Arsenic Question #2 at the IRIS June Bi-Monthly Meeting

Chuck Elkins, Consultant [Pro bono],

Washington, DC

June 25-27, 2014

Suggestions

- Summarize study quality/risk of bias info in evidence tables and provide link to detailed information in DRAGON.
- Distinguish between “study quality” and “risk of bias” & expand the study quality questions.
- Evaluate need for 400 risk of bias evaluations.
- Tailor transparency to emphasize:
  - Providing basis for decisions to public in time so outside experts can ask “Yes, but….” and
  - Fostering clear thinking by NCEA staff.

Chen et al. (2013a) **Study Type:** cross-sectional **Location:** Bangladesh (Araihazar) **Population:** Health Effects of Arsenic Longitudinal Study (HEALS) participants n cases: n/a n control: n/a	**Exposure Surrogate:** drinking water **Exposure Description:** at baseline, water samples from 10,971 tube wells collected and analyzed for total arsenic **Population-Level Exposure:** 81.1 µg/L mean	**Outcome:** carotid intima-media thickness (cIMT) *baseline well water arsenic concentration, µg/L* Exp. Level: continuous; n: NR; adBeta: 5.1; CI: -0.2, 10.3 Stat Method: multiple linear regression
	Exposure Surrogate: urine **Exposure Description:** spot urine samples collected at baseline and at all follow-up visits; total arsenic concentration measured **Population-Level Exposure:** 250.5 µg/g-creatinine mean	**Outcome:** carotid intima-media thickness (cIMT) *baseline urinary arsenic concentration, µg/g-creatinine* Exp. Level: continuous; n: NR; adBeta: 11.7; CI: 1.8, 21.6 Stat Method: multiple linear regression

Use in Assessment

Key study for RfD

Risk of Bias

[+] Unintended exposure
Protocol Deviations
Outcome reporting
[-] Confounding (Analysis)

Link to DRAGON

Comments:

Suggestions

- Summarize study quality/risk of bias info in evidence tables and provide link to detailed information in DRAGON.
- Distinguish between “study quality” and “risk of bias” & expand the study quality questions.
- Evaluate need for 400 risk of bias evaluations.
- Tailor transparency to emphasize:
  - Providing basis for decisions to public in time so outside experts can ask “Yes, but….” and
  - Fostering clear thinking by NCEA staff.

NRC May 2014 Report

- "The committee notes that assessing the quality of the study is not equivalent to assessing the risk of bias in the study. An assessment of study quality evaluates the extent to which the researchers conducted their research to the highest possible standards and how a study is reported. Risk of bias is related to the internal validity of a study and reflects study-design characteristics that can introduce a systematic error (or deviation from the true effect) that might affect the magnitude and even the direction of the apparent effect." (page 6)

Suggestions

- Summarize study quality/risk of bias info in evidence tables and provide link to detailed information in DRAGON.
- Distinguish between “study quality” and “risk of bias” & expand the study quality questions.
- Evaluate need for 400 risk of bias evaluations.
- Tailor transparency to emphasize:
  - Providing basis for decisions to public in time so outside experts can ask “Yes, but….” and
  - Fostering clear thinking by NCEA staff.

Suggestions

- Summarize study quality/risk of bias info in evidence tables and provide link to detailed information in DRAGON.
- Distinguish between “study quality” and “risk of bias” & expand the study quality questions.
- Evaluate need for 400 risk of bias evaluations.
- Tailor transparency to emphasize:
  - Providing basis for decisions to public in time so outside experts can ask “Yes, but….” and
  - Fostering clear thinking by NCEA staff.

NRC May 2014 Report

- The history of subjectivity in science, the arts, and esthetics goes back a long way and still causes tension in scientific discourse (Shapin 2011; Klempe 2012). The only tentative solution is to describe as accurately as possible the methods by which scientific and policy decisions are made, by whom, and with what expertise. (page 26)

Comments re Arsenic Question #2: Risk of Bias

by Chuck Elkins June, 2014

Caveat

I am not very familiar with the arsenic literature and therefore these comments are more generic in nature, looking at much at future assessments as at this assessment of arsenic. These comments are being submitted on my own behalf and not on behalf of any clients.

Terminology

There is confusion with regard to the use of the term "risk of bias" in these arsenic papers and this confusion could hide a serious process problem.

I note that the May 2014 report of the National Research Council distinguishes between "study quality" and "risk of bias," [1] something that the NRC committee that provided the interim report on arsenic did not do. [2]

> "The committee notes that assessing the quality of the study is not equivalent to assessing the risk of bias in the study. An assessment of study quality evaluates the extent to which the researchers conducted their research to the highest possible standards and how a study is reported. Risk of bias is related to the internal validity of a study and reflects study-design characteristics that can introduce a systematic error (or deviation from the true effect) that might affect the magnitude and even the direction of the apparent effect." [1]

This report from the NRC was published, of course, AFTER these arsenic papers were written, so NCEA would not have been expected to follow this latest NRC guidance. In these arsenic papers, NCEA staff have apparently combined these two concepts in its "risk of bias" analysis. This raises two issues: (1) in future assessments shouldn't NCEA separate these two concepts? , and (2) did this combining of the two concepts inadvertently give short shrift to the first of the two concepts, "study quality." In addition, there is the additional concept of fit-for-purpose that is not explained or demonstrated, as far as I could determine, in these papers. All these terms should be clarified, particularly with regard to assessments in the future. In addition, NCEA should make sure that "study quality" as defined by the NRC panel on the IRIS process, is fully examined in all assessments, using a complete set of questions.

---

[1] National Research Council, Review of EPA's Integrated Risk Information System (IRIS) Process, May 2014, page 6.

[2] National Research Council, "Critical Aspects of EPA's IRIS Assessment of Inorganic Arsenic: Interim Report", pages 5,21.

Use of Risk of Bias determinations

NCEA should confirm that the detailed answers regarding study quality/risk of bias analysis for each study, not just the "primary" and "supporting" designations, will be used throughout the development of the assessment whenever a study is considered for actual use in the assessment.

The risk of bias examinations conducted for arsenic contain important detailed information on each study that should be used where appropriate throughout the rest of the development process. One could get the impression from these arsenic papers that the only purpose of the risk of bias determinations will be to determine which studies are designated as primary studies and therefore included in the evidence tables. To the contrary, no respectable assessor, I believe, would rely on a study as supporting a decision in an assessment, without wanting to know the quality of that study in some detail. A simple pass/fail or binary judgment about study quality or risk of bias is just too simple when it comes to actually writing the assessment.

Making more of the detailed Risk of Bias determinations accessible to assessors and reviewers

Because assessors and reviewers should use the detailed risk of bias determination information in their decision making, at least some of the more detailed information needs to be made more accessible to them during the rest of the development process. The obvious place for this information is in the evidence tables.

There may be some initial hesitation to include more information in the evidence tables, lest they become crowded and unwieldy. However, I find the evidence tables as currently constructed, an inadequate substitute for the more detailed prose discussions of study characteristics in pre-Chapter 7 assessments.

We should remember that at least some of the impetus for the use of evidence tables came from the NRC's 2011 review of the formaldehyde IRIS assessment. In that report the NRC recommended that EPA use evidence tables to replace a large volume of text in which NCEA had, in the past, described the pertinent details of particular studies. Certainly those long paragraphs of description were hard to follow particularly because their format and content varied from study to study and paragraph to paragraph. Evidence tables have the benefit of a standard format. However, currently these tables lack some of the detail that was formerly contained in these prose paragraphs. As an example, they completely lack any information about the details of the risk of bias evaluations, and as I have indicated, I do not believe most assessors would find it satisfactory to think of study quality in terms of a binary, good enough, not good enough determination, the only information one can derive from the current evidence tables.

Evidence tables should contain as much of the key information about a study as possible. In its May, 2014 report on the IRIS process, the NRC said:

> "Evidence integration is fundamental in determining whether a chemical poses a hazard. Consequently, the premises and structure of the decision-making process should be as explicit

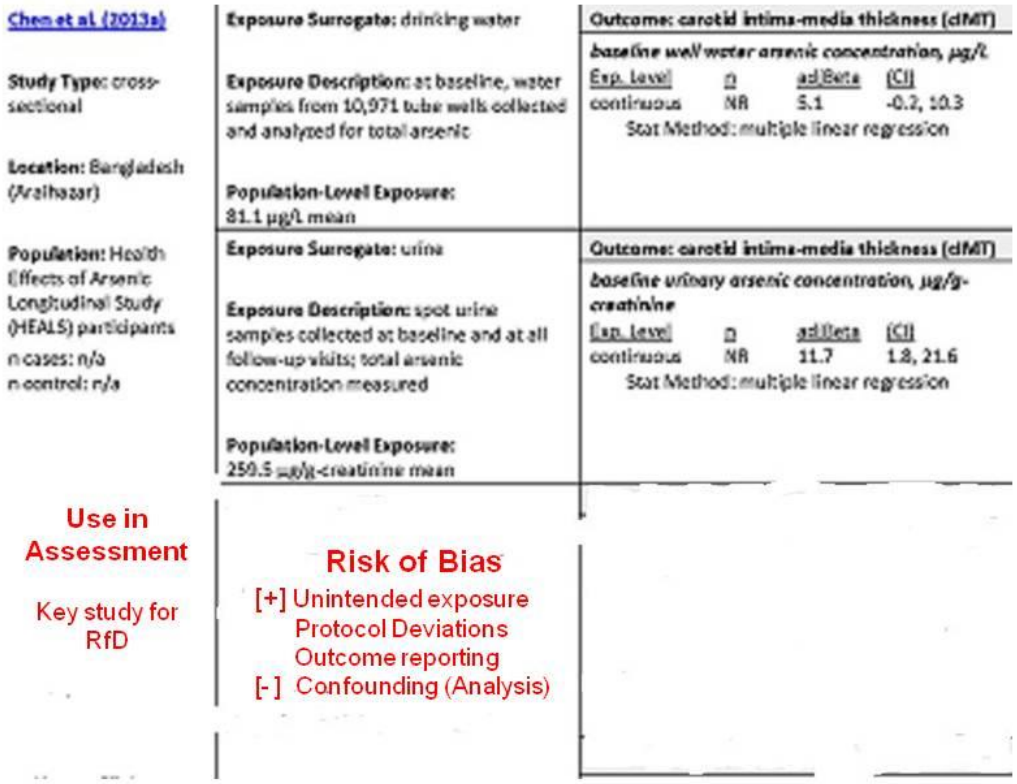
> as possible, and the basis for the determination **needs to be connected explicitly to the evidence tables produced in the IRIS process.** [Emphasis added][3]

Evidence tables can be very useful in bringing data forward about the important studies in a consistent format, and therefore we should not allow the format to unnecessarily limit the kinds of content that are presented. I believe the current format is too restrictive in the kinds of content presented, and therefore are not adequate to replace the extensive paragraphs of prose as the NRC suggested they should be. I believe NCEA and stakeholders need to go back to these kinds of paragraphs in assessments prior to the use of evidence tables and see what kind of data about studies that NCEA felt to be necessary to convey. We should then look for new ways to format that kind of data in a concise manner so that it can be included in the evidence tables. Let us find a way to make these evidence tables more like a Swiss Army knife than like a paring knife.

The value of the evidence tables is not just their brevity in my view, but their uniformity of content and their common format for its presentation. Below in Figure 1 I show one way in which some risk of bias information could be included in the evidence tables without taking up much space. Note that I have included a notation only when the quality rating was less than ++. I have also included, for illustrative purposes, two other places for additional data. One is an indication, to be inserted later in the process, with regard to how the particular study is being used in the actual assessment. Also included is a place for general comments that could not be included elsewhere.

[3] National Research Council, Review of EPA's Integrated Risk Information System (IRIS) Process, May, 2014, page 8

FIGURE #1



Chen et al. (2013a) **Study Type:** cross-sectional **Location:** Bangladesh (Araihazar) **Population:** Health Effects of Arsenic Longitudinal Study (HEALS) participants n cases: n/a n control: n/a	**Exposure Surrogate:** drinking water **Exposure Description:** at baseline, water samples from 10,971 tube wells collected and analyzed for total arsenic **Population-Level Exposure:** 81.1 µg/L mean	**Outcome: carotid intima-media thickness (cIMT)** *baseline well water arsenic concentration, µg/L* Exp. Level: continuous; n: NR; adjBeta: 5.1; (CI): -0.2, 10.3 Stat Method: multiple linear regression
	Exposure Surrogate: urine **Exposure Description:** spot urine samples collected at baseline and at all follow-up visits; total arsenic concentration measured **Population-Level Exposure:** 259.5 µg/g-creatinine mean	**Outcome: carotid intima-media thickness (cIMT)** *baseline urinary arsenic concentration, µg/g-creatinine* Exp. Level: continuous; n: NR; adjBeta: 11.7; (CI): 1.8, 21.6 Stat Method: multiple linear regression
Use in Assessment Key study for RfD	**Risk of Bias** [+] Unintended exposure Protocol Deviations Outcome reporting [-] Confounding (Analysis)	

Comments:

While I am on this topic of bringing more of the details regarding risk of bias forward for later use in the assessment, let me suggest that NCEA find a way to include it, if possible, in the exposure-response arrays. Once again, let us put our heads together to find a way to make this new tool "smarter." Currently, it is too limited to be useful without a risk of misleading people. In short, the assessment of chemical hazard is a complex undertaking, and we need to be careful, as we develop new tools, not to venture too far into simplicity, lest we inadvertently mislead those whom these new tools are designed to help.

For what studies should Risk of Bias be performed?

NCEA needs to do a serious evaluation of whether it is necessary to do this large number of risk of bias evaluations in future assessments.

The arsenic team has conducted a Risk of Bias evaluation on over 400 studies, by my quick count. At least as carried out, this constituted a considerable amount of work. I realize that the arsenic team may have felt that it had to do evaluations on this number of studies because of recommendations in the NRC interim arsenic report. However, it is an obvious point that at the end of the arsenic assessment process, NCEA should look back to evaluate whether this extensive effort was worthwhile. I want to address below the criteria by which I would suggest that NCEA make this judgment with regard to utility.

It is my understanding that NCEA has established a policy that all studies cited in an IRIS assessment must have undergone peer review and, if not, NCEA has the study peer reviewed. This is sound policy. Equally sound, I suggest, would be a policy that all studies citied in an IRIS assessment as well as any study excluded from the assessment because of its quality must undergo a formal evaluation for their quality (including Risk of Bias). By extension, this would mean that studies that are not cited in the assessment or are excluded on the basis of their quality need not undergo a formal evaluation for their quality.

At this stage of the arsenic work, does the arsenic team believe that it needs 400 plus studies (for which they have done Risk of Bias evaluations) for the purpose of hazard evaluation? Perhaps so. If not, then this is a hint that it would be useful to look for ways to move risk of bias evaluations to later in the development process or otherwise find a way to do fewer evaluations, but still meet the overall need for transparency. Decisions about when and how extensively to do risk of bias evaluations depends to some extent on the purposes of transparency, which I address below.

The purposes of transparency

Transparency can serve several possible purposes. Some are important and others are not very important for the IRIS program. This can make a big difference for NCEA's budget.

Transparency is a very popular concept today and its application to the IRIS program is no exception. For example, the National Research Council used the word transparent or transparency over 70 times in its recent report on the IRIS program. I think we can be assured that the IRIS program will adopt a number of policies and produce a considerable amount of material for which the driving force will be the desire to be "transparent."

Therefore, it is important at this juncture to examine WHY transparency is important for the IRIS program so that the IRIS program will be transparent when it is important to be transparent, but not undertake a lot of work for the sake of "transparency" if in fact it returns very little value.

"Integrity as the purpose of transparency

One purpose for transparency that I have heard expressed is the need to foster trust in the integrity of the assessment process. Integrity is important, and I, for one, have always been a fan of Ronald Reagan's slogan, "Trust, but Verify." It has been my impression as I have listened to our colleagues in the National Toxicology Program's OHAT program talk about the need to do risk of bias evaluations and to take all of the other steps of systematic review that there is an undercurrent of "if we can lay out the process at the beginning in all its details, make it difficult to change the process during the development of the assessment, and then have it run (almost as if on automatic pilot), then the public will know that we, the staff of OHAT, did not cherry pick studies or put our thumb on the scale when reaching our conclusions." Of course, I do not remember this being stated so explicitly, so I could be wrong. But grant me that this may be one of the motivations for transparency in systemic review, both for OHAT and for IRIS.

I personally would not put much weight on this purpose of transparency. It is impossible to make systematic review so automatic and so pre-determined that one can eliminate potential bias on the part of those who are running the program. No matter how much one tries to "automate" the development of hazard assessments, there will always be plenty of room for expert judgment.

Of course, if by chance there are staff within the IRIS program who have such strong philosophical views that they lose their objectivity with regard to the data they are being asked to evaluate, then I hope Ken Olden will send them off to another part of the Agency where they can do less harm to the integrity of the scientific process. However, I have my doubts that an excessive amount of transparency is really useful in order to help him or stakeholders identify such instances of bias and unprofessional behavior if there are any. Therefore, if IRIS staff find themselves urged to do extra work in the name of transparency that, in their judgment has no benefit for the decision-making process, I would hope that they will pause and ask what the intended purpose of the transparency is. If it is to enable observers to identify staff bias or unprofessional behavior, then I would recommend that they re-consider the decision to undertake the extra work for this purpose of "integrity."

Fostering constructive input as the purpose of transparency

However, I think there is a very important role of transparency that should receive a lot of attention from the IRIS staff—this is, fostering constructive and timely contributions by experts outside the assessment development team. Expert judgment must necessarily play an important role in the development of any hazard assessment, and this is emphatically the opinion of the NRC panel that just reported on the IRIS program.

People outside the agency, and, for that matter, experts within the agency who are not directly involved in the day-to-day development of a particular assessment, can assist NCEA with their expertise, if the work on a particular assessment is transparent enough and timely.

When is transparency important?

Transparency needs to be in real time, before decisions are made, not after the fact. You have all heard the question, "If a tree falls in the forest and there is no person or animal there to hear it, is there a sound?" At least, under one definition of "sound", the answer is "No." By analogy, I would ask, "If there is an explanation of the basis for a decision in an assessment, but that explanation is not made public until after the IRIS document is finalized or the decision is unchangeable, is there true transparency?" I would say "No." Of course, some historians may be interested in why a decision was made and would benefit from such explanation recorded in the files, but historians were not the reason, I surmise, that the NRC mentioned the word "transparent" more than 70 times in their last report. Transparency is very important in opening up the decision making process WHILE THAT PROCESS IS GOING ON so that experts, both inside and outside the agency can ask "YES, BUT….." before the decision is made. That is the timing when NCEA staff will get the most benefit from other experts. And of course, there is the additional benefit of being transparent, and that is that writing down one's reasoning can foster clearer thinking.

Certainly the draft assessments that are released for public comment at Step 4 of the IRIS process is an important point in time when non-NCEA experts can ask "YES, BUT…" It is terribly important at this step for NCEA staff to have been very transparent in the materials they release at that time about the decisions they have made and their reasoning.

However, I would argue that transparency is NOT needed for all decisions, large and small, that make up the whole of the assessment. Common sense should prevail. I have worked in the area of IRIS assessments for most of the past 19 years since I left EPA, and while at EPA I worked on many hazard assessments, as for example, when I was Director of the Toxic Substances Program that administers TSCA. It has been my experience that stakeholders do not ask about the myriad of decisions that EPA staff make on an assessment. Instead they focus their questions on a much narrower set of issues.

They do ask questions such as:

1. What is so wrong with a particular study that you didn't rely on it?
2. Why do you think this is the strongest study or set of studies to rely on?
3. How did you integrate all of these data in order to come to this conclusion?
4. Why did you choose to focus on this particular health endpoint and not that endpoint?
5. What relationship do your conclusions have to actual human exposures, or to metabolism in humans or to background or endogenous exposures?

They don't tend to ask:

1. What was your evaluation of the quality of this obscure little study (when there are several others that examine the same issue)?
2. Why didn't you find this little-known study in the gray literature?

Yes, if NCEA wants to be really thorough, it does need to look for every little-known study in the literature, and it needs to do risk of bias on every one of studies that it finds, but really I don't hear stakeholders asking you to be that thorough. I believe stakeholders would want you to use common sense and expert judgment in choosing which decisions for which you should strive for full transparency. Quite frankly, I believe they would tell you they want that transparency with regard to the questions

that they have raised (hopefully early on at one of these Step 1 meetings). I fear that if we are not careful, we will move so far toward a mechanistic, formulaic, and automated means of carrying out hazard assessments that we will be in danger of abandoning common sense, expert judgment, and the ability to see both the forest and the trees. Especially during the evidence integration process, NCEA needs to find ways to make the process more transparent on a real-time basis. If done correctly, I believe this will greatly enhance the ability of the staff to integrate the evidence in an objective and efficient manner and not have to engage in frequent re-writing of the assessment in order to deal with "late hits" in the development process.

The key function of Step One meetings

This brings us full circle to what I believe to be the most important function of these Step 1 meetings. For chemicals that are important enough to have caught the interest of stakeholders from industry, the states, NGOs and the academic and research community (usually data-rich chemicals), then you have available to you a set of public experts who can identify issues that they will be asking throughout the development of the assessment and for which, ideally, you would want to take extra pains to be as transparent as possible in your decision making. This also gives you some important direction about which decisions, in that particular assessment, you may not need to expend extra resources on transparency if they are not of particular concern to the body of stakeholders. This does not mean that you can skimp on the rigor of your decision making, but it may mean that you don't need to expend as much time documenting your decisions in these areas as you do in the areas that you know will be of high interest. If, for example, dose response for a particular chemical is a much important issue for stakeholders than hazard assessment, why spend an equal amount of your time being transparent on your decisions on both these elements. NCEA should, in effect, seek guidance and permission from the public to use some common sense in applying its assessment methods to a particular chemical. Otherwise, NCEA may spend too many resources on a particular assessment and thereby fail to produce the large number of assessments needed in the future.

Conclusions

I believe NCEA should:

1. Reduce the confusion about the use of terms related to the evaluation of the quality of a study and make sure that all the important questions are being used.
2. Use risk of bias evaluations in detail through the development of the assessment.
3. Make the details of risk of bias evaluations more accessible to assessors and reviewers, especially in evidence tables and exposure-response arrays.
4. Find ways to do study quality reviews on only those studies that are used in the assessment.
5. Avoid expending excessive resources on transparency designed to ensure the integrity of the process but spend sufficient resources on transparency designed to foster constructive and timely input into the key evidence evaluation integration in the assessment.
6. Maximize the constructive input from experts outside of NCEA, including stakeholders.