



SOT FDA Colloquia on Emerging Toxicological Science Challenges in Food and Ingredient Safety

Speaker Introductions

Brenna Flannery, PhD

Colloquium Co-Chair

US FDA

Brenna.Flannery@fda.hhs.gov



Systematic Review Framework and Dose-Response Methods for Identifying Reference Doses for Inorganic Arsenic

Alexandra Larsen, PhD
US EPA, Research Triangle Park, NC
larsen.alexandra@epa.gov

Jeffrey Gift, PhD
US EPA, Research Triangle Park, NC
gift.jeff@epa.gov

Conflict of Interest Statement

- The presenters declare no conflict of interest.
- The views expressed in this presentation are those of the author(s) and do not necessarily reflect the views or policies of the US EPA



Objectives

- Discuss systematic processes used in the IRIS assessment of inorganic arsenic to identify data for reference dose calculation
 - Literature Review
 - Study Evaluation
 - Judgements on Strength of Evidence
 - Systematic Dose-response methods, including Bayesian meta-regression



Specific Aims–Inorganic Arsenic (iAs) Protocol

- Identified human studies reporting effects of exposure to iAs, focusing on health outcomes suggested by the NRC (2013):
 - Tier 1 (Bladder cancer, lung cancer, skin cancer, skin lesions, ischemic heart disease)
 - Tier 2 (Diabetes, **birth weight**, **neurodevelopmental effects**, immune effects, renal cancer, prostate cancer, nonmalignant respiratory disease)
 - Tier 3 (Hypertension, stroke, **fetal loss/stillbirth/neonatal mortality**, liver cancer, pancreatic cancer, renal disease)
- Conducted study evaluations (risk of bias) using OHAT approach



Specific Aims–Inorganic Arsenic (iAs) Protocol

- Strength of evidence synthesis conclusions across epidemiology studies based on conclusions from other assessments or conducting new systematic review evidence synthesis analyses
 - Because bladder cancer, lung cancer, skin cancer, and skin lesions are accepted hazards, the strength of evidence was considered *robust* and no new evidence synthesis was conducted for hazard identification.
 - For accepted hazards, dose-response systematic reviews conducted to identify studies for dose-response analysis.
 - For other health outcomes, new systematic review evidence synthesis analyses was conducted to characterize the strength of evidence for potential hazard.



Populations, Exposures, Comparators, and Outcomes (PECO)

PECO element	Evidence
Populations	<p>2012–2013: This assessment focused on human studies and considered animal and mechanistic studies (U.S. EPA, 2014). Animal studies may provide supporting evidence for hazard identification. If health effects are reported exclusively in animal studies, mechanistic data will be used to determine human relevance of these effects. Animal and mechanistic studies may also inform susceptibility and dose-response.</p> <p>Post-2013: This assessment focused on human studies and considered animal and mechanistic studies (U.S. EPA, 2015). Animal studies may provide supporting evidence for hazard identification. Animal and AOPn information may also inform susceptibility and dose-response.</p> <p>Post-2017 (current): This assessment focuses on human studies only to include any population and life stage (occupational or general population, including children and other sensitive life stages or populations).</p>
Exposures	<p>Subchronic- or chronic-duration studies of interest provide quantitative estimates of exposure with measurements based on biomonitoring data (e.g., hair, nails, urine, or blood), inhalation (air exposures [$\mu\text{g}/\text{m}^3$]), drinking water exposures ($\mu\text{g}/\text{L}$), cumulative exposures ($\mu\text{g}/\text{m}^3\text{-yr}$; $\mu\text{g}/\text{L}\text{-yr}$), and doses expressed as $\mu\text{g}/\text{d}$ and $\mu\text{g}/\text{kg}\text{-d}$. Studies with episodic or acute exposures will be excluded (i.e., poisonings or other short-term exposures that last up to 30 d). Studies using arsenicals, primarily arsenic trioxide and Fowler's solution will be excluded because chemotherapeutic agents are not within the scope of this review. Studies using arsenide (As^{3-}), an inorganic form of arsenic, also will be excluded. Exposures usually occur via the gas arsine and result in a different, distinctive toxicological profile based on binding to hemoglobin and red blood cell lysis.</p> <p>Post-2019 (current) screening to focus on oral exposure.</p>
Comparators	<p>A comparison or referent population exposed to lower levels (or no exposure/exposure below detection limits) of inorganic arsenic, or exposure to inorganic arsenic for shorter periods of time, or cases vs. controls. Exposure-response quantitative results are presented in sufficient detail (e.g., odds ratios or relative risks with associated confidence intervals, numbers of cases/controls, etc.).</p>
Outcomes	<p>2012–2013 broad problem formulation screening: All health outcomes (both cancer and noncancer) (U.S. EPA, 2014).</p> <p>Post-2013 screening to focus on outcomes classified by the NRC as Tier 1, 2, or 3: Cancers of the bladder, lung, skin, kidney, liver, prostate, pancreas; skin lesions; or noncancer effects of the circulatory system; pregnancy and birth outcomes; neurodevelopmental effects; diabetes; immune system; respiratory disease (nonmalignant); or renal disease (U.S. EPA, 1995).</p> <p>Post-2017 screening of health outcomes prioritized for inclusion in the assessment: cancers of the bladder, lung, kidney, liver, and skin; noncancer effect of inorganic arsenic on the circulatory system (ischemic heart disease, hypertension, and stroke), reproductive system (including pregnancy and birth outcomes), developmental outcomes (including neurodevelopmental toxicity), endocrine system (including diabetes), immune system, respiratory system, and skin</p> <p>Post-2019 (current) screening to further prioritize health outcomes based on hazard judgement, RRB, and potential use for benefit-costs analysis by program offices: bladder cancer, lung cancer, DCS, diabetes, pregnancy outcomes, and neurodevelopmental effects.</p> <p>Note: A broad outcome search strategy was retained during the different phases of outcome prioritization. Epidemiological studies on other health outcomes not prioritized are tagged during screening to monitor for new studies that may affect the problem formulation decisions described above.</p>
PBPK models	<p>Studies describing PBPK models for inorganic arsenic will be included. Studies describing quantitative models or data for understanding kinetics in biological media will be tracked as "potentially relevant supplemental material."</p>

[AOPn](#) = Adverse Outcome Pathway network; PBPK = physiologically based pharmacokinetic.

Note: Animal and mechanistic data are considered supplemental material and not tracked as PECO relevant.

Study Evaluation Overview of Epidemiologic Studies

Individual study level domains (OHAT)
Epidemiological
Selection
Confounding
Performance
Attrition
Detection
Selective reporting bias
Other (internal validity)

Risk of bias evaluation protocol:

- Questions under 6 domains
- Further informed by arsenic-specific clarifications added to OHAT protocol (Appendix C)
- Implemented with 2 independent reviewers

ROB rating	
++	Definitely low
+	Probably low
-	Probably high
--	Definitely high

Rationales and ratings determined for individual questions

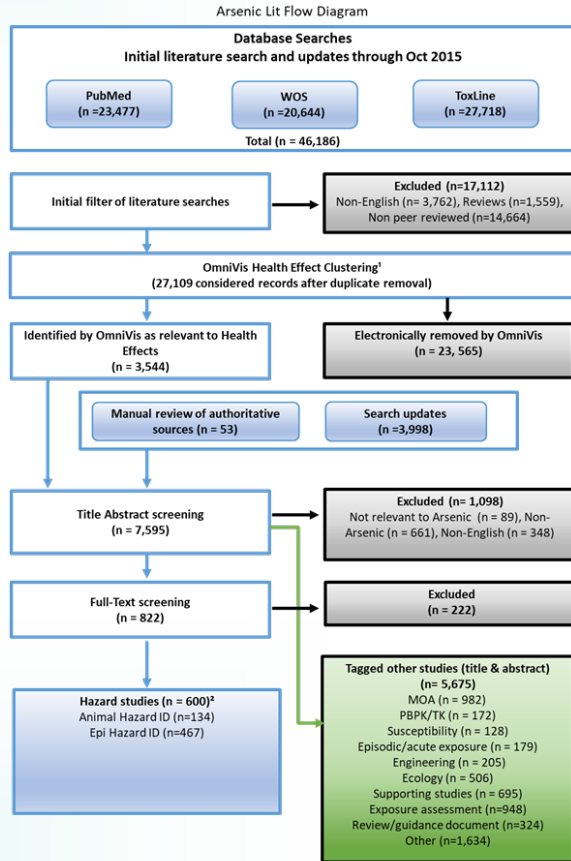
Overall Study Rating
High
Medium
Low
Uninformative

Risk of bias conclusions considered along with strengths and limitations to reach study classification

excluded

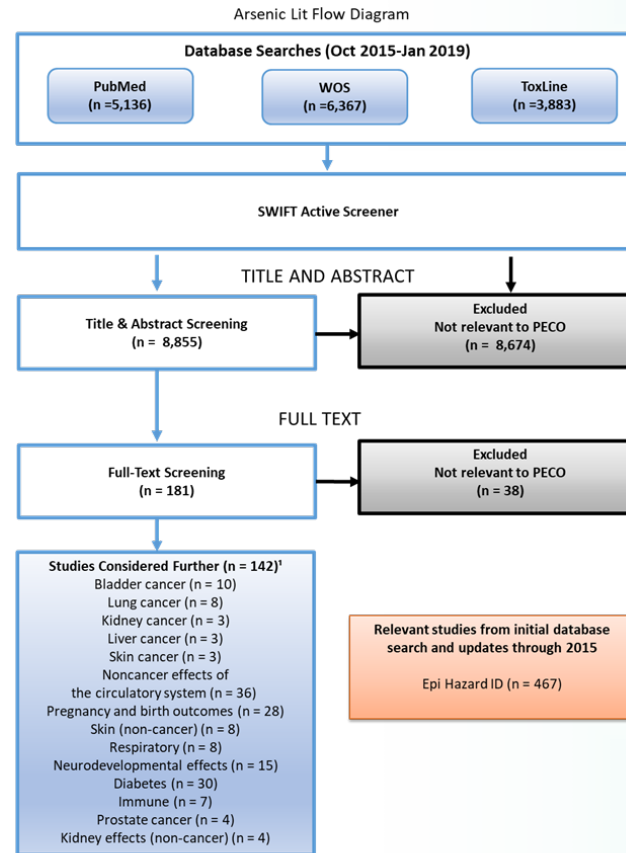


Literature Search and Screening



¹Initial results only

²Studies may be in multiple groups



¹Studies may be in multiple groups

Considerations that Inform Judgements Regarding Strength of the Human Evidence

Consideration	Increased evidence strength	Decreased evidence strength
Evidence synthesis scenarios that do not warrant an increase or decrease in evidence strength will be considered “neutral” and do not need to be described in the evidence profile table.		
Risk of bias (across studies)	An evidence base of <i>high- or medium-</i> confidence studies increases strength.	An evidence base of mostly <i>low-</i> confidence studies decreases strength. Decisions to increase strength for other factors should generally not be made if there are serious concerns for risk of bias.
Study sensitivity (across studies)	An evidence base of studies with mostly <i>good or adequate</i> sensitivity increases strength.	An evidence base of studies with poor sensitivity typically decreases confidence in null conclusions. Conversely, an evidence base of studies with mostly poor sensitivity may increase evidence strength in cases where an association is identified because the most common predicted impact of study insensitivity is towards the null.
Consistency	Similarity of findings for a given outcome (e.g., of a similar magnitude, direction) across independent studies or experiments increases strength, particularly when consistency is observed across populations (e.g., location) or exposure scenarios in human studies.	Unexplained inconsistency (conflicting evidence) decreases strength. Strength should not be decreased if discrepant findings can be explained by study confidence conclusions; variation in population, sex, and life stage; exposure patterns (e.g., intermittent or continuous); levels (low or high); duration; or intensity.
Strength (effect magnitude) and precision	Evidence of a large magnitude effect (considered either within or across studies), can increase strength. Precise results from individual studies or across the set of studies, noting that	The presence of small effects is not typically used to decrease confidence in a body of studies. However, if effect sizes that are small in magnitude are concluded not to be biologically significant, or if there are only a few studies with imprecise results, then strength is decreased.
Biological gradient/dose-response	Evidence of dose-response relationship, which may be demonstrated across studies or within studies.	A lack of dose-response relationship when expected based on biological understanding and having a wide range of doses/exposures evaluated in the evidence base can decrease strength. If the data are not adequate to evaluate a dose-response pattern, then strength is neither increased or decreased.
Coherence	Biologically related findings within an organ system, or across populations (e.g., sex), particularly when a temporal- or dose-dependent progression of related effects is observed within or across studies.	An observed lack of expected coherent changes (e.g., well-established biological relationships), particularly when observed for multiple related endpoints, will typically decrease evidence strength. Decision to decrease depends on the strength of the expected relationship(s), and considers factors (e.g., dose and duration of exposure) across studies of related changes.

Evidence Profile Tables

- Evidence integration conclusions are summarized in evidence profile tables (EPT) for each hazard using the considerations in the previous slide
- Example EPT for neurodevelopment

Evidence Summary and Interpretation				
Studies, outcomes, and confidence	Summary of key findings	Factors that increase certainty	Factors that decrease certainty	Judgment(s) and rationale
Evidence from studies of exposed humans				
<p>Cohort studies</p> <p>Studies that assessed an association between iAs and cognitive and behavior deficits in children and adolescents were generally well-conducted, accounting for appropriate covariates, resulting in interpretations of <i>high</i> confidence.</p>	<p>Though prospective cohort studies are a preferred design, the follow-up in all but one of these prospective studies may have been too short to observe neurodevelopmental effects of iAs that require longer, cumulative arsenic exposure.</p>	<ul style="list-style-type: none"> • <i>Medium or high confidence studies -</i> 	<ul style="list-style-type: none"> • lack of external validity • <i>Lack of expected coherence -</i> 	<p>⊕⊕⊖ <i>Moderate</i></p> <p>Supported primarily by consistent evidence from several well-designed cross-sectional studies of children ranging from 4-16 years of age. The magnitude and strength of the associations were inconsistent across studies, however, this could be due to the variations in the age of the study population or heterogeneity in the assessed outcomes.</p> <p>Additional support is provided by a Bangladesh prospective cohort and a Taiwanese case-control study. In both studies, uncertainty remains due to the short follow-up and high exposure levels, but this coherence across diverse study designs further strengthens the judgment.</p>
<p>Case-control studies</p> <p>Studies were of moderate size (N>50), well-designed, and exposures were well characterized using urine and blood biomarkers; thus, they were generally interpreted with <i>high</i> confidence.</p>	<p>The Taiwan study adds confidence to the association of iAs exposure and neurocognitive effects. The US study does not significantly reduce certainty because it involved lower iAs exposures and may have been confounded by exposure to other metals.</p>	<ul style="list-style-type: none"> • <i>Large or concerning magnitude of effect -</i> 	<ul style="list-style-type: none"> • Co-exposures to other metals 	<p>The judgment is primarily based primarily on evidence at higher iAs exposure levels, generally >100 µg/L.</p>



Prioritized Health Outcomes

Table 2-2. Strength of evidence judgements to help prioritize health outcomes of concern for EPA's inorganic arsenic assessment

Health outcome	NRC tier (NRC, 2013)	EPA strength-of-evidence judgement of human evidence of a causal association
NRC Tiers: Tier 1: Evidence of causality; Tier 2: Other priority outcome; Tier 3: Other endpoints to consider		
Lung cancer	Tier 1	Robust. Based on NRC Tier 1 and conclusions of "carcinogenic" for lung cancer from other assessments (ATSDR, 2016; NTP, 2016; IARC, 2012; WHO, 2011a, b; ATSDR, 2007; IARC, 2004b).
Bladder cancer	Tier 1	Robust. Based on NRC Tier 1 and conclusions of "carcinogenic" for bladder cancer from other assessments or review articles (ATSDR, 2016; NTP, 2016; IARC, 2012; WHO, 2011a, b; ATSDR, 2007; IARC, 2004b).
Skin cancer	Tier 1	Robust. Based on 1995 EPA conclusion of "known carcinogen" based on skin cancer (U.S. EPA, 1995), NRC Tier 1, and conclusions of "carcinogenic" for skin cancer based on other assessments (ATSDR, 2016; NTP, 2016; IARC, 2012; WHO, 2011a, b; ATSDR, 2007).
Ischemic heart disease	Tier 1	Robust. Based on systematic review conducted by EPA on diseases of the circulatory system (ischemic heart disease and hypertension/stroke), which is similar to associations noted in other assessments (ATSDR, 2016; WHO, 2011a, b; ATSDR, 2007) and meta-analysis ^a (Moon et al., 2017a, b; Moon et al., 2013).
Skin lesions	Tier 1	Robust. Based on NRC Tier 1 and conclusions from other assessments (ATSDR, 2016; WHO, 2011a, b; ATSDR, 2007).
Diabetes	Tier 2	Robust. Based on systematic review conducted by EPA, which is similar to associations noted in ATSDR (2016), an expert review conducted as part of an NTP workshop (Maull et al., 2012; Thayer et al., 2012) and a meta-analysis ^a (Wang et al., 2014).
Pregnancy outcomes (fetal and infant morbidity)	Tier 2	Robust. Based on systematic review conducted by EPA on pregnancy and birth outcomes (fetal growth, prematurity, and infant growth in the first 5 yr of life), which is similar to associations noted in ATSDR (2016) and meta-analysis ^a by Quansah et al. (2015).
Pregnancy outcomes (fetal loss, stillbirth, and neonatal mortality)	Tier 3	Robust. Based on systematic review conducted by EPA on pregnancy and birth outcomes (fetal loss and infant mortality in the first 5 yr of life), which is similar to associations noted in ATSDR (2016), review by Bloom et al. (2010), and a meta-analysis ^a by Quansah et al. (2015).
Hypertension/stroke ^b	Tier 3	Robust. Based on systematic review conducted by EPA on diseases of the circulatory system (including ischemic heart disease and hypertension/stroke), which is similar to associations noted in ATSDR (2016), review by Abhyankar et al. (2012), and meta-analysis ^a (Moon et al., 2017a, b; Moon et al., 2013).

Prioritized Health Outcomes (cont.)

Table 2-2. Strength of evidence judgements to help prioritize health outcomes of concern for EPA's inorganic arsenic assessment

Health outcome	NRC tier (NRC, 2013)	EPA strength-of-evidence judgement of human evidence of a causal association
Renal cancer	Tier 2	Moderate. Based on systematic review conducted by EPA, which is similar to associations noted in IARC (2012, 2004b) and ATSDR (2016) .
Nonmalignant respiratory disease	Tier 2	Moderate. Based on systematic review conducted by EPA, which is similar to associations noted in ATSDR (2016) .
Neurodevelopmental toxicity	Tier 2	Moderate. Based on systematic review conducted by EPA, which is similar to associations noted in ATSDR (2016) .
Immune effects	Tier 2	Moderate. Based on systematic review conducted by EPA, which is similar to associations noted in ATSDR (2016) .
Liver cancer	Tier 3	Moderate. Based on systematic review conducted by EPA, which is similar to associations noted in IARC (2012, 2004b) .
Health outcomes considered to have <i>slight</i> evidence		
Prostate cancer	Tier 2	Slight. Based on systematic review conducted by EPA, which is similar to associations noted in IARC (2012, 2004b) .
Pancreatic cancer	Tier 3	Slight. Based on systematic review conducted by EPA and associations noted in IARC (2004b) .
Renal disease	Tier 3	Slight. Based on systematic review conducted by EPA.

Health outcomes with *robust* or *moderate* evidence were identified for potential dose-response analysis

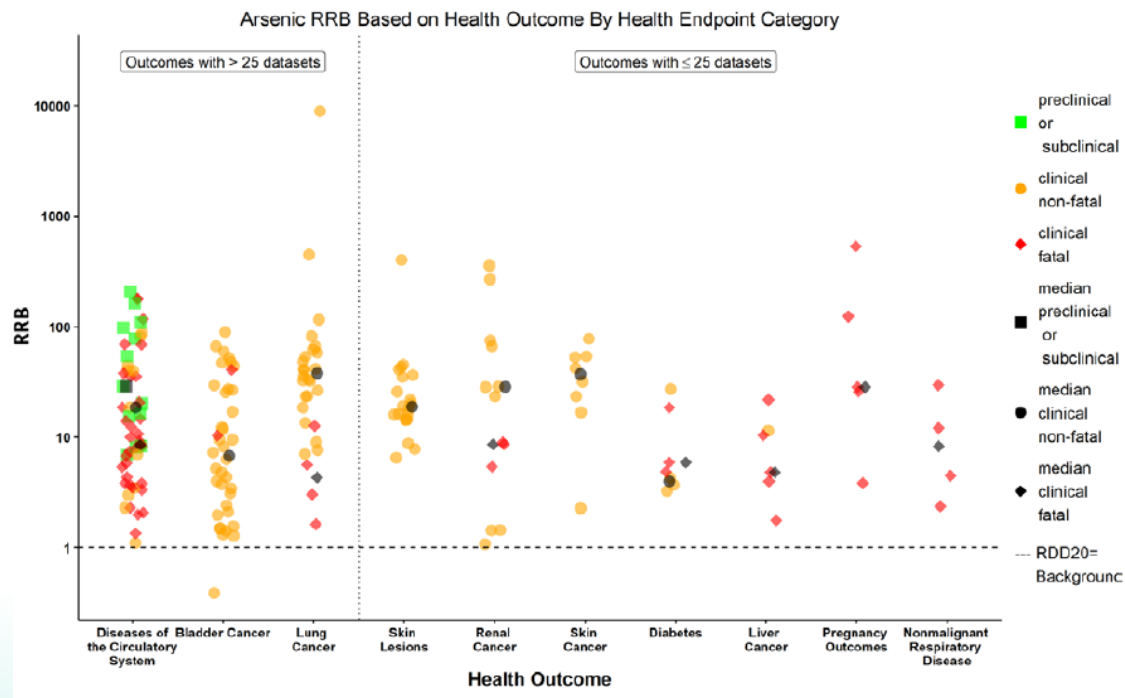
RRB Screening of Epidemiologic Exposure-Response Data

- Relative Risk Exposure to Background (RRB) Screening
 - Purpose of Screening: identify studies and endpoints suitable for dose-response modeling
 - Approach: focus on single study, single best model to derive point of departure
 - EPA's Benchmark Dose Software used to estimate exposure that would increase risk estimate by 20%
 - Attempted to derive RRB values for all 12 robust and moderate endpoints
 - Result is a comparison of iAs potency across hundreds of datasets for multiple endpoints



RRB Screening Conclusions

- All endpoints other than immune and developmental neurocognitive supported RRB modeling
- All modeled health outcomes considered for further dose-response analysis



Bayesian Meta-Regression

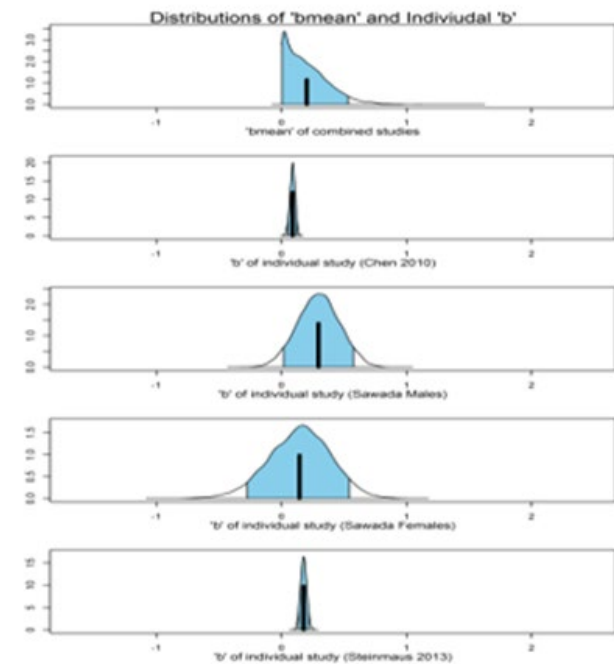
- Hierarchical Bayesian dose-response
 - Purpose: use best possible approach for evaluation of dose-response, pooling information from multiple studies
 - Approach:
 - Convert all exposures to common estimate of intake (ug/kg-day)
 - Multiple study, hierarchical dose-response analysis using the logistic model
 - Investigate low-dose non-linearity with fractional polynomial model (extension of logistic model)
 - Lifetable analysis to estimate lifetime extra risk
 - ***Best possible single study analysis if Bayesian meta-regression not feasible***



Bayesian Meta-Regression

- EPA assumed prospective likelihood is given by a logistic equation
 - Hierarchical structure assumes study-specific β values are normally distributed around mean = β_{mean} with standard deviation = β_{sigma}
 - β_{mean} and β_{sigma} are both assigned priors and updated
 - Hierarchical structure of analysis explicitly accounts for heterogeneity across studies

Posterior Distributions for Pooled and Study-specific Logistic Slope Parameters Using the MLE Dose Estimates.



Bayesian Meta-Regression

- Lifetable analysis includes consideration of background exposure to iAs and is used to estimate extra risk of disease in the target population
 - Background rates of diseased assumed to represent zero extra risk from iAs
 - A mean background iAs dose of 0.071 ug/kg-day assumed
 - Can be used to estimate both risk-at-a-dose and dose-at-a-risk values

Pooled Meta-Regression Estimates of Extra Lifetime Bladder Cancer Incidence Risk at Various Doses (per 10,000) and Drinking Water Exposures using MLE Dose Estimates

	Average Daily Arsenic Dose (ug/kg-day)						
	0.071	0.12	0.19	0.26	0.33	0.75	1.45
	Average Daily Arsenic Drinking Water Concentration (ug/L)						
	1.5	5	10	15	20	50	100
Extra Lifetime Risk ^{a, b}	0	2.0 (0.01 - 6.3)	4.8 (0.02 - 16)	7.7 (0.03 - 25)	11 (0.04 - 35)	29 (0.1 - 106)	64 (0.2 - 271)

^a These extra risk estimates assume a mean U.S. background rate for bladder cancer of 2%. Predicted additional cases in a cohort of size 10,000 for extra risk, x , when the background rate is b , would be $10,000 \cdot (1-b) \cdot x$. Thus, additional cases of bladder cancer at an extra risk of 2/10,000 (0.02%) would be $10,000 \cdot (1-2\%) \cdot 0.02\% = 1.96$.

^b Mean, 2.5% and 97.5% of Bayesian posterior slope distributions were used with US lifetables to estimate mean and credible intervals for extra risk above average background risks.



Alternative Modeling Strategies–Pregnancy Outcomes

- Pregnancy outcomes can be dichotomous (e.g., fetal mortality) or continuous (e.g., birth weight)
- Dichotomous pregnancy outcomes are amenable to our Bayesian meta-regression approach; unclear how to apply lifetable approach
- Continuous pregnancy outcomes would require separate method for dose-response



Alternative Modeling Strategies–Pregnancy Outcomes

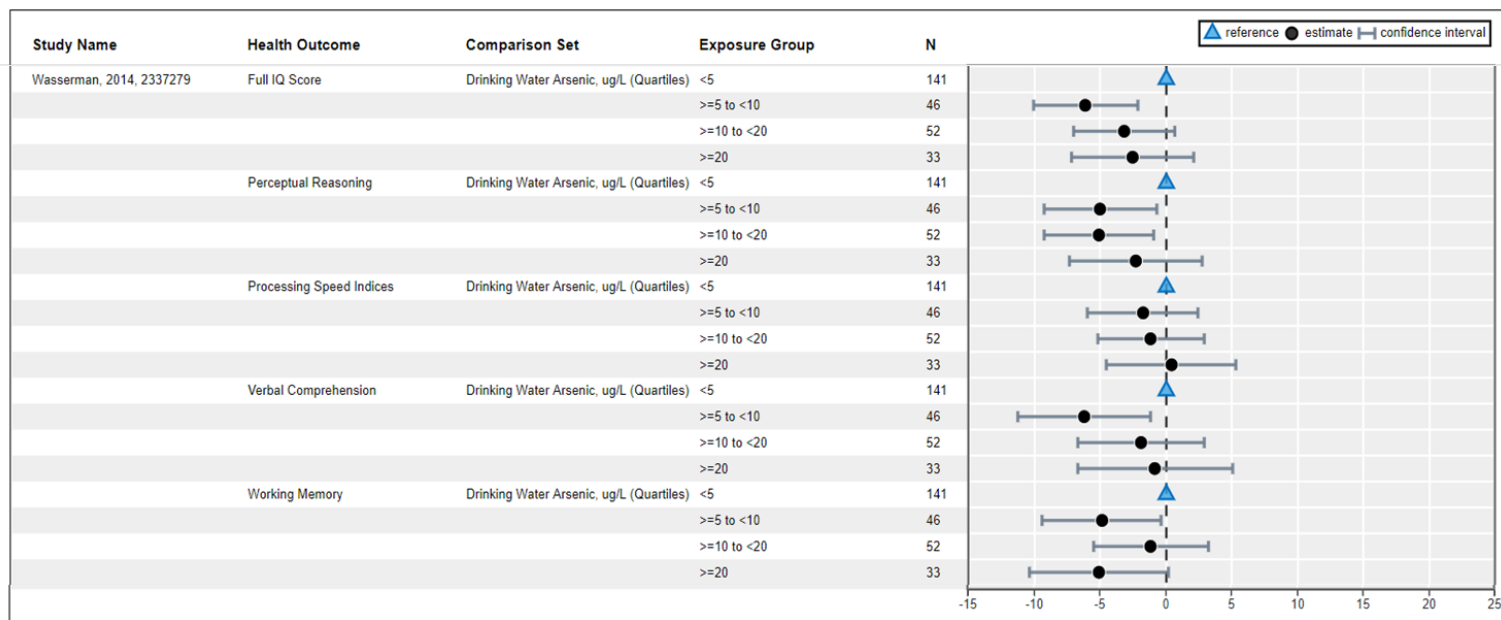
- Alternative meta-regression methods are available (metafor, R, etc.)
- Unclear if urine studies are usable in meta-regression as iAs PBPK model is not parameterized for pregnant women or fetuses
- Relying on single best study may be necessary
- Study-reported beta coefficients can be used to calculate BMDs and BMDLs



Alternative Modeling Strategies – Neurodevelopmental Outcomes

- After dose-response screening, only two studies identified as appropriate for further dose-response analysis

- These studies reported effects as continuous variables



Alternative Modeling Strategies – Neurodevelopmental Outcomes

- Focusing on study performed in the United States will reduce uncertainties in extrapolating effects across diverse study populations
- Dose-response for full IQ score, perceptual reasoning, verbal comprehension, and working memory are non-monotonic
- Raw data requested from study authors to replicate original study findings
- Raw data allows for analysis of effects when iAs treated as continuous variable
- Apparent non-monotonicity may make toxicity value derivation difficult, default to NOAEL/LOAEL approach may be necessary



Summary

- Systematic review methods used to inform hazard identification and dose-response of iAs health outcomes
- Multiple outcomes considered for dose-response analysis with Bayesian methods
- Effects in children or exposed fetuses considered for RfD derivation (developmental neurotoxicity, birth outcomes)
- Not all endpoints suitable for Bayesian modeling, alternative methods may be necessary for RfD derivation



Acknowledgements

- IRIS Inorganic Arsenic Team
 - Janice Lee
 - Allen Davis
 - Jeff Gift
 - Ingrid Druwe
 - Alex Larsen
 - Bevin Blake
 - Martha Powers

