June 16, 2014


Overall

The University of California, Program on Reproductive Health and the Environment (PRHE) strongly commends the USEPA for its efforts to utilize a systematic and transparent method of research synthesis to reach a concise, strength of evidence conclusion about the human health hazard of inorganic arsenic. To our knowledge, this review represents the first time USEPA has conducted an IRIS assessment using internal validity criteria that are explicitly and transparently defined, organized and evaluated as a “risk of bias”. In this regard, the Draft Development Materials fundamentally position USEPA’s IRIS assessment approach in growing alignment with the UCSF Navigation Guide systematic review methodology, [1-5] the National Toxicology Program’s Office of Health Assessment, Translation (OHAT) framework [6] and with recommendations made to USEPA in two recent reports by the National Academies of Science (NAS) National Research Council to utilize risk of bias [7, 8]. As such, the inorganic arsenic review is an historic document and we applaud USEPA for this and other improvements in its hazard assessment methodology, described below.

The protocol provides important infrastructure for future systematic reviews, as this experience is gained and built upon. Moreover, inorganic arsenic is likely one of the largest bodies of evidence to be vetted by USEPA’s IRIS program. Thus, we believe this document demonstrates that uptake of systematic and transparent methods of research synthesis can become the norm for USEPA.

To improve on this groundbreaking document and advance USEPA’s uptake of systematic and transparent methods of research synthesis we recommended that USEPA:

1. Standardize the organization and nomenclature of the protocol;
2. Identify evidence more transparently and efficiently;
3. Develop and implement a more complete methodology for evaluating the quality of human, non-human and mechanistic studies;
4. Develop and implement a complete and transparent methodology for integrating evidence; and
5. Support infrastructure for the advancement of systematic reviews.

Below please find the specific details of each of our recommendations.
**Recommendation 1. Standardize the organization and nomenclature of the protocol.**

Consistent with USEPA’s overarching goal of engaging stakeholders and the public in its decision-making, we recognize an important objective of an IRIS review to be a clear and concise presentation of the information with detail commensurate with full transparency. With a few notable exceptions described below, USEPA’s review meets this goal. Much of the presentation is superb, highlights being, Figure 1.3 *Overall Conceptual Model for Toxicological Review of Inorganic Arsenic* and the Tables in Chapter 3. We recognize and appreciate USEPA’s effort--- it goes a long way in facilitating public input. We believe that while it takes some time initially, USEPA will be able to build upon this approach for more efficient reviews in the future, which will ultimately contribute to expediting the IRIS review process, a critical public health goal.

Systematic review methods are not only new to USEPA, they are new to the field of environmental health overall (4). All interested parties are on a learning curve as to the methodology. We believe USEPA could play a pivotal role in addressing the need to develop a shared understanding of the steps of a systematic review and associated nomenclature.

With that goal in mind, it was not consistently clear how the organization of the inorganic arsenic protocol compares to the scheme of IRIS’ process as portrayed in Figure S-1 in the National Academy of Sciences (NAS) Review of the EPA’s IRIS Process (7) (below).

In addition, the inorganic arsenic protocol uses the term “weight of evidence” which the NAS’s review found to be too vague and of little scientific use (7) and “fit for purpose” which we find to be a vague and confusing way to identify what appear to be unspecified study “inclusion and exclusion criteria” and/or the criteria for evaluating the outcome specific quality of a body of evidence. For example, the IRIS assessment protocol says “evidence tables will serve as an additional method for presenting and evaluating whether the data are fit-for-purpose (i.e., informing hazard identification for inorganic arsenic)” (page 1-56). The NAS review of the IRIS process recommends “Specifically, the protocols should … *explicitly state the inclusion and exclusion criteria for studies* (emphasis added)” (Reference 7, page 131).

Figure S-1 (Source: The National Academy of Sciences (NAS) Review of the EPA’s IRIS Process). Systematic review in the context of the IRIS process. The committee views public input and peer review as integral parts of the IRIS process, although they are not specifically noted in the figure.
We believe it would expedite protocol development and stakeholder engagement if all interested parties knew consistently where to look for the relevant information and could rely on the use of terms consistent with standard methods of systematic review, for example, such as used in Cochrane reviews. In the longer term, research will be needed to empirically test aspects of the methodology that are new to systematic reviews in environmental health. Consistent organization and nomenclature will expedite on-going evaluation of the reviews’ methodologies.

To standardize the organization and nomenclature of the protocol we recommend that USEPA should:

- Organize the presentation of IRIS reviews in a consistent manner. The approach should use elements that have been identified in the schematic in Figure S-1 above.
- Consistently use the term “evidence integration” in place of “weight of evidence” (as recommended by the NAS).
- Consistently use “inclusion and exclusion criteria” in place of “fit for purpose” to describe which studies are being considered in the hazard assessment and why. Consistent with the NAS IRIS review, Inclusion and exclusion criteria should also be explicitly stated as such.

Recommendation 2. Identify evidence more transparently and efficiently.

A critical objective of an IRIS toxicological review should be to find the highest quality evidence on which to base a decision in the shortest period of time. As such, the breadth of the search strategy will depend on USEPA’s anticipation --- based on its scoping activities and topic expertise--- of the quantity and type of data available for decision-making.

Rationale for excluded studies: We infer from the inorganic arsenic protocol that USEPA’s rationale for excluding non-English and non-peer reviewed articles is that there are sufficient high quality data for decision-making in the peer-reviewed English-
language literature. We do not disagree with this rationale, but whatever USEPA’s rationale is for these exclusions should be stated explicitly. Inorganic arsenic is data rich; however excluding non-English papers and/or other non-peer-reviewed original data would be inappropriate for hazard assessments of many if not most chemicals in commerce. As such, these exclusion criteria should not become standard for IRIS assessments.

**Presentation/description of search strategy/results:** Figures 2.1, Table 1-9 and Figure 1.4 and the associated text all relate to identifying the evidence but there is a lack of clarity and/or agreement between them. We recommend USEPA revise to make a consistent and clearer presentation of the search strategy and results.

**“Categorization” of studies versus “inclusion/exclusion criteria”:** Sections 2.4.1 and 2.4.2 describe screening based on “categorization” of studies. We recognize that USEPA is looking for studies for hazard assessment as well as for other purposes. However, we believe it would be clearer if there were explicitly labeled “inclusion and exclusion” criteria for each type of evidence that USEPA is seeking. USEPA’s use of the phrases “setting aside for later review” and “might be reviewed later in development” for papers not considered “primary evidence for hazard evaluation” could be misconstrued as “discretionary” or “somehow still in the running for inclusion” because how and why that would happen is vague. The phrase “relevant for hazard identification” is not explicitly defined by criteria for which studies are in and which are out, (although it is possible to infer this from the text and figures). While one can piece much of this information together, it would be much clearer, and consistent with the NAS recommendations, if USEPA simply re-organized this information to explicitly state inclusion and exclusion criteria for each evidence stream, i.e., human hazard assessment, animal hazard assessment, and mechanistic evidence.

**Stopping date:** We strongly urge USEPA to set an explicit stopping date (added to Table 1.2) for the search in order to ensure timely completion of the review.

**Use of risk of bias to select studies for inclusion:** We fully support USEPA’s efforts to expedite identifying studies relevant for decision-making as it is in the best interests of public health. We understand that to this end, as described on page 2-11, USEPA undertook an additional sorting step “… following the characterization step (described in Section 4) and before full risk of bias evaluation to identify studies less likely to be useful in the overall hazard identification for inorganic arsenic … these studies were not fully “excluded” from consideration for hazard identification; some studies might be used later in this assessment where additional evidence regarding health hazard is needed.”

While we fully support USEPA’s intent, we believe this “tiering” mechanism for sorting studies could: (1) lead to studies entering back into the review in a less than transparent manner, i.e., open the review to “cherry picking” studies. It is not clear how and why USEPA would actually use high risk of bias studies – from which data will not be extracted --- to “support findings from low risk of bias studies;” and (2) be a less efficient
mechanism for meeting the objective of finding the highest quality data needed for a decision in the shortest period of time.

Therefore, we recommend that USEPA utilize a PECO question --- Population, Exposure, Comparator, and Outcome --- in concert with explicit inclusion/exclusion criteria, to limit the review to only the highest quality studies in cases such as inorganic arsenic where there is a very large body of evidence.

Specifically, we interpret the very excellent Figure 1.3 Overall Conceptual Model for Toxicological Review of Inorganic Arsenic to effectively be the PECO question for the inorganic arsenic review. We recommend that USEPA consider creating a formal PECO question based on Figure 1.3. We believe this would yield two benefits: 1. It would align USEPA’s approach with formal methods of systematic review such as Cochrane and thus increase recognition of USEPA’s methodology among stakeholders from other disciplines; this would in turn benefit uptake of environmental health science among clinical and other decision-makers; and 2. It would provide a potentially more efficient mechanism to incorporate the exposure and outcome parameters that are now being applied through risk of bias assessment to tier studies.

For example, (additional) inclusion criteria for human studies could be: 1. Exposure was assessed on an individual level; 2. Outcome was assessed using a direct measure; and 3. There was a comparator group. This would limit the review to higher quality observational studies in the more rapid screening phase of the literature search rather than the much more labor-intensive risk of bias phase. It would also make it clearer what studies will be included in USEPA decision-making.

We make the above recommendation based on experience conducting multiple systematic reviews using the Navigation Guide methodology and can verify that the screening based on a well-written PECO question is very efficient and that assessing risk of bias for included studies is more labor intensive. However, we do not know how these two methods of balancing efficiency with sufficiency of evidence for making a decision compare in terms of the final result. In this early stage of implementing systematic reviews in environmental health, we believe it would be incredibly beneficial for USEPA to consider conducting the review using its proposed method and also by using screening criteria to identify the high quality studies from the start. This could help empirically establish what the impacts of these approaches are on the “efficiency” of the review and “sufficiency” of the included studies. USEPA should also conduct a sensitivity analysis to explore how including or excluding studies with varying degrees of bias impacts later steps of the assessment process.

Exclusion of observational studies of “real world” exposure scenarios: While we concur with USEPA that mixed exposure scenarios, i.e., simultaneous exposure to inorganic arsenic and other substance(s), are potentially confounded by health impacts of the other exposure(s), these epidemiologic studies represent the true conditions of peoples’ lives – that people are exposed to multiple chemicals simultaneously. Data on the relationship of arsenic to health outcomes from these studies are critical to
understanding risks in the context of multiple exposures and consistent with the recommendations from the 2009 NAS report “Science and Decisions. Advancing Risk Assessment.” We recommend against these exclusion criteria and that these study limitations be addressed in the evaluation of the included evidence phase of the review.

To Identify evidence more transparently and efficiently we recommend that USEPA should:

- Explicitly state its rationale is for excluding non-English and non-peer-reviewed studies. These exclusion criteria should not become standard for IRIS assessments.
- Improve the presentation of the search strategy and results to make it more consistent and clear.
- Re-organize and expand the study “categorization” information to explicitly state “inclusion and exclusion” criteria for each evidence stream, i.e., human hazard assessment, animal hazard assessment and mechanistic evidence.
- Set an explicit stopping date for the search and add it to Table 1.2 in order to ensure timely completion of the review.
- Utilize a PECO question in concert with explicit inclusion/exclusion criteria to limit the review to only the highest quality studies in cases such as inorganic arsenic where there is a very large body of evidence. We anticipate that this would be more efficient than USEPA’s proposed tiering of studies based on risk of bias assessment.
- Evaluate the impact of identifying high quality studies through the literature screening phase compared to through a tiered risk of bias approach and conduct sensitivity analysis to explore how including or excluding studies with varying degrees of bias impacts the results.
- Do not exclude otherwise high quality studies because the exposure to inorganic arsenic occurred at the same time as exposure to other chemicals. These study limitations should be addressed in the evaluation of the included evidence phase of the review.

Recommendation 3. Develop and implement a more complete methodology for evaluating the quality of human, non-human and mechanistic studies.

Risk of bias assessment: As stated in our introductory comments, we highly commend USEPA’s use of OHAT’s risk of bias assessment criteria (6) in concert with its DRAGON software to assess risk of bias. (See: http://www.icfi.com/insights/products-and-tools/dragon-dose-response). This section is extremely clear and exemplary of a highly efficient and effective method for addressing internal validity of individual studies. We also strongly support the level of quality assurance and control in USEPA’s assessment, i.e., independent assessments by multiple reviewers and written rationale recorded for every decision, etc. These are truly transformative decisions on the part of USEPA that will advance its public health mission.
Risk of bias for mechanistic studies: USEPA provides a detailed risk of bias protocol based on OHAT’s method (6) for both human and animal evidence streams. However, USEPA will also be utilizing a third evidence stream in its decision-making on inorganic arsenic, i.e., mechanistic data. We recognize that there is no existing standard method for assessing risk of bias for mechanistic studies. Nevertheless, USEPA will be incorporating these data into its hazard assessment and in doing so making decisions as to what it thinks are quality data. The NAS IRIS review states:

“As in other experiments, risk of bias should be considered in evaluating mechanistic toxicology data” (Reference 7, Chapter 5, page 70).

The NAS report also encouraged USEPA to advance methods in this nascent field stating:

“Although additional methodologic work might be needed to establish empirically supported criteria for animal or mechanistic studies, an IRIS assessment needs to include a transparent evaluation of the risk of bias of studies used by USEPA as a primary source of data for the hazard assessment. EPA should specify the empirically based criteria it will use to assess risk of bias for each type of study design in each type of data stream” (Reference 7, Chapter 8, page 131).

Given the import of mechanistic studies in the evidence integration phase, we strongly recommend that the criteria USEPA will use to judge the quality of mechanistic studies be explicitly stated a priori in the form of a risk of bias assessment for this evidence stream.

Risk of bias for funding source: We strongly recommend that USEPA include a risk of bias domain for funding source. The NAS review of the IRIS process states: “Funding sources should be considered in the risk-of-bias assessment conducted for systematic reviews that are part of an IRIS assessment” (Reference 7, Chapter 5, 75). In addition the Cochrane Collaboration is currently considering formally adopting this domain (9).

Other evidence quality assessment parameters: As illustrated in Figure 1, below, risk of bias is but one quality assessment factor leading to evidence integration. USEPA's plan for dose response is well described. But we could not otherwise discern how USEPA plans to move systematically and transparently from risk of bias for each individual study to the quality of evidence across all studies, i.e., where and how it will incorporate factors such as indirectness, inconsistency, imprecision, publication bias, large magnitude of effect, and confounding, etc. These are standard features of systematic review methods developed by UCSF (1-5) and OHAT (6) that were derived from GRADE. The NAS IRIS review found:
“EPA has not developed procedures that describe how the evidence evaluation for individual studies will be incorporated, either qualitatively or quantitatively, into an overall assessment. ... The risk-of-bias assessment of individual studies should be carried forward and incorporated into the evaluation of evidence among data streams” (Reference 7, Chapter 5, page 75).

We recommend that USEPA incorporate a systematic and transparent method to evaluate the quality of evidence for each evidence stream it considers in its review in order to transparently carry these ratings into the evidence integration step of the review.

Figure 1. Evaluating Study Quality and Strength of Evidence


Reporting quality: USEPA’s inorganic arsenic protocol does not provide for contacting study authors for missing data needed for study quality and data analysis. We believe this is an important shortcoming of its review for two reasons: 1. Based on our experience with conducting systematic reviews, there will likely be a lot of missing data
and this lack of data will constrain USEPA’s ability to conduct its review; and 2. It is critical for USEPA to proactively identify and take steps to address reporting quality in environmental health studies in order to advance hazard assessment. NAS’ review of the IRIS process recommended:

“EPA should contact investigators to obtain missing information that is needed for the evaluation of risk of bias and other quality characteristics of included studies. The committee expects that, as happened in the clinical literature in which additional reporting standards for journals were implemented (Turner et al. 2012), the reporting of toxicologic research will eventually improve as risk-of-bias assessments are incorporated into the IRIS program. However, a coordinated approach by government agencies, researchers, publishers, and professional societies will be needed to improve the completeness and accuracy of the reporting of toxicology studies in the near future.” (Reference 7, Chapter 5, page 75)

EPA is in a position to address this problem in an overarching manner and thus make a fundamental contribution to improving hazard assessments. We recommend that USEPA add contacting study authors for missing data to its inorganic arsenic protocol. At a minimum USEPA should compile missing data to systematically record what information was unavailable for risk of bias and for data analysis and evaluate the impact on these missing data on the inorganic arsenic review to gather empirical evidence for broader change in environmental health.

**To develop and implement a complete methodology for evaluating the quality of human, non-human and mechanistic studies USEPA should:**

- Develop and implement *a priori* risk of bias for mechanistic studies;
- Include funding source as a risk of bias domain;
- Implement a systematic and transparent method to evaluate the quality of the body of evidence for each evidence stream considered in its review; and
- Contact study authors for missing data.

**Recommendation 4. Develop and implement a complete and transparent methodology for integrating evidence.**

*Strength of the evidence for hazard:* Based on Section 1.5.5.4 of the inorganic arsenic review, we understand USEPA’s plan for evidence integration to be to combine the five-level hierarchy used by USEPA in its Integrated Science Assessment for Lead methodology (10) with the descriptors of causality in USEPA’s 2005 Cancer Guidelines (11). We understand that this framework would be implemented by developing evidence tables for included studies by outcome or endpoint, and that at a minimum these tables will include key features such as study design, exposure metrics, and dose-response information. Further, these tables will be used to assess whether the study is fit-for-purpose, as above, what we find to be a vague and not useful term
described by USEPA as "informing hazard identification for inorganic arsenic." USEPA also states, "For each health effect domain, a series of specific questions or criteria will be developed to help inform the fit-for-purpose, based upon NRC recommendations.”

The inorganic arsenic document later discusses factors that influence characterization of the strength of the evidence within the outlined causality framework (pages 1-57 to 1-60).

We find the description of what USEPA is proposing to lack transparency and coherence as to what criteria will be used to rate each body of evidence separately and how the evidence streams will be combined and linked to the definitions in Table 1-5. For the sake of efficiency and transparency, we strongly recommend that IRIS develop or utilize an OHAT/Navigation Guide/GRADE- like framework for rating the quality of the body of each evidence stream and connecting those quality ratings to an overall strength of evidence conclusion on the health hazards of inorganic arsenic. The results of our first case study of applying the Navigation Guide demonstrate that this structured approach is feasible and desirable (1-4).

We note that the NAS found: “If EPA does move to a structured evidence-integration process, it should combine resources with NTP to leverage the intellectual resources and scientific experience in both organizations” (Reference 7, Chapter 6, page 100).

We strongly support USEPA combining resources with OHAT to advance evidence integration methodologies in environmental health sciences.

Strength of the evidence for susceptibility: We highly support USEPA’s a priori consideration that early life of human development is a susceptible life stage (page 1-28). We understand that USEPA then plans to “identify other susceptible life stages” by conducting a strength of evidence framework for assessing the populations, the potential impact of life stage, human variability, and environmental factors which may influence susceptibility. However there is no protocol presented for how that review would be conducted, except for Table 1-4 (pages 1-28 to 1-29) Strength of Evidence Framework for Susceptibility. The criteria for causality in Table 1-4 were adopted from USEPA’s Integrated Science Assessment for Lead methodology (10). We recommend that USEPA extend its a priori consideration of susceptibility to adolescence. We believe this would be consistent with USEPA’s 2005 Cancer Guidelines which:

… “view childhood as a sequence of life stages rather than viewing children as a subpopulation, the distinction being that a subpopulation refers to a portion of the population whereas a life stage is inclusive of the entire population. Exposures that are of concern extend from conception through adolescence and also include preconception exposures of both parents. These cancer guidelines use the term “childhood” in this more inclusive sense” (Reference 11 pages 1-15 – 1-16).
Use of a default approach for assuming susceptibility based on developmental lifestage is also supported by the 2009 NAS report Science and Decisions, Advancing Risk Assessment. If USEPA decides to conduct a strength-of-evidence assessment for life stage susceptibility for inorganic arsenic, it should utilize a systematic and transparent method for its review. However, if information is found lacking, current science and the NAS recommends using a default approach to account for developmental susceptibilities.

To develop and implement a complete and transparent methodology for integrating evidence USEPA should:

- Develop or utilize an OHAT/Navigation Guide/GRADE- like framework for rating the quality of the body of each evidence stream and connecting those quality ratings to an overall strength of evidence conclusion on the health hazards of inorganic arsenic;
- Extend its a priori consideration of susceptibility to adolescence; and
- Utilize a systematic and transparent method to review life stage susceptibility if such data are needed to support the inorganic arsenic review or apply a default for developmental susceptibility as recommended by the NAS.

Recommendation 5. Support infrastructure for the advancement of systematic reviews.

We commend USEPA for its demonstration of the use of DRAGON software developed by contractors to the National Toxicology Program (NTP) for its risk of bias assessment and for its health effects characterization using a pre-determined standard vocabulary. We believe USEPA’s use of DRAGON will help ensure consistent and transparent presentation of its reviews. We urge USEPA to work with NTP in an ongoing manner to develop this and other open source tools and to train scientists in their use. We believe such infrastructure development will be critical to increasing the efficiency of IRIS assessments and to expediting uptake of systematic reviews in environmental health.

For the same reason, we commend USEPA for its proposed implementation of natural language processing to expedite the literature search process. However, we strongly recommend that concurrent with its implementation USEPA develop empirical evidence of its sensitivity and specificity compared to a search conducted by well-trained humans in order to support and improve on its use.

To support infrastructure for the advancement of systematic reviews we recommend that USEPA:

- Support the development of infrastructure and training of scientists in systematic review methods, including but not limited to NTP’s DRAGON and natural language processing.
• Develop empirical evidence of the reliability and efficiency of natural language processing in comparison to well-trained humans order to support and improve on its use.

In summary, we recognize and appreciate the critical advances that USEPA has made in increasing the transparency of its IRIS assessment. The inorganic arsenic review is a milestone in USEPA’s efforts to advance evidence-based-decision making in environmental health. In support of this goal we have proposed many recommendations to improve this and future reviews and we thank you in advance for your consideration of our comments.

Sincerely,

Tracey J. Woodruff, PhD, MPH
Professor and Director
UCSF Program on Reproductive Health and the Environment
woodrufft@obgyn.ucsf.edu

Patrice Sutton, MPH
Research Scientist
UCSF Program on Reproductive Health and the Environment
suttonp@obgyn.ucsf.edu
References


