March 1, 2021

Comments on the ORD Staff Handbook for Developing IRIS Assessments

Submitted online via Regulations.gov to docket EPA-HQ-ORD-2018-0654

These comments are submitted on behalf of the undersigned scientists. We declare collectively that we have no direct or indirect financial or fiduciary interest in the subject of these comments. The co-signers' institutional affiliations are included for identification purposes only and do not imply institutional endorsement or support unless indicated otherwise.

We appreciate the opportunity to provide written comments on the *ORD Staff Handbook for Developing IRIS Assessments,*[1] hereafter referred to as the *Handbook*. Over the last decade, the assessments produced by the Integrated Risk Information System (IRIS), U.S. Environmental Protection Agency (EPA), have undergone multiple reviews by the National Academy of Sciences (NAS). The NAS has recommended changes to improve IRIS' approach to evaluating scientific evidence, including implementation of systematic review. In 2014 and 2018, the NAS released reports evaluating whether the IRIS program had been responsive to its past recommendations.[2,3] Both review committees were impressed with IRIS' progress, including steps to develop and implement systematic review methods and that there is "*a commitment to use systematic-review methods to conduct IRIS assessments."*[4] We commend the EPA on the substantial progress it has made in adopting and implementing systematic review methods in conducting IRIS assessments.

The draft *Handbook* has been in preparation for several years, and it is unfortunate that its public release was long delayed by previous EPA senior leadership. The IRIS program is critically important to EPA's mission of protecting human health, and the *Handbook* is an important milestone in the program's adoption of systematic review methods. However, there are methodological flaws in the current *Handbook* that need to be addressed; these methodological flaws are not consistent with NAS recommendations from the 2014, and 2018, or with the methods that EPA has stated to the NAS it is using in the IRIS program. Further, these methods are inconsistent with a number of the recommendations made recently by the NAS in its consensus report *"The Use of Systematic Review in EPA's Toxic Substances Control Act Risk Evaluations (2021)."*[5] For example, the risk of bias method presented in the *Handbook* is not validated and excludes studies based on one "critically deficient" domain, which could significantly reduce the available evidence to identify the harms caused by the toxic substances it evaluates. Use of this type of 'scoring' has been identified consistently as not appropriate in systematic reviews. For example, in the recent consensus report the NAS explicitly recommends that EPA "Do not exclude studies based on risk of bias, study quality, or reporting quality."[6]

[1] US EPA (2020) ORD Staff Handbook for Developing IRIS Assessments EPA/600/R-20/137. Available: https://cfpub.epa.gov/ncea/iris_drafts/recordisplay.cfm?deid=350086

[2] National Research Council. 2014. Review of EPA's Integrated Risk Information System (IRIS) Process. Washington, DC: The National Academies Press. https://doi.org/10.17226/18764.

[3] National Academies of Sciences, Engineering, and Medicine. 2018. Progress Toward Transforming the Integrated Risk Information System (IRIS) Program: A 2018 Evaluation. Washington, DC: The National Academies Press

[4] National Academies of Sciences, Engineering, and Medicine. 2018. Progress Toward Transforming the Integrated Risk Information System (IRIS) Program: A 2018 Evaluation. Washington, DC: The National Academies Press. Pp 1

[5] National Academies of Sciences, Engineering, and Medicine 2021. The Use of Systematic Review in EPA's Toxic Substances Control Act Risk Evaluations. Washington, DC: The National Academies Press. https://doi.org/10.17226/25952.

[6] National Academies of Sciences, Engineering, and Medicine 2021. The Use of Systematic Review in EPA's Toxic Substances Control Act Risk Evaluations. Washington, DC: The National Academies Press. Pp 40. https://doi.org/10.17226/25952.

Further, many of the concerns we highlight below in how the study evaluation step is conducted in the *Handbook* are also reflected in comments we previously submitted on the Systematic Review Protocol for the PFBA, PFHxA, PFHxS, PFNA, and PFDA IRIS Assessment and the Systematic Review Protocol for the Methylmercury Assessment.[7,8]

We are therefore concerned that implementation of the current methods and processes outlined in the *Handbook* can lead to biased assessments of the evidence. It is highly likely that relevant studies will be excluded from the final evaluations, which would in turn underestimate the true harms of the chemicals assessed by IRIS. Although there are several areas needing improvement, the *Handbook* provides a strong foundation for conducting IRIS assessments in the years to come. The *Handbook* should also serve as the basis for planning and conducting systematic reviews not just for IRIS but across all of EPA's programs.

Our comments address the following main issues:

1. **The Overview in the *Handbook* is a helpful outline of the key steps in conducting an IRIS assessment. It would be significantly improved with a few minor edits/insertions that would demonstrate and reinforce the use of key foundations of Systematic Review, including the Protocol and PECO statement.**
2. **The *Handbook* states that revisions to the initial protocol may be necessary as the assessment team develops a greater understanding of scientific/technical issues that arise during assessment development. However, some points require greater emphasis to ensure that the purpose of having a protocol is not compromised.**
3. **The *Handbook's* epidemiology study evaluation is incompatible with validated best practice methods already being implemented in environmental health in fundamental ways:**
   a) **Use of an overall risk of bias rating is inappropriate, not recommended by the National Academy of Sciences, and is not used by the National Toxicology Program's Office of Health Assessment and Translation or UCSF's Navigation Guide.**
   b) **The *Handbook's* evaluation of epidemiology studies wrongly conflates how well a study is reported with how well the underlying research was conducted.**
   c) ***The Handbook's* risk of bias method should not be used to exclude research based on one "critically deficient" methodological limitation.**
   d) **The *Handbook* should require at least two reviewers to make risk of bias study determinations.**
4. **The *Handbook* is unclear on the distinction between "Synthesis" and "Integration" of evidence. The *Handbook* would be improved by merging Chapter 9 ("Analysis and Synthesis of Human and Experimental Animal Data") into Chapter 11 ("Evidence Integration"), or alternately by moving the Chapter 11 content (regarding conclusions by evidence stream) into Chapter 9.**
5. **The Handbook is not clear regarding important aspects of how the IRIS program derives Toxicity Values and should incorporate updated methods (Chapters 12 and 13).**
6. **The *Handbook* should consider financial conflicts of interest as a potential source of bias in research.**

[7] US EPA (2019) Systematic Review Protocol for the PFBA, PFHxA, PFHxS, PFNA, and PFDA IRIS Assessments. EPA/635/R-19/049. Available: https://www.regulations.gov/document?D=EPA-HQ-ORD-2019-0275-0002

[8] US EPA (2020) Systematic Review Protocol for the Methylmercury IRIS Assessment (Preliminary Assessment Materials) EPA/635/R-19/243. Available: https://cfpub.epa.gov/ncea/iris_drafts/recordisplay.cfm?deid=345309

We appreciate the opportunity to provide public input. Please do not hesitate to contact us with any questions regarding these comments.

Sincerely,

Nicholas Chartres, PhD
Associate Director, Program on Reproductive Health and the Environment
Department of Obstetrics, Gynecology and Reproductive Sciences
University of California, San Francisco

Tracey Woodruff, PhD, MPH
Professor and Director, Program on Reproductive Health and the Environment
Department of Obstetrics, Gynecology and Reproductive Sciences
University of California, San Francisco

Swati Rayasam, MSc
Science Associate, Program on Reproductive Health and the Environment
Department of Obstetrics, Gynecology and Reproductive Sciences
University of California, San Francisco

Max Aung, PhD, MPH
Research Scientist, Program on Reproductive Health and the Environment
Department of Obstetrics, Gynecology and Reproductive Sciences
University of California, San Francisco

Aditi Shah, MPH
Policy Analyst, Program on Reproductive Health and the Environment
Department of Obstetrics, Gynecology and Reproductive Sciences
University of California, San Francisco

Stephanie Eick, PhD, MPH
Postdoctoral Scholar, Program on Reproductive Health and the Environment
Department of Obstetrics, Gynecology and Reproductive Sciences
University of California, San Francisco

Dana Goin, PhD, MA
Postdoctoral Scholar, Program on Reproductive Health and the Environment
Department of Obstetrics, Gynecology and Reproductive Sciences
University of California, San Francisco

Phil Brown, PhD
University Distinguished Professor of Sociology and Health Sciences
Director, Social Science Environmental Health Research Institute
Northeastern University

Courtney Cooper, BS
Research Coordinator, Program on Reproductive Health and the Environment
Department of Obstetrics, Gynecology and Reproductive Sciences
University of California, San Francisco

Mary Gant, MS
Policy Analyst (retired), National Institute of Environmental Health Sciences
National Institutes of Health

Robert M. Gould, MD
Associate Adjunct Professor, Department of Obstetrics, Gynecology and Reproductive Sciences, University of California, San Francisco
President, San Francisco Bay Physicians for Social Responsibility

Juleen Lam, PhD MHS MS
Assistant Professor
California State University, East Bay

Sharyle Patton
Director, Health and Environment Program
Commonweal

Ted Schettler, MD, MPH
Science Director
Science and Environmental Health Network

Patrice Sutton, MPH
Research Scientist, Program on Reproductive Health and the Environment
Department of Obstetrics, Gynecology and Reproductive Sciences
University of California, San Francisco

DETAILED COMMENTS

1. **The Overview in the *Handbook* is a helpful outline of the key steps in conducting an IRIS assessment. It would be significantly improved with a few minor edits/insertions that would demonstrate and reinforce the use of key foundations of Systematic Review, including the Protocol and PECO statement.**

We recommend the following changes be made:

- Figure O-1 in the draft handbook (shown below) completely omits the preparation of a protocol as a step in conducting an assessment.[9] The original version of the figure from the 2014 NAS report, *Review of EPA's Integrated Risk Information System (IRIS) Process* explicitly includes the protocol as the first step of systematic review, immediately following from problem formulation.[10] The figure should be revised to include the protocol.

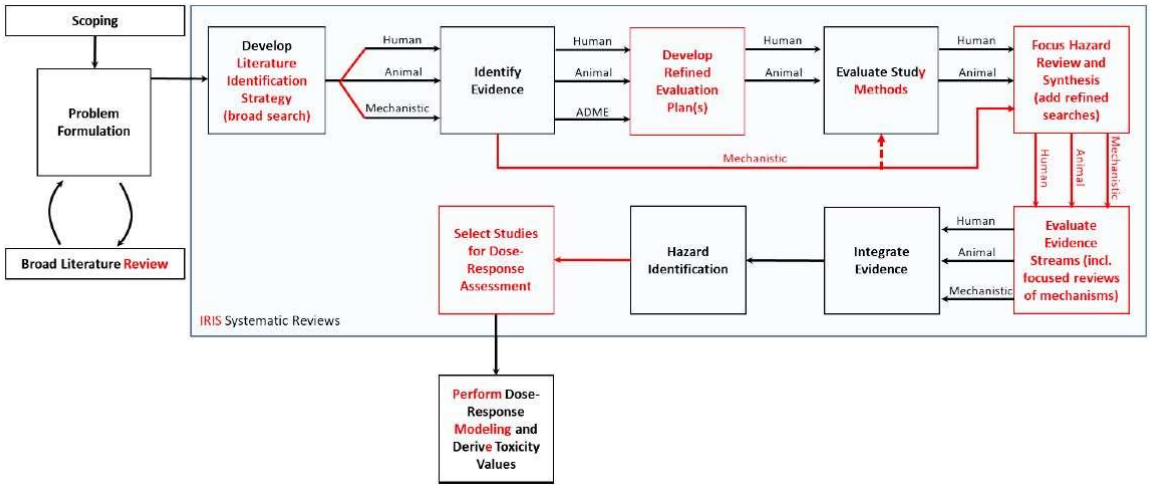


Figure O-1. Integrated Risk Information System (IRIS) assessment draft development process. [11]

- **Page xv, line 16**: IRIS should provide a better explanation of a PECO statement by revising the bullet point as: "*Populations, Exposures, Comparators, and Outcomes (PECO) criteria that define the objectives of the assessment and the scope of studies considered relevant to the assessment*."
- **Pages xix-xx**: IRIS should reference the protocol where relevant in conducting the subsequent steps: for Literature search and screening ("*Perform comprehensive literature search(es) as specified in the protocol*"); and for Study evaluation ("*Evaluate individual human and animal health effect studies, considering bias and sensitivity, following the instructions specified in the protocol*").

[9] US EPA (2020) ORD Staff Handbook for Developing IRIS Assessments EPA/600/R-20/137. Pp xviii Available: https://cfpub.epa.gov/ncea/iris_drafts/recordisplay.cfm?deid=350086

[10] National Research Council. Review of EPA's Integrated Risk Information System (IRIS) Process. Washington, DC: The National Academies Press; 2014.

[11] US EPA (2020) ORD Staff Handbook for Developing IRIS Assessments EPA/600/R-20/137. Pp xviii Available: https://cfpub.epa.gov/ncea/iris_drafts/recordisplay.cfm?deid=350086

- **Page xxi, line 10**: An important part of the IRIS Assessment Plan is to present a draft PECO; this should be stated here.
- **Page xxi, line 12**: This text should include explanation of why a protocol is important (including pre-specification of methods, reducing bias, and increasing transparency) and the function it serves in conducting an assessment. EPA should insert the statement that is already in Chapter 3: "*The protocol is a central component of a systematic review. It is intended to improve transparency and reduce bias in the conduct of the review by describing the review question and methods in advance (CRD, 2013; Higgins and Green, 2011a; IOM, 2011)*." [12]
- **Page xxii, line 13**: IRIS should expand on screening to say "*screened <u>for inclusion/exclusion using the criteria specified in the protocol</u> to compile a literature inventory of studies <u>relevant to the PECO, which will therefore be</u> included in the assessment*."

2. The *Handbook* states that revisions to the initial protocol may be necessary as the assessment team develops a greater understanding of scientific/technical issues that arise during assessment development. However, some points require greater emphasis to ensure that the purpose of having a protocol is not compromised.

First, as changes or refinements to the methods of conducting the assessment are made, they should be documented in protocol revisions which are made available to the public before the revised/refined step of the systematic review is conducted. Second, although IRIS has the ability/requirement to revise a protocol, this should not be treated as an excuse to issue an initial protocol that is vague and generic. IRIS needs to engage in assessment-specific issues in the initial protocol so that it is truly informative about how the assessment will be conducted. Greater efforts in earlier steps such as problem formulation (including, for example, conducting a preliminary literature survey organized by endpoints, or "evidence mapping") may reduce the need for revisiting or refining methods after the initial protocol has been issued. [13]

Third, the *Handbook* text regarding protocol revisions should be substantially expanded.[14] It is important to convey that revisions and/or refinements to the methods made after the initial protocol is completed must be documented in a revised protocol made available to the public. Protocol revisions should not remain out of public view until a draft assessment is made public; this is contrary to the principle of reporting planned methods in a protocol before those aspects of the assessment are conducted. Each protocol revision should be explicit about what elements of the protocol have changed and the reasons for the changes. This section should also point to places in the *Handbook* where protocol revisions (or "refinements" of the approach that should be recorded as protocol revisions) are discussed. These include: *"Supplemental Literature Searches"* (Pp. 4-15), "*Updating the Literature" Search*" (Pp. 4-16) and "*Refined Evaluation Plan*" (Pp. 5-1).

Fourth, it is not clear why "*Refined Evaluation Plan*" would be a separate section in the protocol. Any refinements should be integrated into the other relevant sections of the protocol (e.g. *"Study evaluation*

[12] US EPA (2020) ORD Staff Handbook for Developing IRIS Assessments EPA/600/R-20/137. Pp 3-1 Available: https://cfpub.epa.gov/ncea/iris_drafts/recordisplay.cfm?deid=350086

[13] US EPA (2020) ORD Staff Handbook for Developing IRIS Assessments EPA/600/R-20/137. Pp 3-1, lines 11-17 (also Overview page xvi, line 19) Available: https://cfpub.epa.gov/ncea/iris_drafts/recordisplay.cfm?deid=350086

[14] US EPA (2020) ORD Staff Handbook for Developing IRIS Assessments EPA/600/R-20/137. Pp 3-2, line 8. Available: https://cfpub.epa.gov/ncea/iris_drafts/recordisplay.cfm?deid=350086

strategy").[15] Fifth, in "*Supplemental Literature Searches,*" methods for conducting any supplemental searches must be incorporated into a revised protocol before they are conducted – same as the original search.[16] Finally, Figure 6-1 should be modified to reference the protocol as an important part of the study evaluation process.[17]

3. The *Handbook's* epidemiology study evaluation is incompatible with validated best practice methods already being implemented in environmental health in fundamental ways:

a) Use of an overall risk of bias rating is inappropriate, not recommended by the National Academy of Sciences and is not used by the National Toxicology Program's Office of Health Assessment and Translation or UCSF's Navigation Guide.

It is vital that the internal validity or risk of bias of the primary studies which underpin evidence-based decision making in environment health are assessed with transparent and accepted methods.[18] The approach to risk of bias in the *Handbook* is inconsistent with two previously validated methods that evaluate the risk of bias in human epidemiological studies. Additionally, the approach the IRIS method is using for their risk of bias tool is not consistent with recommendations in all three NAS reports on using systematic reviews in the IRIS program. Further, in their most recent report *"The Use of Systematic Review in EPA's Toxic Substances Control Act Risk Evaluations (2021)*, the NAS recommends using [19] risk of bias tools similar to the University of California San Francisco's (UCSF) Navigation Guide, and the National Toxicology's Office of Health Assessment and Translation Approach (OHAT).[20,21] The *Handbook's* risk of bias evaluations of epidemiological evidence is based off of the principles of the Cochrane Risk of Bias in Nonrandomized Studies of Interventions [ROBINS-I] tool but "*modified for use with the types of studies more typically encountered in environmental and occupational epidemiology rather than clinical interventions*."[22] As shown in Figure 6-1 below, reproduced from the *Handbook*, there are seven domains for epidemiology studies and reviewers would need to assign a consensus judgment of *good*, *adequate*, *deficient*, *not reported*, or *critically deficient* for each domain (except for sensitivity, which only has *adequate* or *deficient*).[23] The domains assessed are similar to OHAT and Navigation Guide, with the important exception of not including financial conflict of interest, discussed more in Point 6 below in these comments.

[15] US EPA (2020) ORD Staff Handbook for Developing IRIS Assessments EPA/600/R-20/137. Pp 3-3 (also page 5-1). Available: https://cfpub.epa.gov/ncea/iris_drafts/recordisplay.cfm?deid=350086

[16] US EPA (2020) ORD Staff Handbook for Developing IRIS Assessments EPA/600/R-20/137. Pp. 4-15. Available: https://cfpub.epa.gov/ncea/iris_drafts/recordisplay.cfm?deid=350086

[17] US EPA (2020) ORD Staff Handbook for Developing IRIS Assessments EPA/600/R-20/137. Pp. 6-3. Available: https://cfpub.epa.gov/ncea/iris_drafts/recordisplay.cfm?deid=350086

[18] A.A. Rooney, G.S. Cooper, G.D. Jahnke, et al. How credible are the study results? Evaluating and applying internal validity tools to literature-based assessments of environmental health hazards
Environ. Int., 92-93 (Supplement C) (2016), pp. 617-629

[19] National Academies of Sciences, Engineering, and Medicine 2021. The Use of Systematic Review in EPA's Toxic Substances Control Act Risk Evaluations. Washington, DC: The National Academies Press. Pp 40 https://doi.org/10.17226/25952.

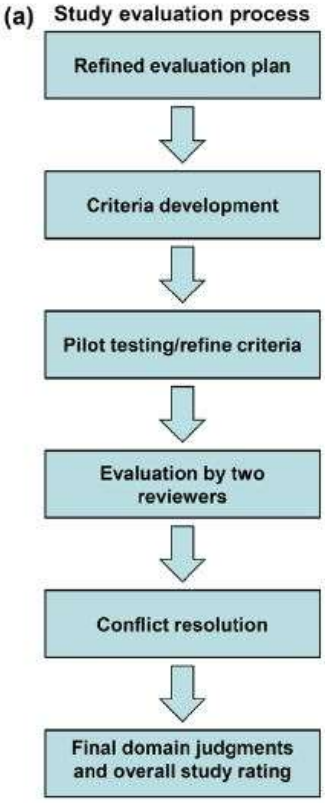
[20] Woodruff TJ, Sutton P. The Navigation Guide systematic review methodology: a rigorous and transparent method for translating environmental health science into better health outcomes. Environ Health Perspect. 2014;122(10):1007-1014.

[21] National Toxicology Program (NTP). Handbook for Conducting a Literature-Based Health Assessment Using OHAT Approach for Systematic Review and Evidence Integration. Research Triangle Park, NC: National Toxicology Program; 2019.

[22] US EPA (2020) ORD Staff Handbook for Developing IRIS Assessments EPA/600/R-20/137. Pp 6-10 Available: https://cfpub.epa.gov/ncea/iris_drafts/recordisplay.cfm?deid=350086

[23] US EPA (2020) ORD Staff Handbook for Developing IRIS Assessments EPA/600/R-20/137. Pp 6-3 Available: https://cfpub.epa.gov/ncea/iris_drafts/recordisplay.cfm?deid=350086

The biggest difference comes in the next step, where reviewers then assign an overall study rating of *high*, *medium*, or *low* confidence, or *uninformative* for a specific health outcome.[24] This step is not part of the risk of bias evaluation in either Navigation Guide or OHAT and has several fundamental scientific flaws described below.



(b) Individual evaluation domains

Animal	Epidemiology
Selection and performance • Allocation • Observational bias/blinding	Participant selection
Confounding/variable control	Confounding
Selective reporting and attrition	Selective reporting
Exposure methods sensitivity • Chemical administration and characterization • Exposure timing, frequency, and duration	Exposure measurement
Outcome measures and results display • Endpoint sensitivity and specificity • Results presentation	Outcome ascertainment Analysis
Reporting quality	Other sensitivity

Domain judgments

Judgment	Interpretation
Good	Appropriate study conduct relating to the domain and minor deficiencies not expected to influence results.
Adequate	A study that may have some limitations relating to the domain, but they are not likely to be severe or to have a notable impact on results.
Deficient	Identified biases or deficiencies interpreted as likely to have had a notable impact on the results or prevent reliable interpretation of study findings.
Critically Deficient	A serious flaw identified that makes the observed effect(s) uninterpretable. Studies with a critical deficiency are considered "uninformative" overall.

Overall study rating for an outcome

Rating	Interpretation
High	No notable deficiencies or concerns identified; potential for bias unlikely or minimal; sensitive methodology.
Medium	Possible deficiencies or concerns noted, but resulting bias or lack of sensitivity is unlikely to be of a notable degree.
Low	Deficiencies or concerns were noted, and the potential for substantive bias or inadequate sensitivity could have a significant impact on the study results or their interpretation.
Uninformative	Serious flaw(s) makes study results unusable for hazard identification or dose response but may be used to highlight possible research gaps.

Figure 6-1. Overview of Integrated Risk Information System (IRIS) study evaluation approach. (a) An overview of the evaluation process. (b) The evaluation domains and definitions for ratings (i.e., domain and overall judgments, performed on an outcome-specific basis.[25]

[24] US EPA (2020) ORD Staff Handbook for Developing IRIS Assessments EPA/600/R-20/137. Pp 6- A6/7 Available: https://cfpub.epa.gov/ncea/iris_drafts/recordisplay.cfm?deid=350086

[25] US EPA (2020) ORD Staff Handbook for Developing IRIS Assessments EPA/600/R-20/137. Pp 6-3 Available: https://cfpub.epa.gov/ncea/iris_drafts/recordisplay.cfm?deid=350086

The *Handbook* provides guidance in addition to Figure 6-1 for rating the overall confidence, noting that:

> "*…low confidence studies would have a deficient evaluation for one or more domains, although some medium confidence studies may have a deficient rating in domain(s) considered to have less influence on the magnitude or direction of effect estimates.* ***Low confidence results*** *are given less weight compared to high or medium confidence results during evidence synthesis and integration and* ***are generally not used as the primary sources of information for hazard identification or derivation of toxicity values*** *unless they are the only studies available…* [26] *Studies with critically deficient judgments in any evaluation domain will almost always be classified as uninformative…****Uninformative studies will not be considered further in the synthesis and integration of evidence for hazard identification or dose-response****."*(emphasis added) [27]

The *Handbook's* system for assigning an overall study rating is confusing, ambiguous and not empirically based. Firstly, the *Handbook* states that studies rated as 'low confidence' "*have a deficient evaluation for one or more domains*" but at the same time it allows studies to be classified as 'medium-confidence' if they "*have a deficient rating in domain(s) considered to have less influence on the magnitude or direction of the outcome-specific results.*" However, the *Handbook* does not define what those domains are and provides no scientific evidence to support EPA's judgments of these domains as being more influential than other domains, or to support the magnitude or direction of the results. For example, there is empirical evidence that inadequate application of randomization and blinding results in overestimation of efficacy of drug effects.[28,29] However, such empirical examinations of the association between the methods and results for each risk of bias domain in the ROBINS-I, and the *Handbook's* subsequent adaptation of ROBINS-I, have not been conducted and it is unclear whether these tools would stand up to such an empirical assessment. Therefore, to rate a study as overall 'low' or 'medium' confidence based on arbitrary measures is not validated and concerning and would likely result in exclusion of studies that are informative to the risk assessment.

Further, although the *Handbook's* risk of bias evaluation does not explicitly use scores, the use of a rating system that generates an overall rating based on an individual domain or several domains combined, essentially acts as a score and assumes that we know empirically how much each risk of bias domain should contribute to the overall rating. The use of 'quality scores' has not been able to distinguish between studies with a high and low risk of bias in meta-analyses [30] and empirical evidence is lacking to establish how each risk of bias item should be weighted.[31] The use of scores falsely implies a relationship between scores (i.e. high vs low) and effect or association and therefore the use of only 'high' quality studies (or only 'high' and 'medium' quality studies) will lead to a biased evaluation of the evidence. The NAS in its 2014 review of the EPA's IRIS program's method for systematic review, strongly supported a methodology that did not incorporate quantitative scoring:

---

[26] US EPA (2020) ORD Staff Handbook for Developing IRIS Assessments EPA/600/R-20/137. Pp 6-6 Available: https://cfpub.epa.gov/ncea/iris_drafts/recordisplay.cfm?deid=350086

[27] US EPA (2020) ORD Staff Handbook for Developing IRIS Assessments EPA/600/R-20/137. Pp 6-6/7 Available: https://cfpub.epa.gov/ncea/iris_drafts/recordisplay.cfm?deid=350086

[28] Wood L, Egger M, Gluud LL, Schulz KF, Juni P, Altman DG, Gluud C, Martin RM, Wood AJ, Sterne JA. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. BMJ. 2008;336(7644):601–5.

[29] Page MJ, Higgins JP, Clayton G, Sterne JA, Hrobjartsson A, Savovic J. Empirical evidence of study design biases in randomized trials: systematic review of meta-epidemiological studies. PLoS One. 2016;11(7):e0159267.

[30] Juni P, Witschi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trials for meta-analysis. Jama. 1999;282(11):1054–60.

[31] Higgins J, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions. Chichester: The Cochrane Collaboration and Wiley-Blackwell; 2008.

> *"there is no empirical basis for weighting the different criteria in the scores. Reliability and validity of the scores often are not measured. Furthermore, quality scores have been shown to be invalid for assessing risk of bias in clinical research (Juni et al. 1999)."*[32]

Further, in the recent consensus report "*The Use of Systematic Review in EPA's Toxic Substances Control Act Risk Evaluations (2021)"* the NAS stated that:

> "***For example, tools are preferred that rely on the evaluation of individual domains rather than the creation of overall quality scores*** *(Eick et al. 2020). Such tools provide a structured framework within which to make qualitative decisions on the overall quality of studies and to identify potential sources of bias. Overall quality scores may not adequately distinguish between studies with high and low risk of bias in meta-analyses (Herbison et al. 2006).* ***Importantly, there is also a lack of empirical evidence on the use of quality scores*** *(Jüni et al. 1999)*." [33] (emphasis added)

The report continued:

> "*The reliance on numeric quality scores is problematic because scores do not distinguish between high- and low-quality studies, and the relationship between quality scores and an association or effect is inconsistent and unpredictable (Greenland and O'Rouke 2001; Herbison et al. 2006; Jüni et al. 1999). More generally, the use of numerical scoring in critical appraisal does not follow standards for the conduct of systematic reviews. Additionally,* ***there was no justification provided for the weighting of specific metrics within the domains to create the overall quality score, making it difficult to determine if the weights are appropriate****. The committee notes that many public comments also discussed these problems with using numeric scores to evaluate studies*." [34] (emphasis added)

The citation the NAS uses by Eick et al. 2020 is a study that recently conducted at UCSF to understand the implications of applying different methods available to assess risk of bias. [35] We examined how three systematic review methods (OHAT, IRIS and TSCA) compare and how they could lead to different conclusions, which could have important policy implications as regulators often use systematic reviews when determining the toxicity of a chemical. We compared and assessed the risk of bias methods in the tools using 15 studies that were previously included in a systematic review using UCSF's Navigation Guide. Using the Navigation Guide to review the studies, scientists found that there is sufficient evidence supporting an association between polybrominated diphenyl ethers (PBDE) exposure and reduced IQ. [36] Critically, in a NAS 2017 report that uses the PBDE systematic review, the NAS found there was "no evidence of risk of bias in the assessment" and the NAS committee used the Navigation Guide

[32] National Research Council. Review of EPA's Integrated Risk Information System (IRIS) Process. Washington, DC: National Academies Press; 2014

[33] National Academies of Sciences, Engineering, and Medicine 2021. The Use of Systematic Review in EPA's Toxic Substances Control Act Risk Evaluations. Washington, DC: The National Academies Press. Pp 36 https://doi.org/10.17226/25952.

[34] National Academies of Sciences, Engineering, and Medicine 2021. The Use of Systematic Review in EPA's Toxic Substances Control Act Risk Evaluations. Washington, DC: The National Academies Press. Pp 39 https://doi.org/10.17226/25952.

[35] Eick SM, Goin DE, Chartres N, Lam J, Woodruff TJ. Assessing risk of bias in human environmental epidemiology studies using three tools: different conclusions from different tools. Systematic reviews. 2020;9(1):249.

[36] Lam J, Lanphear BP, Bellinger D, et al. Developmental PBDE Exposure and IQ/ADHD in Childhood: A Systematic Review and Meta-analysis. Environmental health perspectives. 2017;125(8):086001.

review of PBDEs and IQ as a basis for its own assessment.[37] However, in our 2020 study of risk of bias methods we found two of the other tools we applied to these PBDE studies, including the IRIS tool, would have prevented us from reaching this conclusion.

Across the three tools and the Navigation Guide, we found that the risk of bias is rated similarly when measured for some domains categories. However, in contrast to the Navigation Guide and OHAT approach, the IRIS and TSCA tools both included a measure of overall study quality and ultimately that has implications for the overall body of evidence. Although we applied the IRIS tool instructions from *the Handbook for Developing IRIS Assessments Version 1.0 April 2019,* it operates in exactly the same way as the tool described in the current draft *Handbook* and therefore the findings of this study are transferable and relevant to how the adapted ROBINS-I tool could be misused/downgrade a body of evidence. In our main analysis, **all PBDEs studies had overall study quality confidence ratings of "low" or "uninformative" using the *Handbook* method, in contrast to the 2017 NAS findings.** These ratings were consistent with the guidance provided in the *Handbook*.

The *Handbook states*: "***Low confidence results*** *are given less weight compared to high or medium confidence results during evidence synthesis and integration and* ***are generally not used as the primary sources of information for hazard identification or derivation of toxicity values*** *unless they are the only studies available…"*[38] *and "****Uninformative studies will not be considered further in the synthesis and integration of evidence for hazard identification or dose-response***"(emphasis added) [39] A large proportion of the evidence may not be included in evidence synthesis based on one methodological or reporting issue in a study when using the current approach for study quality evaluation in the IRIS program. If an IRIS assessment of PBDEs was conducted using the *Handbook*, it would come to significantly different conclusions than the 2017 NAS report because the epidemiological studies, having been rated by IRIS as "low confidence" or "uninformative," would not be used as primary sources of evidence for hazard assessment or toxicity value derivation.

Finally, in the recent consensus report "*The Use of Systematic Review in EPA's Toxic Substances Control Act Risk Evaluations (2021)"* the NAS recommended that EPA:

> "*Do not use numeric scores to evaluate studies; replace them with domain-based scoring as is done in the tools used in the Navigation Guide and OHAT*." [40]

Overall, there is no scientific justification for EPA to assign these scoring measures to the individual domains that will lead to a biased evaluation of the studies. We therefore strongly recommend against the use of an overall score and instead recommend that the ratings of each domain of the risk of bias tool are reported for each study to clearly highlight the different sources of bias in the study, similar to the approaches used in the Navigation Guide and OHAT.[41,42]

---

[37] National Academies of Sciences Engineering, and Medicine. (2017). Application of Systematic Review Methods in an Overall Strategy for Evaluating Low-Dose Toxicity from Endocrine Active Chemicals. Page. 8.Washington, D.C.: The National Academies Press; 2011

[38] US EPA (2020) ORD Staff Handbook for Developing IRIS Assessments EPA/600/R-20/137. Pp 6-6 Available: https://cfpub.epa.gov/ncea/iris_drafts/recordisplay.cfm?deid=350086

[39] US EPA (2020) ORD Staff Handbook for Developing IRIS Assessments EPA/600/R-20/137. Pp 6-6/7 Available: https://cfpub.epa.gov/ncea/iris_drafts/recordisplay.cfm?deid=350086

[40] National Academies of Sciences, Engineering, and Medicine 2021. The Use of Systematic Review in EPA's Toxic Substances Control Act Risk Evaluations. Washington, DC: The National Academies Press. Pp 40 https://doi.org/10.17226/25952.

[41] Woodruff TJ, Sutton P. The Navigation Guide systematic review methodology: a rigorous and transparent method for translating environmental health science into better health outcomes. Environ Health Perspect. 2014;122(10):1007-1014.

[42] National Toxicology Program (NTP). Handbook for Conducting a Literature-Based Health Assessment Using OHAT Approach for Systematic Review and Evidence Integration. Research Triangle Park, NC: National Toxicology Program; 2019.

b) The *Handbook's* evaluation of epidemiology studies wrongly conflates how well a study is reported with how well the underlying research was conducted.

The *Handbook* states:

> *"Study evaluation, as operationalized in the IRIS program, is analogous to other approaches that evaluate "study quality" or "utility" in that a wider set of issues are addressed in addition to risk of bias, including the rigor of study execution, study sensitivity, and* ***reporting****."*[43] (emphasis added)

Table 6-1, "Key concerns for study evaluation of health effect studies," defines "Reporting Quality" as "*Assess[ing] whether enough information is provided to understand how the study was designed and conducted*"[44]. Additionally, the text states "*Reporting quality and risk of bias are considered during the evaluation of each domain, and* ***the rating may be lowered when information needed to evaluate a domain is not available***."[45] (emphasis added)

As shown in Figure 6-1 above, reviewers would need to assign a consensus judgment of *good*, *adequate*, *deficient*, *not reported*, or *critically deficient* for each domain. The *Handbook* states:

> *"Not reported indicates that the information necessary to evaluate the domain was not available in the study.* ***Generally, this term carries the same functional interpretation as deficient for the purposes of the study confidence classification****. Depending on the number and severity of other limitations identified in the study, it may or may not be worth reaching out to the study authors for this information."* [46] *(emphasis added)*

As highlighted above in point 3 (a), the *Handbook* provides guidance in addition to Figure 6-1 for rating the overall confidence, noting that:

> "***low confidence studies would have a deficient evaluation for one or more domains****, although some medium confidence studies may have a deficient rating in domain(s) considered to have less influence on the magnitude or direction of effect estimates.* ***Low confidence results*** *are given less weight compared to high or medium confidence results during evidence synthesis and integration, and* ***are generally not used as the primary sources of information for hazard identification or derivation of toxicity values*** *unless they are the only studies available…* [47] *Studies with critically deficient judgments in any evaluation domain will almost always be classified as uninformative…****Uninformative studies will not be considered further in the***

[43] US EPA (2020) ORD Staff Handbook for Developing IRIS Assessments EPA/600/R-20/137. Pp 6-1 Available: https://cfpub.epa.gov/ncea/iris_drafts/recordisplay.cfm?deid=350086
[44] US EPA (2020) ORD Staff Handbook for Developing IRIS Assessments EPA/600/R-20/137. Pp 6-2 Available: https://cfpub.epa.gov/ncea/iris_drafts/recordisplay.cfm?deid=350086
[45] US EPA (2020) ORD Staff Handbook for Developing IRIS Assessments EPA/600/R-20/137. Pp 6-10 Available: https://cfpub.epa.gov/ncea/iris_drafts/recordisplay.cfm?deid=350086
[46] US EPA (2020) ORD Staff Handbook for Developing IRIS Assessments EPA/600/R-20/137. Pp 6-5/6 Available: https://cfpub.epa.gov/ncea/iris_drafts/recordisplay.cfm?deid=350086
[47] US EPA (2020) ORD Staff Handbook for Developing IRIS Assessments EPA/600/R-20/137. Pp 6-6 Available: https://cfpub.epa.gov/ncea/iris_drafts/recordisplay.cfm?deid=350086

synthesis and integration of evidence for hazard identification or dose-response."(emphasis added) [48]

Therefore studies with only one deficiency in reporting may not be used *as the primary sources of information for hazard identification or derivation of toxicity values*. This is very concerning and should be removed from the *Handbook*.

Study reporting addresses how well research findings are written up, i.e., whether there is a complete and transparent description of what was planned, what was done, what was found, and what the results mean. Guidelines and checklists for authors have been developed to help ensure all information pertinent to assessing the quality and meaning of research is included in the report. The "Strengthening of Reporting of Observational Studies in Epidemiology" or "STROBE" Initiative is an example of a checklist of items that should be included in articles reporting such research.[49]

How completely and clearly a study is reported is not a scientifically valid measure of the quality of the underlying research.[50,51,52,53] As GRADE methodologists have succinctly stated, "… just because a safeguard against bias is not reported does not mean it was neglected." [54] Research has documented that important information is often missing or unclear in published studies,[55] as word limits, styles, and other specifications are highly variable, and also non-standardized among peer-reviewed journals. As such, efforts to improve reporting are focused on uptake of reporting guidelines by journal editors and researchers.[56,57,58]

The Cochrane Handbook for conducting a systematic review clearly distinguishes reporting and bias, the latter which is defined as "a systematic error, or deviation from the truth, in results or inferences".[59] The Cochrane Handbook is explicit about not conflating reporting with bias, stating:

---

[48] US EPA (2020) ORD Staff Handbook for Developing IRIS Assessments EPA/600/R-20/137. Pp 6-6/7 Available: https://cfpub.epa.gov/ncea/iris_drafts/recordisplay.cfm?deid=350086

[49] See Strobe statement at: https://www.strobe-statement.org/index.php?id=strobe-aims

[50] Higgins JPT, Green S. Cochrane Handbook for Systematic Reviews of Interventions. Version 5.1.0 [Updated March 2011]: The Cochrane Collaboration. Available from http://www.cochrane-handbook.org.; 2011.

[51] Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, Norris S, Falck-Ytter Y, Glasziou P, DeBeer H, Jaeschke R, Rind D, Meerpohl J, Dahm P, Schünemann HJ. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. Journal of Clinical Epidemiology. 2011;64(4):383-94. doi: 10.1016/j.jclinepi.2010.04.026

[52] Devereaux PJ, Choi PT, El-Dika S, Bhandari M, Montori VM, Schünemann HJ, Garg AX, Busse JW, Heels-Ansdell D, Ghali WA, Manns BJ, GH. G. An observational study found that authors of randomized controlled trials frequently use concealment of randomization and blinding, despite the failure to report these methods. J Clin Epidemiol. 2004;57(12):1232-6; PMCID: 15617948

[53] Soares HP, Daniels S, Kumar A, Clarke M, Scott C, Swann S, B; D, Group. RTO. Bad reporting does not mean bad methods for randomised trials: observational study of randomised controlled trials performed by the Radiation Therapy Oncology Group. BMJ. 2004;328((7430)):22-4.; PMCID: PMC313900.

[54] Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, Norris S, Falck-Ytter Y, Glasziou P, DeBeer H, Jaeschke R, Rind D, Meerpohl J, Dahm P, Schünemann HJ. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. Journal of Clinical Epidemiology. 2011;64(4):383-94. doi: 10.1016/j.jclinepi.2010.04.026

[55] Lee W, Bindman J, Ford T, Glozier N, Moran P, Stewart R, M H. Bias in psychiatric case-control studies: literature survey. Br J Psychiatry. 2007;190:204-9.; PMCID: 17329739.

[56] Vandenbroucke JP, von Elm E, Altman DG, Gøtzsche PC, Mulrow CD, Pocock SJ, Poole C, Schlesselman JJ, Egger M, Initiative. S. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. Int J Surg. 2014;12(12):1500-24. doi: 10.1016/j.ijsu.2014.07.014

[57] Kilkenny C, Browne W, Cuthill IC, Emerson M, Altman DG. Animal research: reporting in vivo experiments--the ARRIVE guidelines. J Cereb Blood Flow Metab. 2011;31(4):991-3. Epub 2011/01/06. doi: 10.1038/jcbfm.2010.220. PubMed PMID: 21206507; PMCID: 3070981.

[58] Shamseer L, Moher D, Clarke M, Ghersi D, Liberati A, Petticrew M, Shekelle P, Stewart LA, Group. P-P. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) :elaboration and explanation. BMJ. 2015;350:(g7647). doi: 10.1136/bmj.g7647

[59] Higgins JPT, Green S. Cochrane Handbook for Systematic Reviews of Interventions. Version 5.1.0 [Updated March 2011]: The Cochrane Collaboration. Available from http://www.cochrane-handbook.org.; 2011.

> "*Bias may be distinguished from quality. The phrase 'assessment of methodological quality' has been used extensively in the context of systematic review methods to refer to the critical appraisal of included studies. The term suggests an investigation of the extent to which study authors conducted their research to the highest possible standards*." [60]

The Cochrane Handbook draws a distinction between assessment of methodological quality and assessment of risk of bias, and recommends a focus on the latter. The reasons for this distinction include:

1. The key consideration in a Cochrane review is the extent to which results of included studies should be *believed*. Assessing risk of bias targets this question squarely.
2. A study may be performed to the highest possible standards yet still have an important risk of bias. For example, in many situations it is impractical or impossible to blind participants or study personnel to an intervention group. It is inappropriately judgmental to describe all such studies as of "low quality", but that does not mean they are free of potential bias resulting from knowledge of intervention status.
3. Some markers of quality in medical research, such as obtaining ethical approval, performing a sample size calculation and reporting a study in line with the CONSORT Statement (Moher 2001d), are unlikely to have direct implications for risk of bias.
4. An emphasis on risk of bias overcomes ambiguity between the quality of reporting and the quality of the underlying research (although it does not overcome the problem of having to rely on reports to assess the underlying research).

Finally, in the recent consensus report "*The Use of Systematic Review in EPA's Toxic Substances Control Act Risk Evaluations (2021)"* the NAS stated that:

> "*The committee notes that TSCA's "fit-for-purpose evaluation framework" may not produce the desired results. It includes items that do not assess risk of bias, such as relevance and incomplete reporting…. **Incomplete reporting can be a challenge in evaluating a study, but it is not a marker of the validity of the study findings***." [61]

Therefore, **IRIS should only assess risk of bias and not reporting** when evaluating the quality of studies using, validated best practices risk of bias tools that have been recommended by the NAS, the Navigation Guide and OHAT.[62, 63, 64]

[60] Higgins JPT, Green S. Cochrane Handbook for Systematic Reviews of Interventions. Version 5.1.0 [Updated March 2011]: The Cochrane Collaboration. Available from http://www.cochrane-handbook.org.; 2011.
[61] National Academies of Sciences, Engineering, and Medicine 2021. The Use of Systematic Review in EPA's Toxic Substances Control Act Risk Evaluations. Washington, DC: The National Academies Press. Pp 39 https://doi.org/10.17226/25952.
[62] National Academies of Sciences, Engineering, and Medicine 2021. The Use of Systematic Review in EPA's Toxic Substances Control Act Risk Evaluations. Washington, DC: The National Academies Press. Pp 40 https://doi.org/10.17226/25952.
[63] Woodruff TJ, Sutton P. The Navigation Guide systematic review methodology: a rigorous and transparent method for translating environmental health science into better health outcomes. Environ Health Perspect. 2014;122(10):1007-1014.
[64] National Toxicology Program (NTP). Handbook for Conducting a Literature-Based Health Assessment Using OHAT Approach for Systematic Review and Evidence Integration. Research Triangle Park, NC: National Toxicology Program; 2019.

c) The *Handbook's* risk of bias method should not be used to exclude research based on one "deficient" or "critically deficient" methodological limitation.

As we have noted in point 3 a) above, the *Handbook* states that:

> *"**low confidence studies would have a deficient evaluation for one or more domains….Low confidence results** are given less weight compared to high or medium confidence results during evidence synthesis and integration and **are generally not used as the primary sources of information for hazard identification or derivation of toxicity values** unless they are the only studies available…* [65] *and "**Studies with critically deficient judgments in any evaluation domain** will almost always be classified as uninformative. **Uninformative studies will not be considered further in the synthesis and integration of evidence for hazard identification or dose-response**"(emphasis added)* [66]

This approach to exclude studies based on only one methodological limitation when conducting risk of bias assessments in the *Handbook* is again inconsistent with two previously validated methods recommended by the NAS, the Navigation Guide[67] and OHAT, [68] that are used to evaluate the risk of bias in human epidemiological studies. Neither method recommends excluding a study based on single measure. While the Navigation Guide does not exclude any studies based on the risk of bias assessment, the OHAT approach "*favors inclusion of studies unless they are problematic in multiple key aspects of study quality, an approach that offsets concerns about potentially excluding studies based on a single measure, which could seriously limit the evidence base available for an evaluation, given the type of studies available in environmental health*."[69]

While we understand that there will be variation in the internal validity and thus quality across studies, it is more appropriate to exclude studies based on pre-defined inclusion/exclusion criteria when there is a large database (such as only evaluating cohort studies), rather than an arbitrary rating of the evidence, based off one domain that is not empirically supported. Further, there are various strategies that IRIS should use to evaluate quantitatively the influence of the levels of bias across the studies via meta-analysis. These strategies include: restricting the primary analysis to those studies with a lower risk of bias and demonstrating how conclusions might be affected by the inclusion of high risk of bias studies; performing a sensitivity analysis; presenting multiple (stratified) analyses, or presenting every included study and summarizing the risk of bias; and using structured approaches like GRADE. [70]

This ensemble of approaches is consistent with how the NAS recommended EPA should conduct risk of bias assessments in the consensus report "*The Use of Systematic Review in EPA's Toxic Substances Control Act Risk Evaluations (2021)"*:

---

[65] US EPA (2020) ORD Staff Handbook for Developing IRIS Assessments EPA/600/R-20/137. Pp 6-6 Available: https://cfpub.epa.gov/ncea/iris_drafts/recordisplay.cfm?deid=350086

[66] US EPA (2020) ORD Staff Handbook for Developing IRIS Assessments EPA/600/R-20/137. Pp 6-6/7 Available: https://cfpub.epa.gov/ncea/iris_drafts/recordisplay.cfm?deid=350086

[67] Woodruff TJ, Sutton P. The Navigation Guide systematic review methodology: a rigorous and transparent method for translating environmental health science into better health outcomes. Environ Health Perspect. 2014;122(10):1007-1014.

[68] National Toxicology Program (NTP). Handbook for Conducting a Literature-Based Health Assessment Using OHAT Approach for Systematic Review and Evidence Integration. Research Triangle Park, NC: National Toxicology Program; 2019.

[69] National Toxicology Program (NTP). Handbook for Conducting a Literature-Based Health Assessment Using OHAT Approach for Systematic Review and Evidence Integration. Pp38. Research Triangle Park, NC: National Toxicology Program; 2019.

[70] National Toxicology Program (NTP). Handbook for Conducting a Literature-Based Health Assessment Using OHAT Approach for Systematic Review and Evidence Integration. Pp38. Research Triangle Park, NC: National Toxicology Program; 2019.

> "*While there is inevitably variation in the internal validity and risk of bias across individual studies, **it is standard practice to include all studies, even the studies with a high risk of bias into the evidence synthesis**. The most appropriate method to exclude studies from evidence synthesis is based on predefined exclusion criteria that should preclude an irrelevant study from being evaluated*[71]*……..Another problematic element of TSCA's "fit-for-purpose evaluation framework" is that the unacceptable studies are excluded from further analyses. Any fatal flaws in the methodology or conduct should be included in the exclusion criteria applied during the screening process. **Once a study is determined to be eligible, the study could be included in the synthesis and the risk-of-bias assessment and its limitations accounted for in any qualitative or quantitative synthesis**…..In the synthesis step, **low-quality studies may be excluded as a sensitivity analysis, but it is inappropriate to leave them out of synthesis completely**." [72]

As highlighted in point 3 (a), Eick et al. found that all epidemiological studies of PBDEs and neurodevelopment were downgraded when assessing risk of bias using the IRIS study evaluation tool, and the use of overall study rating led to rating of low/uninformative and thus each PBDEs study would be excluded.[73] Therefore, a large proportion of the evidence may not be included based on one methodological or reporting issue in a study when using the current approach to evaluation study quality in the IRIS program.

In its final recommendations in the recent consensus report "*The Use of Systematic Review in EPA's Toxic Substances Control Act Risk Evaluations (2021)"* the NAS recommended that EPA:

> "***Do not exclude studies based on risk of bias, study quality, or reporting quality***…. [74] and *"Use established tools for assessing risk of bias and study quality such as those developed for use by OHAT or the Navigation Guide, or, at a minimum, remove inappropriate appraisal criteria from the current tools*."[75]

We therefore strongly recommend against the exclusion of a study based on one "deficient" or "critically deficient" domain and support the use of either of the two methods recommended by the NAS, in order to better evaluate the quality of human epidemiological evidence before making final determinations on the hazards assessed in IRIS assessments.

d) The *Handbook* should require at least two reviewers to make risk of bias study determinations.

The Navigation Guide and OHAT both recommend the use of two assessors to conduct risk of bias assessments, independently, with conflicts resolved by consensus and the use of a third member for

[71] National Academies of Sciences, Engineering, and Medicine 2021. The Use of Systematic Review in EPA's Toxic Substances Control Act Risk Evaluations. Washington, DC: The National Academies Press. Pp 36 https://doi.org/10.17226/25952.
[72] National Academies of Sciences, Engineering, and Medicine 2021. The Use of Systematic Review in EPA's Toxic Substances Control Act Risk Evaluations. Washington, DC: The National Academies Press. Pp 40 https://doi.org/10.17226/25952.
[73] Eick SM, Goin DE, Chartres N, Lam J, Woodruff TJ. Assessing risk of bias in human environmental epidemiology studies using three tools: different conclusions from different tools. Systematic reviews. 2020;9(1):249.
[74] National Academies of Sciences, Engineering, and Medicine 2021. The Use of Systematic Review in EPA's Toxic Substances Control Act Risk Evaluations. Washington, DC: The National Academies Press. Pp 40 https://doi.org/10.17226/25952.
[75] National Academies of Sciences, Engineering, and Medicine 2021. The Use of Systematic Review in EPA's Toxic Substances Control Act Risk Evaluations. Washington, DC: The National Academies Press. Pp 40 https://doi.org/10.17226/25952.

arbitration if required.[76,77] This is based on the Institute of Medicine (IOM) highlighting the need of two assessors as *"Quality assurance and control are critical during data collection and extraction because of the substantial potential for errors."*[78]

The *Handbook* states that:

> *"However, based on assessment needs, the assessment team should make decisions about how many reviewers are needed. While more than one reviewer is ideal, there may be rare instances when one reviewer is acceptable, such as when the assessment needs to be conducted under a rapid time frame and the outcome being reviewed is unlikely to be a driver for the assessment."* [79]

In the 2014 report, *Review of EPA's Integrated Risk Information System (IRIS) Process* the NAS recommended (item 3 Chapter 2):

> "*When extracting data for evidentiary tables, EPA should use **at least two reviewers** to assess each study independently for risk of bias. The reliability of the independent coding should be calculated; if there is good agreement, multiple reviewers might not be necessary*. (emphasis added)" [80]

Further, the 2018 NAS evaluation of the IRIS Program, *Progress Toward Transforming the Integrated Risk Information System (IRIS) Program: A 2018 Evaluation* reported that "EPA also uses two people to complete the risk-of-bias evaluation." [81] It is concerning that IRIS is now proposing to deviate from what they reported to the NAS in 2018.

Therefore, IRIS needs to clearly state: 1) the reliability of any independent coding that is to be calculated, demonstrating good agreement to justify multiple reviewers are not necessary in their approach and 2) the time frame that IRIS would consider to be rapid enough to only use one reviewer. As it stands, the definition is arbitrary and could be misused. Further, it is unclear how EPA can identify any instances when *"the outcome being reviewed is unlikely to be a driver for the assessment"* [82] when EPA clearly states that "*Study evaluation occurs before extracting results and characterizing hazards associated with exposure to the chemical of interest.*"[83] Therefore, how will EPA know if the outcome will be a driver of the assessment if they have not extracted the study results?

---

[76] Woodruff TJ, Sutton P. The Navigation Guide systematic review methodology: a rigorous and transparent method for translating environmental health science into better health outcomes. Environmental Health Perspectives. 2014;122(10):A283.

[77] National Toxicology Program (NTP). Handbook for Conducting a Literature-Based Health Assessment Using OHAT Approach for Systematic Review and Evidence Integration. Pp38. Research Triangle Park, NC: National Toxicology Program; 2019.

[78] Institute of Medicine. 2011. Finding What Works in Health Care: Standards for Systematic Reviews. Washington, DC: The National Academies Press. https://doi.org/10.17226/13059.

[79] US EPA (2020) ORD Staff Handbook for Developing IRIS Assessments EPA/600/R-20/137. Pp 6-4 Available: https://cfpub.epa.gov/ncea/iris_drafts/recordisplay.cfm?deid=350086

[80] National Research Council. Review of EPA's Integrated Risk Information System (IRIS) Process. Washington, DC: The National Academies Press; 2014.

[81] National Academies of Sciences, Engineering, and Medicine. 2018. Progress Toward Transforming the Integrated Risk Information System (IRIS) Program: A 2018 Evaluation. Washington, DC: The National Academies Press

[82] US EPA (2020) ORD Staff Handbook for Developing IRIS Assessments EPA/600/R-20/137. Pp 6-4 Available: https://cfpub.epa.gov/ncea/iris_drafts/recordisplay.cfm?deid=350086

[83] US EPA (2020) ORD Staff Handbook for Developing IRIS Assessments EPA/600/R-20/137. Pp 6-2 Available: https://cfpub.epa.gov/ncea/iris_drafts/recordisplay.cfm?deid=350086

We therefore strongly recommend that IRIS use more than one reviewer to assess the risk of bias of the studies as recommended by two previously validated methods recommended by the NAS, the Navigation Guide[84] and OHAT.[85] Neither method recommends the use of a single reviewer to assess the risk of bias of the included studies.

4. The *Handbook* is unclear on the distinction between "Synthesis" and "Integration" of evidence. The *Handbook* would be improved by merging Chapter 9 ("Analysis and Synthesis of Human and Experimental Animal Data") into Chapter 11 ("Evidence Integration"), or alternately by moving the Chapter 11 content (regarding conclusions by evidence stream) into Chapter 9.

Chapter 9 of the *Handbook* is confusing because it is represented as presenting the approach to synthesis of evidence within an evidence stream, but the actual process for synthesis of evidence is not presented until Chapter 11. In addition, there is substantial overlap in content between Chapters 9 and 11. The discussion in Chapter 9 is very unstructured (beyond the presentation of considerations in Table 9-1) and does not indicate any particular output of the synthesis. The *Handbook* would be improved by merging Chapter 9 into Chapter 11, or alternately by moving the Chapter 11 content (regarding conclusions by evidence stream) into Chapter 9.

The *Handbook* should clarify how *"synthesis of an evidence stream"* is different from *"evidence integration within...the evidence stream."*[86] It would be much clearer to say that "Synthesis" is the development of conclusions for a particular evidence stream (corresponding to what is currently described as Step 1 on page 11-2, line 28); and "Integration" is the development of conclusions combining the evidence streams. This is exactly the delineation between the two terms recommended in the recent NAS consensus report on the use of systematic review in TSCA assessments "*The Use of Systematic Review in EPA's Toxic Substances Control Act Risk Evaluations (2021).*"[87] The *Handbook* seems to use "Integration" for both of these steps, which leaves the meaning of "Synthesis" unclear – and also leaves unclear what is done at the "Synthesis" stage of the assessment.

Table 9-1[88] and Table 11-2[89] present very similar information – descriptions of various considerations for "Evidence Integration" such as consistency, precision and coherence. The two tables should be combined. Further, for Table 9-1, for each consideration, there should be a clear statement regarding the nature of evidence that increases confidence in an effect (Table 11-2 does provide such statements, which are not provided in Table 9-1). Examples include:

[84] Woodruff TJ, Sutton P. The Navigation Guide systematic review methodology: a rigorous and transparent method for translating environmental health science into better health outcomes. Environ Health Perspect. 2014;122(10):1007-1014.

[85] National Toxicology Program (NTP). Handbook for Conducting a Literature-Based Health Assessment Using OHAT Approach for Systematic Review and Evidence Integration. Research Triangle Park, NC: National Toxicology Program; 2019.

[86] US EPA (2020) ORD Staff Handbook for Developing IRIS Assessments EPA/600/R-20/137. Pp 9-1, line 5. Available: https://cfpub.epa.gov/ncea/iris_drafts/recordisplay.cfm?deid=350086

[87] National Academies of Sciences, Engineering, and Medicine 2021. The Use of Systematic Review in EPA's Toxic Substances Control Act Risk Evaluations. Washington, DC: The National Academies Press. https://doi.org/10.17226/25952.

[88] US EPA (2020) ORD Staff Handbook for Developing IRIS Assessments EPA/600/R-20/137. Pp 9-3-5 Available: https://cfpub.epa.gov/ncea/iris_drafts/recordisplay.cfm?deid=350086

[89] US EPA (2020) ORD Staff Handbook for Developing IRIS Assessments EPA/600/R-20/137. Pp 11-10-12. Available: https://cfpub.epa.gov/ncea/iris_drafts/recordisplay.cfm?deid=350086

- Biological gradient/dose-response: include a statement such as "evidence of a dose-response gradient increases confidence in an effect (but lack of such a gradient does not indicate reduced confidence)."
- Coherence: include a statement that greater coherence increases confidence in an effect.
- Mechanistic evidence: include a statement such as "Mechanistic evidence that helps explain the biological process by which the target chemical may produce the adverse outcome increases confidence in that outcome."

Several additional revisions should be made to Table 9-1. For "*Strength and precision*," [90] why are the considerations "magnitude of effect" and "precision" combined? These are treated as separate considerations in GRADE, the OHAT Handbook and the Navigation Guide. [91, 92,93] They are also treated as separate considerations in Table 11-1 (but again combined in Table 11-2). Also, in "*Strength and precision*", p-values are not the best indicator of precision and should not be used. Page 9-8 more appropriately refers to "narrow confidence bounds or smaller standard errors (SEs)" as indicators of precision and does not mention p-values. For "*Mechanistic evidence*,"[94] "apical" is not the correct term to use and it should be deleted; it implies that only apical effects may be the focus of IRIS assessments. The IRIS definition of an adverse health effect is:

> "*A biochemical change, functional impairment, or pathologic lesion that affects the performance of the whole organism, or reduces an organism's ability to respond to an additional environmental challenge*."[95]

The outcomes stated in IRIS PECO statements, and the critical effects in numerous IRIS assessments, are often not apical effects but precursors or early biological changes that are more sensitive indicators of toxicity. (This use of "apical" should also be changed on page 4-35.)

Further, clarification is needed regarding the outcome/product of the synthesis stage. [96] It appears that it may be only a statement of strengths and weaknesses of the studies, as well as a summary of their results – but it lacks any conclusions regarding strength/confidence of an evidence stream for a given endpoint.

---

[90] US EPA (2020) ORD Staff Handbook for Developing IRIS Assessments EPA/600/R-20/137. Pp 9-4. Available: https://cfpub.epa.gov/ncea/iris_drafts/recordisplay.cfm?deid=350086

[91] Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, Norris S, Falck-Ytter Y, Glasziou P, DeBeer H, Jaeschke R, Rind D, Meerpohl J, Dahm P, Schünemann HJ. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. Journal of Clinical Epidemiology. 2011;64(4):383-94. doi: 10.1016/j.jclinepi.2010.04.026

[92] Woodruff TJ, Sutton P. The Navigation Guide systematic review methodology: a rigorous and transparent method for translating environmental health science into better health outcomes. Environ Health Perspect. 2014;122(10):1007-1014.

[93] National Toxicology Program (NTP). Handbook for Conducting a Literature-Based Health Assessment Using OHAT Approach for Systematic Review and Evidence Integration. Research Triangle Park, NC: National Toxicology Program; 2019.

[94] US EPA (2020) ORD Staff Handbook for Developing IRIS Assessments EPA/600/R-20/137. Pp 9-5. Available: https://cfpub.epa.gov/ncea/iris_drafts/recordisplay.cfm?deid=350086

[95] US EPA Integrated Risk Information System (IRIS) Glossary. Available: https://sor.epa.gov/sor_internet/registry/termreg/searchandretrieve/glossariesandkeywordlists/search.do?details=&glossaryName=IRIS%20Glossary

[96] US EPA (2020) ORD Staff Handbook for Developing IRIS Assessments EPA/600/R-20/137. Pp 9-7, line 3. Available: https://cfpub.epa.gov/ncea/iris_drafts/recordisplay.cfm?deid=350086

In Table 11-3, an *"exposure response gradient is demonstrated"* should not be a requirement for a judgment of "robust," though it certainly increases confidence in the evidence. [97] In Tables 11-3[98] and 11-4[99] the category of "Slight" evidence is stated to include "*Strong mechanistic evidence… in the absence of other substantive data."* Strong mechanistic evidence often provides a strong basis to infer that an adverse outcome is expected to occur. Strong evidence for a mechanistic event that has a well-established relationship to an adverse outcome should be considered "Moderate" evidence, which would be consistent with treatment of strong mechanistic evidence in the OHAT Handbook. [100] A footnote to each table indicates that, due to further development of mechanistic assays, "strong mechanistic evidence" in the future would constitute "moderate" evidence. The footnote does not provide a defensible rationale for discounting current "strong mechanistic evidence."

Finally, in Table 11-5[101] the *Handbook* proposes use of entirely new terms as hazard descriptors: "evidence demonstrates" and "evidence indicates." Given that there are existing sets of hazard descriptors in use that are already familiar to the environmental health community, it is unclear why IRIS proposes to use these new, unfamiliar terms. The similarity of the terms "evidence demonstrates" and "evidence indicates" is likely to lead to substantial confusion and ineffective risk communication. We recommend using the terms "known to be toxic" and "probably toxic" for the two strongest categories of evidence as used in the Navigation Guide.[102]

5. The Handbook is not clear regarding important aspects of how the IRIS program derives Toxicity Values and should incorporate updated methods (Chapters 12 and 13).

Chapter 12 of the *Handbook* states that *"High or medium confidence studies…are highly preferred over low confidence studies"* for derivation of toxicity values. The domain judgments that may lead to classification of a study as "low confidence" are varied, and not all of these are necessarily related to the utility of a study for deriving toxicity values. [103] It is therefore inappropriate to exclude "low confidence" studies for toxicity value derivation without examination of the underlying reasons for a study being classified as "low confidence." Further, as discussed in Point 3 (c), if meta-analysis or other techniques for combining data for multiple studies are employed, it is not appropriate to exclude any of the evidence; rather, limitations such as potential risks of bias may be considered in sensitivity analyses. The recent NAS review of the TSCA systematic review methods *"The Use of Systematic Review in EPA's Toxic Substances Control Act Risk Evaluations (2021)."* recommends "not exclud[ing]e studies based on risk of bias, study quality, or reporting quality." [104]

---

[97] US EPA (2020) ORD Staff Handbook for Developing IRIS Assessments EPA/600/R-20/137. Pp 11-13-15 Available: https://cfpub.epa.gov/ncea/iris_drafts/recordisplay.cfm?deid=350086

[98] US EPA (2020) ORD Staff Handbook for Developing IRIS Assessments EPA/600/R-20/137. Pp 11-14 Available: https://cfpub.epa.gov/ncea/iris_drafts/recordisplay.cfm?deid=350086

[99] US EPA (2020) ORD Staff Handbook for Developing IRIS Assessments EPA/600/R-20/137. Pp 11-16 Available: https://cfpub.epa.gov/ncea/iris_drafts/recordisplay.cfm?deid=350086

[100] National Toxicology Program (NTP). Handbook for Conducting a Literature-Based Health Assessment Using OHAT Approach for Systematic Review and Evidence Integration. Pp 68-9. Research Triangle Park, NC: National Toxicology Program; 2019.

[101] US EPA (2020) ORD Staff Handbook for Developing IRIS Assessments EPA/600/R-20/137. Pp 22-24 Available: https://cfpub.epa.gov/ncea/iris_drafts/recordisplay.cfm?deid=350086

[102] Woodruff TJ, Sutton P. The Navigation Guide systematic review methodology: a rigorous and transparent method for translating environmental health science into better health outcomes. Environ Health Perspect. 2014;122(10):1007-1014.

[103] US EPA (2020) ORD Staff Handbook for Developing IRIS Assessments EPA/600/R-20/137. Pp 12-4 (also Table 12-2). Available: https://cfpub.epa.gov/ncea/iris_drafts/recordisplay.cfm?deid=350086

[104] National Academies of Sciences, Engineering, and Medicine 2021. The Use of Systematic Review in EPA's Toxic Substances Control Act Risk Evaluations. Washington, DC: The National Academies Press. Pp 40. https://doi.org/10.17226/25952.

For Table 12-2, the entry for *"Exposure route*" should acknowledge exposure metrics (e.g. a biomarker measured in blood or urine) independent of any particular route of exposure, in addition to oral and inhalation routes.[105] For *"Subject selection,"* IRIS should consider modifying this to *"Studies that provide risk estimates in the most susceptible groups or in diverse populations are preferred."*[106] For "Health outcomes", the recommended preference for outcomes with *"greater biological significance"* is confusing given that hazard assessments by IRIS and other risk assessment programs often prefer more subtle outcomes that may be more sensitive indicators of toxicity (including early biological changes and precursor events).[107]

In the Chapter 13 section on *"Developing candidate toxicity values"* states that for human variation it states

> "*The assessment accounts for variation in susceptibility across the human population and the possibility that the available data may not be representative of individuals who are most susceptible to the effect. If population-based data for the effect or for characterizing the internal dose are available, the potential for data-based adjustments for toxicodynamics or toxicokinetics is considered (U.S. EPA, 2014b).22 Further, "when sufficient data are available, an intraspecies UF either less than or greater than 10× may be justified (U.S. EPA, 2002b). However, a reduction from the default (10) is only considered in cases when there are dose-response data for the most susceptible population" (U.S. EPA, 2002b). This factor is reduced only if the POD is derived or adjusted specifically for susceptible individuals [not for a general population that includes both susceptible and nonsusceptible individuals; (U.S. EPA, 2002b), §4.4.5; (U.S. EPA, 1998), §4.2; (U.S. EPA, 1996b), §4; (U.S. EPA, 1994), §4.3.9.1; (U.S. EPA, 1991), §3.4].* ***Otherwise, a factor of 10 is generally used to account for this variation.***" (emphasis added)

A critical element of chemical risk assessment is taking into account the full range of variability in response to environmental chemical exposures across the entire human population. This ensures the protection of everyone, including those most susceptible (due to biological sensitivity) and/or vulnerable (increased sensitivity due to external factors like cumulative exposure to multiple chemical and non-chemical stressors). Many factors, including genetics, age/life stage of development, socioeconomic status (SES), and/or underlying health status can affect population variability in response to chemical exposures. However, current chemical risk assessment approaches do not adequately account for susceptible and vulnerable subgroups.

[105] US EPA (2020) ORD Staff Handbook for Developing IRIS Assessments EPA/600/R-20/137. Pp 12-6 Available: https://cfpub.epa.gov/ncea/iris_drafts/recordisplay.cfm?deid=350086

[106] US EPA (2020) ORD Staff Handbook for Developing IRIS Assessments EPA/600/R-20/137. Pp 12-6 Available: https://cfpub.epa.gov/ncea/iris_drafts/recordisplay.cfm?deid=350086

[107] US EPA (2020) ORD Staff Handbook for Developing IRIS Assessments EPA/600/R-20/137. Pp 12-7 Available: https://cfpub.epa.gov/ncea/iris_drafts/recordisplay.cfm?deid=350086

This issue has been reviewed by the National Academy of Sciences (NAS)[108,109,110] and scientific articles[111,112,113] which find that approaches used in the past need to be upgraded to better incorporate human variability into chemical risk assessment. For example, current methods do not account for *in utero* susceptibility to chemical exposures, despite ample scientific literature demonstrating increased sensitivity among developing fetuses and the potential for fetal origins of disease.[114,115,116] Modern approaches have also failed to consider multiple environmental, social, and/or economic stressors in human health risk assessment, and have neglected to account for additional vulnerabilities faced by marginalized communities, such as environmental injustice, psychosocial stress, racism and/or discrimination, lack of economic opportunity, food insecurity, and poorer access to education and health care, which can increase individual sensitivity to chemical exposures and shift the overall distribution of population health risks to the higher end of the spectrum.[117,118] Upgraded approaches would provide more adequate protection for sensitive subgroups, such as pregnant women, developing fetuses/neonates, children/adolescents, low SES communities, those with preexisting disease and/or genetic susceptibility, and those burdened by additional occupational and/or environmental exposures.[119,120,121] This would also align with the 2016 Frank Lautenberg Chemical Safety for the 21st Century Act (Lautenberg TSCA), which amended the 1976 Toxic Substances Control Act (TSCA), mandates protection of susceptible and highly exposed populations.[122] The adoption and use of

[108] National Research Council. Toxicity Testing in the 21st Century: A Vision and a Strategy [Internet]. Washington, D.C.: The National Academies Press; 2007. Available from: https://www.nap.edu/download/11970#

[109] National Research Council. Phthalates and Cumulative Risk Assessment: The Task Ahead [Internet]. Washington, D.C.: The National Academies Press; 2008 [cited 2011 Oct 15]. Available from: https://doi.org/10.17226/12528

[110] National Research Council. Science and Decisions: Advancing Risk Assessment [Internet]. Washington, D.C.: The National Academies Press; 2009 [cited 2011 Oct 15]. Available from: http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=202175

[111] Chiu WA, Rusyn I. Advancing chemical risk assessment decision-making with population variability data: challenges and opportunities. Mamm Genome. 2018;29:182–9.

[112] Axelrad DA, Setzer RW, Bateson TF, DeVito M, Dzubow RC, Fitzpatrick JW, et al. Methods for evaluating variability in human health dose–response characterization. Human and Ecological Risk Assessment: An International Journal. 2019;0:1–24.

[113] Bhat VS, Meek ME (Bette), Valcke M, English C, Boobis A, Brown R. Evolution of chemical-specific adjustment factors (CSAF) based on recent international experience; increasing utility and facilitating regulatory acceptance. Critical Reviews in Toxicology. 2017;47:733–53

[114] Koman PD, Singla V, Lam J, Woodruff TJ. Population susceptibility: A vital consideration in chemical risk evaluation under the Lautenberg Toxic Substances Control Act. PLoS Biol [Internet]. 2019 [cited 2020 Apr 30];17. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6715167/

[115] US EPA O. Review of the Reference Dose and Reference Concentration Processes Document [Internet]. 2002. Available from: https://www.epa.gov/risk/review-reference-dose-and-reference-concentration-processes-document

[116] Wignall JA, Shapiro AJ, Wright FA, Woodruff TJ, Chiu WA, Guyton KZ, et al. Standardizing Benchmark Dose Calculations to Improve Science-Based Decisions in Human Health Assessments. Environmental Health Perspectives. 2014;122:499–505.

[117] National Research Council. Science and Decisions: Advancing Risk Assessment [Internet]. Washington, D.C.: The National Academies Press; 2009 [cited 2011 Oct 15]. Available from: http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=202175

[118] Morello-Frosch R Zuk, M, Jerrett, M, Shamasunder, B, Kyle, AD. Racial & Ethnic Disparities: Understanding The Cumulative Impacts Of Inequalities In Environmental Health: Implications For Policy. Health Affairs. 2011;30:5879–87.

[119] National Research Council. Toxicity Testing in the 21st Century: A Vision and a Strategy [Internet]. Washington, D.C.: The National Academies Press; 2007. Available from: https://www.nap.edu/download/11970#

[120] National Research Council. Phthalates and Cumulative Risk Assessment: The Task Ahead [Internet]. Washington, D.C.: The National Academies Press; 2008 [cited 2011 Oct 15]. Available from: https://doi.org/10.17226/12528

[121] National Research Council. Science and Decisions: Advancing Risk Assessment [Internet]. Washington, D.C.: The National Academies Press; 2009 [cited 2011 Oct 15]. Available from: http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=202175

[122] Koman PD, Singla V, Lam J, Woodruff TJ. Population susceptibility: A vital consideration in chemical risk evaluation under the Lautenberg Toxic Substances Control Act. PLoS Biol [Internet]. 2019 [cited 2020 Apr 30];17. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6715167/

upgraded recommendations have been limited, likely due to a combination of corporate interference and inadequate resources to assess biological and social determinants of health in combination. [123,124,125]

There are multiple examples from the scientific literature showing that a default factor of 10 is not sufficient to ensure protection of sensitive subpopulations.[126] Additionally, as the California Office of Environmental Health and Hazards Assessment (OEHHA) recommends the default use of a 30-fold adjustment factor for intra-species variability (based on a higher default of 10 for toxicokinetic (TK) variability multiplied by the default of 3 for toxicodynamic (TD) variability, since 10 x 3 = 30), indicating increased TK susceptibility to benzo[a]pyrene, tetrachloroethylene (TCE), and other industrial chemicals among sensitive subgroups (*i.e.*, young infants, children, and highly exposed workers with specific gene interactions).[127] The CA default factor of 30 can be modified upwards or downwards depending on chemical specific information (*e.g.*, for benzene because of variability in metabolism and other sensitivities the non-cancer variability is 100). We therefore strongly recommend the IRIS program uses a default factor of 30-fold unless there is empirical evidence suggesting a lower factor is appropriate.

In the Chapter 13 section on *"Characterizing Uncertainty and Confidence in Toxicity Values,"* the *Handbook* states that *"EPA will continue to seek improvements in its dose-response methods."* Among the important improvements that should be discussed in the *Handbook* would be use of established methods (e.g., probabilistic assessment) to quantify the level of risk associated with each noncancer toxicity value (e.g., RfD) in an EPA IRIS assessment,[128,129,130,131] as well as estimates of the population exposure levels associated with different levels of risk (e.g., 1 in 100,000; 1 in 10,000; 1 in 1,000) for all identified health effects.[132] These calculations will give decision-makers better information about how exposures in the population translate into population risks for different health endpoints, and enable quantification of health benefits for risk management alternatives that may reduce exposure. [133] The 2009 NAS report *Science and Decisions* presented several recommendations to EPA on adopting methods for calculating risk-specific doses for non-cancer health effects, as an improvement on traditional approaches to reference value derivation.[134] In addition, the NAS 2013 report providing advice for the planned IRIS assessment of inorganic arsenic recommended derivation of risk-specific

[123] Janssen S, Sass J, Solomon G, Schettler T. Strengthening toxic chemical risk assessments report [Internet]. NRDC; 2012 [cited 2012 Feb 12]. Available from: https://www.nrdc.org/sites/default/files/strengthening-toxic-chemical-risk-assessments-report.pdf

[124] Morello-Frosch R Zuk, M, Jerrett, M, Shamasunder, B, Kyle, AD. Racial & Ethnic Disparities: Understanding The Cumulative Impacts Of Inequalities In Environmental Health: Implications For Policy. Health Affairs. 2011;30:5879–87.

[125] Koman PD, Singla V, Lam J, Woodruff TJ. Population susceptibility: A vital consideration in chemical risk evaluation under the Lautenberg Toxic Substances Control Act. PLoS Biol [Internet]. 2019 [cited 2020 Apr 30];17. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6715167/

[126] National Research Council. Science and Decisions: Advancing Risk Assessment [Internet]. Washington, D.C.: The National Academies Press; 2009 [cited 2011 Oct 15]. Table 4-1Available from: http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=202175

[127] Air Toxicology and Epidemiology Branch, Office of Environmental Health Hazard Assessment, California Environmental Protection Agency. Technical Support Document For the Derivation of Noncancer Reference Exposure Levels [Internet]. 2008. Available from: https://oehha.ca.gov/media/downloads/crnr/noncancertsdfinal.pdf

[128] Ginsberg GL. Cadmium risk assessment in relation to background risk of chronic kidney disease. Toxicol Environ Health A. 2012;75(7):374-90.doi: 10.1080/15287394.2012.670895.

[129] Chiu WA, Axelrad DA, Dalaijamts C, Dockins C, Shao K, Shapiro AJ, Paoli G. Beyond the RfD: broad application of a probabilistic approach to improve chemical dose-response assessment for non-cancer effects. Environmental Health Perspectives, 2018 June;126(6):067009. doi:10.1289/EHP3368

[130] Hattis D, Baird S, Goble R. A straw man proposal for a quantitative definition of the RfD. Drug Chem Toxicol. 2002;25:403–36.

[131] WHO. Guidance document on evaluating and expressing uncertainty in hazard characterization. Harmonization project document 11. [Internet]. WHO; 2014. Available from: http://www.inchem.org/documents/harmproj/harmproj/harmproj11.pdf

[132] McGartland, A., Revesz, R., Axelrad, D. A., Dockins, C., Sutton, P., & Woodruff, T. J. (2017). Estimating the health benefits of environmental regulations. Science, 357(6350), 457-458.

[133] McGartland, A., Revesz, R., Axelrad, D. A., Dockins, C., Sutton, P., & Woodruff, T. J. (2017). Estimating the health benefits of environmental regulations. Science, 357(6350), 457-458.

[134] National Research Council. Science and Decisions: Advancing Risk Assessment. Washington, D.C.:National Academies Press; 2009. Ch 5

doses instead of an RfD for that assessment.[135] We note that IRIS has previously committed, in a presentation to the NAS, to apply the International Programme in Chemical Safety (IPCS) probabilistic methodology in IRIS assessments for these purposes,[136] and the NAS 2018 report on IRIS considered this commitment to be responsive to earlier NAS recommendations concerning uncertainty analysis for IRIS toxicity values.[137] Further, IRIS staff have published an application of the IPCS method to acrolein.[138] The *Handbook* should follow through on this previous IRIS commitment by adding text indicating that derivation of risk-specific doses for non-cancer effects will become a standard part of assessing uncertainty in IRIS toxicity values.

The Chapter 13 section on "Characterizing Uncertainty and Confidence in Toxicity Values" also includes a brief mention of model averaging and other "more model-robust approaches" to dose-response assessment.[139] Model averaging and related techniques have been recommended by the NAS and by members of EPA's Science Advisory Board (SAB),[140,141,142] so it would be useful to expand on this brief text to better inform the public regarding when these approaches are considered by IRIS.
Chapter 13 also states *"Basis of the POD: A modeled BMDL is preferred over a NOAEL, which is in turn preferred over a LOAEL."*[143] This statement requires clarification; presuming that the choice is between values derived from separate studies, a NOAEL should not necessarily be preferred to a LOAEL when the LOAEL is at a lower dose. The NOAEL may be from a relatively insensitive study that misses effects occurring at lower doses; use of such a NOAEL as a point of departure in preference to a lower LOAEL from a different, more sensitive study may result in a reference value that is not health-protective for effects observed at the LOAEL.

Finally, the concluding section of Chapter 13 on "Overall Toxicity Values" is incomplete.[144] It should reiterate that a reference value is intended to be protective for all endpoints (and thus would point to selection of a candidate value for the endpoint found to provide the lowest reference value), and that the cancer potency should represent risk from all tumor sites combined.

6. The *ORD Staff Handbook for Developing IRIS Assessments* should consider financial conflicts of interest as a potential source of bias in research.

In section 9.4.3 Reporting or Publication Bias, the *Handbook* states:

---

[135] National Research Council. 2013. *Critical Aspects of EPA's IRIS Assessment of Inorganic Arsenic: Interim Report*. Washington, DC: The National Academies Press. https://doi.org/10.17226/18594.

[136] National Academies of Sciences, Engineering, and Medicine. 2018. Progress Toward Transforming the Integrated Risk Information System (IRIS) Program: A 2018 Evaluation. Washington, DC: The National Academies Press. Pp 90-91

[137] National Academies of Sciences, Engineering, and Medicine. 2018. Progress Toward Transforming the Integrated Risk Information System (IRIS) Program: A 2018 Evaluation. Washington, DC: The National Academies Press. Pp 119

[138] Blessinger T, Davis A, Chiu WA, et al. Application of a unified probabilistic framework to the dose-response assessment of acrolein. *Environment international.* 2020;143:105953.

[139] US EPA (2020) ORD Staff Handbook for Developing IRIS Assessments EPA/600/R-20/137. Pp 124, line 2 Available: https://cfpub.epa.gov/ncea/iris_drafts/recordisplay.cfm?deid=350086

[140] National Research Council. Review of EPA's Integrated Risk Information System (IRIS) Process. Pp 124. Washington, DC: The National Academies Press; 2014.

[141] National Academies of Sciences, Engineering, and Medicine. 2018. Progress Toward Transforming the Integrated Risk Information System (IRIS) Program: A 2018 Evaluation. Pp 119 Washington, DC: The National Academies Press

[142] SAB: Consultation on EPA's Consolidated Human Toxicity Assessment Guideline. EPA-SAB-20-008. July 31, 2020. Available https://yosemite.epa.gov/sab/sabproduct.nsf/LookupWebProjectsCurrentBOARD/5C5E7E088099A0D7852585B90062D739/$File/EPA-SAB-20-008.pdf

[143] US EPA (2020) ORD Staff Handbook for Developing IRIS Assessments EPA/600/R-20/137. Pp 13-19, line 23 Available: https://cfpub.epa.gov/ncea/iris_drafts/recordisplay.cfm?deid=350086

[144] US EPA (2020) ORD Staff Handbook for Developing IRIS Assessments EPA/600/R-20/137. Pp 13-20, line 11 Available: https://cfpub.epa.gov/ncea/iris_drafts/recordisplay.cfm?deid=350086

> *"A potential conflict of interest (COI) by one or more authors of a study may contribute to reporting or publication bias (Guyatt et al., 2011b). While IRIS does not formally include COIs as a component in the evaluation of bias and sensitivity of study outcomes, funding source and a report of a COI by the authors can be noted for a study in Health Assessment Workspace Collaborative (HAWC). When there is evidence that a conflict of interest is may be present, a more careful assessment of the consistency of study results, publication and reporting bias may be merited for a health effect*.[145]

Firstly, this paragraph needs to be corrected as it does not make sense "*When there is evidence that a conflict of interest is may be present"* Secondly, the NAS in its 2014 *Review of EPA's Integrated Risk Information System (IRIS) Process* found that "Funding sources should be considered in the risk-of-bias assessment conducted for systematic reviews that are part of an IRIS assessment."[146] The NAS recommendation is based on empirical evidence that non-methodological characteristics, including author conflicts of interest (COI) and industry sponsorship, can also influence the findings of a study. It has been demonstrated across several areas of research that even when studies have the same methodological risk of bias or internal validity, studies with industry sponsorship are associated with more favorable outcomes towards the study sponsor.[147,148] In studies of harmful exposures such as chemicals, this funding bias would be expected to be associated with a bias towards the null (finding that the chemical does not have a toxic effect). The need to account for this potential bias is empirically supported. Thus, it is quite concerning that the *Handbook* only states that *"funding source and a report of a COI by the authors* ***<u>can be noted for a study</u>***" and there is not a standard practice recommended in the IRIS program. Further, industry sponsorship can bias research through various mechanisms, including how they frame the research questions, design and conduct a study, selectively report the results, code events, analyze the study data and spin conclusions.[149,150,151,152] Thus limiting the assessment of how industry can bias a study to only publication and reporting is insufficient.

The ROBINS-I tool on which EPA's IRIS Assessments study evaluation of epidemiological evidence is based, focuses on a narrow definition of bias based on a methodological flaw that may lead to an error in quantitative effect estimates.[153] The Navigation Guide assesses financial conflicts of interest as a separate domain within its risk of bias evaluation. [154] OHAT, however, *"collects information about funding source during data extraction and considers it at multiple points in the evaluation"* as financial COI can be accounted for at various time points throughout the review process including in an

[145] US EPA (2020) ORD Staff Handbook for Developing IRIS Assessments EPA/600/R-20/137. Pp 9-14 Available: https://cfpub.epa.gov/ncea/iris_drafts/recordisplay.cfm?deid=350086

[146] National Research Council. Review of EPA's Integrated Risk Information System (IRIS) Process. Page. 79. Washington, DC: National Academies Press; 2014

[147] Lundh A, Lexchin J, Mintzes B, Schroll JB, Bero L. Industry sponsorship and research outcome. Cochrane Database Syst Rev. 2017;2:MR000033.

[148] White J, Bero LA. Corporate manipulation of research: strategies are similar across five industries. Stanford Law Policy Rev. 2010;21(1):105–34.

[149] Odierna DH, Forsyth SR, White J, et al. The cycle of bias in health research: a framework and toolbox for critical appraisal training. Account Res. 2013;20(2):127-41.

[150] Fabbri A, Lai A, Grundy Q, et al. The Influence of Industry Sponsorship on the Research Agenda: A Scoping Review. Am J Public Health. 2018;108(11):e9-e16

[151] Psaty BM, Prentice RL. Minimizing bias in randomized trials: the importance of blinding. JAMA. 2010;304(7):793-4

[152] Psaty BM, Kronmal RA. Reporting mortality findings in trials of rofecoxib for Alzheimer disease or cognitive impairment: a case study based on documents from rofecoxib litigation. JAMA. 2008;299(15):1813-7

[153] Bero, L., Chartres, N., Diong, J. et al. The risk of bias in observational studies of exposures (ROBINS-E) tool: concerns arising from application to observational studies of exposures. Syst Rev 7, 242 (2018) doi:10.1186/s13643-018-0915-2

[154] Woodruff TJ, Sutton P. The Navigation Guide systematic review methodology: a rigorous and transparent method for translating environmental health science into better health outcomes. Environmental Health Perspectives. 2014;122(10):A283.

assessment of the selective reporting of results, publication bias, and in assessing inconsistency in a body of evidence. [155,156,157] However, it is not always possible to identify such biases due to a lack of study registries and lack of publication of protocols for the types of evidence used in systematic reviews to assess the harms of chemicals. Therefore, the simplest way to identify such potential biases is by assessing funding source and author COI as a specific risk of bias domain as recommended by the Navigation Guide.[158]

Importantly, including funding as a risk of bias domain *does not lead to the exclusion of industry sponsored studies*, it only means identifying it as a domain of potential bias and then evaluating its impact on the overall quality of the body of evidence. Therefore, we again support the recommendation made by the NAS to IRIS that "*Funding sources should be considered in the risk-of-bias assessment conducted for systematic reviews that are part of an IRIS assessments*."

[155] National Toxicology Program (NTP). Handbook for Conducting a Literature-Based Health Assessment Using OHAT Approach for Systematic Review and Evidence Integration. Pp38. Research Triangle Park, NC: National Toxicology Program; 2019.

[156] Viswanathan M, Ansari M, Berkman N, Chang S, Hartling L, McPheeters L, Santaguida P, Shamliyan T, Singh K, Tsertsvadze A, Treadwell J. Assessing the Risk of Bias of Individual Studies

[157] National Toxicology Program. Handbook for Conducting a Literature-Based Health Assessment Using OHAT Approach for Systematic Review and Evidence Integration. In: U.S. Department of Health and Human Services, editor.: Office of Health Assessment and Translation, Division of National Toxicology Program, National Institute of Environmental Health Sciences; 2019.

[158] Woodruff TJ, Sutton P. The Navigation Guide systematic review methodology: a rigorous and transparent method for translating environmental health science into better health outcomes. Environmental Health Perspectives. 2014;122(10):A283.