognitive effects are more sensitive.
- If it is determined that neurodevelopmental effects are a sensitive endpoint but the RfD for these effects is too uncertain for use as the final RfD, EPA should consider application of a database uncertainty factor (UF_D) to protect for potentially more sensitive neurodevelopmental effects.
- The more recently published literature that identifies neurodevelopmental health impairments associated with iAs exposure should be evaluated.

b. To estimate candidate values for DCS and diabetes, the meta-regression pooled slope and upper confidence limit were used to calculate a 5% response level BMD_{05} and $BMDL_{05}$, respectively. Please comment on whether this approach and the organ-specific candidate values below are scientifically justified and clearly described.

Regarding the BMD calculations that are relevant to both DCS and diabetes, the EPA needs to clarify and provide more support for the methods used to calculate the BMD and BMDL. BMD and BMDL estimates were obtained using the following equations:

$$BMD = \frac{\ln\left({odds\ at\ P(d)}/{odds\ at\ P(0)}\right)}{\beta\ mean}$$

$$BMDL = \frac{\ln\left({odds\ at\ P(d)}/{odds\ at\ P(0)}\right)}{95^{th} upper\ bound\ on\ \beta\ mean}$$

where P(d) and P(0) are the probabilities associated with 5% and 0% extra risk, respectively, and $\beta\ mean$ is the pooled slope from the Bayesian meta-regression.

The EPA should explain the basis of these equations in a manner that is understandable to scientists unfamiliar with the statistical approaches used in this assessment. Also, the terms "odds at P(d)" and "odds at P(0)" should be replaced with notation that more clearly represents the intended quantities, e.g., "Odds(d_{05})" and "Odds(d_0)" which represent the odds of disease at the dose corresponding to 5% (d_{05}) and the dose corresponding to 0% extra risk (d_0), respectively. The SAB believes it would be easier to understand the BMD and BMDL if the equation for the extra risk was presented in the same section. Additionally, it would be helpful to include an explanation of extra risk, such the one found in Section 2.2.2.2 of "EPA's Approach for Assessing the Risks Associated with Chronic Exposure to Carcinogens," which is posted at https://www.epa.gov/iris/epas-approach-assessing-risks-associated-chronic-exposure-carcinogens.

EPA should also refer to animal data to provide more plausibility for the BMDs. For instance, these data could provide more confidence in the conclusion that the BMD for diabetes is higher than the BMD for DCS. However, the SAB recognizes that the animal literature for diabetes is not as clear and definitive as that for DCS. It is evolving and has been improved by using the humanized As3MT mouse that metabolizes arsenic like humans. Thus, this is a consideration that the EPA should make for future assessments.

Future assessments should make better use of the posterior distribution of the pooled slope in estimating the BMD. The EPA can use the values of the slope sampled at each iteration of the Markov Chain Monte Carlo algorithm in the assessment to estimate a posterior distribution of the BMD. Although this will not change the estimated BMDL, it will provide a more complete picture of the uncertainty of the BMD.

i. For DCS, an organ-specific candidate $BMDL_{05}$ value of 0.094 µg/kg-day was derived based on increased CVD incidence.

The EPA should reconsider the endpoint used for DCS. The SAB believes that IHD incidence (both fatal and non-fatal events) or IHD mortality (fatal events) is preferable to CVD incidence as the basis for the RfD because IHD is a well-defined endpoint while CVD may include several different diseases and may not be consistently defined among studies (see response to charge question 3a). Additionally, as discussed in more detail in the response to charge question 9, it is not clear that incidence is preferable to fatalities as the basis for the RfD for DCS, and EPA should consider the benefits and limitations of using incidence versus fatalities. Selecting IHD mortality as the key outcome could be easily justified as

it is an important outcome of severity and is clearly identified in the population and in research studies. The number of studies available could play a role for selecting IHD incidence or IHD mortality as the key heart disease outcome to select for the dose-response meta-analysis and derivation of the $BMDL_{05}$. Also, the acronym DCS should be removed from this portion of the document because it causes confusion about what endpoint is being discussed.

EPA should also reconsider the choice of a BMR of 5% rather than 1% for IHD incidence (or IHD mortality) and CVD incidence. The draft document (p. 4-59, paragraph beginning on line 28) states that a $BMDL_{05}$ rather than a $BMDL_{01}$ was used because clinically diagnosed type II diabetes, CVD, and IHD are not frank effects that "warrant a lower BMR [than 5%] on the basis of severity." However, CVD and IHD incidence include both fatal and non-fatal occurrences of these diseases, and mortality is unquestionably a frank and severe effect, which could favor the focus on IHD mortality instead of IHD incidence. Additionally, part of EPA's rationale for concluding that these are not frank effects that warrant a lower BMR is that these diseases occur with high frequency in the U.S general population. However, the frequency of incidence and mortality for these diseases in the U.S. general population does not appear to be relevant to whether they are considered frank and severe effects.

Finally, the SAB notes that the $BMDL_{05}$ value of 0.140 µg/kg-day for IHD incidence reported in Tables C-81 and 4-22 differed from the value (0.128 µg/kg-day) reported in Table 4-17. This discrepancy should be corrected.

ii. For diabetes, an organ -specific candidate $BMDL_{05}$ value of 0.13 µg/kg-day was derived.

The SAB agrees with the approach used by the EPA to calculate the BMDL for diabetes, including an extra risk of 5%. However, a few inconsistencies and typos were identified by the SAB. The $BMDL_{05}$ value reported in Tables C-81 and 4-22 differed from the value reported in Tables 4-17 and 4-21. In Table 4-11, the posterior mean of the slope for the Grau-Perez et al. (2017) is equal to the 5^{th} percentile, which we assume is a typo. In addition, lines 9-11 of p. 4-45 contain the posterior mean and 5^{th} percentile of the variance of the prior distribution for the pooled slope (b_sigma), which are not included in Table 4-12 as suggested.

Recommendations:

Tier 1:

- Explain, in language accessible to a general scientific audience, the basis for the BMD and BMDL equations used for DCS and diabetes.
- Conduct analyses of both IHD mortality (fatal events only) and IHD incidence (both fatal and non-fatal events) instead of CVD incidence as the DCS endpoint for computation of the RfD.
- A BMR of 1% instead of 5% (i.e., $BMDL_{01}$ instead of $BMDL_{05}$) should be considered in light of the severity (substantial incidence of mortality) of IHD.

Tier 2:

- Revise the presentation of the odds in the BMD and BMDL equations to provide a better representation of the intended quantities, and which is more in line with standard epidemiological and biostatistical notation.

- Define extra risk (in words and via an equation) in the section containing the BMD and BMDL.
- Remove the DCS acronym.
- Review Tables C-81 and 4-17, 4-21, and 4-22 to ensure the correct BMD and BMDL values are reported.
- In Table 4-11, provide the correct posterior mean of the slope for the Grau-Perez et al. (2017) study.

Tier 3:

- Use animal data to support the plausibility of the estimated BMDs.
- Use the results from the Bayesian meta-regression to generate a complete posterior distribution for the BMD.

c. *For pregnancy outcomes, decreased birth weight was selected for benchmark dose modeling and the study-reported linear regression slope was used to estimate an organ-specific candidate $BMDL_{05}$ value of 0.23 µg/kg-day. Please comment on whether this approach is scientifically justified and clearly described.*

The SAB has several comments regarding the approach for candidate values for pregnancy; these comments included (a) the selection and reliance of a single study, and (b) the dose-response modeling of this study.

Selection and reliance of a single study (Kile et al 2016)

Regarding the use of a single-study approach (vs. a combined data approach as was used for other non-cancer outcomes), generally, the SAB understood the rationale provided as to why a single study approach was used and agreed that it was clearly explained that it was the only study which utilized a dose metric based on drinking water concentrations (expressed as µg/L). Other studies which used blood, urine, or drinking water exposure (expressed as µg/day) were excluded because no PBPK models for converting biomarker metrics were validated for pregnant women. While this rationale was generally acceptable from an analytical perspective (e.g., it was the only metric readily suitable for use in dose-response), it was not found to be sufficient for the purposes of risk assessment in the context of describing if or why the study was representative or protective for this outcome relative to the overall evidence base. The EPA identified 102 studies, of which 68 were high and medium confidence; it is not clear if the single study selected (Kile et al., 2016, a study of a Bangladeshi population) adequately represents the overall evidence base for this outcome, and specifically if it was the most sensitive. Often, the Agency utilizes dose-response arrays or similar techniques to ensure that the most sensitive study has been selected, but no such exercise appears to have been conducted for this outcome.

As has been discussed for other outcomes, use of human data is preferred for dose-response modeling. The human data for decreased birthweight are sufficient to support the conclusions for dose-response presented in the draft document. Specifically, as stated in the charge question, the point of departure ($BMDL_{05}$) for the RfD is based directly on the linear regression slope reported in Kile et al. (2016). However, information from other study types that characterize MOA, such as the studies

discussed on p. 3-96 – 3-97, could be considered to improve the plausibility, confidence, and protectiveness of the RfD.

The SAB also finds that the potential for threats to both internal and external validity had not been adequately assessed via critical appraisal, and, given such, the rationale for relying on this study requires additional explanation. Specifically, it was noted that there are differences in nutrition between the U.S. and Bangladeshi populations; it was not made clear in the assessment how this important factor could impact the findings reported in the study, as well as the generalizability of the results. Though the EPA has defined a benchmark response on the CDC Wonder database (https://wonder.cdc.gov/natality.html**),** they did not address its appropriateness, particularly given the differences in the distribution of iAs exposures between the two populations, nor did they address any potential sensitivities and uncertainties in using the CDC Wonder database as it related to impact on the dose-response analysis. Although the drinking water concentrations in the Bangladeshi population evaluated in Kile et al. (2016) and the U.S. population (Nigra et al., 2020) overlap, the proportion of individuals with high exposures was larger in Kile et al. (2016) than in the U.S. population (Nigra et al., 2020). Specifically, in Kile et al. (2016), the median drinking water iAs level was 2.3 µg/L, the 67th percentile was 18 µg/L, and the 75th percentile was 36 µg/L (Kile et al., 2016). In U.S. community water systems in 2009-10, the mean iAs concentration was 1.6 µg/L and iAs exceeded 10 µg/L in just 2.3% of systems. As these are important factors that could influence the dose-response analysis, the SAB finds it important to both acknowledge and assess these differences to enhance the transparency and reliability of the benchmark dose modeling for pregnancy outcomes.

It was also not clear why use of U.S.-based adjustments from the CDC Wonder database were acceptable for birthweight but were unacceptable for neurodevelopmental. Providing additional information on this apparent inconsistency would be helpful.

The EPA further describes the confidence in the POD for birth weight as being high (Table 4-16), but it does not provide an integrative rationale between the evidence judgement conclusion of 'evidence indicates' relative to the single study used in dose response. It is not clear how such a judgment can be determined without addressing the above aspects, as confidence should be based on the protectiveness of a value and not only its suitability based on dose-metric. Further, the lack of validated PBPK models and subsequent ability to assess and compare points of departure for pregnancy outcomes seems to be an important uncertainty associated with the assessment that warrants acknowledgement.

Dose-response modeling:

From the modeling perspective, the use of a single study to derive the BMDL is justified since that was the only study with data that reported drinking water exposure levels in µg/L. Studies which used blood, urine, or drinking water exposure (expressed as µg/day) were excluded because no PBPK models for converting biomarker metrics were validated for pregnant women. The SAB finds that the use of the hybrid approach that harmonizes binary and continuous outcome BMRs was reasonable, as was the use of a 5% BMR level instead of 1% since birthweight is not a severe developmental effect. Several comments were offered regarding the transparency and clarity:

- The EPA had to re-estimate the slope from the Kile et al. (2016) study to be in units of g decrease in birthweight per µg /L iAs in drinking water, with the methods explained on p. 4-52. It is unclear how a 95% CI was derived for the re-expressed slope.
- On p. 4-52, it is stated that 8.27% of live births fell below the public health definition of low birth weight (i.e., 2,500 g, or 5.5 lb), and on p. 4-53, it is stated that: "Therefore, given a background response and a BMR = 5% extra risk, the BMD would be the dose that results in 12.86% of the responses falling below the 2,500 g cut-off value." It is suggested that the equation used to derive the value of 12.86% from 8.27% of births being of low birth weight and a 5% extra risk of low birth rate be provided for those who are unfamiliar with these calculations.
- It is stated on p. 4-53 that: "Using this value for the mm term results in a BMDL value of 17.3 µg/L maternal serum concentration." It appears that this should be "drinking water concentration" rather than "maternal serum concentration."

Recommendations:

Tier 1:

- Provide rationale for the selection of the Kile et al. (2016) study beyond the suitability of the dose-metric, and specifically the adequacy this single study for determining a point of departure for this outcome. Such rationale should be provided both in the selection of candidate studies as well as carried through to the descriptions of confidence in the point of departure as well as uncertainties in the assessment.
- Provide a description and rationale for the inconsistency in use of U.S.-based comparisons to adjust/interpret data based on the Bangladeshi population for the pregnancy outcome (vs. neurodevelopmental).

Tier 2:

- Qualitatively address potential uncertainties in the dose-response assessment introduced by utilizing the CDC Wonder database.
- Clearly describe the process to re-express β and BMDL calculations (described in comments; p. 4-52, 4-53).

Tier 3:

- Consider using animal and/or mechanistic data to support and improve confidence and plausibility of the selected POD at low levels of iAs exposure when dose-response is based on observational studies in humans.
- Quantitatively assess the potential impact and sensitivities of residual/uncontrolled biases in observational data, as well as impact of the use of surrogates (e.g., application of U.S.-based standard to other geographies) or related aspects when deviations or amendments to author-reported data are required for dose-response analyses.

2.8. Charge Question 8. Uncertainty Factors

EPA has evaluated and applied where appropriate uncertainty factors to account for intraspecies variability (UF_H), interspecies differences (UF_A), database limitations (UF_D),

duration (UF_S), and LOAEL-to-NOAEL extrapolation (UF_L) for inorganic Arsenic.

General Comments:

Consideration of "uncertainty" applies to several topics throughout the document: dose conversions, systematic review steps, and others. While the SAB acknowledges that consideration of uncertainty in general is not the same as selection of uncertainty factors (UFs), the SAB recommends adding a separate section for overall uncertainty assessment to the document. Most of these areas of uncertainty are already discussed within the document, but compilation into a section that addresses "uncertainty" in the assessment would be helpful.

a. *Has uncertainty been adequately accounted for in the derivation of the toxicity values? Please describe and provide suggestions, if needed.*

The only place the UF application is described is in Table 4-18 which makes it difficult to find and evaluate. If the rationale for UF selection is endpoint or study specific, then more text is needed in the *RfD* derivation section of the document about the basis for UF selection to ensure clarity.

The SAB agrees that a UF_A, UF_S, or UF_L of >1 (i.e., adjustment for animal-to-human, subchronic-to-chronic, or LOAEL-to-NOAEL uncertainties) is not needed. Comments on the rationale for a UF_H of 3 for interindividual variability (sensitive subpopulations) are in the response to 8.b below.

A UF_D for database limitations of 1 (i.e., no adjustment) was used in the draft document. An adjustment for database limitations (i.e., $UF_D > 1$) is made when the available scientific information, or lack thereof, indicates that another effect may occur at lower doses resulting in a lower point of departure (POD) than the critical effect for the RfD (USEPA, 2002 https://www.epa.gov/sites/default/files/2014-12/documents/rfd-final.pdf**;** USEPA, 2022: Document Display | NEPIS | US EPA **(**https://ordspub.epa.gov/ords/eims/eimscomm.getfile?p_download_id=545991**)**

It is not known whether neurodevelopmental toxicity is a more sensitive endpoint than the critical effect used for the RfD in the draft document, and this needs to be more fully evaluated. This concern is described in detail under charge question7.*a.* This additional dose-response evaluation of neurodevelopmental effects recommended in the response to charge question 7.*a.*. will inform the need for application of a $UF_D > 1$.

Specifically, this dose-response assessment will indicate whether neurodevelopmental toxicity is the most sensitive effect. If developmental neurotoxicity is the most sensitive effect, the evaluation will also determine if an RfD based on this effect is appropriate for use as the final RfD. If it is determined that neurodevelopmental effects are the most sensitive effect, but an RfD for this effect is too uncertain for use as the primary RfD, an uncertainty factor >1 (e.g., 3) for database limitations (UF_D) should be considered.

The SAB applauds EPA for the use of dose-response meta-analyses. However, the novel use of sensitivity analyses (i.e., which studies impacted – or not – the analyses) requires more detailed discussion of how this sensitivity/variability is considered in selection of UF_H. Stronger scientific support for the choice of UF_H would be obtained by considering the full database of studies rather than

focusing on a single study. Studies identified through the systematic review process for hazard characterization and dose-response analysis potentially provide insight into human variability and selection of an appropriate value for this UF. In the future, additional meta regression analyses could address this more quantitatively. If EPA chooses to focus on individual studies, the study that is the focus should be clearly identified, and more explanation is needed about how "the studies that had the largest impact on the pooled slope" are used to inform UF_H and how this is related to the populations examined in those particular studies.

Recommendations:

Tier 1:

- Additional dose-response evaluation is recommended to assess whether or not neurodevelopmental toxicity is the most sensitive effect.
- The SAB recommends consideration of whether a $UF_D > 1$ is appropriate based on the dose-response evaluation of neurodevelopmental effects.

Tier 2:

- The novel use of sensitivity analyses requires more detailed discussion of how sensitivity/variability is considered in selection of UF_H.
- To the extent possible, better utilize the systematic review process and findings to support data-driven selection and application of uncertainty factors.

b. For DCS, diabetes, EPA applied a $UF_H = 3$ to account for potential interindividual differences in toxicokinetics and toxicodynamics related to iAs exposure in humans. EPA determined that a higher UF_H is not necessary given that studies that investigated non-cancer effects in sensitive subpopulations were included in the meta-regressions for CVD incidence, IHD incidence, and diabetes. For all three endpoints, the study that had the largest impact on the final pooled slope value was one that investigated effects in a sensitive subpopulation. Please comment on whether this approach is scientifically justified and clearly described.

For CVD, IHD and diabetes, the first sentence of the justification in Table 4-18[2] for using a UF_H of 3 instead of the default values of 10 is sufficient. The additional justifications in the table for application of $UF_H = 3$ are not clearly supportable, as described below. The overall recommendation is to remove these additional justifications, since they are not necessary and may not be scientifically supportable.

CVD and diabetes:

It is stated that the studies that had the largest impact were from an American Indian population that is 30-50% more likely than other populations in the U.S. to have these diseases. Although these diseases are more frequent in American Indians than other U.S. residents, it is not clear whether this

[2] "A UF_H of 3 is applied to account for potential interindividual differences in pharmacokinetics and pharmacodynamics relating to iAs exposure in humans. A higher UFH is not necessary for DCS and diabetes endpoints because the meta-regressions investigated a heterogeneous mix of multiple study populations, each of which included and adjusted for many sensitive subpopulations including smokers, sex, nutritional status, lifestage, genetic variability, and methylation capacity."

indicates that American Indians are more sensitive to the effects of iAs on development of CVD and diabetes than other populations. Thus, it is not clear that it is appropriate to reduce the UF for interindividual differences based on this rationale. As noted above, it is suggested that this discussion be removed. Citations are not provided in Table 4-18 for the studies of CVD and diabetes referred to in the sensitivity analyses. Although the recommendation is to delete this discussion, a general comment is that study citations for key statements such as this one should be provided throughout the document.

IHD incidence:

For IHD incidence, Table 4-18 stated that the study that had the most influence on the pooled slope had "a study population with a much higher water consumption rate than the average U.S. population (0.017 L/kg-day vs. 0.011 L/kg-day) and thus represents a sensitive subpopulation with respect to degree of exposure." The consumption rate for the study population is only 1.5 times higher than the default rate. Importantly, a more highly exposed population (due to greater water intake) is not a sensitive population, and this does not appear to be an appropriate rationale for reducing the UF_H to 3 from the default value of 10. Additionally, if study-specific water consumption data are available for a study in which drinking water iAs concentration is the exposure metric, the study-specific consumption rate should have been used to determine the dose in µg/kg-day.

As a general comment, citations for key studies and the basis for important conclusions such as this one should be clearly provided in the main document.

Recommendations:

Tier 1:

- Remove the justifications in Table 4-18 for use of a UF_H 3 rather than the default value of 10 that do not appear to be scientifically supportable.

Tier 2:

- Citations for key studies and the basis for important conclusions (such as the basis for water consumption of 0.017 L/kg-day and studies of CVD and diabetes referred to in the sensitivity analyses) should be provided throughout the document.
- For IHD incidence, the SAB recommends providing a different rationale for reducing the UF_H to 3 from the default value of 10 because a more highly exposed population by virtue of a greater water intake is not a sensitive population.

c. *For pregnancy outcomes, EPA applied a UF_H = 3 to account for potential interindividual differences in toxicokinetics and toxicodynamics related to iAs exposure in humans. EPA determined that a higher UF_H is not necessary as the Bangladeshi population that formed the basis of the POD for birth weight experiences low birth weight at a much greater rate than U.S. populations and represents a sensitive subpopulation. Please comment on whether this approach is scientifically justified and clearly described.*

While it was clear which UFs were applied, the rationale for selection and application could be better substantiated. The limited discussion in Tables 4-18 and 4-19 provides transparency in the numbers applied. However, given the extensive systematic review efforts, it was not clear as to why the EPA did not provide more support to "data-driven" selection of the UFs, with emphasis on describing the UFs selected for each outcome (vs addressing the UFs across outcomes as is generally done in Table 4-18).

For the pregnancy outcome, the EPA indicated that the UF_H of 3, which is lower than the default UF_H of 10, is sufficient to account for interindividual differences in the Bangladeshi population. EPA also indicated that a higher UF_H is not necessary because the Bangladeshi population that formed the basis of the POD is known to have major public health problem with low birth weight, citing notable differences in background prevalence. The assessment further stated, "Overall, a 3-fold UF is warranted to account for potential interindividual differences in pharmacokinetics and toxicodynamics within these sensitive subpopulations studied and the fact that a limited set of sensitive populations have been studied and may not represent the total spectrum of sensitive groups. However, when considering population variability, specifically within the context of the entire U.S. population, use of data from these sensitive groups of individuals largely predisposed to developing such effects (as compared to U.S. individuals) does not warrant a higher than 3-fold UF."

The SAB suggests EPA should provide additional rationale for conclusion to reduce the UFH to 3 from the default values of 10. It was not clear whether the higher prevalence of low birth weight in Bangladesh indicates that the Bangladeshi population is also more sensitive to the effects of iAs on birth weight than other populations, or what other factors are responsible for the high prevalence of low birth weight in the Bangladeshi population (noting that in this case, the lack of consideration of the impact of other risk modifiers or confounding factors would also impact the confidence in the dose-response and thus should be accounted for both in the modeling as well as in selection of uncertainty factors).

Further, the Agency indicates that a limited set of sensitive populations has been studied, and that they may not represent the total spectrum of sensitive groups. The EPA identified 102 studies for this outcome, including 68 high and medium studies. Given the availability of such data, the substantiation of the UF for this outcome may be better informed with support from the underlying evidence base already reviewed. As is discussed in response to charge question 7c (see Section 2.7), the selection of the Kile et al. (2016) study was not based on sensitivity, but rather availability of a suitable dose-metric, and thus it was not clear if the candidate study and POD were protective or representative of the outcome overall.

When these aspects are considered collectively, it was not completely clear whether it is appropriate to reduce the UF for interindividual differences based on this rationale. Additional discussion and substantiation are needed to clarify both the sensitivity as well as the generalizability of the study with respect to interindividual variability in toxicokinetics and toxicodynamics.

Recommendations:

Tier 2:

- It is recommended that EPA further substantiate the selected UF_H for the pregnancy outcome – and specifically provide rationale for departing from the default value of 10 - using data from

the evidence base reviewed, as well as clarifying the existing rationale around sensitivity of the Bangladeshi population to the effects of iAs (vs. other factors) and of the specific study selected from the 102 screened studies.

Tier 3:

- Better utilize the systematic review process and findings to support data-driven selection and application of uncertainty factors.
- Because guidance around UFs was generally drafted prior to the use of observational data in developing toxicity reference values, it would be helpful for the Agency to consider refining the recommendations for choice of UFs to be more aligned with describing and characterizing uncertainties in using observational data, including guidance for determining when populations are considered sensitive and/or representative of toxicokinetic and toxicodynamic variability.

2.9. Charge Question 9. RfD

*From the identified human health effects of iAs and the derived organ-specific toxicity values for diabetes, DCS effects and pregnancy outcomes, an RfD of **0.031 µg/kg-day based on increased CVD incidence in humans** was selected. This RfD is expected to be protective against all noncancer adverse health effects associated with iAs and across all lifestages. Please comment on whether the selected RfD is scientifically justified and clearly described.*

The RfDs presented in the draft document are based on human epidemiology studies, including studies with low exposures in the background range. The endpoints considered for RfD development were CVD incidence, fatal CVD, IHD incidence, and fatal IHD (all four of which are categorized as DCS), as well as diabetes, birth weight, and neurodevelopmental effects. Of these endpoints, RfDs were developed for CVD incidence (0.031 µg/kg-day; rated as high confidence), IHD incidence (0.043 µg/kg-day; no confidence rating provided), diabetes (0.042 µg/kg-day; rated as high confidence), and birth weight (0.077 µg/kg-day; rated as medium-low confidence). The RfD of 0.031 µg/kg-day for CVD incidence was selected as the overall RfD because it was the lowest and was thus assumed to be protective of the other endpoints.

While the methodology used to develop the proposed RfD based on CVD incidence in humans was clearly described, many of the details needed to understand the methodology are difficult to understand because they are found only in other documents or in footnotes. The section on RfD derivation in the main document should clearly present all information needed to understand the methodology used to develop the RfD.

Regarding scientific justification of the proposed RfD, the overall RfD based on increased CVD incidence should be reconsidered for several reasons, which fall into two general categories discussed separately below. First, it is unclear that an RfD developed using traditional methods (i.e., estimation of a POD and application of uncertainty factors) is scientifically supportable and health protective for non-cancer effects with the dose-response relationship identified in the draft document. Additionally, it is unclear whether the health endpoint, point of departure, and uncertainty factors selected for development of the RfD are scientifically supportable and sufficiently health protective.

Is development of an RfD for inorganic arsenic scientifically supportable?

The definition of an RfD in EPA guidance is: "an estimate (with uncertainty spanning perhaps an order of magnitude) of a daily oral exposure to the human population (including sensitive subgroups) that is likely to be without appreciable risk of deleterious effects during a lifetime" (USEPA, 2002 https://www.epa.gov/sites/default/files/2014-12/documents/rfd-final.pdf).

Although EPA (2002) defines an RfD as the dose that is not expected to cause an "appreciable" risk, it is widely understood that an RfD is set below a threshold dose below which no deleterious effects are expected. Relevant to this point, IRIS concluded that it is not appropriate to develop an RfD for neurodevelopmental effects of lead because no threshold for these effects has been identified.

The assessment of non-cancer effects in the draft IRIS iAs document differs from most other IRIS non-cancer assessments (except for lead, as mentioned above) in that the results of the dose-response meta-analysis indicate a risk for several non-cancer endpoints (CVD, IHD, diabetes) from any daily dose of iAs, including daily doses below the proposed RfD. These analyses harmonize the approaches used for cancer and non-cancer effects, in accordance with previous recommendations to EPA from several advisory groups. Consistent evaluation of the several noncancer and cancer endpoints occurring with iAs exposure is scientifically appropriate based upon the results of the dose-response analysis. The inability to identify a threshold for DCS (IHD and CVD) and diabetes caused by iAs exposure suggests that development of an RfD for iAs may not be appropriate, like the conclusion mentioned above for lead.

Development of an RfD for an effect for which no threshold is identified requires implicitly or explicitly establishing an acceptable risk level (i.e., a risk level that is not "appreciable"). Such a judgment is fundamentally not determined by health science and therefore is beyond the scope of the IRIS program, absent clear direction from others in the Agency and Federal Government. While the risks at the proposed RfD (0.031 μg/kg-day) are not presented in the assessment, health risks for adverse DCS (IHD and CVD) and diabetes are provided at the very similar estimated average background level (0.0365 μg/kg-day) and can easily be calculated at other doses including the RfD. If an RfD is established, the risks of exposure at the RfD for each health effect for which a dose-response meta-analysis was performed should be clearly presented. Given the definition of the RfD from EPA (2002), those risks can be assumed to represent an absence of appreciable risk or an acceptable risk level.

An option for assessing noncancer health effects as an alternative to an RfD is to utilize the fitted equations currently shown only in Figures (e.g., Fig 4-11 - IHD incidence, Fig 4-18 - diabetes which are copied below) as they allow calculation of average or upper bound risk estimates across a range of doses.

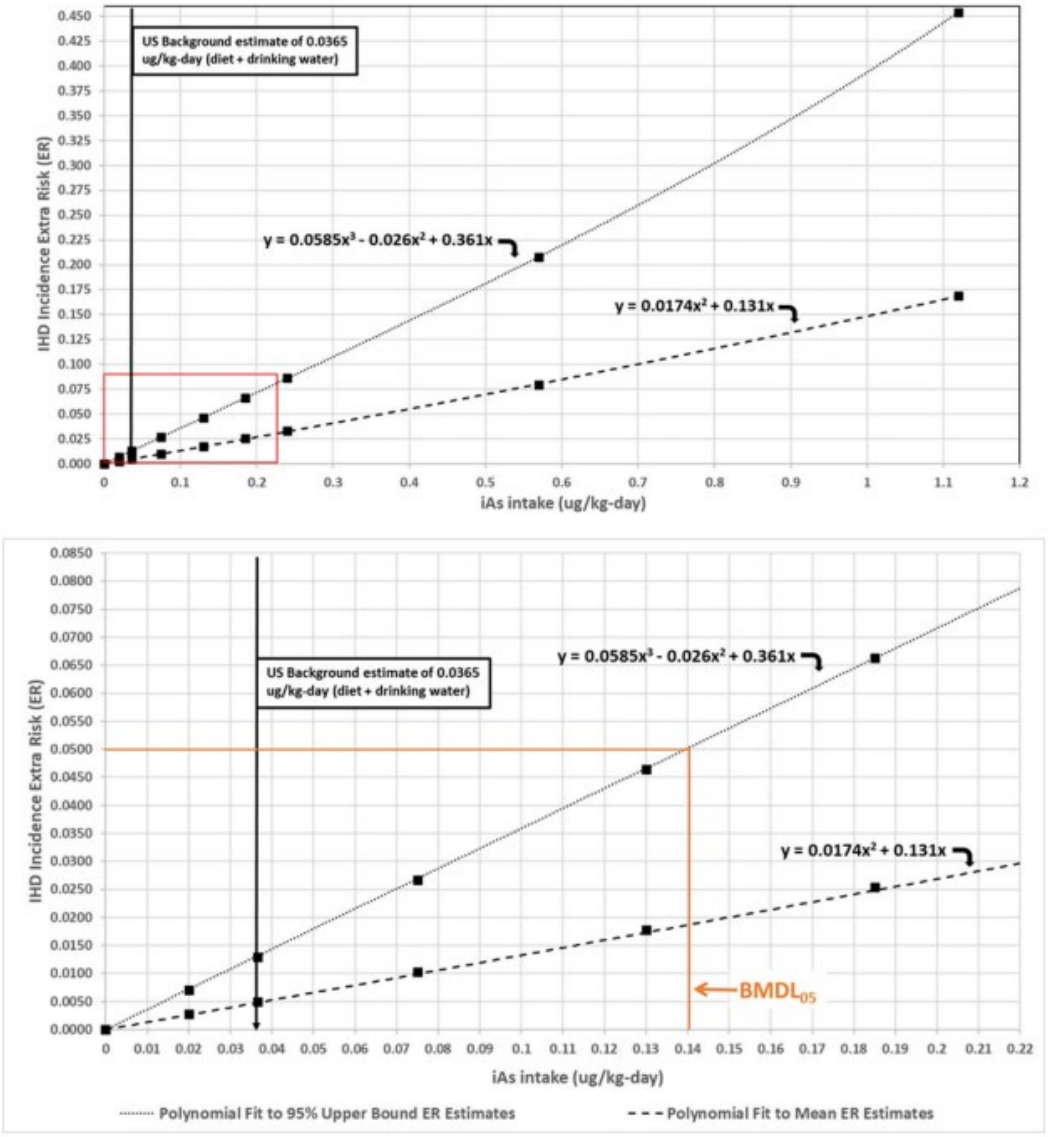

Figure 4-11. U.S. IHD incidence (all studies) lifetime extra risk versus µg/kg-d MLE doses for all doses (top plot) and low doses (bottom plot). See Section 4.3.4 for discussion of 0.0365 ug/kg-day U.S. background dose estimate.

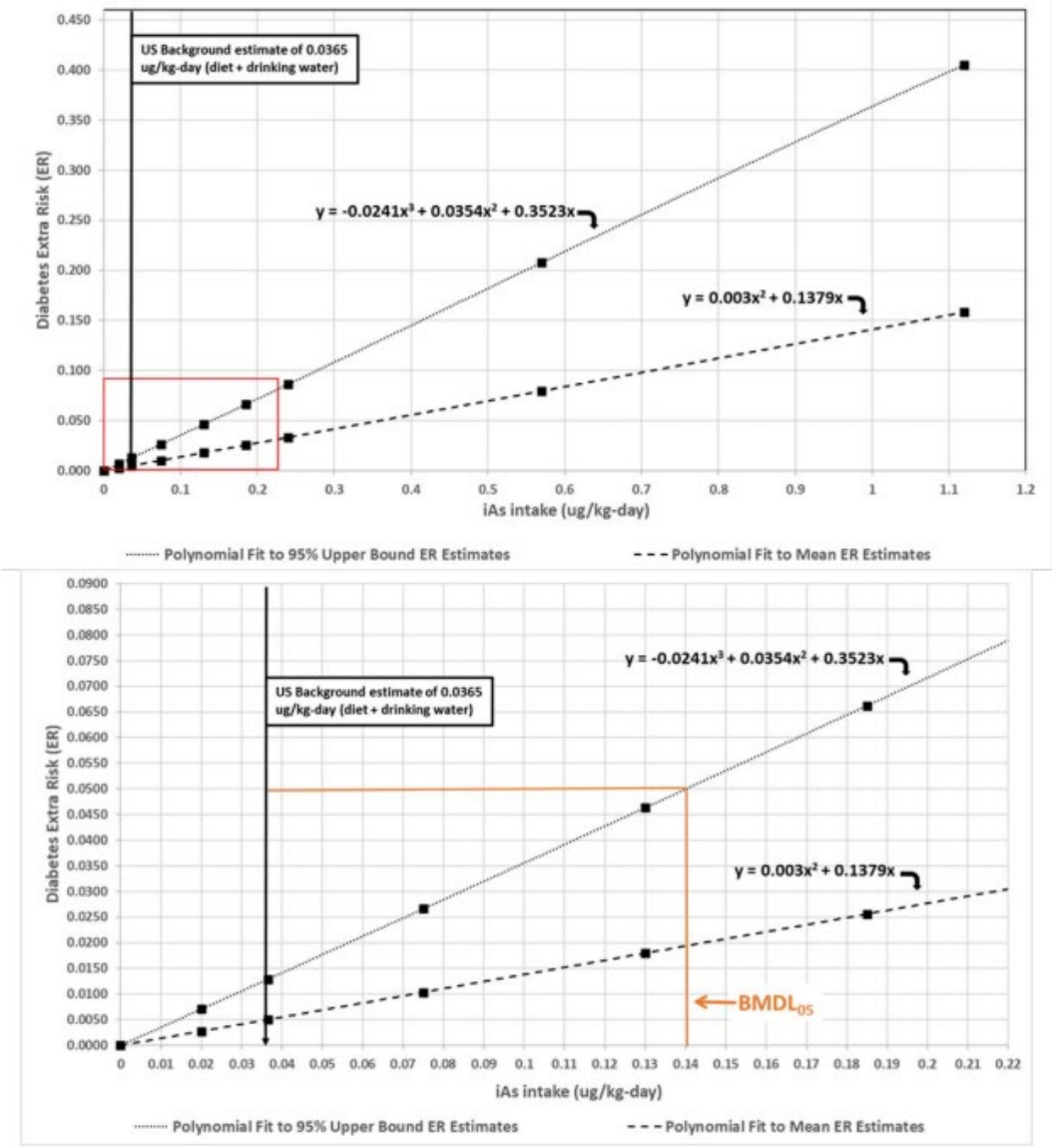

Figure 4-18. U.S. diabetes (all studies) lifetime extra risk versus µg/kg-d MLE doses for all doses (top plot) and low doses (bottom plot). See Section 4.3.4 for discussion of 0.0365 µg/kg-day U.S. background dose estimate.

Another option is to present a slope factor that relates dose and risk, analogous to a CSF, based on the results of the dose-response meta-analysis. Since the fitted equations presented in the figures are polynomials, it would be necessary to determine the dose below which a linear slope factor could be applied. It is recognized that such risk-based approaches for non-cancer effects would require that EPA make new science-policy decisions, including identification of a target risk level for non-cancer effects. This would necessitate revising the approaches that have traditionally been used for assessment of non-cancer effects based on a threshold assumption.

The discussion of how background exposure was considered in the analyses presented in the draft document should be expanded, including a clear presentation of the information needed for application of the assessment's conclusions.

Calculations can also be done for the two cancer endpoints (Fig 4-7 - bladder cancer, Fig 4-9 - lung cancer – copied from the draft toxicological review of iAs and placed below) allowing direct comparison of the risks of several major health effects of iAs. These fitted polynomial equations are nonlinear at higher doses. It is noted that the analyses presented in the draft IRIS documents indicate that the risks of non-cancer effects of iAs are 3- to10-fold higher than the risk of cancer at the same iAs dose. It is important to recognize that serious non-cancer health effects (e.g., CVD, IHD, diabetes) that result from low-dose exposure to iAs are of concern, not just cancer.

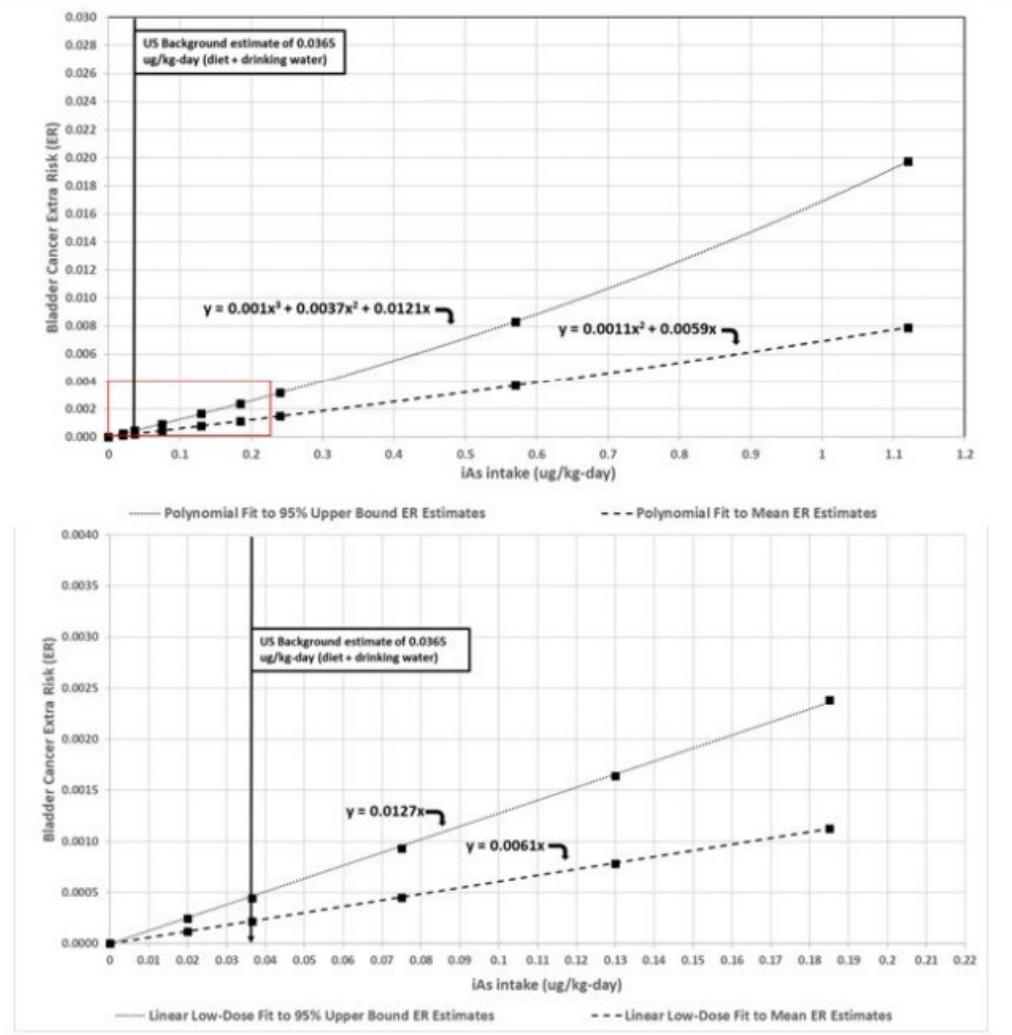

Figure 4-7. U.S. bladder cancer lifetime extra risk versus µg/kg-d iAs doses for all doses (top plot) and low doses (bottom plot). The polynomial equations can be used to approximate high dose extra risk. The linear equations can be used to approximate low dose (< 0.22 µg/kg-day) extra risk. The linear slope of 1.27×10^{-2} for the 95% upper bound is analogous to an EPA cancer slope factor (CSF). See Section 4.3.4 for discussion of 0.0365 µg/kg-day U.S. background dose estimate.

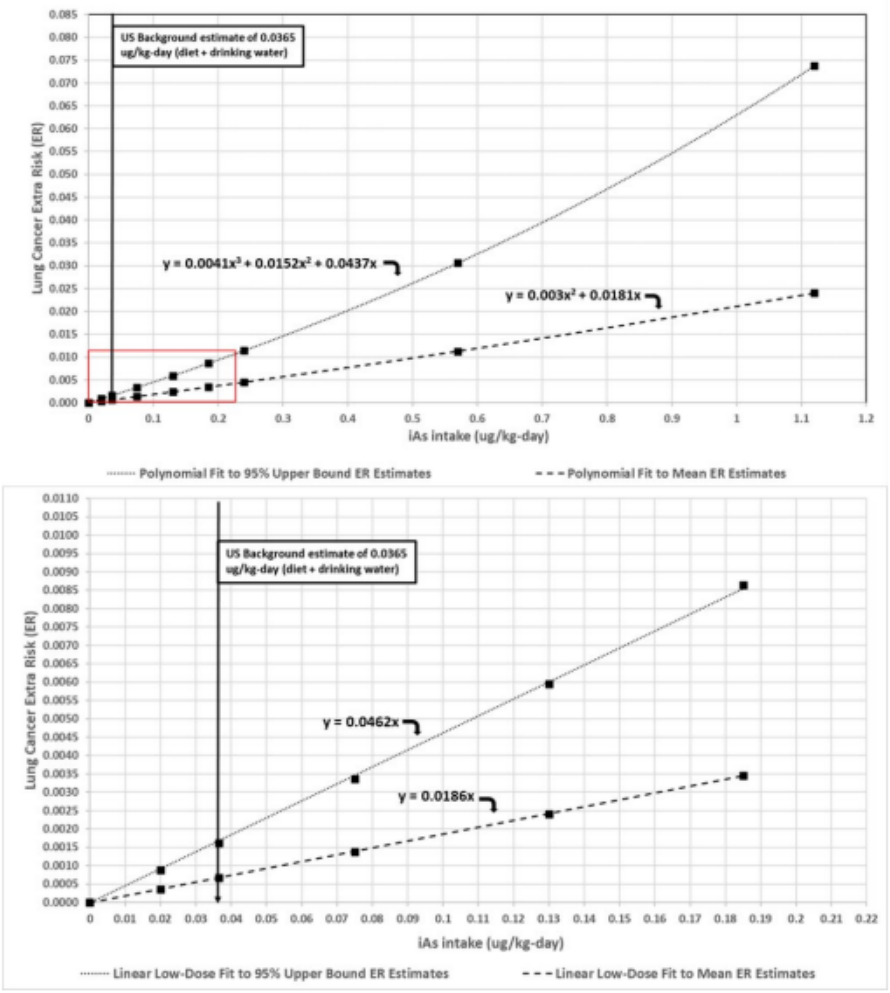

Figure 4-9. U.S. lung cancer lifetime extra risk versus µg/kg-d iAs doses for all doses (top plot) and low doses (bottom plot). The polynomial equations can be used to approximate high dose extra risk. The linear equations can be used to approximate low dose (< 0.22 µg/kg-day) extra risk. The linear slope of 4.62×10^{-2} for the 95% upper bound is analogous to an EPA cancer slope factor (CSF). See Section 4.3.4 for discussion of 0.0365 µg/kg-day U.S. background dose estimate.

In summary, it is important that the IRIS iAs document clearly presents the approach used for assessment of risks of non-cancer effects, including examples of calculations of risk levels. If EPA chooses to establish an RfD, it is highly recommended that the IRIS document discusses the important concepts and their implications reviewed above.

Is the proposed RfD based upon an appropriate health endpoint, point of departure, and uncertainty factors?

Consideration of neurodevelopmental effects in RfD development

The proposed overall RfD is based on the RfD for CVD incidence, which was lower than the RfDs that were derived for other endpoints (IHD incidence, diabetes, and decreased birth weight), with the intention to be protective for all these endpoints. As noted in the response to charge question 7, it may be feasible to perform dose-response for neurodevelopmental effects to determine whether it is a more sensitive endpoint that *is* sufficiently strongly supported to be considered as the basis of the overall RfD. Alternatively, as discussed in the response to charge question 8, consideration of a database uncertainty factor greater than 1 is recommended if quantitative analysis of neurodevelopmental effects is not feasible.

Consideration of IHD rather than CVD in RfD development

The responses to charge question 4a and 7b provide a strong recommendation that RfD development for DCS focus on IHD rather than CVD. As noted in the responses to the other charge questions, IHD represents a more homogeneous disease process and endpoint in contrast to the diversity of effects included in CVD. Thus, IHD, rather than CVD, is the preferred endpoint for evaluating DCS.

Consideration of mortality as well as incidence in RfD development

Although both incidence and mortality data are available from studies of iAs and IHD and CVD, PODs and RfDs were developed for incidence but not mortality for these endpoints. The rationale provided for not developing RfDs based on mortality (Table 4-16) is that "according to EPA's *A Review of the Reference Dose and Reference Concentration Processes* (USEPA, 2002), studies investigating mortality endpoints are not preferred for reference value derivation," which appears to be based on USEPA (2002) stating that "a study showing only effect levels for mortality or other extremely severe toxicity would not be sufficient to set a reference value." However, it appears likely that USEPA (2002) intended this statement to refer to death or very severe toxicity observed in animal toxicology studies, not to increased risks of fatal disease in humans caused by chronic exposure to environmental contaminants. This is especially true because incidence data for IHD and CVD, which was used as the basis for the RfDs for these endpoints, includes both fatal and non-fatal occurrences of the disease.

As noted in the response to charge question 7.b.i, it is more technically challenging to accurately assess incidence than mortality. For IHD, there are four studies of incidence and four for fatal IHD. EPA should consider the strengths and limitations of the database supporting dose-response meta-analysis to determine whether incidence or mortality is more appropriate as the basis for the POD and RfD. In contrast to IHD, CVD has a larger database of mortality studies, five, with only two for incidence.

Choice of BMR for RfD development

The PODs for the noncancer endpoints in the draft IRIS documents were derived as the $BMDL_{05}$ for each endpoint. As shown on Figures 4-10 (CVD), 4-11 (IHD) and 4-18 (diabetes), the extra risk at these PODs is 5% (500 per 10,000 people) based on the upper bound estimates, and a substantial fraction of that risk for cardiovascular disease represents fatal disease outcomes. The draft IRIS document (p. 4-59 – 4-60) states that a $BMDL_{05}$ was used instead of a $BMDL_{01}$ because CVD, IHD, and diabetes are not "frank effects and do not support a lower BMR on the basis of severity" and that, for this reason, the fact that "BMD_{01} estimates are well below the low end of observable dose range, whereas BMD_{05} estimates are well within, but still close to the low end of the observable range" becomes more important in selecting the BMR. However, as discussed in the response to charge question 7, the SAB does not agree that these are not frank effects and concludes that the choice of the $BMDL_{05}$ rather

than the $BMDL_{01}$ should be reconsidered in light of the severity of these effects and the disease burden that it represents in the U.S. population. It should be clarified that the acceptable increase in risk from exposure to an environmental contaminant should not be greater for a health effect with a high background rate than for a health effect with a low background rate.

Discussion of uncertainty

Additional discussion of the qualitative and quantitative uncertainties associated with the RfD is needed (in addition to the characterization of confidence that is already included), consistent both with standard practice in risk assessment and the IRIS handbook (USEPA, 2022). This discussion would be useful to risk managers, particularly regarding cost-benefit analysis. Specifically, while the assessment of uncertainty due to confounding is discussed in Section C.1.1, it is not clear how uncertainties related to factors such as systematic bias, residual bias, and uncontrolled confounding were considered for each outcome.

Recommendations:

Tier 1:

- The IRIS document should clearly present the issues relevant to the assessment of risks of non-cancer effects, including examples of calculations of risk levels for non-cancer endpoints at various iAs doses.
- Risk levels for non-cancer and cancer effects at the same dose of iAs should be presented and compared, with recognition that both types of health effects from low-dose exposure to iAs are of concern.
- EPA should reconsider whether setting an RfD for iAs is scientifically supportable, and a detailed discussion of this issue should be presented in the IRIS document. This discussion should address the following issues:
    - Whether an RfD is appropriate for a contaminant with non-cancer effects that exhibit a low-dose linear (non-threshold) dose-response.
    - The need for a science-policy decision on a risk level that does not represent appreciable risk if an RfD for iAs is developed.
    - Consideration of an alternative approach for assessment of non-cancer effects of iAs based on the polynomial equations or a slope factor that relates risk and dose, with recognition of the need to select a target risk level for non-cancer effects if this approach is used.
- Among the other outcomes discussed in the previous sections, use of neurodevelopmental effects in RfD development should be reconsidered. The feasibility of performing dose-response evaluation and identifying a POD for neurodevelopmental effects should be evaluated. If it is determined that it is not feasible to develop a POD for neurodevelopmental effects, application of a database UF > 1 to account for potentially more sensitive neurodevelopmental effects should be considered.
- Consistent with the IRIS handbook, an uncertainty section should be added to the document summarizing uncertainties in the non-cancer assessment and resulting reference values.

Tier 2:

- The discussion of how background exposure was considered in the analyses presented in the draft document should be expanded, including a clear presentation of the information needed for application of the assessment's conclusions.
- The following recommendations apply if a decision is made to develop an RfD for arsenic:
    - The section on RfD derivation in the main document should clearly present all information needed to understand the methodology used to develop the RfD.
    - The predicted risk at the RfD should be presented for each non-cancer effect for which no threshold is identified, and it should be indicated that this risk level is not considered to represent an appreciable risk.
    - EPA should evaluate whether fatal occurrences of IHD and/or incidence of IHD (including both fatal and non-fatal occurrences) is most appropriate as the basis for the RfD. The document should include a discussion of qualitative and quantitative uncertainties associated with the RfD including how uncertainties related to factors such as systematic bias, residual bias, and uncontrolled confounding were considered for each outcome.

REFERENCES

Abuawad, A.K., Bozack, A.K., Navas-Acien, A., Goldsmith, J., Liu, X., Hall, M.N., Ilievski, V., Lomax-Luu, A.M., Parvez, F., Shahriar, H., Uddin, M.N., Islam, T., Graziano, J.H.,and Gamble, M.V. (2023). The Folic Acid and Creatine Trial: Treatment Effects of Supplementation on Arsenic Methylation Indices and Metabolite Concentrations in Blood in a Bangladeshi Population. *Environ Health Perspect*. 131(3):37015.

Abuawad, A, Goldsmith, J., Herbstman, J.B., Parvez, F**.,** Islam, T., LoIacono, N., Graziano, J.H., Navas-Acien, A., and Gamble, M.V. (2022). Urine Dilution Correction Methods Utilizing Urine Creatinine or Specific Gravity in Arsenic Analyses: Comparisons to Blood and Water Arsenic in the FACT and FOX Studies in Bangladesh. *Water* 14(9):1477

Allen, B; Shao, K; Hobbie, K; Mendez, W, Jr; Lee, J**.,** Cote, I., Druwe, I; Gift, J; Davis, J.A. (2020a). Systematic dose-response of environmental epidemiologic studies: Dose and response pre-analysis. *Environ Int* 142: 105810. http://dx.doi.org/10.1016/j.envint.2020.105810

Allen, B., Shao, K., Hobbie, K., Mendez, W., Lee, J.S., Cote, I., Druwe, I., Gift, J., Davis, J. A. (2020b). Bayesian hierarchial dose-response meta-analysis of epidemiological studies: Modeling and target population prediction methods. *Environ Int* 145: 10611. http://dx.doi.org/10.1016/j.envint.2020.106111**.**

Avanasi, R., Shin, H.M., Vieira, V.M., & Bartell, S.M. (2016a). Impacts of geocoding uncertainty on reconstructed PFOA exposures and their epidemiological association with preeclampsia. *Environmental Research*, 151, 505–512. 10.1016/j.envres.2016.08.019

Avanasi, R., Shin, H.M., Vieira, V.M., & Bartell, S.M. (2016b). Variability and epistemic uncertainty in water ingestion rates and pharmacokinetic parameters, and impact on the association between perfluorooctanoate and preeclampsia in the C8 Health Project population*. Environmental Research*, 146, 299–307. 10.1016/j.envres.2016.01.011

Bulka, C.M., Scannell, B.M., Lombard, M.A. et al. (2022). Arsenic in private well water and birth outcomes in the United States. *Environ Int* 163: 107176.

Baris, D., Waddell, R., Beane Freeman, L.E. et al. (2016). Elevated bladder cancer in northern New England: The role of drinking water and arsenic. *J. Natl Cancer Inst* 108: A3.

Chen, Y., Wu, F., Liu, M. et al. (2013b). A prospective study of arsenic exposure, arsenic methylation capacity, and risk of cardiovascular disease in Bangladesh. *Environ Health Perspect.* 121: 832-838.

Crump, K.S. and Allen, B. (2011). Toward making epidemiologic data more useful in quantitative risk assessment. *The Open Epidemiology Journal* 4: 30-44.

De la Ossa, C.A., Ramfrez-Giraldo, A.F., Arroyo-Alvis, K. et al. (2023). Neuropsychological effects and cognitive deficits associated with exposure to mercury and arsenic in children and adolescents of the Mojana region, Colombia. *Environmental Research* 216: 114467.

De Water, E, Curtin, P., Jennings, C. et al. (2022). Prenatal metal mixture concentrations and reward motivation in children. *Neurotoxicology* 88:124.

Dheer, R., Patterson, J., Dudash, M., Stachler, E.N., Bibby, K.J., Stolz, D.B., Shiva, S., Wang, Z., Hazen, S.L., Barchowsky, A., and Stolz, J.F. (2015). Arsenic induces structural and compositional colonic microbiome change and promotes host nitrogen and amino acid metabolism. *Toxicol Applied Pharmacol* 289: 397

Doherty, B.T., Romano, M.E., Gui, J. et al. (2020). Periconceptional and prenatal exposure to metal mixtures in relation to behavioral development at 3 years of age. *Environmental Epidemiology* 4(4) p e0106 *DOI:* 10.1097/EE9.0000000000000106

El-Masri, HA; Kenyon, EM. (2008). Development of a human physiologically based pharmacokinetic (PBPK) model for inorganic arsenic and its mono- and di-methylated metabolites. *J Pharmacokinet Pharmacodyn* 35: 31-68. http://dx.doi.org/10.1007/s10928-007-9075-z.

El-Masri, H.A., Hong, T., Henning,C. et al. (2018a). Erratum: "Evaluation of a physiologically based pharmacokinetic (PBPK) model for inorganic arsenic exposure using data from two diverse human populations" [Erratum]. *Environ Health Perspect*.126: 99002.

El-Masri, H.A., Hong, T., Henning, C. et al. (2018b). Evaluation of a physiologically based pharmacokinetic (PBPK) model for inorganic arsenic exposure using data from two diverse human populations. *Environ Health Perspect.*126*:* 077004.

Freire, C., Amaya, E., Gil, F. et al. (2018). Prenatal co-exposure to neurotoxic metals and neurodevelopment in pre-school children: The Environment and Childhood (INMA) Project. *Science and the Total Environment* 621: 340.

Freire, C., Amaya, E., Gil, F. et al. (2019). Placental metal concentrations and birth outcomes_ The Environment and Childhood (INMA) project*. International J. of Hygiene and Environ Health Perspect* 222: 468.

Gilbert-Diamond, D., Cottingham, K.L., Grubber, J.F. et al. (2011). Rice consumption contributes to arsenic exposure in US women. *Proc Natl Acad Sci* 108(51):20656-20660

Gilbert-Diamond, D., Emond, J.A., Baker, E.R., Korrick, S.A., Karagas, M.R. (2016). Relation between in Utero Arsenic Exposure and Birth Outcomes in a Cohort of Mothers and Their Newborns from New Hampshire. *Environ Health Perspect* 124: 1299-1307. http://dx.doi.org/10.1289/ehp.1510065

Glassmeyer, S.T., Burns, E.E., Focazio, M.J. et al. (2023). Water, Water, Everywhere, but Every Drop Unique: Challenges in the Science to Understand the Role of Contaminants of Emerging Concern in the Management of Drinking Water Supplies. *GeoHealth*, 7(12): e2022GH000716.

Grau-Perez, M., Kuo, C.C., Gribble, M.O. et al. (2017). Association of low-moderate arsenic exposure and arsenic metabolism with incident diabetes and insulin resistance in the strong heart family study. *Environ. Health Perspect* 125:127004

Hays, A.M., Clark Lantz, R., Rodgers, L.S., Sollome, J.S., Vaillancourt, R.R., Andrew, A.S., Hamilton, J.W., and Camenish, C.D. (2008). Arsenic-induced decreases in the vascular matrix. *Toxicol Pathol* 36(6): 805.

Herrera, A., Pineda, J., Antonio, M.T. (2013). Toxic effects of perinatal arsenic exposure on the brain of developing rats and the beneficial role of natural antioxidants. *Environ Toxicol Pharmacol* 36: 73-79.

Hoover, J.H., Coker, E.S., Erdei, E., Luo, L., Begay, D., MacKenzie, D. and J. Lewis. (2023). Preterm Birth and Metal Mixture Exposure among Pregnant Women from the Navajo Birth Cohort Study. *Environ Health Perspect* 131(12):127014. doi: 10.1289/EHP10361. Epub 2023 Dec 18. PMID: 38109118; PMCID: PMC10727039.

Howe, C.G., Farzan, S.F., Garcia, E., Jursa, T., Iyer, R., Berhane, K., Chavez, T.A., Hodes, T.L., Grubbs, B.H., Funk, W.E., Smith, D.R., Bastain, T.M., Breton, C.V. (2020). Arsenic and birth outcomes in a predominately lower income Hispanic pregnancy cohort in Los Angeles*. Environ Res* 184: 109294. http://dx.doi.org/10.1016/j.envres.2020.109294

Howe, C.G., Claus Henn, B, Farzan S.F., Habre, R, Eckel, S.P., Grubbs, B.H., Chavez, T.A., Faham, D., Al-Marayati, L., Lerner, D., Quimby, A., Twogood, S., Richards, M.J., Meeker, J.D., Bastain, T.M., Breton, C.V. (2021). Prenatal metal mixtures and fetal size in mid-pregnancy in the MADRES study. *Environ Res*:196:110388. doi: 10.1016/j.envres.2020.110388. Epub 2020 Oct 28. PMID: 33129852; PMCID: PMC8079562.

Huang, H, Wei, L., Chen, X., Zhang, R., Su, L. et al. (2021). Cord serum elementomics profiling of 56 elements depicts risk of preterm birth: Evidence from a prospective birth cohort in rural Bangladesh. *Environ Int*. 156:106731. doi: 10.1016/j.envint.2021.106731. Epub 2021 Jun 28. PMID: 34197971.

Jiang, E.X., Domingo-Relloso, A., Abuawad, A. et al. (2023). Arsenic exposure and epigenetic aging: The association with cardiovascular disease and all-cause mortality in the strong-heart study. *Environ Health Perspectives* 131(12). https://doi.org/10.1289/EHP11981**.**

Karagas, M.R., Stukel, T.A. and Tosteson, T.D. (2002). Assessment of Cancer risk and environmental levels of arsenic in New Hampshire. *Int. J. Hyg. Env. Health* 205(1-2):85-94

Kile, M.L., Cardenas, A., Rodrigues, E. et al. (2016). Estimating effects of arsenic exposure during pregnancy on perinatal outcomes in a Bangladeshi cohort. *Epidemiology* 29: 173.

Kuo, C.C., Howard, B.V., Umans, J.G. et al. (2015). Arsenic exposure, arsenic metabolism, and incident diabetes in the strong heart study. *Diabetes Care* 38: 620.

Laue, H.E., Coker, M.O., Madan, J.C. (2022). The developing microbiome from birth to 3 years: the gut–brain axis and neurodevelopmental outcomes. *Front Pediatr* 10:815885

Laue, H., Moroishi, Y., Jackson, B.P. et al. (2023). Bacterial modification of the association between arsenic and austism-related social behavior scores. *Exposure and Health* 15: 343

Leening, M.J., Ferket, B.S., Steverberg, E.W. et al. (2014). Sex differences in lifetime risk and first manifestation of cardiovascular disease: Prospective population based cohort study. *BMJ* 349: g5992.

Lewis, D.R., Southwick, J.W., Ouellet-Hellstrom, R., Rench, J., Calderon, R.L. (1999). Drinking water arsenic in Utah: A cohort mortality study. *Environ Health Perspect* 107: 359-365. http://dx.doi.org/10.2307/3434539

Lewis, J.V., Knapp, E.A., Bakre, S. et al. (2023). Associations between area-level arsenic exposure and adverse birth outcomes: An Echo-wide cohort analysis. *Environmental Research* 236:116772.

Luo, Y.W., McCullough, L.E., Tzeng, L.E. et al. (2017). Maternal blood cadmium, lead and arsenic levels, nutrient combinations, and offspring birthweight. *BMC Public Health* 17: 354.

Makhani, K.C., Chiavatti, C., Plourde, D., Silva, L.F.N., Lemaire, M., Lemarie, C.A., Lehoux, S. and K.K. Mann. (2018). Using the Apolipoprotein E Knock-Out Mouse Model to define atherosclerotic plaque changes induced by low dose arsenic. *Toxicological Science* 166(1): 213.

Meliker, J.R., Slotnick, M.J., Avruskin, G.A. et al. (2010). Lifetime exposure to arsenic in drinking water and bladder cancer: A population-based case-control study in Michigan, USA. *Cancer Causes Control* 21: 745-757.

Monrad, M, Erboll, A.K., Sorensen, M. et al. (2017). Low-level arsenic in drinking water and risk of incident myocardial infarction: A cohort study. *Environ Res* 154: 318-324.

Moon, K.A., Guallar, E., Umans, J.G. et al. (2013). Association between exposure to low to moderate arsenic levels and incident cardiovascular disease: A prospective cohort study. *Ann Intern Med* 159: 649-659.

NASEM (National Academies of Sciences, Engineering, and Medicine). (2019). Review of EPA's updated problem formulation and protocol for the inorganic arsenic IRIS assessment. Washington, DC: The National Academies Press. http://dx.doi.org/10.17226/25558.

Nigra, A.E., Chen, Q., Chillrud, S.N. et al. (2020). Inequalities in public water arsenic concentrations in counties and community water systems across the United States, 2006-2011. *Environ Perspect* 128: 127001.

Nigra, A.E., Moon, K. A., Jones, M.R., Sanchez, T.R. (2021). Urinary arsenic and heart disease mortality in NHANES 2003- 2014. *Environ Research* 200: 111387.

Nino, S.A., Morales-Martinez, A., Chi-Ahumada, E. et al. (2019). Arsenic exposure contributes to the bioenergetic damage in an Alzheimer's disease model. *J. Chemical and Computation* 10:323.

NRC (National Research Council). (2013). Critical aspects of EPA's IRIS assessment of inorganic arsenic: interim report. Washington, DC: The National Academies Press.

NRC (National Research Council). (2014). Review of EPA's Integrated Risk Information System (IRIS) process. Washington, DC The National Academies Press. http://dx.doi.org/10.17226/18764.

Osorio-Yanez, C., Ayllon-Vergara, J.C., Aguilar-Madrid, G., Arreola-Mendoza, L., Hernandez-Castellano, E., Barrera-Hernandez, A., Vizcaya-Ruiz, A.D., and Del Razo, L.M. (2013). Carotid intima-media thickness and plasma asymmetric dimethylarginine in Mexican children exposed to inorganic arsenic. *Environ Health Perspect* 121(9): 1090.

Signes-Pastor, A.J., Gutierrez-Gonzalez, E., Garcia-Villarino, M., et al. (2021). Toenails as a biomarker of exposure to arsenic: A review. *Environ Research* 195: 110286. b

Sohel, N, Persson, LÅ, Rahman, M, Streatfield, PK, Yunus, M, Ekström, E, Vahter, M. (2009). Arsenic in drinking water and adult mortality: A population-based cohort study in rural Bangladesh. Epidemiology 20: 824-830. http://dx.doi.org/10.1097/EDE.0b013e3181bb56ec.

Soucy, N.V., Ihnat, M.A., Chandrashekhar, D.K, Hess, L., Post, M.J., Klei, L.R., Clark, C., and Barchowsky, A. (2003). Arsenic stimulated angiogenesis and tumorigenesis in vivo. *Toxicol Science* 76(2): 271.

Soucy, N.V, Mayka, D., Klei, L.R., Nemec, A.A., Bauer, J.A.,and Barchowsky, A. (2005). Neovascularization and angiogenic gene expression following chronic arsenic exposure in mice. *Cardiovas Toxicol* 5(1): 29.

Stein, C.R., Wu, H., Bellinger, D.C., Smith, D.R., Wolff, M.S. and Savitz, D.A. (2022). Exposure to metal mixtures and neuropsychological functioning in middle childhood. *Neurotoxicology* 93:84.

Steinmaus, C.M., Ferreccio, C, Romo, J.A. et al. (2013). Drinking water arsenic in northern Chile: High cancer risks 40 years after exposure cessation. *Cancer epidemiol Biomarkers Prev* 22: 623-630.

Straub, A.C., Clark, K.A., Ross, M.A., Chandrea, A.G., Li, S., Gao, X., Pagano, P.J., Stolz, D.B., and Barchowsky, A. (2008). Arsenic-stimulated liver sinusoidal capillarization in mice requires NADPH oxidase-generated superoxide. *J. Clin Invest* 118(12): 3980.

Tsiatis, A.A. and Davidian, M. (2004). Joint modeling of longitudinal and time-to-event data: an overview. *Statistica Sinica* 14(3):809-834.

U.S. EPA (U.S. Environmental Protection Agency). (1995). Integrated Risk Information System (IRIS) on arsenic, inorganic [EPA Report]. Cincinnati, OH.

U.S. EPA (U.S. Environmental Protection Agency). (2002). A review of the reference dose and reference concentration processes. (EPA630P020002F). Washington, DC.

U.S. EPA (U.S. Environmental Protection Agency). (2012). Benchmark dose technical guidance. (EPA100R12001). Washington, DC: U.S. Environmental Protection Agency. Risk Assessment Forum.

U.S. EPA (U.S. Environmental Protection Agency). (2022). ORD staff handbook for developing IRIS assessments. (EPA 600/R-22/268). Washington, DC: U.S. Environmental Protection Agency, Office of Research and Development, Center for Public Health and Environmental Assessment.

Valeri, L., Mazumda, M.M., Bobb, J.F., Henn, B.C., Rodrigues, E.,Sharif, O.M.A., Kile, M.L., and Wright, R.O..(2017). The joint effect of prenatal exposure to metal mixtures on neurodevelopmental outcomes at 20-40 months of age: Evidence from rural Bangladesh. *Environmental Health Perspe*ctives 125(6) https://doi.org/10.1289/EHP614

Vaidya, N, Holla, B., Heron, J. et al. (2023). Neurocognitive Analysis of Low-level Arsenic Exposure and Executive Function Mediated by Brain Anomalies Among Children, Adolescents, and Young Adults in India. *JAMA Network Open* 6(5):e2312810. doi:10.1001/jamanetworkopen.2023.12810

Wang, S.X, Wang, Z.H., Chang, X.T. et al. (2007). Arsenic and Fluoride Exposure in Drinking Water: Children's IQ and Growth in Shanyin County, Shanxi Province, China *Environmental Health Perspectives* 115(4):643-647

Wasserman, GA, Liu, X, Parvez, F, Ahsan, H, Factor-Litvak, P, van Geen, A, Slavkovich, V, Lolacono, N.J; Cheng, Z, Hussain, I, Momotaj, H, Graziano, J.H. (2004). Water arsenic exposure and children's intellectual function in Araihazar, Bangladesh*. Environ Health Perspect* 112: 1329-1333. http://dx.doi.org/10.1289/ehp.6964

Wasserman, G.A., Liu, X., Parvez, F. et al. (2007). Water arsenic exposure and intellectual function in 6-year old children in Araihazar, Bangladesh. *Environmental Health Perspectives* 115(2):285-289

Wasserman, G.A., Liu, X, Loiacono, N.J., Kline, J, Factor-Litvak, P., van Geen, A., Mey, J.L., Levy, D., Abramson, R., Schwartz, A., Graziano, J.H. (2014). A cross-sectional study of well water arsenic and child IQ in Maine schoolchildren. *Eviron Health* 13: 23. http://dx.doi.org/10.1186/1476-069X-13-23**.**

Weisskopf, M. G., & Webster, T. F. (2017). Trade-offs of personal vs. more proxy exposure measures in environmental epidemiology. *Epidemiology*, 28(5), 635–643. 10.1097/ede.0000000000000686

APPENDIX A: EDITIORIAL CORRECTIONS

The SAB recommends that the following editorial corrections be addressed for the final draft.

Charge Question 2b

- Units are sometimes missing e.g., mg/L or % for the exposure group.

On Figure 4-1 (showing the RRBs across outcomes and datasets),

- It does not seem necessary to include the symbols for the medians in the legend. However, it could be helpful to have a distinction between US and non-US studies such that for example – the solid-colored circles could be US studies and lighter-filled circles could represent non-US studies.
- RRB should be spelled out on the y-axis and the formula should be provided in the legend (rather than as footnote 23 on page 4-3). The legend should say something like: "Relative Risk over the US Background exposure (RRB-US) for individual study datasets (shown as solid circles) and the medians (black outlined shapes). The RRB was derived by …."

On Figure 3-3, the top outcome for diseases of the circulatory system in HAWC is "Mortality". The SAB questions whether this is correct as it includes case-control and cohort studies that are not based on mortality data.

On Figure 3-4 (and similar figures), the SAB suggests including: "high confidence" and "medium confidence" in the legend. In the figure title, please add that these were studies rated with high or medium confidence to make this clear.

Other Notes: There are a few minor typos throughout the document (eg, P3-8 L8; P3-19 L11; P3-89 L9; 3-99 L8; P3-118 L10 – 246.5 g creatinine/L; P3-110 L1…). Are sample sizes (N) missing from several of the Thumbnail schematics (e.g., Fig 3-10a, etc).

Charge Question 3c

Page 3-69, line 4: Please reword the sentence on "fetal and infant loss (stillbirth and spontaneous abortion)" to "fetal loss and infant death (spontaneous abortion, stillbirth, and infant death)."

Page 3-70, lines 6-8: The iAs levels in drinking water wells were commonly exceeding 200 µg/L, while most studies reported positive associations from 100 µg/L to >2,000 µg/L. The lower limit of the exposure with positive associations needs to be explained better if not providing an iAs level comparison of the different studies.

Page 3-71, line 23: Please correct the typo of "cohort "tudy'" to "cohort study."

Page 3-73, b) OR—continuous, drinking water: it is not clear why confidence intervals were not given.

Page 3-80, lines 3-8: Comparing the birth weight reduction by iAs exposure in preterm and full-term infants needs the context of the association between iAs and preterm birth in this study.

Page 3-83, line 13: The Davis et al., 2015 study is described as observing no association with fetal growth, but it reports an inverse association with head circumference that was stronger among girls.

Page 3-83, lines 4-6: The sentence is confusing and may need to be re-worded.

Page 3-89, line 9: Please correct the typo of "six stud3-89epresenting" to "six studies representing."

Page 3-99, line 8: The last sentence is incomplete by ending with "postnatal."

Charge Question 3d

Page 3-109, Lines 10-13: In the Calderon et al. paper, Martinez had lower mean urinary As concentration and lower FSIQ and verbal IQ scores. The statement needs to be revised for "Martinez (high iAs exposure) compared with Morales (lower iAs exposure)."

Page 3-111, Lines 5-31: The MINIM (MINIMat) study was mentioned in two paragraphs, with slightly different name of the study. It is better for EPA to consolidate the study findings into one paragraph.

Table 3-8 P3-134 Summary subdivides neurodevelopmental outcomes differently from presentation of Database Overview Section 3.4.2 P3-114 and associated figures. Consistency is recommended. here). Such omissions should be presented in more details and rationalized as to how they impact or not impact the determination: *available **evidence indicates** that inorganic arsenic likely causes neurodevelopmental effects.*

The summary statement (P3-81 L15-18) could be sharpened. Presumably the studies reporting nonsignificant trends (inverse relationship between iAs and birth weight) are indeed null results and should be concluded to not support the association.

Charge Question 4a

Page 3-11 line 23 indicates that hair and nails may be susceptible to external contamination, but they are not equally susceptible. Per Karagas et al., 2000, *American Journal of Epidemiology,* 152(11):84-90:

"Another important feature of nail tissue is that it is generally considered less susceptible to external contamination than is hair. Agahian et al. (**23**) estimated that 98 percent of external iAs was removed by washing fingernail samples exposed to particulate iAs. We had similar results in our own laboratory using a high-specific-activity radiotracer. After exposure of toenail samples for 15 hours to even the highest concentrations of water iAs we encountered, less than 1 percent of the iAs detected in the lowest concentration toenails was due to absorption (data not shown). The association between dietary intake and toenail iAs concentrations further argues that the source of iAs is probably from ingestion rather than from exogenous sources."

On page 4-7, lines 28-29, mentions that studies were excluded if the EPA determined that exposure metric "was not sufficiently reliable (e.g., toenails)". Earlier in the document it says that the reason to exclude studies of nails, hair, and blood was the lack of suitable models to compute standardized exposure variables for these biomarkers, which seems more reasonable. The committee recommends correcting pages 4-7, lines 28-29.

Page 3-52 Lines 15-21 seem a bit contradictory. "As summarized in supporting materials from the 2011 NTP workshop evaluating the association between iAs and diabetes (Maull et al., 2012), adjustment of urinary arsenic concentrations by creatinine may lead to bias because creatinine excretion is reduced in diabetics, and the direction of the overall bias cannot be predicted. Studies with creatinine corrected urinary intake biomarker data were preferred, as urine creatinine is one practical approach to correct arsenic concentrations for urine dilution as compared to 24-h or 12-h urine samples (Hsieh *et al.*, 2019)." The exclusion of diabetes studies that did not adjust for creatinine may not be appropriate if the adjustment does not affect the risk estimates. Also, specific gravity is an alternate strategy to control for urinary dilution, but this was not mentioned. See more on the strategies used to correct for urine dilution when using urinary arsenic as the biomarker of arsenic exposure in epidemiological studies in the answer to question 4b.

Page 3-56 Lines 23-26. Was the association only in women but not in men or was the cohort comprised of only women? "Evidence from retrospective cohort studies also largely reported a positive association between arsenic exposure and diabetes. ''Ippoliti (2015) reported an association between cumulative arsenic (CAI) exposure levels >804.0 µg with diabetes mortality in females (Hazard 25 Ratio (HR) of 2.56 CI 95% 1.43, 4.57 p<0.001)."

Page 3-61 lines 14-15. The study would fit better in the section on birth outcomes.

Charge Question 4b

Editorial comments for conversion Equations and Tables

Cumulative exposure units for the product of concentration times years

The nomenclature used for cumulative units is confusing and should be changed. Cumulative drinking water exposure is correctly described on p C-5 as having units of (µg iAs/L drinking water) × years. The units of concentration are multiplied by years of exposure to give the cumulative value. Describing calculation of well water concentration (µg/L) the text on p C-5 says "cumulative exposure (µg/L-yr)" is divided by years of exposure indicating that "-yr" in the units is in the numerator. However, the IRIS assessment consistently describes daily dose, e.g., background exposure on p ES-2, in units of µg iAs/kg-day, in which "-day" is in the denominator; this value would be multiplied by the number of days to get the cumulative dose over time. Cumulative drinking water concentration exposure is analogous to the area under the concentration curve routinely used in the pharmaceutical industry specified by units such as µg/L x hours or mg/L*days to insure it is understood as the product of concentration and time. The EPA should revise its specification of the cumulative units. This change needs to be made in all text, tables, figures, and supplementary materials including EXCEL or other computer files.

P C-6 lines 33-4 "rather than cumulative exposure in units of (µg iAs/L drinking water) years." Add "x" before years to make it consistent with text on p C-5 line 22.

Tables C-18, C-28, C-36, C-44:

Equations for converting water concentrations

Water concentration (µg/L) is indicated in the Exposure or Dose Metric column for 12 studies (13 rows because Mostafa et al 2008 has analyses for smokers and nonsmokers). In Table C-44 the six studies (Chen et al. (2013), D'Ippoliti et al. (2015), James et al. (2015), Sohel et al. (2009), Wade et al. (2009), Wade et al. (2015)) are all shown with the water concentration (µg/L) labeled "WE" in the Exposure or Dose Metric column and the equation is the same as Eq. 4 on p. C-5 using the term "WE" and multiplied by the drinking water consumption rate. This presentation is clear to a reader. Table C-28 three studies (D'Ippoliti et al. (2015), Ferreccio et al. (2000), Mostafa et al. (2008) – two rows for smokers and nonsmokers) and Table C-36 two studies (James et al. (2013, Pan et al. (2013)) have equations in which CE-Val has been substituted for WE. CE-Val is undefined. Either these equations should be consistent with those for the six studies in Table C-44 or, if there is some difference justifying CE-Val as the dose metric in the equation it should be defined, and the difference specified. Finally, in Table C-18 the study Bates et al 2004 has the wrong equation.

Equations for converting daily dose

Daily dose (µg /d) is indicated in the Exposure or Dose Metric column for four studies, three in Table C-18 (Baris et al. 2016, Meliker et al 2010, Steinmaus et al 2013) and one in Table C-28 (Steinmaus et al. 2013). For studies with daily dose (DD) shown as the dose metric, it should be made clear that this refers to daily dose from drinking water with elevated iAs levels, since overall daily dose is calculated by adding the daily doses from drinking water with elevated iAs, diet, and drinking water with background iAs concentration. The nomenclature for dose in the equations is inconsistent showing µg /d in two and CE-Val in one (CE-Val is undefined) and the equation in Baris et al 2016 is for cumulative dose in mg.

Equations for converting cumulative intake in (µg/L) x years

Cumulative exposure expressed as the product of concentration times years is indicated in the Exposure or Dose Metric column for four studies, one in Table C-18 (Chen et al. 2010b) and three in Table C-28 (Chen et al. 2010a, Dauphine et al 2013, Chen et al 1996). The equations for these papers do not use consistent nomenclature for the exposure dose metric or the reported duration of exposure. The dose metric in the equation is given as CE (cumulative exposure) or CE-Val (undefined). The reported duration of exposure is given as RD or RDWE. The equation in Table C-18 for Chen et al 2010 is for cumulative exposure expressed as total dose (mg).

Equations for converting cumulative intake in mg (two studies) and urinary biomarkers (10 studies) are presented consistently in the Tables.

Table C-18 p C-56

- Baris et al 2016 - Exposure column indicates daily dose (µg/d) but Equation column shows equation for dose conversion for cumulative exposure. The paper reported analyses for three metrics (µg/L concentration, µg/d average daily dose, and mg cumulative intake. The Conversion Factor Validation Spreadsheet_v4 notes the latter two metrics in the Exposure Metric column. The exposure and equation columns in this table need to be consistent and the EXCEL spreadsheet, Baris2016_CE5-Ln_mg-08-08-22-unlagged, appears to be for the cumulative exposure.
- Bates et al 1995 - Values for AGE are missing from the column (61.1, 13.1)

- Bates et al 2004 - Equation is for a cumulative intake rather than the drinking water concentration indicated in the Exposure or Dose Metric column and in Conversion Factor Validation Spreadsheet_v4. Correct the equation.
- Chang et al 2016 - Add country. Table appears correct for urinary biomarker.
- Chen et al 2010b - Equation is for a cumulative intake in units of mg but the cumulative intake indicated in the Exposure or Dose Metrics column and in Conversion Factor Validation Spreadsheet_v4 is in units of (µg/L) x years. Correct the equation. See comment above under Equations for converting cumulative intake in (µg/L) x years.
- Huang et al 2018 - Add country. Table appears correct for urinary biomarker.
- Lin et al 2018 - Add country. Table appears correct for urinary biomarker.
- Meliker et al 2010 - See comments above under Equations for converting daily dose.
- Steinmaus et al 2003 - Table appears correct for cumulative exposure in mg.
- Steinmaus et al 2013 - See comments above under Equations for converting daily dose.
- Wu et al 2013 - Table appears correct for urinary biomarker.

Table C-28 p C-82

- Argos et al 2014 Table appears correct for urinary biomarker.
- Chen et al 2010 See comment above under Equations for converting cumulative intake in (µg/L) x years.
- Dauphine et al 2013 RD column reports SD of 10, but that is 3xSD and it should be 3.3. See comment above under Equations for converting cumulative intake in (µg/L) x years.
- D'Ippoliti et al 2015 See comments above under Equations for converting water concentration.
- Ferreccio et al 2000 See comments above under Equations for converting water concentration.
- Garcia-Esquinas et al 2013 Table appears correct for urinary biomarker.
- Mostafa et al 2008 (smokers and nonsmokers) See comments above under Equations for converting water concentration. The tab in Conversion Factor Validation Spreadsheet_v4 has the year of publication incorrectly as 2009.
- Steinmaus et al 2013. See comments above under Equations for converting daily dose.

Table C-36 p C-103

- Coronado-González et al. (2007) - Add USA as country. Equation – change Doseb to Dose^{b}
- Grau-Perez et al. (2017) - Add USA as country. Equation – change Doseb to Dose^{b}
- James et al. (2013) - Add USA as country. LE column change 25. to 2.5. Exposure or Dose metric is not consistent with Conversion Factor Validation Spreadsheet_v4, which indicates cumulative exposure rather than well water concentration WE. See comments above under Equations for converting water.
- Pan et al. (2013) - Add Bangladesh as country. See comments above under Equations for converting water concentration.

Table C-44 page C-120

This table gives 3SD (three times the standard deviation) and should be changed to give SD consistent with Tables C18, C-28, and C-36.

- Chen et al 1996 - See comment above under Equations for converting cumulative intake in (µg/L) x years.
- James et al 2015 - The value for low exposure (LE) is different in the table and Conversion Factor Validation Spreadsheet_v4. Clarify.

- Sohel et al 2009 - The value for low exposure (LE) is different in the table and Conversion Factor Validation Spreadsheet_v4. Clarify.
- Wade et al 2009 - The value for low exposure (LE) is different in the table and Conversion Factor Validation Spreadsheet_v4. Clarify.
- Wade et al 2015 - There appears to be a footnote "g" for the low exposure (LE) value, but no footnote exists.

The EPA refer to a non-existent Section 5.3 in Appendix A for more information on the dose-conversions (p. 4-14, line 18). This sentence should be revised to include the appropriate section reference.

Charge Question 5

- Fix the following typos:
    a. Incorrect posterior means for the pooled slope and the Pan et al. slope in Table 4-11.
    b. Page 4-17, Line 15: "U.S. lifetime risk of 70%" lacks "for CVD incidence."

Charge Question 7a

Two apparent errors that need correction were noted in the section of the document relevant to this charge question:

- Table 4-16 shows Wasserman et al. (2018) (dose related decreased IQ in adolescents in Bangladesh) as the primary study, not Wasserman et al. (2014) (elementary school students in Maine). It is assumed this is an error because Wasserman et al. (2014) is stated to be the key study in both the draft document and the charge question, while Wasserman et al. (2018) is not discussed in detail.

- The text (p. 4-56, lines 5-6) describing the results of the study shown in Table 4-14 says: "The reduction in average Working Memory score in the highest exposure group (–5.07) was highly significant, but the change in the 5–10 µg/L group (–1.13) was not." This should say instead "…but the change in the 10-20 µg/L group (-1.13) was not…"

Charge Question 7b

On lines 9-11 of p. 4-45, the EPA should provide the correct table number for the posterior mean and 5^{th} percentile of b_sigma.

APPENDIX B: ADDITIONAL COMMENTS

For many of the charge questions, the SAB has provided Tier 2 recommendations that can be readily addressed to improve clarity and transparency. These are listed below:

- The titling and labeling of tables and figures need to be clarified and elaborated in several places. The SAB recommends denoting the studies that were selected for dose-response analyses.
- The report needs clarification on the key confounders, covariates, and risk modifiers and how these were incorporated into the assessment and synthesis.
- Acknowledge some studies that did not find significant associations with adverse fetal, newborn, and infant health outcomes when summarizing the studies.
- All the calculation spreadsheets should be reviewed as an additional QA/QC step to ensure they are correct, and the terminology is consistent between the documents and the spreadsheets in light of the numerous editorial comments made on the document.
- Additional detail is needed on specific aspects of results. For studies using urinary arsenic, EPA should indicate if arsenic was speciated and if arsenic was not speciated or does not include arsenobetaine, if seafood intake is rare in the population, in which case the use of total urinary arsenic is an adequate biomarker of iAs exposure.
- Revise the definition of expected effective counts in cohort studies to be consistent with the equation provided.
- Where appropriate (e.g., equations 6 and 7 above), replace RR with adjusted RR to clarify how effective count calculations were performed in cohort studies.
- In Table 4-11, provide the correct posterior mean of the slope for the Grau-Perez et al. (2017) study.
- Citations for key studies and the basis for important conclusions (such as the basis for water consumption of 0.017 L/kg/d and studies of CVD and diabetes referred to in the sensitivity analyses) should be provided throughout the document.
- EPA should clarify why the Spanish INMA population-based cohort was deemed high quality and low bias and included in the birth health outcomes assessment Freire et al. (2019), whereas Freire et al. (2018), which analyzed the same birth cohort and identified a correlation between cord blood iAs and cognitive impairments (lower MSCA scores in children ages 4-5 years) was excluded from consideration or reference. This exemplifies the more general concerns raised about transparency and details about excluded studies.
- Given concerns raised above, an assessment of the more recently published literature identifying neurodevelopmental health impairments associated with iAs exposure seems warranted. Examples include:
    - Vaidya et al. 2023: Presents evidence of impairments in executive function associated with low level exposures to iAs (urinary measure) among children, adolescents, and young adults. The study addressed risk factors including poor nutrition and poverty as risk factors.
    - De la Ossa et al. 2023: Presents data of associations in blood iAs and impairments in verbal measures and executive function.