February 18, 2022

Comments on the Draft Toxic Substances Control Act (TSCA) Systematic Review Protocol

Submitted online via Regulations.gov to docket EPA-HQ-OPPT-2021-0414

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 *Draft Toxic Substances Control Act (TSCA) Systematic Review Protocol, [1]* hereafter referred to as the *2021 Draft TSCA Method.* As one of the leading academic experts in environmental health systematic review, the University of California San Francisco's Program on Reproductive Health and the Environment (PRHE) has previously provided detailed comments to EPA and to the National Academies of Sciences, Engineering, and Medicine (NASEM) about the TSCA systematic review methodology.[2]

We want to start by acknowledging and appreciating the challenges that EPA is facing, with rapidly approaching deadlines under the Frank R. Lautenberg Chemical Safety for the 21st Century Act, which amended the Toxic Substances Control Act (amended TSCA). Rather than being able to start with a clean slate, current EPA leadership has inherited much of its work and thus has had to adapt.

In May 2018, EPA released a document[3] (referred to in these comments as the *2018 TSCA Method*) outlining its systematic review process approximately 17 months[4] after initiating risk evaluations for the first 10 chemicals mandated by amended TSCA. Amended TSCA requires EPA to make decisions about chemical risks based on the "best available science" and the "weight of the scientific evidence."[5] EPA defines "weight of the scientific evidence" as "…a systematic review method, applied in a manner suited to the nature of the evidence or decision, that uses a pre-established protocol to comprehensively, objectively, transparently, and consistently identify and evaluate each stream of evidence, including strengths, limitations, and relevance of each study and to integrate evidence as necessary and appropriate based upon strengths, limitations, and relevance.'' [6]

[1] US EPA. (2021). Draft systematic review protocol supporting TSCA risk evaluations for chemical substances version 1.0: A generic TSCA systematic review protocol with chemical-specific methodologies. https://www.epa.gov/system/files/documents/2021-12/draft-systematic-review-protocol-supporting-tsca-risk-evaluations-for-chemical-substances_0.pdf

[2] US EPA. (2018). Application of systematic review in TSCA risk evaluations. https://www.epa.gov/sites/default/files/2018-06/documents/final_application_of_sr_in_tsca_05-31-18.pdf

[3] US EPA. (2018). Application of systematic review in TSCA risk evaluations. https://www.epa.gov/sites/default/files/2018-06/documents/final_application_of_sr_in_tsca_05-31-18.pdf

[4] EPA. (2022). Risk evaluations for existing chemicals under TSCA. https://www.epa.gov/assessing-and-managing-chemicals-under-tsca/risk-evaluations-existing-chemicals-under-tsca#:~:text=Initiation%20of%20the%20Risk%20Evaluation,-A%20risk%20evaluation&text=With%20the%20passage%20of%20the,announced%20on%20December%2019%2C%202016.

[5] 15 USC §2625 (h)-(i)

[6] 40 CFR 702.33

In August 2018, PRHE submitted comments to EPA identifying a number of fundamental problems with the *2018 TSCA Method* and how, if left unresolved, it would jeopardize the ongoing risk evaluations and harm public health. [7] Despite this, EPA continued to use the *2018 TSCA Method* and we continued to identify how its flaws would negatively impact the risk evaluation process in public comments[8,9,10,11,12,13,14] and in peer reviewed publications.[15,16] Several of the fundamental systematic review deficiencies we identified in the *2018 TSCA Method* have also been identified by the EPA's Science Advisory Committee on Chemicals (SACC) in its peer review of the draft risk evaluations of Pigment Violet 29 (PV29),[17,18] 1,4-dioxane and the cyclic aliphatic bromide cluster (HBCD),[19,20] 1-bromopropane (1-BP)[21,22] and n-methylpyrrolidone.[23] In 2020, the NASEM announced it would be undertaking a review of the *2018 TSCA Method*. In June 2020, we provided comments showing how the *2018 TSCA Method* did not represent the "best available science" as required by TSCA and recommending EPA use a validated systematic review methodology.[24] In February 2021, the NASEM

[7] US EPA. (2018). Problem Formulations for Risk Evaluations To Be Conducted Under Toxic Substances Control Act, and General Guiding Principles To Apply Systematic Review in TSCA Risk Evaluations. Comment submitted by Veena Singla, Associate Director, Science and Policy, Program on Reproductive Health and the Environment, University of California, San Francisco et al. Available: https://www.regulations.gov/document?D=EPA-HQ-OPPT-2018-0210-0107

[8] US EPA. (2019). Draft Toxic Substances Control Act Risk Evaluations: Color Index Pigment Violet 29. Comment submitted by Veena Singla, Associate Director, Science and Policy, Program on Reproductive Health and the Environment, University of California, San Francisco (UCSF) et al. Available: https://www.regulations.gov/document?D=EPA-HQ-OPPT-2018-0604-0014

[9] US EPA. (2019). Draft Toxic Substances Control Act Risk Evaluations: 1,4-Dioxane and Cyclic Aliphatic Bromide Cluster (HBCD). Comment submitted by Swati Rayasam, Science Associate, Program on Reproductive Health and the Environment, University of California, San Francisco et al. Available: https://www.regulations.gov/document?D=EPA-HQ-OPPT-2019-0237-0059

[10] US EPA. (2019). Draft Toxic Substances Control Act Risk Evaluations: 1-Bromopropane. Comment submitted by Swati Rayasam, Science Associate, Program on Reproductive Health and the Environment, University of California, San Francisco et al. Available: https://www.regulations.gov/document?D=EPA-HQ-OPPT-2019-0235-0053

[11] US EPA. (2019). Draft Toxic Substances Control Act Risk Evaluations: Methylene Chloride. Comment submitted by Swati Rayasam, Science Associate, Program on Reproductive Health and the Environment, University of California, San Francisco et al. Available: https://www.regulations.gov/document?D=EPA-HQ-OPPT-2019-0437-0069

[12] US EPA. (2020). Draft Toxic Substances Control Act Risk Evaluations: N-Methyl-2-pyrrolidone (NMP). Comment submitted by Swati Rayasam, Science Associate, Program on Reproductive Health and the Environment, University of California, San Francisco et al. Available: https://www.regulations.gov/document?D=EPA-HQ-OPPT-2019-0236-0048

[13] US EPA. (2020). Draft Toxic Substances Control Act Risk Evaluations: Carbon Tetrachloride. Comment submitted by Swati Rayasam, Science Associate, Program on Reproductive Health and the Environment, University of California, San Francisco et al. Available: https://www.regulations.gov/document?D=EPA-HQ-OPPT-2019-0499-0041

[14] US EPA. (2020). Draft Toxic Substances Control Act Risk Evaluations: Carbon Tetrachloride. Comment submitted by Nicholas Chartres, Associate Director, Program on Reproductive Health and the Environment, University of California, San Francisco et al. Available: https://www.regulations.gov/document?D=EPA-HQ-OPPT-2019-0501-0087

[15] Singla V, Sutton P, Woodruff TW. (2019) The Environmental Protection Agency Toxic Substances Control Act Systematic Review Method May Curtail Science Used to Inform Policies, With Profound Implications for Public Health. Am J Public Health. doi: 10.2105/AJPH.2019.305068

[16] Eick, S.M., Goin, D.E., Chartres, N. et al. (2020). Assessing risk of bias in human environmental epidemiology studies using three tools: different conclusions from different tools. Syst Rev 9, 249. Available: https://doi.org/10.1186/s13643-020-01490-8

[17] US EPA. (2019). Peer Review for EPA Draft Risk Evaluation of C.I. Pigment Violet 29 (PV29). Available: https://www.regulations.gov/docket?D=EPA-HQ-OPPT-2018-0604 D=EPA-HQ-OPPT-2018-0604

[18] US EPA. (2019). Draft Toxic Substances Control Act Risk Evaluations: 1-Bromopropane. Comment submitted by Swati Rayasam, Science Associate, Program on Reproductive Health and the Environment, University of California, San Francisco et al. Available: https://www.regulations.gov/document?D=EPA-HQ-OPPT-2019-0235-0053

[19] US EPA. (2019). Peer Review for EPA Draft Risk Evaluations for 1,4-Dioxane and Cyclic Aliphatic Bromide Cluster (HBCD Available: https://www.regulations.gov/document?D=EPA-HQ-OPPT-2019-0237-0064

[20] US EPA. (2019). Proposed High-Priority Substance Designation Under the Toxic Substances Control Act. Comment submitted by Swati Rayasam, Science Associate, Program on Reproductive Health and the Environment, University of California, San Francisco et al. Available: https://www.regulations.gov/document?D=EPA-HQ-OPPT-2018-0430-0015

[21] US EPA. (2019). Peer Review for the United States Environmental Protection Agency (EPA) Draft Risk Evaluation for 1-Bromopropane (1-BP). Available: https://www.regulations.gov/document?D=EPA-HQ-OPPT-2019-0235-0061

[22] US EPA. (2019). Peer Review for the United States Environmental Protection Agency (EPA) Draft for Risk Methylene Chloride. Available: https://www.regulations.gov/document?D=EPA-HQ-OPPT-2019-0437-0080

[23] US EPA. (2019). Peer Review for the United States Environmental Protection Agency (EPA) Draft for N-Methylpyrrolidone (NMP). Available: https://www.regulations.gov/document?D=EPA-HQ-OPPT-2019-0236-0066

[24] UCSF PRHE. (2020). Comments from Academics, Scientists and Clinicians on: The Application of Systematic Review in TSCA Risk Evaluations. Available: https://prhe.ucsf.edu/sites/g/files/tkssra341/f/wysiwyg/NAS_TSCA%20SR%20Method%20comments.pdf

provided extensive comments and recommendations, including that there was "strong consensus," that the *2018 TSCA Method* "did not meet the standards of systematic review methodology."[25] This prompted EPA to make a statement that it would no longer be using this systematic review method and was in the process of developing a new approach incorporating recommendations from the NASEM.[26] However, in EPA's new *2021 Draft TSCA Method*, EPA has failed to adhere to the NASEM recommendations in fundamental ways as we review below.

We continue to highlight, similarly to the NASEM, that there are existing methods that have been reviewed and tested for environmental chemicals that could be implemented by EPA right now to improve the basis of their decision making and reduce the burden and time to making decisions. Both PRHE's Navigation Guide and the National Toxicology Program's Office of Health Assessment and Translation *OHAT Approach for Systematic Review and Evidence Integration for Health Effects Evaluations* (OHAT) method have been used or recommended by the NASEM[27,28,29] and demonstrated in case studies in the peer-reviewed literature.[30,31,32,33,34, 35,36,37] Further, EPA's Office of Research and Development has adopted a systematic review methodology as part of its Integrated Risk Information System (IRIS) program.[38] While we have provided comments around some inadequacies in the current IRIS method, we maintain that it is a fundamentally stronger method than the *2021 Draft TSCA Method*. We urge EPA to cease use of the *2021 Draft TSCA Method* and follow the NASEM recommendation that

[25] NASEM. (2021). The use of systematic review in EPA's Toxic Substances Control Act Risk Evaluations. https://www.nap.edu/catalog/25952/the-use-of-systematic-review-in-epas-toxic-substances-control-act-risk-evaluations

[26] EPA Press Office. (2021). EPA commits to strengthening science used in chemical risk evaluations. https://www.epa.gov/newsreleases/epa-commits-strengthening-science-used-chemical-risk-evaluations

[27] 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. Washington, D.C.: The National Academies Press; 2011

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

[29] 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; 2018.

[30] Johnson PI, Sutton P, Atchley DS, Koustas E, Lam J, Sen S, Robinson KA, Axelrad DA, Woodruff TJ. The Navigation Guide - evidence-based medicine meets environmental health: systematic review of human evidence for PFOA effects on fetal growth. Environ Health Perspect. 2014;122(10):1028-39. Epub 2014/06/27. doi: 10.1289/ehp.1307893. PubMed PMID: 24968388; PMCID: 4181929.

[31] Koustas E, Lam J, Sutton P, Johnson PI, Atchley DS, Sen S, Robinson KA, Axelrad DA, Woodruff TJ. The Navigation Guide - evidence-based medicine meets environmental health: systematic review of nonhuman evidence for PFOA effects on fetal growth. Environ Health Perspect. 2014;122(10):1015-27. Epub 2014/06/27. doi: 10.1289/ehp.1307177. PubMed PMID: 24968374; PMCID: 4181920.

[32] Lam J, Koustas E, Sutton P, Johnson PI, Atchley DS, Sen S, Robinson KA, Axelrad DA, Woodruff TJ. The Navigation Guide - evidence-based medicine meets environmental health: integration of animal and human evidence for PFOA effects on fetal growth. Environ Health Perspect. 2014;122(10):1040-51. Epub 2014/06/27. doi: 10.1289/ehp.1307923. PubMed PMID: 24968389; PMCID: 4181930

[33] Vesterinen H, Johnson P, Atchley D, Sutton P, Lam J, Zlatnik M, Sen S, Woodruff T. The relationship between fetal growth and maternal glomerular filtration rate: a systematic review. J Maternal Fetal Neonatal Med. 2014:1-6. Epub Ahead of Print; PMCID: 25382561.

[34] Johnson PI, Koustas E, Vesterinen HM, Sutton P, Atchley DS, Kim AN, Campbell M, Donald JM, Sen S, Bero L, Zeise L, Woodruff TJ. Application of the Navigation Guide systematic review methodology to the evidence for developmental and reproductive toxicity of triclosan. Environ Int. 2016;92-93:716-28. doi: 10.1016/j.envint.2016.03.009. PubMed PMID: 27156197.

[35] Lam J, Sutton P, Halladay A, Davidson LI, Lawler C, Newschaffer CJ, Kalkbrenner A, Joseph J. Zilber School of Public Health, Windham GC, Daniels N, Sen S, Woodruff TJ. Applying the Navigation Guide Systematic Review Methodology Case Study #4: Association between Developmental Exposures to Ambient Air Pollution and Autism. PLoS One. 2016;21(11(9)). doi: 10.1371/journal.pone.0161851.

[36] Lam J, Lanphear B, Bellinger D, Axelrad D, McPartland J, Sutton P, Davidson LI, Daniels N, Sen S, Woodruff TJ. Developmental PBDE exposure and IQ/ADHD in childhood: A systematic review and meta-analysis. Environmenal Health Perspectives. 2017;125(8). doi: doi: 10.1289/EHP1632.

[37] Lam J, Koustas E, Sutton P, Cabana M., Whitaker E., Padula A, Vesterinen H, Daniels N, Woodruff TJ. Applying the Navigation Guide: Case Study #6. Association Between Formaldehyde Exposures and Asthma. In preparation. 2019.

[38] 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; 2018.

"the methods for developing IRIS assessments can serve as a model for other EPA programs that are implementing systematic review methods."[39]

With so much at stake, we continue to be concerned that EPA is using its limited time and resources to maintain a deeply flawed systematic review method instead of applying one of the three existing and validated methods.[40]

The *2021 Draft TSCA Method* indicates that this updated method is in line with EPA's statutory requirements under amended TSCA and consistent with the best available science, but it is not. The *2021 Draft TSCA Method* is inconsistent with current, established, best available empirical methods for systematic review, and while this updated protocol incorporates pieces of the three aforementioned methods, partial incorporation of a validated systematic review methodology does not equate to using the "best available science" as we detail below. To ensure that the 23 ongoing risk evaluations will not continue to be biased and underestimate risk, and to decrease EPA's resource and time burden, it should immediately adhere to the NASEM recommendation to use an established method (such as the IRIS Method after making necessary improvements). If EPA wants to use the "best available science" for its risk evaluations under amended TSCA, it will adopt one method and adhere to it from start to finish. The continued use of the *2021 Draft TSCA Method*, even with the updates to the prior TSCA method incorporated, continues to underestimate and undervalue risks from industrial chemicals – leaving the public at risk from harmful chemical exposures.

Our comments address the following main issues:

1. **EPA still does not adhere to best practices on protocol development for systematic reviews.**
    a. **EPA failed to make to make the *2021 Draft TSCA Method* available while it continued conducting systematic reviews for the current risk evaluations, which casts reasonable doubt on the validity of these evaluations.**
    b. **EPA fails to include protocol development in the *2021 Draft TSCA Method* as an explicit step in the systematic review process.**
    c. **EPA still needs to create chemical-specific protocols for each systematic review it conducts for TSCA risk evaluations.**

    We recommend:
    - **Making draft systematic review protocols available prior to conducting risk evaluations.**
    - **Including protocol development as an explicit step in the systematic review process.**
    - **Developing chemical-specific protocols for each of the 23 ongoing risk evaluations conducted under amended TSCA.**

2. **EPA's PECO statements do not provide a clear, consistent, reasonable, and appropriate basis for determining which studies are relevant and which studies are not relevant for hazard assessment.**

[39] NASEM. (2022). Review of US EPA's ORD staff handbook for developing IRIS assessments: 2020 version. pp 12. https://www.nap.edu/catalog/26289/review-of-us-epas-ord-staff-handbook-for-developing-iris-assessments

[40] Cruden, EA. (2021). Burnout, expertise gaps plague EPA chemicals office. *E&E News.* Available: https://www.eenews.net/articles/burnout-expertise-gaps-plague-epa-chemicals-office/

a. **Most of EPA's PECO statements appear to inappropriately exclude important toxicity endpoints from the TSCA hazard assessments.**
b. **EPA must use the same PECO statement for both title-abstract screening and full-text review.**
c. **There are unexplained and unwarranted inconsistencies across the PECO statements for different chemicals.**
d. **PECOs should explicitly list metabolites of target chemicals where appropriate.**

We recommend:

- **Including studies of biochemical changes and other early biological indicators of hazards as relevant to the human and animal evidence streams, removing the restriction that limits relevant health effects studies to those with apical or organ-level outcomes.**
- **Applying a single, peer-reviewed PECO statement throughout the process for each risk evaluation, including for both title-abstract screening and full-text review.**

3. **The method outlined in the *2021 Draft TSCA Method* fails to adhere to best practice for conducting data extraction for systematic review.**
    a. **The draft document provides insufficient detail concerning data extraction for hazard studies.**
    b. **Data extraction should be conducted independently by two separate reviewers.**

We recommend:

- **Publishing details on data extraction, such as a template, for health effects studies.**
- **Conducting data extraction by two reviewers operating independently in parallel.**

4. **The *2021 Draft TSCA Method's* approach to study evaluation is incompatible with validated best practice methods already being implemented in environmental health and fails to adhere to the NASEM recommendations to EPA in its review *of Application of Systematic Review in TSCA Risk Evaluations* in fundamental ways, including:**
    a. **Continues to use an inappropriate quantitative scoring method almost identical to the approach in the *2018 TSCA Method.*[41]**
    b. **Continues to inappropriately use an overall study quality rating.**
    c. **Continues to exclude or downgrade research based on one "critically deficient" methodological limitation.**
    d. **Continues to conflate how well a study is reported with how well the underlying research was conducted and continues to include inappropriate appraisal criteria, such as statistical power.**
    e. **Continues to ignore evaluation of funding bias.**
    f. **Continues to be infeasible to implement.**

We recommend:

- **Using existing, validated methods for assessing study quality.**

[41] US EPA. (2018). Application of systematic review in TSCA risk evaluations. https://www.epa.gov/sites/default/files/2018-06/documents/final_application_of_sr_in_tsca_05-31-18.pdf

- **Considering all relevant studies in evidence synthesis and evidence integration, with no exclusion of studies based on study quality ratings.**
- **Following NASEM recommendations regarding appropriate considerations for study quality evaluation.**

5. **The *2021 Draft TSCA Method's* approach to evidence integration is unclear and inconsistent.**
    a. **The *2021 Draft TSCA Method* fails to draw a clear distinction between evidence synthesis and evidence integration.**
    b. **The TSCA approach to within-stream evidence synthesis needs improvement.**
    c. **EPA's example using six human neurotoxicity studies reaches a poorly-justified conclusion that does not appropriately apply the identified considerations for evidence synthesis.**
    d. **The "Evidence Profile Figure Template" (*Table 7-15*) introduces new considerations for within-stream evidence synthesis that are inconsistent with the earlier discussion of within-stream considerations.**
    e. **EPA has imposed an inappropriate limitation on evidence conclusions that can be drawn from mechanistic evidence.**
    f. **The concept presented in the evidence integration chapter of mechanistic data as a third stream of evidence, alongside human and animal evidence, is unclear and not carried out consistently in other sections of the document.**
    g. **The relationship of the section on characterizing key sources of uncertainty to the rest of the evidence integration chapter is unclear.**
    h. **The *2021 Draft TSCA Method* provides insufficient guidance on toxicokinetic modeling.**

    We recommend:
    - **Identifying evidence synthesis and evidence integration as two distinct steps in the systematic review process, with synthesis conducted for each evidence stream separately, and integration combining the findings across evidence streams.**
    - **Clarifying considerations that are applied in conducting within-stream evidence synthesis and applying a consistent set of considerations throughout the process.**
    - **Clarifying the role of mechanistic evidence in evidence integration.**

6. **Dose-response assessment should be conducted for all identified and suspected health hazards.**

    We recommend:

    - **Conducting dose-response assessment for all outcomes with conclusions of "evidence demonstrates," "evidence indicates likely," and "evidence suggests but is not sufficient to conclude."**

We appreciate the opportunity to provide input. Please do not hesitate to contact us with any questions regarding this request.

Sincerely,

Courtney Cooper, MPH
Science Associate, Program on Reproductive Health and the Environment
Department of Obstetrics, Gynecology and Reproductive Sciences
University of California, San Francisco

Chanese Forté, PhD-PhD, MPH
Assistant Research Scientist, Program on Reproductive Health and the Environment
Department of Obstetrics, Gynecology and Reproductive Sciences
University of California, San Francisco

Max Aung, PhD, MPH
Associate Research Scientist, 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

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

Daniel Axelrad, MPP
Independent Consultant
Washington, DC

Linda S. Birnbaum, PhD
Scientist Emeritus and Scholar in Residence
National Institute of Environmental Health Sciences and Duke University
Chapel Hill, North Carolina

Courtney Carignan
Assistant Professor
Michigan State University
East Lansing, Michigan

Kimberly Gehle MD, MPH
Senior Medical Advisor
Atlanta, Georgia

Robert M. Gould, MD
Associate Adjunct Professor
Program on Reproductive Health and the Environment
Department of Obstetrics, Gynecology and Reproductive Sciences
University of California, San Francisco

Carol Kwiatkowski, PhD
Science and Policy Senior Associate
Green Science Policy Institute
Paonia, Colorado

Jennifer McPartland, PhD
Senior Scientist
Environmental Defense Fund
Washington, DC

Alberto Modenese, PhD, MD
Post-Doc Researcher, Occupational Physician
Department of Biomedical, Metabolic and Neural Sciences
University of Modena & Reggio Emilia
Modena, Italy

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

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

Laura N. Vandenberg, PhD
Professor of Environmental Health Sciences
University of Massachusetts Amherst, School of Public Health & Health Sciences
Amherst, Massachusetts

Paul Whaley, PhD
Honorary Researcher
Lancaster Environment Centre, Lancaster University
Lancaster, United Kingdom

DETAILED COMMENTS

1. EPA still does not adhere to best practices on protocol development for systematic reviews.

a. EPA failed to make to make the *2021 Draft TSCA Method* available while it continued conducting systematic reviews for the current risk evaluations, which casts reasonable doubt on the validity of these evaluations.

EPA's *2021 Draft TSCA Method* indicates that it addresses the issue of conducting risk evaluations prior to releasing their systematic review methods, stating "Previously, EPA did not have a complete clear and documented TSCA systematic review (SR) Protocol. EPA is addressing this lack of a priori protocol by releasing this TSCA SR Protocol."[42] However, between February 2021 and the time this updated protocol was publicly released in December 2021 (approximately ten months), EPA has continued to develop population, exposure, comparator, and outcome (PECO) statements, create eligibility criteria for study inclusion, screen studies, and evaluate study quality for the 23 ongoing risk evaluations without publicly releasing the systematic review methods being used, rendering this statement inaccurate. This has cast reasonable doubt on validity of the chemical assessments and demonstrates that EPA is repeating the same errors that were made with the *2018 TSCA Method.* Although the assessment process began under the previous admistration, the current one is failing to incorporate lessons learned from these mistakes.

Conducting chemical risk evaluations prior to publicly releasing a systematic review method is in direct contradiction to the recommendations from the NASEM, which has touched on this point on multiple occasions:

(1) NASEM 2021: *The use of systematic review in EPA's Toxic Substances Control Act Risk Evaluations*
   a. "Key elements of systematic review include the following:…**Developing a protocol, which a priori describes the specific criteria and approaches that will be used throughout the review process**"[43] (emphasis added).
   b. "The research questions and the approach should inform **the first step of the systematic review, the development of the protocol**"[44] (emphasis added).

(2) NASEM 2021: *Review of US EPA's ORD staff handbook for developing IRIS assessments: 2021 version*
   a. "In a systematic review, **the protocol is a complete account of planned methods, which should be registered prior to conduct of the review. The term 'registration,'in this context, is generally understood to mean the public release of the protocol** in a time-stamped, read-only format"[45] (emphasis added).

[42] US EPA. (2021). Draft systematic review protocol supporting TSCA risk evaluations for chemical substances version 1.0: A generic TSCA systematic review protocol with chemical-specific methodologies, pp. 144. https://www.epa.gov/system/files/documents/2021-12/draft-systematic-review-protocol-supporting-tsca-risk-evaluations-for-chemical-substances_0.pdf

[43] NASEM. (2021). The use of systematic review in EPA's Toxic Substances Control Act Risk Evaluations, pp. 10. Washington, DC: The National Academies Press. https://doi.org/10.17226/25952

[44] NASEM. (2021). The use of systematic review in EPA's Toxic Substances Control Act Risk Evaluations, pp. 3. Washington, DC: The National Academies Press. https://doi.org/10.17226/25952

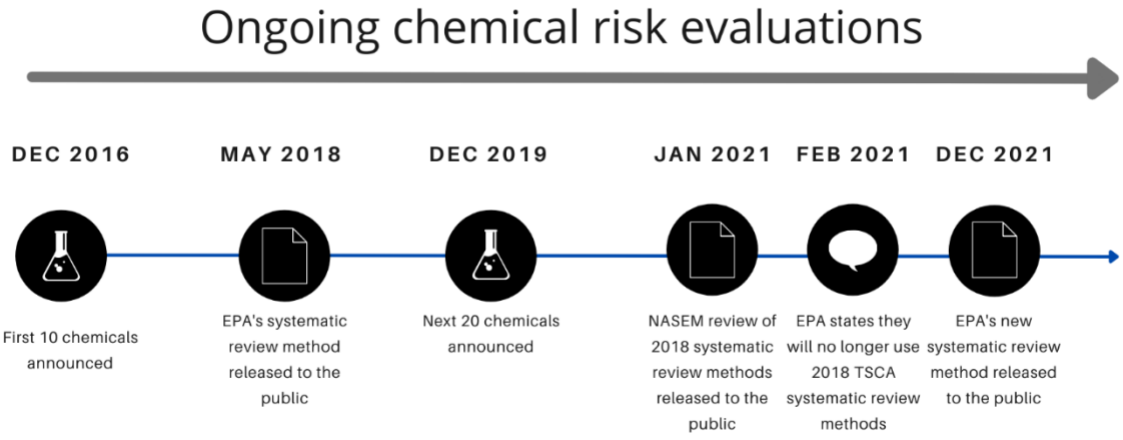
[45] NASEM. (2021). Review of US EPA's ORD staff handbook for developing IRIS assessments: 2021 version, pp. 5. https://www.nap.edu/catalog/26289/review-of-us-epas-ord-staff-handbook-for-developing-iris-assessments

This current approach that EPA has taken is also in conflict with requirements of the statute and the risk evaluation framework rule: TSCA risk evaluations are required to rely on the weight of the scientific evidence [15 U.S.C. § 2625(i)] that is defined in the framework rule as "**a systematic review method, applied in a manner suited to the nature of the evidence or decision, that uses a pre-established protocol** to comprehensively, objectively, transparently, and consistently, identify and evaluate each stream of evidence, including strengths, limitations, and relevance of each study and to integrate evidence as necessary and appropriate based upon strengths, limitations, and relevance"[46] (emphasis added).

Additionally, by incorporating only certain aspects of other systematic review methods, EPA is failing to adhere to its statutory mandate outlined in Section 26(h) of amended TSCA: *"*In carrying out sections 4, 5, and 6, to the extent that the Administrator makes a decision based on science, the Administrator shall use scientific information, technical procedures, measures, methods, protocols, methodologies, or models, **employed in a manner consistent with the best available science**" (emphasis added).

There are several established systematic review methods, endorsed by the NASEM and considered to be "consistent with the best available science (TSCA Section 26[h])," that could be adopted for use in conducting TSCA risk evaluations, including the National Toxicology Program (NTP) Office of Health Assessment and Translation (OHAT) method,[47] the University of California, San Francisco Program on Reproductive Health and the Environment (UCSF PRHE) Navigation Guide method,[48] and the Integrated Risk Information System (IRIS) method, [49] if recommendations from the NASEM on the draft *IRIS Handbook* are incorporated. EPA could have used these approaches to rapidly develop a publicly available method prior to initiating the "next 20" chemical risk evaluations instead of developing a protocol and doing the evaluations concurrently, as outlined in ***Figure 1***. As it stands, the systematic review method described in the *2021 Draft TSCA Method* fails to meet their statutory requirements.

Figure 1. EPA's Current Risk Evaluation Timeline



[46] US EPA. (2017). Procedures for chemical risk evaluation under the amended Toxic Substances Control Act. https://www.govinfo.gov/content/pkg/FR-2017-07-20/pdf/2017-14337.pdf

[47] OHAT (Office of Health Assessment and Translation). 2019. Handbook for Conducting a Literature-Based Health Assessment Using OHAT Approach for Systematic Review and Evidence Integration. https://ntp.niehs.nih.gov/ntp/ohat/pubs/handbookmarch2019_508.pdf

[48] Woodruff, T. J., and P. Sutton. 2014. The Navigation Guide systematic review methodology: A rigorous and transparent method for translating environmental health science into better health outcomes. Environmental Health Perspectives 122(10):1007–1014.

[49] US EPA. (2021). ORD staff handbook for developing IRIS assessments. https://cfpub.epa.gov/ncea/iris_drafts/recordisplay.cfm?deid=350086

b. EPA fails to include protocol development in the *2021 Draft TSCA Method* as an explicit step in the systematic review process.

The NASEM recommended that the TSCA systematic review process include publication of a protocol before the systematic review is initiated:

> In order to improve these issues with TSCA's approach to problem formulation and protocol development, the committee recommends the following…A systematic review protocol that details the prespecified methods, including eligibility and critical appraisal criteria, and that is peer-reviewed and publicly posted before the review commences should be prepared.[50]

EPA still lacks protocol development as a specific step in the systematic review process in the *2021 Draft TSCA Method*, failing to include a public protocol as part of the systematic review process – as is shown in several figures outlining the steps of a systematic review method, specifically ***Figures 2-1*** (included below),[51] ***3-1***,[52] *and* ***7-3***[53] in the document. This is a critical missing piece because creating protocols for all review components prior to conducting the review minimizes bias and ensures transparency in decision-making, and thus is specified as a best practice by all established systematic review methods.[54,55] Furthermore, "Both the OHAT and Navigation Guide methods include the key steps recommended by a previous National Academies committee (NRC 2014)[56] for problem formulation, **protocol development,** specifying a study question, developing a PECO statement, identifying and selecting the evidence, evaluating the evidence, and integrating the evidence"[57] (emphasis added). ***Figure S-1*** (copied below) from the NASEM 2014 review of the EPA IRIS program's systematic review method,[58] presents all the components of a science or evidence-based systematic review. The goal of the protocol is to ensure that

[50] US EPA. (2021). Draft systematic review protocol supporting TSCA risk evaluations for chemical substances version 1.0: A generic TSCA systematic review protocol with chemical-specific methodologies, pp. 26-27. https://www.epa.gov/system/files/documents/2021-12/draft-systematic-review-protocol-supporting-tsca-risk-evaluations-for-chemical-substances_0.pdf

[51] US EPA. (2021). Draft systematic review protocol supporting TSCA risk evaluations for chemical substances version 1.0: A generic TSCA systematic review protocol with chemical-specific methodologies, pp. 30. https://www.epa.gov/system/files/documents/2021-12/draft-systematic-review-protocol-supporting-tsca-risk-evaluations-for-chemical-substances_0.pdf

[52] US EPA. (2021). Draft systematic review protocol supporting TSCA risk evaluations for chemical substances version 1.0: A generic TSCA systematic review protocol with chemical-specific methodologies, pp. 33. https://www.epa.gov/system/files/documents/2021-12/draft-systematic-review-protocol-supporting-tsca-risk-evaluations-for-chemical-substances_0.pdf

[53] US EPA. (2021). Draft systematic review protocol supporting TSCA risk evaluations for chemical substances version 1.0: A generic TSCA systematic review protocol with chemical-specific methodologies, pp. 92. https://www.epa.gov/system/files/documents/2021-12/draft-systematic-review-protocol-supporting-tsca-risk-evaluations-for-chemical-substances_0.pdf

[54] National Research Council. (2014). Review of EPA's Integrated Risk Information System (IRIS) process. https://www.nap.edu/catalog/18764/review-of-epas-integrated-risk-information-system-iris-process

[55] Institute of Medicine. (2011). Finding what works in health care: Standards for systematic reviews. https://www.ncbi.nlm.nih.gov/books/NBK209518/

[56] National Research Council. (2014). Review of the Environmental Protection Agency's state-of-the-science evaluation of nonmontonic dose-response relationships as they apply to endocrine disruptors. https://doi.org/10.17226/18608.

[57] NASEM. (2017). Application of systematic review methods in an overall strategy for evaluating low-dose toxicity from endocrine active chemicals, pp. 119.Washington, D.C.: The National Academies Press. https://doi.org/10.17226/24758

[58] National Research Council. (2014). Review of EPA's Integrated Risk Information System (IRIS) process, pp. 31. https://www.nap.edu/catalog/18764/review-of-epas-integrated-risk-information-system-iris-process

judgements regarding evidence evaluation are made prior to reviewing the evidence to decrease the likelihood of biased recommendations.

Protocols, once developed, must be published a priori as they ensure the integrity of other steps in the systematic review process and provide time for peer review. Without this process, other crucial components of the systematic review collapse[59,60,61,62,63,64]. Importantly, there is no way to easily determine if the 23 ongoing risk evaluations are biased when there is a lack of clarity around the systematic review methodology applied. The NASEM reiterates the importance of this stating:

> A protocol makes the methods and the process of the review transparent, provides the opportunity for peer review of the methods, and stands as a record of the review process. Having a protocol minimizes the potential for bias in many steps throughout the systematic review, such as in evidence identification, by ensuring that inclusion of studies in the review does not depend on the findings of the studies (NRC 2014).[65]

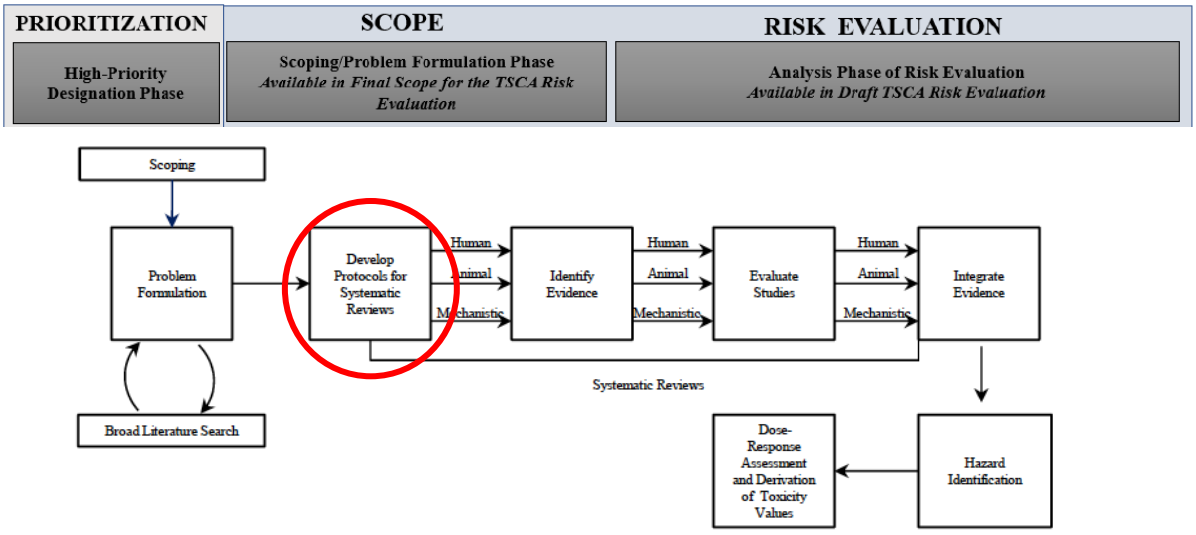


FIGURE S-1 Systematic review in the context of the IRIS process. The committee views public input and peer review as integral parts of the IRIS process, although they are not specifically noted in the figure.

[59] NASEM. (2021). The use of systematic review in EPA's Toxic Substances Control Act Risk Evaluations, pp. 6. Washington, DC: The National Academies Press. https://doi.org/10.17226/25952

[60] NASEM. (2021). The use of systematic review in EPA's Toxic Substances Control Act Risk Evaluations, pp. 34. Washington, DC: The National Academies Press. https://doi.org/10.17226/25952

[61] NASEM. (2021). The use of systematic review in EPA's Toxic Substances Control Act Risk Evaluations, pp. 40. Washington, DC: The National Academies Press. https://doi.org/10.17226/25952

[62] NASEM. (2021). The use of systematic review in EPA's Toxic Substances Control Act Risk Evaluations, pp. 41. Washington, DC: The National Academies Press. https://doi.org/10.17226/25952

[63] NASEM. (2021). The use of systematic review in EPA's Toxic Substances Control Act Risk Evaluations, pp. 49. Washington, DC: The National Academies Press. https://doi.org/10.17226/25952

[64] National Research Council. (2014). Review of EPA's Integrated Risk Information System (IRIS) process, pp. 76. https://www.nap.edu/catalog/18764/review-of-epas-integrated-risk-information-system-iris-process

[65] NASEM. (2021). The use of systematic review in EPA's Toxic Substances Control Act Risk Evaluations, pp. 18. Washington, DC: The National Academies Press. https://doi.org/10.17226/25952

Conducting risk evaluations without first developing and publishing a completed systematic review protocol could have serious public health implications. The public relies on studies, or evidence, included in the systematic review process to indicate whether the chemical in question may pose serious health risks. Without a clear, pre-published, publicly available protocol there is a risk that relevant studies may be excluded or that the evaluation of studies included may be biased, which could lead to underestimation of the true risk of chemical exposures. Said differently, failing to have a systematic review protocol in place and available for public peer review prior to conducting risk evaluations could substantially affect the results of the risk evaluation. This in turn could affect the risk determination and any subsequent risk management, potentially with implications on future exposure levels and health outcomes.

c. EPA still needs to create chemical-specific protocols for each systematic review it conducts for TSCA risk evaluations.

In the *2021 Draft TSCA Method*,[66] EPA states that the document "also provides specific details of the systematic reviews for the individual chemicals listed in Table 1-1."[67] These chemical-specific details, which are in the appendices, contain information on PECO statements, chemical search terms, potentially relevant supplemental materials, and links to interactive literature trees and evidence tables. However, these appendices and the *2021 Draft TSCA Method* do not substitute the need for chemical specific systematic review protocols that should be made publicly available for each TSCA risk evaluation. Although "the application of systematic review principles is generally expected to be consistent across risk evaluations, as outlined in this generic protocol, with customized criteria and approaches applied, as necessary, to meet the assessment needs of individual risk evaluations as noted in chemical specific appendices"[68] this is insufficient. EPA needs to publish a stand-alone complete systematic review protocol for every chemical that will be assessed under TSCA, accounting for chemical-specific considerations in conducting the risk evaluation and provide adequate time for peer review and public comments on each protocol.

The NASEM 2021 review of the *2018 TSCA Method* recommended that "Registering the protocol for **each** risk evaluation is important: That protocol should include an explicit search strategy, and search strategies for each database should be consistently listed in the appendix to the risk evaluation"[69] (emphasis added). They went a step further to state that when there are not protocols for reviews **"…the transparency of the entire risk evaluation is compromised** because in addition to not developing clear questions for the systematic reviews, there are no protocols for the reviews or to guide the

[66] US EPA. (2021). Draft systematic review protocol supporting TSCA risk evaluations for chemical substances version 1.0: A generic TSCA systematic review protocol with chemical-specific methodologies, pp. 144. https://www.epa.gov/system/files/documents/2021-12/draft-systematic-review-protocol-supporting-tsca-risk-evaluations-for-chemical-substances_0.pdf

[67] US EPA. (2021). Draft systematic review protocol supporting TSCA risk evaluations for chemical substances version 1.0: A generic TSCA systematic review protocol with chemical-specific methodologies, pp. 24. https://www.epa.gov/system/files/documents/2021-12/draft-systematic-review-protocol-supporting-tsca-risk-evaluations-for-chemical-substances_0.pdf

[68] US EPA. (2021). Draft systematic review protocol supporting TSCA risk evaluations for chemical substances version 1.0: A generic TSCA systematic review protocol with chemical-specific methodologies, pp. 31. https://www.epa.gov/system/files/documents/2021-12/draft-systematic-review-protocol-supporting-tsca-risk-evaluations-for-chemical-substances_0.pdf

[69] NASEM. (2021). The use of systematic review in EPA's Toxic Substances Control Act Risk Evaluations, pp. 34. Washington, DC: The National Academies Press. https://doi.org/10.17226/25952

synthesis step. Consequently, the review process is not documented or prespecified from its start, and clarity is lacking when the review is finished and published"[70] (emphasis added).

Chemical specific protocols published a priori offer the optimal approach to identify the best empirical evidence for each chemical. Failing to do so opens the potential to harm public health by missing key evidence or providing biased recommendations based upon the evaluations of the evidence.

We recommend:

- **Making draft systematic review protocols available prior to conducting risk evaluations.**
- **Including protocol development as an explicit step in the systematic review process. Developing chemical-specific protocols for each of the 23 ongoing risk evaluations conducted under amended TSCA.**

2. EPA's PECO statements do not provide a clear, consistent, reasonable, and appropriate basis for determining which studies are relevant and which studies are not relevant for hazard assessment.

a. Most of EPA's PECO statements appear to inappropriately exclude important toxicity endpoints from the TSCA hazard assessments.

EPA presents its PECO statements for environmental and human health hazards in Appendix H.5. PECO statements play a critical role in conducting a systematic review as they provide criteria for screening the literature search results to identify which studies are relevant (included in the risk evaluation) and not relevant (excluded from further consideration). The *2021 Draft TSCA Method* indicates that certain types of studies that do not satisfy the PECO criteria may be tagged as "potentially relevant supplemental material" that "may be reviewed, evaluated for data quality, and incorporated into risk evaluations as needed for each chemical assessment."[71] Supplemental material may be used to inform the hazard assessments but would not serve as the basis of hazard findings or be used to derive a point of departure for risk characterization; for example, ***Figure 4-2*** says that supplemental studies are not considered in dose-response assessment.

For the outcome element, EPA's PECO statements come in two main versions (with underlining added to show key differences; noted below is which version is used for each ongoing risk evaluation):

Version A
Human: All health outcomes (both cancer and non-cancer)
Animal and Plants: All biological effects (including bioaccumulation from laboratory studies with concurrently measured water and tissue concentrations).
Screener note: Measurable biological effects relevant for humans, animals and plants may include but are not limited to: mortality, behavioral, population, cellular, physiological, growth, reproduction, systemic, point of contact effects.

[70] NASEM. (2021). The use of systematic review in EPA's Toxic Substances Control Act Risk Evaluations, pp. 26. Washington, DC: The National Academies Press. https://doi.org/10.17226/25952

[71] US EPA. (2021). Draft systematic review protocol supporting TSCA risk evaluations for chemical substances version 1.0: A generic TSCA systematic review protocol with chemical-specific methodologies, pp. 345. https://www.epa.gov/system/files/documents/2021-12/draft-systematic-review-protocol-supporting-tsca-risk-evaluations-for-chemical-substances_0.pdf

Version B
Human: All health outcomes (cancer and non-cancer) at the organ level or higher.
Animal and Plants: All apical biological effects (effects measured at the organ level or higher) and bioaccumulation from laboratory studies with concurrently measured media and/or tissue concentrations. Apical endpoints include but are not limited to reproduction, survival, and growth.
Screener note:

- Measurable biological effects relevant for humans, animals and plants may include but are not limited to: mortality, behavioral, population, physiological, growth, reproduction, systemic, point of contact (irritation and sensitization) effects.
- Effects measured at the cellular level of biological organization and below are to be tagged as supplemental, mechanistic.

Version B incorporates several limitations on health effects studies that are considered relevant for hazard identification. For human studies, Version B specifies that only studies "at the organ level or higher" are to be included. For animal studies, Version B specifies that only "apical" effects "measured at the organ level or higher" are to be included. The "screener note" for Version B deletes "cellular" from the list of relevant measurable biological effects and indicates that "Effects measured at the cellular level of biological organization and below are to be tagged as supplemental, mechanistic."

Version A is used in the PECO statements for both title and abstract screening and full-text screening for the following chemicals:

- *p*-Dichlorobenzene and *o*-Dichlorobenzene (***Tables Apx H-7*** and ***H-9***)
- Ethylene Dibromide (***Tables Apx H-15*** and ***H-17***)
- TPP (***Tables Apx H-35*** and ***H-37***) (The screener note re "measurable biological effects" is omitted from the PECO statements for this chemical.)

Version B is used in the PECO statements for both title and abstract screening and full-text screening for the following chemicals:

- 1,3-Butadiene (***Tables Apx H-19*** and ***H-21***)
- Formaldehyde (***Tables Apx H-39*** and ***H-41***)
- Phthalic Anhydride and Phthalic Acid (***Tables Apx H-43*** and ***H-45***)
- Phthalates (7 chemicals) (***Tables Apx H-47*** and ***H-49***)
- D4 (***Tables Apx H-51*** and ***H-53***)

For the following chemicals, Version A is used in the PECO statement for title and abstract screening, and the PECO is then changed to use Version B for full-text screening:

- Chlorinated Solvents (5 chemicals) (***Tables Apx H-11*** and ***H-13***)
- HHCB (***Tables Apx H-23*** and ***H-25***)
- TBBPA (***Tables Apx H-27*** and ***H-29***)
- TCEP (***Tables Apx H-31*** and ***H-33***)

The differences between Version A and Version B may significantly affect the scope of hazardous outcomes that are considered in a TSCA risk evaluation. By limiting the relevant human and animal studies to those with "apical" effects or those with effects at the "organ level or higher," EPA appears to be excluding studies of important biochemical markers of effect and other outcomes at the cellular level

that are strong indicators of hazards and which have commonly been used as critical effects in previous EPA hazard assessments.

EPA's PECO statements provide very limited guidance for screeners on what effects are to be considered "apical" or "organ-level." The PECOs say: "Apical endpoints include but are not limited to reproduction, survival, and growth" and "Measurable biological effects relevant for humans, animals and plants may include but are not limited to: mortality, behavioral, population, physiological, growth, reproduction, systemic, point of contact (irritation and sensitization) effects."[72] The *2021 Draft TSCA Method* provides no further guidance on which outcomes are to be considered apical or organ-level, and which outcomes are to be considered cellular-level.

The National Academy of Sciences (NAS) has defined an apical end point as "An observable outcome in a whole organism, such as a clinical sign or pathologic state, that is indicative of a disease state that can result from exposure to a toxicant,"[73] and identified "tumors, birth defects, and neurologic impairments"[74] as examples. No biochemical measures or early biological changes were mentioned among the examples.

The definition of an apical effect appears to be narrower than the definition of an adverse effect provided by the EPA IRIS program: "a biochemical change, functional impairment, or pathologic lesion that affects the performance of the whole organism, or reduces an organism's ability to response to an additional environmental challenge."[75] The definition of adverse effect includes, for example, "a biochemical change;" such changes appear to be excluded from Version B as they would likely be considered cellular-level effects rather than organ-level or apical effects.

Biochemical and/or cellular-level outcomes have been identified as critical effects in numerous past EPA hazard assessments, including some of the completed TSCA risk evaluations. Examples of these outcomes and past assessments include:

- reduced thyroid hormone levels (2020 TSCA risk evaluation of HBCD; 2021 toxicity assessment of PFBS)[76,77]
- decreased erythrocyte counts and hemoglobin (2020 TSCA risk evaluation of perchloroethylene)[78]
- Measures of immune function, such as increases in immunoglobulin E, white blood cells, lymphocytes, natural killer cells, and interlukin-4 levels (2020 TSCA risk evaluation of perchloroethylene)[79]

[72] US EPA. (2021). Draft systematic review protocol supporting TSCA risk evaluations for chemical substances version 1.0: A generic TSCA systematic review protocol with chemical-specific methodologies, pp. 354. https://www.epa.gov/system/files/documents/2021-12/draft-systematic-review-protocol-supporting-tsca-risk-evaluations-for-chemical-substances_0.pdf

[73] National Academies of Sciences, Engineering, and Medicine. (2007). Toxicity Testing in the 21st Century: A Vision and a Strategy, pp. 38. Washington, DC: The National Academies Press. Available: https://doi.org/10.17226/11970.

[74] National Academies of Sciences, Engineering, and Medicine. (2007). Toxicity Testing in the 21st Century: A Vision and a Strategy, pp. 177. Washington, DC: The National Academies Press. Available: https://doi.org/10.17226/11970.

[75] US EPA. IRIS Glossary. Available: https://www.epa.gov/iris/iris-glossary

[76] US EPA. (2020). Risk evaluation for cyclic aliphatic bromide cluster (HBCD). https://www.epa.gov/sites/default/files/2020-09/documents/1._risk_evaluation_for_cyclic_aliphatic_bromide_cluster_hbcd_casrn25637-99-4_casrn_3194-5_casrn_3194-57-8.pdf

[77] US EPA. (2021). Human health toxicity values for perfluorobutane sulfonic acid and related compound potassium perflueorobutane sulfonate. https://cfpub.epa.gov/ncea/risk/recordisplay.cfm?deid=350888

[78] US EPA. (2020). Risk evaluation for percholorethylene (Ethene, 1,1,2,2-Tetrachloro-). https://www.epa.gov/sites/default/files/2020-12/documents/1_risk_evaluation_for_perchloroethylene_pce_casrn_127-18-4_0.pdf

[79] US EPA. (2020). Risk evaluation for percholorethylene (Ethene, 1,1,2,2-Tetrachloro-). https://www.epa.gov/sites/default/files/2020-12/documents/1_risk_evaluation_for_perchloroethylene_pce_casrn_127-18-4_0.pdf

- decreased sperm quality or concentration (2020 TSCA risk evaluations of trichloroethylene and perchloroethylene; 2018 and 2019 IRIS staff published systematic reviews of health effects of phthalates)[80, 81, 82,83]
- cholinesterase inhibition (numerous assessments of pesticides, including cumulative risk assessments of organophosphate and carbamate pesticides)[84,85]
- reduced male fetal testosterone or adult male testosterone (2018 and 2019 IRIS staff published systematic reviews of health effects of phthalates) [86,87]

EPA must either document that it has considered outcomes like altered thyroid hormone levels and other biochemical changes or cellular-level effects to be included in the animal and human evidence streams in the ongoing TSCA risk evaluations or provide a justification for why these outcomes should not serve as the basis for hazard assessments conducted under TSCA – especially given such outcomes have been identified hazards in previous TSCA risk evaluations.

Tagging biochemical and cellular-level outcomes as "supplemental, mechanistic" is not a substitute for including the studies within the animal and human evidence streams. This approach instead constrains the role of biochemical outcomes and other cellular changes to *possibly* providing biological support for studies of apical outcomes. The *2021 Draft TSCA Method* describes how the use of information identified as mechanistic is focused on informing interpretation of separate studies of apical outcomes:

> mechanistic data evaluations consider the support for and involvement of specific events or sets of events within the context of a broader research question (*e.g.,* support for a hypothesized mechanism, consistency with known biological processes), rather than evaluations of individual apical endpoints considered in relative isolation…while mechanistic data is evaluated on its own, the mechanistic evidence is most useful in demonstrating the relevance and reliability of apical outcome findings in animal and human studies.[88]

If no studies have been conducted of apical outcomes related to a biochemical outcome that has been studied, it is unclear whether the biochemical outcome will be considered at all. As noted above, supplemental studies "**may** be reviewed, evaluated for data quality, and incorporated into risk

[80] Yost EE, Euling SY, Weaver JA, Beverly BEJ, Keshava N, Mudipalli A, Arzuaga X, Blessinger T, Dishaw L, Hotchkiss A, Makris SL. Hazards of diisobutyl phthalate (DIBP) exposure: A systematic review of animal toxicology studies. Environ Int. 2019 Apr;125:579-594. doi: 10.1016/j.envint.2018.09.038. Epub 2018 Dec 25. PMID: 30591249; PMCID: PMC8596331.

[81] Radke EG, Braun JM, Meeker JD, Cooper GS. Phthalate exposure and male reproductive outcomes: A systematic review of the human epidemiological evidence. Environ Int. 2018 Dec;121(Pt 1):764-793. doi: 10.1016/j.envint.2018.07.029. Epub 2018 Oct 16. Erratum in: Environ Int. 2019 Apr;125:606-607. PMID: 30336412.

[82] US EPA. (2020). Risk evaluation for percholorethylene (Ethene, 1,1,2,2-Tetrachloro-). https://www.epa.gov/sites/default/files/2020-12/documents/1_risk_evaluation_for_perchloroethylene_pce_casrn_127-18-4_0.pdf

[83] US EPA. (2020). Risk evaluation for trichloroethylene. https://www.epa.gov/sites/default/files/2020-11/documents/1._risk_evaluation_for_trichloroethylene_tce_casrn_79-01-6.pdf

[84] US EPA. (2006). Organophosphorus cumulative risk assessment. https://www.regulations.gov/document/EPA-HQ-OPP-2006-0618-0002

[85] US EPA. (2008). Revised *N*-methyl carbamate cumulative risk assessment. https://www.regulations.gov/document/EPA-HQ-OPP-2008-0347-0029

[86] Yost EE, Euling SY, Weaver JA, Beverly BEJ, Keshava N, Mudipalli A, Arzuaga X, Blessinger T, Dishaw L, Hotchkiss A, Makris SL. Hazards of diisobutyl phthalate (DIBP) exposure: A systematic review of animal toxicology studies. Environ Int. 2019 Apr;125:579-594. doi: 10.1016/j.envint.2018.09.038. Epub 2018 Dec 25. PMID: 30591249; PMCID: PMC8596331.

[87] Radke EG, Braun JM, Meeker JD, Cooper GS. Phthalate exposure and male reproductive outcomes: A systematic review of the human epidemiological evidence. Environ Int. 2018 Dec;121(Pt 1):764-793. doi: 10.1016/j.envint.2018.07.029. Epub 2018 Oct 16. Erratum in: Environ Int. 2019 Apr;125:606-607. PMID: 30336412.

[88] US EPA. (2021). Draft systematic review protocol supporting TSCA risk evaluations for chemical substances version 1.0: A generic TSCA systematic review protocol with chemical-specific methodologies, pp. 112. https://www.epa.gov/system/files/documents/2021-12/draft-systematic-review-protocol-supporting-tsca-risk-evaluations-for-chemical-substances_0.pdf

evaluations **as needed** for each chemical assessment"[89] (emphasis added) but it is unclear how a determination would be made to incorporate these studies into the risk evaluation, particularly in the absence of a related apical outcome study. Even if included to support a hazard conclusion based on apical outcomes, it appears such studies may be used only to support the qualitative step of selecting a hazard descriptor (e.g., "evidence suggests") and would not be considered as a potential point of departure for the risk characterization.

A further problem with exclusive reliance on studies of apical endpoints is that it is not consistent with current science. An important theme of the NAS 2007 *Toxicity Testing in the 21st Century* report was that the field of toxicity testing should move away from reliance on testing of apical outcomes, and EPA's research programs have invested heavily in this new direction. Government and academic toxicology labs now rarely conduct studies of apical endpoints because the science has moved towards a much greater focus on more sensitive endpoints representing upstream biological changes ("key events") that lead to apical outcomes. A restriction to consider only apical or organ-level studies will substantially bias the evidence base of TSCA risk evaluations toward inclusion of industry-funded guidelines studies and away from non-industry studies.

b. EPA must use the same PECO statement for both title-abstract screening and full-text review.

Unusual and contrary to standard systematic review practices, Appendix H.5 provides two PECO statements for health hazards for each assessment: a PECO applied in title-abstract screening, then a revised PECO applied at the full-text review stage. The *2021 Draft TSCA Method* explains: "Although EPA may use the same criteria statements at full-text screening, the criteria may be revised as needed based on the screeners' experiences during title/abstract screening."[90] EPA also states that "As each screening project progressed the PECO and supplemental material criteria for each chemical evolved. Therefore, wording may vary among PECO statements."[91]

These revisions to PECO statements appear to be in conflict with NASEM advice. In its 2021 review of the TSCA SR method, the NASEM was critical of changes in inclusion/exclusion criteria: "In the TSCA evaluation process, eligibility criteria are not predefined in the protocols and shift during the systematic review process."[92] The NASEM recommended "Eligibility criteria need to be based on PECO statements that are formulated in a standard way and need to be predefined in the protocol. The eligibility of outcomes needs to be carefully considered a priori to prevent a systematic exclusion of outcomes that could bias the results, such as excluding studies that have findings counter to those anticipated for the included outcomes."[93]

[89] US EPA. (2021). Draft systematic review protocol supporting TSCA risk evaluations for chemical substances version 1.0: A generic TSCA systematic review protocol with chemical-specific methodologies, pp. 345. https://www.epa.gov/system/files/documents/2021-12/draft-systematic-review-protocol-supporting-tsca-risk-evaluations-for-chemical-substances_0.pdf

[90] US EPA. (2021). Draft systematic review protocol supporting TSCA risk evaluations for chemical substances version 1.0: A generic TSCA systematic review protocol with chemical-specific methodologies, pp. 41. https://www.epa.gov/system/files/documents/2021-12/draft-systematic-review-protocol-supporting-tsca-risk-evaluations-for-chemical-substances_0.pdf

[91] US EPA. (2021). Draft systematic review protocol supporting TSCA risk evaluations for chemical substances version 1.0: A generic TSCA systematic review protocol with chemical-specific methodologies, pp. 345. https://www.epa.gov/system/files/documents/2021-12/draft-systematic-review-protocol-supporting-tsca-risk-evaluations-for-chemical-substances_0.pdf

[92] US EPA. (2021). Draft systematic review protocol supporting TSCA risk evaluations for chemical substances version 1.0: A generic TSCA systematic review protocol with chemical-specific methodologies, pp. 34. https://www.epa.gov/system/files/documents/2021-12/draft-systematic-review-protocol-supporting-tsca-risk-evaluations-for-chemical-substances_0.pdf

[93] US EPA. (2021). Draft systematic review protocol supporting TSCA risk evaluations for chemical substances version 1.0: A generic TSCA systematic review protocol with chemical-specific methodologies, pp. 35. https://www.epa.gov/system/files/documents/2021-12/draft-systematic-review-protocol-supporting-tsca-risk-evaluations-for-chemical-substances_0.pdf

We are not aware of any other systematic review approach that allows for changes in the PECO while an assessment is being conducted. If changes to a PECO are deemed necessary or useful by the systematic review authors, the literature screening process should be re-started with the revised PECO, with complete documentation in a revised protocol of the change in PECO and the rationale for change, rather than applying the revised PECO in the middle of the process.

Changing the PECO mid-stream can result in inappropriate and unreasonable study inclusions and exclusions, as whether a study is included or excluded can depend on the differences between the two PECOs. With a changing PECO, it is possible that studies are excluded at title-abstract screening that would be eligible for inclusion at full-text screening – undermining the entire process by removing eligible studies before they reach the full-text stage. This scenario is illustrated by differences in the two PECO statements for HHCB (***Tables Apx H-23*** and ***H-25***). The exposure element of the HHCB PECO statement used for title and abstract screening identifies the "relevant forms and isomers" of the chemical with two bullet points: first, the full chemical name of HHCB and its CAS Number (1222-05-5); and second, a statement that "No isomers were included for HHCB." The PECO for full-text screening, however, provides a significant modification to the "relevant forms and isomers" by adding a list of 15 different isomers, each with its own CAS number. The statement that ""No isomers were included for HHCB," applied in title-abstract screening, may have resulted in exclusion of studies conducted with any of the 15 isomers that were identified only in the PECO for full-text review.

Additional portions of the HHCB PECO's exposure element were changed for full-text screening. The PECO for title-abstract screening says:

> **Human:** Any exposure to HHCB (CASRN 1222-05-5)
> **Animal:** Any exposure to HHCB (CASRN 1222-05-5) including via water (including environmental aquatic exposures), soil or sediment, diet, gavage, injection, dermal, and inhalation.

For full-text screening, the text for human studies has been expanded and the text for animal studies is unchanged (added text is underlined):

> **Human:** Any exposure to HHCB (CASRN 1222-05-5) <u>singularly or in mixture, including exposure as measured by internal concentrations of these chemicals or metabolites of these chemicals in a biological matrix (*i.e.,* urine, blood, semen, etc.)</u>.
> **Animal:** Any exposure to HHCB (CASRN 1222-05-5) including via water (including environmental aquatic exposures), soil or sediment, diet, gavage, injection, dermal, and inhalation.

These changes in the PECO statement, implemented for full-text review but after title-abstract screening was completed, could further adversely affect the identification of relevant studies for this risk evaluation. First, the revised statement for human studies at the full-text stage that relevant studies could include mixture studies and studies of internal concentrations, or their metabolites is useful, but failure to apply this explanation at the title-abstract screening stage could again have resulted in exclusion of relevant studies of mixtures or internal concentrations. Second, the PECO statement fails to identify any metabolites of the target chemical, making it challenging for assessors conducting the full-text review to apply this additional consideration (it is unclear whether the listed isomers are the metabolites referred to in the amended PECO statement). Third, the bullet point for animal studies makes no mention of mixture studies, internal concentrations or metabolites, potentially leading to exclusion of studies that are relevant. Fourth, the comparators element of the PECO statement (both versions) makes no mention of the 15 isomers that have been added for the full-text PECO, potentially

creating further confusion for assessors about whether studies with isomer comparators should be included or excluded.

It appears that EPA's motivation in revising the PECO statements for the full-text stage was to take the opportunity to incorporate what it considered to be improvements or clarifications. However, incorporating improvements to the PECO statement only for the full-text review stage will be of limited effectiveness if studies covered by these additional considerations have already been excluded at the title-abstract screening stage.

The majority of PECO statements presented in Appendix H.5 incorporated substantive changes between title-abstract screening and full-text screening. As noted above, four PECO statements (applied to eight chemicals) were modified at the full-text stage to limit relevant studies to those of apical outcomes or organ-level effects, when no such limitation was applied at the title-abstract stage.

PECO statements for several other chemicals (5 chlorinated solvents, ***Table Apx H-13***; TBBPA, ***Table Apx H-29***; TCEP, ***Table Apx H-33***) were similarly revised at the full-text stage to include the statement (quoted above for HHCB) regarding mixtures and internal concentrations of the chemicals or their metabolites. In addition, the PECO for TPP adds a different statement regarding mixture at the full-text stage (***Table Apx H-37***) but does not incorporate the statement regarding internal concentrations and metabolites.

The exposure, comparator, and outcome elements of the PECO statement for formaldehyde each include a "Screener Note" at the title-abstract stage (***Table Apx H-39),*** but these screener notes are omitted at the full-text stage (***Table Apx H-41***). No explanation is provided for why the screener notes were deleted. The screener note for exposure included a useful expanded list of occupations that would be considered proxies for exposures to formaldehyde in occupational epidemiology studies ("pathologists, funeral directors, embalmers, wood workers, garment/textile workers, hospital nurses, cosmetologists, and other workers in industries"). A much shorter list of occupations is provided in the human studies bullet point ("pathologists, funeral directors, embalmers"), but the expanded list of occupations from the screener note is absent at the full-text review stage.

For ethylene dibromide (***Tables Apx H-15*** and ***H-17***) and 1,3-butadiene (***Tables Apx H-19*** and ***H-21***), there were no substantive differences between the two PECO statements (changes were limited to formatting). It is not clear why two separate PECO statements are provided in these cases.

c. There are unexplained and unwarranted inconsistencies across the PECO statements for different chemicals.

EPA has appropriately developed individual PECO statements for each assessment, to allow for incorporation of chemical-specific considerations in the PECOs. Although it is important to have a separate PECO for each assessment, there should be consistency across PECOs in format, content and approach. Any differences across PECOs should only be those necessary to account for chemical-specific considerations, and there should not be unexplained differences that are not due to chemical-specific considerations.

Among the chemicals currently being assessed under TSCA are two organophosphate flame retardants, TCEP and TPP. Given their similarities in uses and chemistry, it might be expected that the PECO

statements for these two chemicals would be very similar. But there are important differences between the PECOs applied to full-text screening of TCEP and TPP (***Tables Apx H-33*** and ***H-37***):

- Exposures: the TCEP PECO statement for human studies includes the following text that does not appear in the TPP PECO statement: "including exposure as measured by internal concentrations of these chemicals or metabolites of these chemicals in a biological matrix (*i.e.,* urine, blood, semen, etc.)."
- Exposures: the TPP PECO statement for animal studies includes the following text that does not appear in the TCEP PECO statement: "Studies involving exposures to mixtures will be included only if they include exposure to triphenyl phosphate alone."
- Outcomes: for TCEP, outcomes are restricted to apical or organ-level effects (see above); for TPP this restriction is not applied.

These differences in PECO statements between two similar chemicals could result in different decisions between the two assessments regarding the types of studies that are included and excluded, without any explanation of why they should be different. The variation in whether the text regarding mixture studies and internal concentrations are included or omitted is found throughout the entire set of PECO statements for studies of hazards in Appendix H.5, with no clear pattern or explanation for why these points are included in PECOs for some chemicals and not for the others.

As discussed above, there are differences in the outcome element of the PECO statements, with either of two different versions seen in almost all cases. However, the PECO statement for asbestos (***Tables Apx H-55*** and ***H-57***) introduces a third version in which outcomes in human studies are restricted to system-level or higher outcomes, rather than organ-level or higher: "Health outcomes including lung cancer, mesothelioma, laryngeal cancer, and ovarian cancer and all non-cancer at the system level (*e.g.,* immune, cardiovascular, respiratory) or higher." No explanation is provided for why the outcomes considered relevant for asbestos alone have this restriction.

The comparator element for human studies in most PECO statements includes text indicating that exposure "for a shorter period of time" constitutes an appropriate comparison; however, this text is not provided for TBBPA (***Tables Apx H-27*** and ***H-29***) or asbestos (***Tables Apx H-55*** and ***H-57***), and no explanation for the omission is provided.

The exposure element in the formaldehyde PECO includes a unique component, identifying occupational categories that can serve as proxies for exposure: "Studies on occupations known to use or produce formaldehyde (*e.g.,* pathologists, funeral directors, embalmers) should be considered a relevant proxy for formaldehyde exposure." This is a useful addition to the PECO, and EPA should evaluate whether similar listings of occupational categories would be useful for other chemicals.

In addition, the exposure element of the phthalates PECO for title-abstract screening (***Table Apx H-47***) includes two conflicting bullet points in the "Screener Note:"

- "Studies involving exposures to mixtures will be included only if they also include exposure to DBP, BBP, DEHP, DIBP, dicyclohexyl phthalate, DIDP or DINP alone. Otherwise, mixture studies will be tagged as mixture studies will be tagged as supplemental."
- "Animal and plant studies involving exposures to mixtures will be **included** only if they also include exposure to DBP, BBP, DEHP, DIBP, dicyclohexyl phthalate, DIDP and/or DINP alone. Otherwise, animal and plant mixture studies will be tagged as **Supplemental**. Human mixture studies are ***included."***

According to the first bullet point, human mixture studies lacking exposures to individual phthalates are "supplemental," and according to the second bullet point such studies are "included." At the full-text stage, only the first bullet is provided, resolving the internal conflict to exclude mixture studies but tag them as supplemental. However, this is different from full-text screening PECOs for several other chemicals, which include the statement that "human mixture studies are included" (five chlorinated solvents, HHCB, TBBPA, TCEP). No explanation is provided for why human mixture studies are included for some chemicals and are tagged as supplemental for other chemicals.

d. PECOs should explicitly list metabolites of target chemicals where appropriate.

The exposure elements of several PECO statements mention studies of chemical metabolites as being relevant/included without naming those metabolites. Inclusion of chemical metabolite studies in the PECOs is important to clarify that these studies are relevant for hazard assessment, but particularly in cases where there are known studies of metabolites it is important to reduce uncertainty in the screening process by explicitly naming those metabolites in the PECO statement. This is particularly important for phthalates. As in other PECOs, the phthalates PECO (***Tables Apx H-47*** and ***H-49***) for human studies notes that exposure may be represented by "metabolites of these chemicals in a biological matrix (*i.e.,* urine, blood, semen, etc.). See list of common metabolites for each phthalate below." However, no list of metabolites is provided in the document. This omission is critical because there is a large body of human epidemiological evidence related to multiple health outcomes based on measurements of urinary phthalate metabolites. Failure to provide a list of the metabolites to the screeners could lead to studies being inappropriately excluded from the TSCA risk evaluation. In addition, the comparators element of the PECO statement lists the relevant phthalates but omits the metabolites, again potentially creating confusion for inclusion/exclusion decisions.

We recommend:

- **Including studies of biochemical changes and other early biological indicators of hazards as relevant to the human and animal evidence streams, removing the restriction that limits the relevant health effects studies to those with apical or organ-level outcomes.**
- **Applying a single, peer-reviewed PECO statement throughout the process for each risk evaluation, including for both title-abstract screening and full-text review.**

3. The method outlined in the *2021 Draft TSCA Method* fails to adhere to best practice for conducting data extraction for systematic review

a. The draft document provides insufficient detail concerning data extraction for hazard studies.

The *2021 Draft TSCA Method* presents only minimal information on data extraction for human health studies:

- Human health hazard studies of animal toxicity data
    - Target organ system, species/strain/sex, dose/concentrations, study duration, hazard value, and effect.
- Human health hazard studies of epidemiological data
    - Endpoint, study population, exposure, and results specific to individual target organs/systems.

No further details are provided in the document. It is difficult to evaluate the completeness of EPA's approach to data extraction for hazard studies based on this limited text. The *2021 Draft TSCA Method* provides data extraction templates for other disciplines, but there are no tables/templates for health effects studies.

b. Data extraction should be conducted independently by two separate reviewers.

The TSCA approach to data extraction is for two individuals to participate in sequence, rather than independently, "For quality control, data extraction is performed by one member of the evaluation team and independently verified by at least one other member. Discrepancies in data extraction are resolved by discussion or consultation with a third member of the evaluation team."[94]

This process is likely to produce more errors in data extraction than a process in which data extraction is conducted independently and in parallel by two individuals. The *Cochrane Handbook for Systematic Reviews of Interventions* considers data extraction by two people working independently to be "highly desirable" for extracting study characteristics, and "mandatory" for extracting study outcome data:

> Because errors that occur at the data extraction stage are rarely detected by peer reviewers, editors, or users of systematic reviews, it is recommended that more than one person extract data from every report to minimize errors and reduce introduction of potential biases by review authors… As a minimum, information that involves subjective interpretation and information that is critical to the interpretation of results (e.g., outcome data) should be extracted independently by at least two people… Evidence in support of duplicate data extraction comes from several indirect sources. One study observed that independent data extraction by two authors resulted in fewer errors than data extraction by a single author followed by verification by a second (Buscemi et al 2006). A high prevalence of data extraction errors (errors in 20 out of 34 reviews) has been observed (Jones et al 2005). A further study of data extraction to compute standardized mean differences found that a minimum of seven out of 27 reviews had substantial errors (Gøtzsche et al 2007).[95]
>
> **We recommend:**
>
> - **Publishing details on data extraction, such as a template, for health effects studies**
> - **Conducting data extraction by two reviewers operating independently in parallel.**

4. The *2021 Draft TSCA Method's* approach to study evaluation is incompatible with validated best practice methods already being implemented in environmental health and fails to adhere to the NASEM recommendations to EPA in its review *of Application of Systematic Review in TSCA Risk Evaluations* in fundamental ways.

[94] US EPA. (2021). Draft systematic review protocol supporting TSCA risk evaluations for chemical substances version 1.0: A generic TSCA systematic review protocol with chemical-specific methodologies, pp. 81. https://www.epa.gov/system/files/documents/2021-12/draft-systematic-review-protocol-supporting-tsca-risk-evaluations-for-chemical-substances_0.pdf

[95] Li T, Higgins JPT, Deeks JJ (editors). Chapter 5: Collecting data. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). *Cochrane Handbook for Systematic Reviews of Interventions* version 6.2 (updated February 2021). Cochrane, 2021. Available from www.training.cochrane.org/handbook.

The approach outlined below to conducting evaluations of human epidemiology and animal toxicology studies, is incompatible with validated best practice methods already being implemented in environmental health and critically, **has to not adhered to the recommendations to USEPA** in the NASEM review of *Application of Systematic Review in TSCA Risk Evaluations* in fundamental ways, including that it:

a. Continues to use an inappropriate quantitative scoring method almost identical to the approach in the *2018 TSCA Method.*[96]

EPA states that:

> In response to a variety of commenters, including NASEM and SACC, the TSCA SR Protocol does not include a quantitative/weighted scoring system for data evaluation. Rather, the TSCA SR Protocol applies ordinal rankings to guide the qualitative categorization of high, medium, low, or critically deficient for each data evaluation metric. [97]

This statement is misleading because EPA is still using a quantitative scoring method to derive the ordinal rankings. While the ordinal ranking appears to be qualitative, it is actually based on instructions (described below) for translating a quantitative score into a corresponding qualitative category (high, medium or low). EPA has simply obscured the scoring with the language regarding ordinal rankings. Although EPA has removed unequal weighting of metrics, it is still applying quantitative scoring with all metrics weighted equally. According to EPA:

> The individual metric rankings are based on an ordinal scale and are used to obtain an overall study ranking. All metrics have an equal weight in determining the overall study ranking, and domains within which the metrics are nested (*e.g.,* test substance) are strictly used to categorize the individual metrics. Therefore, each domain may have different numbers of metrics. The resulting values are converted to an ordinal quality level (*high, medium, low, critically deficient*, or *not rated/not applicable*). [98]

EPA goes on to state that:

> The data quality evaluation metrics (Appendix K through Appendix T) each receive a ranking (1, 2, 3, 4, or 0 for metrics that are N/A). The ordinal ranking of each metric provides consistency and transparency to the evaluation process and is translated to an overall quality rating which informs the characterization of studies during the evidence integration phase. See Appendix Q for an example of how metric rankings are used to obtain the overall study ranking for animal toxicity studies. This categorical ranking of the data evaluation system is not intended to imply a false sense of precision and/or accuracy implicit in other numerical scoring systems previously employed in the first 10 risk evaluations. Therefore, the qualitative study quality rankings are not the numerical scores used in the evidence integration of previous risk evaluations and is a

[96] US EPA. (2018). Application of systematic review in TSCA risk evaluations. https://www.epa.gov/sites/default/files/2018-06/documents/final_application_of_sr_in_tsca_05-31-18.pdf

[97] US EPA. (2021). Draft systematic review protocol supporting TSCA risk evaluations for chemical substances version 1.0: A generic TSCA systematic review protocol with chemical-specific methodologies, pp. 27. https://www.epa.gov/system/files/documents/2021-12/draft-systematic-review-protocol-supporting-tsca-risk-evaluations-for-chemical-substances_0.pdf

[98] US EPA. (2021). Draft systematic review protocol supporting TSCA risk evaluations for chemical substances version 1.0: A generic TSCA systematic review protocol with chemical-specific methodologies, pp. 68-69. https://www.epa.gov/system/files/documents/2021-12/draft-systematic-review-protocol-supporting-tsca-risk-evaluations-for-chemical-substances_0.pdf

> significant change from previous publication in EPA's Application of Systematic Review in TSCA Risk Evaluations (U.S. EPA, 2018a). [99]

This passage, however, overstates the extent of change from the scoring method presented in the previous approach described in *Application of Systematic Review in TSCA Risk Evaluations* (*2018 TSCA Method*). Quantitative scoring is still applied, with calculation of an average metric score for each study and The average metric score is then used to assign the study an overall ranking of high, medium, or low – for example an average metric value of less than 1.7 results in a "high" ranking. **This and other cutoff values used for binning studies into high, medium, or low rankings are identical to those used in the first 10 TSCA existing chemical risk evaluations.** In the previous approach described the*2018 TSCA Method,* EPA states:

> A numerical scoring method is used to convert the confidence level for each metric into the overall quality level for the data/information source. The overall study score is equated to an overall quality level (*High*, *Medium*, or *Low*) using the level definitions and scoring scale shown in Table A-1. The scoring scale was obtained by calculating the difference between the highest possible score of 3 and the lowest possible score of 1 (i.e., 3-1= 2) and dividing into three equal parts (2 ÷ 3 = 0.67). This results in a range of approximately 0.7 for each overall data quality level, which was used to estimate the transition points (cut-off values) in the scale between *High* and *Medium* scores, and *Medium* and *Low* scores. These transition points between the ranges of 1 and 3 were calculated as follows: Cut-off values between *High* and *Medium*: 1 + 0.67= 1.67, rounded up to 1.7 (scores lower than 1.7 will be assigned an overall quality level of *High*) Cut-off values between *Medium* and *Low*: 1.67 + 0.67= 2.34, rounded up to 2.3 (scores between 1.7 and lower than 2.3 will be assigned an overall quality level of *Medium)*.[100]

Illustrative of how the *2021 Draft TSCA Method* scoring method is almost identical to the previous scoring method can be seen at the bottom of ***Table_Apx R-4. Summary of Domains, Metrics, and Range of Metric Rankings for Studies with Biomarkers*** (see Table below):

> For the purposes of calculating an overall study ranking, the range of possible metric rankings is 1 (high) to 3 (low) for each metric. No calculations will be conducted if a study receives an "critically deficient" rating (="4") for any metric. The sum of rankings will differ if some metrics are not ranked (*i.e.,* they are not applicable). The cutoffs between categories were defined by calculating the difference between the highest possible ranking of 3 and the lowest possible ranking of 1 (*i.e.,* 3 – 1 = 2) and dividing it into three equal parts (2 ÷ 3 = 0.67). This results in a range of approximately 0.7 for each overall study data quality ranking, which is used to estimate the transition points (cut-off values) in the scale between high and medium rankings, and medium and low rankings. These transition points between the ranges of 1 and 3 are determined as follows: cut-off values between high and medium: 1 + 0.67 = 1.67, rounded to 1.7 (rankings lower than 1.7 are assigned an overall quality level of high); and cut-off values

[99] US EPA. (2021). Draft systematic review protocol supporting TSCA risk evaluations for chemical substances version 1.0: A generic TSCA systematic review protocol with chemical-specific methodologies, pp. 71. https://www.epa.gov/system/files/documents/2021-12/draft-systematic-review-protocol-supporting-tsca-risk-evaluations-for-chemical-substances_0.pdf

[100] EPA (2018) Application of Systematic Review in TSCA Risk Evaluations. Page. 30.

between medium and low: 1.67 + 0.67 = 2.34, rounded to 2.3 (rankings between 1.7 and lower than 2.3 are assigned an overall quality level of medium). [101]

Table_Apx R-4. Summary of Domains, Metrics, and Range of Metric Rankings for Studies with Biomarkers

Domain	Metric	Range of Metric Rankings
Study Participation	Participant Selection	1 to 3
	Attrition	1 to 3
	Comparison Group	1 to 3
Exposure Characterization	Measurement of Exposure	1 to 3
	Exposure Levels	1 to 3
	Temporality	1 to 3
Outcome Assessment	Outcome Measurement or Characterization	1 to 3
	Reporting Bias	1 to 3
Potential Confounding/ Variable Control	Covariate Adjustment	1 to 3
	Covariate Characterization	1 to 3

[101] US EPA. (2021). Draft systematic review protocol supporting TSCA risk evaluations for chemical substances version 1.0: A generic TSCA systematic review protocol with chemical-specific methodologies, pp. 644. https://www.epa.gov/system/files/documents/2021-12/draft-systematic-review-protocol-supporting-tsca-risk-evaluations-for-chemical-substances_0.pdf

	Co-exposure Confounding/Moderation/Mediation	1 to 3
Analysis	Study Design and Methods	1 to 3
	Statistical Power	1 to 3
	Reproducibility of Analyses	1 to 3
	Statistical Models	1 to 3
Other (if applicable) Considerations for Biomarker Selection and Measurement (Lakind et al., 2014)	Use of Biomarker of Exposure	1 to 3
	Effect Biomarker	1 to 3
	Method Sensitivity	1 to 3
	Biomarker Stability	1 to 3
	Sample Contamination	1 to 3
	Method Requirements	1 to 3
	Matrix Adjustment	1 to 3
Range of sums (if all metrics ranked)[b]		22 to 66
Range of overall rankings[c] (if all metrics ranked)		1 to 3 (22/22 to 66/24)

High	Medium	Low
≥1 and <1.7	≥1.7 and <2.3	≥ 2.3 and 3

[a] For the purposes of calculating an overall study ranking, the range of possible metric rankings is 1 (high) to 3 (low) for each metric. No calculations will be conducted if a study receives an "critically deficient" rating (= "4") for any metric.

[b] The sum of rankings will differ if some metrics are not ranked (*i.e.*, they are not applicable).

[c] The cutoffs between categories were defined by calculating the difference between the highest possible ranking of 3 and the lowest possible ranking of 1 (*i.e.*, 3 – 1 = 2) and dividing it into three equal parts (2 ÷ 3 = 0.67). This results in a range of approximately 0.7 for each overall study data quality ranking, which is used to estimate the transition points (cut-off values) in the scale between high and medium rankings, and medium and low rankings. These transition points between the ranges of 1 and 3 are determined as follows:

- cut-off values between high and medium: 1 + 0.67 = 1.67, rounded to 1.7 (rankings lower than 1.7 are assigned an overall quality level of high); and
- cut-off values between medium and low: 1.67 + 0.67 = 2.34, rounded to 2.3 (rankings between 1.7 and lower than 2.3 are assigned an overall quality level of medium).

Table_Apx R-4. Summary of Domains, Metrics, and Range of Metric Rankings for Studies with Biomarkers[102]

As shown in ***Table Apx R-4***, to calculate an overall quality score, EPA assigns a numerical rating of 1 (high) to 3 (low) for each metric, with no calculation being conducted if a study receives a 4 (critically deficient) rating for **any metric.** EPA then sums all of the metric scores (numerator) and divides it by the total number of metrics in the tool (denominator) to determine the overall quality score for the study. EPA has assigned arbitrary cut offs to designate studies 'Low', 'Medium' or 'High' quality. In the example shown above here in ***Table Apx R-4***, EPA has also incorrectly calculated the denominator, which should be 22 based off the 22 metrics in this tool, and not 24 as currently stated.

Despite the NASEM recommendation in the review of the *2018 TSCA Method* explicitly telling 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."[103] EPA has disregarded the NASEM and continued with an almost identical scoring method as shown above. Further, the approach described by EPA in the *2021*

[102] US EPA. (2021). Draft systematic review protocol supporting TSCA risk evaluations for chemical substances version 1.0: A generic TSCA systematic review protocol with chemical-specific methodologies, pp. 644. https://www.epa.gov/system/files/documents/2021-12/draft-systematic-review-protocol-supporting-tsca-risk-evaluations-for-chemical-substances_0.pdf

[103] 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.

Draft TSCA Method directly contradicts the recommendation of House Committee on Science, Space & Technology Chair Rep. Eddie Bernice Johnson in a letter sent to EPA Administrator Michael Regan that "In its updated methodology, EPA must also wholly supplant the "numerical scoring system" from the 2018 Systematic Review"[104] as "This unusual method for data synthesis is out of step with common scientific practice."[105]

As we have stated in previous comments to EPA, 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. "Quality scores" have not been able to distinguish between studies with a high and low risk of bias in meta-analyses [106] and empirical evidence is lacking to establish how each risk of bias item should be weighted.[107] The use of scores falsely implies a relationship between scores (i.e. high vs low) and effect or reliability of a study, and therefore the use of only "high" and "medium" quality studies will lead to a biased body of evidence.

The NASEM review of the *2018 TSCA Method* highlighted:

> 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. [108]

EPA must comply with the NASEMs recommendations and House Committee on Science, Space & Technology Chair Rep. Eddie Bernice Johnson's recommendations and use a domain approach to conducting study quality evaluations, in place of the numerical scoring system described in the *2021 Draft TSCA Method*. This quantitative scoring method is applied to all streams of evidence, and our comments reflect our objection to EPA's applying scoring to any and all streams of evidence.

In addition, aspects of the *2021 Draft TSCA Method's* quantitative approach are arbitrary. In ***Tables Apx Q-9*** and ***Apx R-7*****,** details are provided on the metrics used for scoring toxicology and epidemiology studies. The tables show that for many metrics, three choices are available (high=1, medium=2, or low=3). For other metrics, only two choices are available. In these latter instances, there is variability in treatment across metrics: for some metrics, the choices are (high=1 or low=3), and for other metrics the choices are (medium=2 or low=3). No explanation is given for why a choice of "high" is available for some of these dichotomous metrics and not for others, and no explanation is given for the selection of dichotomous metrics for which "high" is an option and the selection of metrics offering "medium"

[104] House Science, Space, and Technology Committee Chairwoman Eddie Bernice Johnson. Letter to The Honorable Michael Regan Administrator U.S. Environmental Protection Agency, October 20, 2021. https://science.house.gov/news/press-releases/chairwoman-johnson-asks-epa-administrator-regan-to-review-toxic-substances-control-act-systematic-review-methodology#:~:text=(Washington%2C%20DC)%20%E2%80%93%20Today,risk%20evaluations%20conducted%20under%20the

[105] House Science, Space, and Technology Committee Chairwoman Eddie Bernice Johnson. Letter to The Honorable Michael Regan Administrator U.S. Environmental Protection Agency, October 20, 2021. https://science.house.gov/news/press-releases/chairwoman-johnson-asks-epa-administrator-regan-to-review-toxic-substances-control-act-systematic-review-methodology#:~:text=(Washington%2C%20DC)%20%E2%80%93%20Today,risk%20evaluations%20conducted%20under%20the

[106] 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.

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

[108] 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.

instead. The *2021 Draft TSCA Method* seems to indicate that the difference was intentional, at least for epidemiology studies, and justified by the resulting study scores:

> EPA removed the *High* criterion from some metrics, particularly in dichotomous metrics (*e.g., High/Low*) that were primarily being binned as *High* by reviewers across the majority of the studies. These dichotomous metrics were contributing to the overall quality scores being skewed towards *High.* To address this, EPA shifted some of the dichotomous metrics such that the highest metric ranking possible (for all studies) is a *Medium*. The change led to the dichotomous metrics having less significant impact to the ordinal ranking and the overall quality rating for each study.[109]

This differential treatment of dichotomous metrics introduces an arbitrary element into the overall score for a study, and also implicitly gives greater weight to dichotomous metrics in which a selection of "high" is available relative to those dichotomous metrics for which "medium" is the option available for studies deemed to satisfy the criteria. In addition, the information in ***Tables Apx Q-6*** and ***Apx R-4***, showing that scores of 1-3 are available for all metrics, is incorrect since a score of "high=1" cannot be assigned for several metrics.

A further indication of problems created by assigning an overall score, and the method used for making this assignment, is that the *2021 Draft TSCA Method* allows the study quality reviewers to override the results:

> Although the overall data quality level is derived from the summary of individual metric rankings and serves as the baseline for consideration of the study quality, it may not be the final decision. The primary reviewer and QC reviewer may agree to provide an updated ranking; in such cases, they must provide a justification for the ranking adjustment to ensure transparency for the decision.[110]
>
> Any single set of data quality criteria, even for a given category of studies (*e.g.,* animal toxicity studies), cannot necessarily address all aspects of quality relevant for an individual study in the category. Thus, EPA allowed reviewers the ability to adjust the final rankings based on professional judgment.[111]

The *2021 Draft TSCA Method* is unclear on what criteria are applied in determining an overall ranking should be adjusted, and what documentation is required in such instances.

b. Continues to inappropriately use an overall study quality rating.

EPA states that:

> The data quality evaluation domains are evaluated by assessing two to seven unique metrics. Each metric is binned into a data quality level of high, medium, low, or critically deficient. Each

[109] US EPA. (2021). Draft systematic review protocol supporting TSCA risk evaluations for chemical substances version 1.0: A generic TSCA systematic review protocol with chemical-specific methodologies, pp. 150. https://www.epa.gov/system/files/documents/2021-12/draft-systematic-review-protocol-supporting-tsca-risk-evaluations-for-chemical-substances_0.pdf

[110] US EPA. (2021). Draft systematic review protocol supporting TSCA risk evaluations for chemical substances version 1.0: A generic TSCA systematic review protocol with chemical-specific methodologies, pp. 171. https://www.epa.gov/system/files/documents/2021-12/draft-systematic-review-protocol-supporting-tsca-risk-evaluations-for-chemical-substances_0.pdf

[111] US EPA. (2021). Draft systematic review protocol supporting TSCA risk evaluations for chemical substances version 1.0: A generic TSCA systematic review protocol with chemical-specific methodologies, pp. 151. https://www.epa.gov/system/files/documents/2021-12/draft-systematic-review-protocol-supporting-tsca-risk-evaluations-for-chemical-substances_0.pdf

> data quality ranking is assigned a value (i.e., 1 through 4) that is used in the method of assessing the overall quality of the study.[112]

The *2021 Draft TSCA Method* continues:

> The individual metric rankings are based on an ordinal scale and are used to obtain an overall study ranking. All metrics have an equal weight in determining the overall study ranking, and domains within which the metrics are nested (*e.g.,* test substance) are strictly used to categorize the individual metrics. Therefore, each domain may have different numbers of metrics. The resulting values are converted to an ordinal quality level (*high, medium, low, critically deficient*, or *not rated/not applicable*). [113]

In contrast, the NASEM review of the *2018 TSCA Method* recommended *not* using an overall study confidence rating:

> 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).[114]

EPA has again failed to provide any scientific evidence to support deriving an overall study rating from equal weighting of the selected metrics. The use of overall rankings derived from study scores falsely implies a relationship between the rankings (i.e. high vs low) and effect or reliability of a study and therefore the use of only "high" and "medium" quality studies will lead to a biased evaluation of the evidence.

Rating a study as overall "low" or "uninformative for dose-response" based on arbitrary measures is not validated. The approach for assessing quality of epidemiological studies consists of 22 metrics; among these, there are 12 metrics that can be rated as "Critically Deficient (Ranking = 4)" (**Table_Apx R-7. Evaluation Criteria for Epidemiological Studies**). [115] Assigning a ranking of "uninformative for dose-response" occurs automatically when any metric is scored as "Critically Deficient." As a result, a study is judged uninformative based on only one metric that is "Critically Deficient" without any consideration of the other aspects of the study. Further, the overall study confidence ratings of the TSCA tool are subject to reviewer judgement and expert opinion. Thus, it is expected that different groups with differing levels

[112] US EPA. (2021). Draft systematic review protocol supporting TSCA risk evaluations for chemical substances version 1.0: A generic TSCA systematic review protocol with chemical-specific methodologies, pp. 642. https://www.epa.gov/system/files/documents/2021-12/draft-systematic-review-protocol-supporting-tsca-risk-evaluations-for-chemical-substances_0.pdf

[113] US EPA. (2021). Draft systematic review protocol supporting TSCA risk evaluations for chemical substances version 1.0: A generic TSCA systematic review protocol with chemical-specific methodologies, pp. 68-69. https://www.epa.gov/system/files/documents/2021-12/draft-systematic-review-protocol-supporting-tsca-risk-evaluations-for-chemical-substances_0.pdf

[114] 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.

[115] US EPA. (2021). Draft systematic review protocol supporting TSCA risk evaluations for chemical substances version 1.0: A generic TSCA systematic review protocol with chemical-specific methodologies, pp. 647. https://www.epa.gov/system/files/documents/2021-12/draft-systematic-review-protocol-supporting-tsca-risk-evaluations-for-chemical-substances_0.pdf

of expertise would reach different conclusions. As such it demonstrates the essential problem with including an overall study rating, which may lead to the exclusion of a study or could inappropriately downgrade study findings in the overall assessment.

Overall, there is no scientific justification for EPA to derive overall study rankings by aggregating scores across individual domains. This approach will lead to a biased evaluation of the studies. We therefore strongly recommend against the use of an overall ranking 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.[116,117]

c. Continues to exclude or downgrade research based on one "critically deficient" methodological limitation.

EPA states that:

> Critically deficient - Serious flaws are noted related to the metric that consequently make **the study unusable for quantitative analyses**."[118] and that "EPA relies more heavily on studies with overall study quality rankings of high, medium, or low to quantitatively or qualitatively support risk evaluations (overall study quality rankings are defined below in ***Table 5-1***). Should any metric be rated as critically deficient [ranking of 4], the study will automatically have an overall study ranking of uninformative. The **quantitative use of an uninformative study may be inappropriate** (i.e., dose-response for hazard studies), however the information within an uninformative study may be used to qualitatively provide contextual or supportive information for the risk evaluation...The terminology of critically deficient and uninformative was made in collaboration with EPA/ORD and matches the terminology in the IRIS Handbook (U.S. EPA, 2020).[119]

EPA then goes on to state in ***Table 5-1, Definition of Overall Study Quality Rankings*** that:

> Uninformative: Serious flaw(s) are identified and therefore**, the data cannot be used or have strict limits on use** (e.g., it will not be used for dose-response assessment in hazard assessments).[120]

[116] 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.

[117] 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.

[118] US EPA. (2021). Draft systematic review protocol supporting TSCA risk evaluations for chemical substances version 1.0: A generic TSCA systematic review protocol with chemical-specific methodologies, pp. 69. https://www.epa.gov/system/files/documents/2021-12/draft-systematic-review-protocol-supporting-tsca-risk-evaluations-for-chemical-substances_0.pdf

[119] US EPA. (2021). Draft systematic review protocol supporting TSCA risk evaluations for chemical substances version 1.0: A generic TSCA systematic review protocol with chemical-specific methodologies, pp. 68-69. https://www.epa.gov/system/files/documents/2021-12/draft-systematic-review-protocol-supporting-tsca-risk-evaluations-for-chemical-substances_0.pdf

[120] US EPA. (2021). Draft systematic review protocol supporting TSCA risk evaluations for chemical substances version 1.0: A generic TSCA systematic review protocol with chemical-specific methodologies, pp. 68-69. https://www.epa.gov/system/files/documents/2021-12/draft-systematic-review-protocol-supporting-tsca-risk-evaluations-for-chemical-substances_0.pdf

Finally, in ***Table_Apx Q-9. Data Quality Criteria for Animal Toxicity Studies***[121], EPA states that for 19 out of the 24 metrics, studies with a "Critically Deficient" (Ranking = 4) rating have " a serious flaw that makes the study unusable" (Metrics 4-20, 22, and 23).

Although EPA says that it has dropped its previous approach of excluding studies ranked as "uninformative," the way the *2021 Draft TSCA Method* describes how EPA will use these studies is inconsistent, ambiguous and confusing. In section *5 DATA EVALUATION*, EPA begins by stating that for each evidence stream, a "critically deficient" rating in any metric **"**make the study unusable for quantitative analysis.*"* EPA then says, however, that any study with a metric rated as "critically deficient" will be rated as "uninformative" overall, and "the quantitative use of an uninformative study may be inappropriate.*"* EPA then changes the language again in Table 5-1 on how it will handle studies rated overall "uninformative" with a "critically deficient rating" and states that the data from these "cannot be used or have strict limits on use (e.g., it will not be used for dose-response assessment in hazard assessments)." Finally, in ***Table_Apx Q-9***, EPA explicitly states that for the majority of metrics in animal toxicity studies, a "critically deficient" rating "makes the study unusable," directly contradicting EPA's statements elsewhere in the document that studies with this rating are not excluded from further consideration. . It is unclear which of these inconsistent statements actually applies to use of the "critically deficient" studies, and it is unclear whether the treatment of these studies is the same across evidence streams, i.e. equivalent for human and animal studies.

EPA's approach in the *2021 Draft TSCA Method* is not the approach recommended by the NASEM in *The Use of Systematic Review in EPA's Toxic Substances Control Act Risk Evaluations (2021)*:

> 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[122]…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**[123] (emphasis added).

In its final recommendations in *The Use of Systematic Review in EPA's Toxic Substances Control Act Risk Evaluations (2021),* the NASEM recommended that EPA "**Do not exclude studies based on risk of bias, study quality, or reporting quality**"[124] (emphasis added).

---

[121] US EPA. (2021). Draft systematic review protocol supporting TSCA risk evaluations for chemical substances version 1.0: A generic TSCA systematic review protocol with chemical-specific methodologies, pp. 599. https://www.epa.gov/system/files/documents/2021-12/draft-systematic-review-protocol-supporting-tsca-risk-evaluations-for-chemical-substances_0.pdf

[122] 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.

[123] 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.

[124] 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.

The NASEM again emphasized these points in its review of the *ORD Staff Handbook for Developing IRIS Assessments,* saying that *"*A major concern is that studies that are judged as 'critically deficient' or 'deficient' in one evaluation domain are typically rated as 'uninformative' or 'low confidence' studies that are generally not considered further in the IRIS assessment."[125] This concern was validated when the NASEM highlighted that "EPA provided data from recent IRIS assessments showing that the proportion of human studies rated as 'uninformative' and excluded from further consideration ranged from 0 to 50 percent, and 0 to 41.5 percent for animal studies. Thus, depending on the IRIS assessment, excluding studies at the study evaluation stage could lead to a substantial proportion of excluded studies due to a critically deficient rating in one domain." [126] Therefore, the same approach as it is described in the *2021 Draft TSCA Method* may reduce the available evidence to assess the harms of environmental exposures in TSCA risk evaluation by erroneously excluding studies, which leads to inaccurate conclusions about the quality of the body of evidence. The NASEM went on to highlight that "study evaluation ratings should not be used to exclude studies."[127]

This approach to exclude studies based on only one methodological limitation when conducting study quality evaluations in the *2021 Draft TSCA Method* is inconsistent with two previously validated methods recommended by the NASEM, the Navigation Guide[128] and OHAT,[129] that are used to evaluate the risk of bias in human epidemiological studies and animal toxicological 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."[130]

While 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 EPA 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. [131]

[125] National Academies of Sciences, Engineering, and Medicine 2021. Review of U.S. EPA's ORD Staff Handbook for Developing IRIS Assessments: 2020 Version. Washington, DC: The National Academies Press. pp 49 https://doi.org/10.17226/26289.

[126] National Academies of Sciences, Engineering, and Medicine 2021. Review of U.S. EPA's ORD Staff Handbook for Developing IRIS Assessments: 2020 Version. Washington, DC: The National Academies Press. pp 49 https://doi.org/10.17226/26289.

[127] National Academies of Sciences, Engineering, and Medicine 2021. Review of U.S. EPA's ORD Staff Handbook for Developing IRIS Assessments: 2020 Version. Washington, DC: The National Academies Press. pp 49 https://doi.org/10.17226/26289.

[128] 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.

[129] 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.

[130] 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.

[131] 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.

d. Continues to conflate how well a study is reported with how well the underlying research was conducted and continues to include inappropriate appraisal criteria such as statistical power.

EPA states that:

> During data evaluation, EPA assesses the risk of bias, methodological quality, sensitivity and reporting of individual data sources…Each domain contains a unique set of metrics, or sub-categories of attributes, which are intended to assess an aspect of the methodological quality, sensitivity, risk of bias, or lack of reporting of the study…Reporting is the completeness in which study methods and/or results are described. [132]

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.[133]

However, how completely and clearly a study is reported is not a scientifically valid measure of the quality of the underlying research.[134,135,136,137] As GRADE methodologists have succinctly stated, "… just because a safeguard against bias is not reported does not mean it was neglected." [138] Research has documented that important information is often missing or unclear in published studies,[139] as word limits, styles, and other specifications are highly variable, and non-standardized among peer-reviewed

[132] US EPA. (2021). Draft systematic review protocol supporting TSCA risk evaluations for chemical substances version 1.0: A generic TSCA systematic review protocol with chemical-specific methodologies, pp. 66. https://www.epa.gov/system/files/documents/2021-12/draft-systematic-review-protocol-supporting-tsca-risk-evaluations-for-chemical-substances_0.pdf

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

[134] 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.

[135] 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

[136] 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

[137] 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.

[138] 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

[139] 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.

journals. As such, efforts to improve reporting are focused on uptake of reporting guidelines by journal editors and researchers.[140,141,142]

EPA goes on to state that:

> Among the aspects of study quality evaluated by EPA, reporting quality (*i.e.,* how completely an element was reported in a study) is an important consideration. Other frameworks (*e.g.,* the IRIS Program) use a separate reporting quality domain for animal toxicology studies. EPA's TSCA method integrates reporting quality metrics within each domain because reporting contributes to the evaluation of each facet of the data source. If sufficient methodological details are not reported, specific metrics that incorporate reporting quality cannot be rated as *high* because it is not possible to tell whether the methods were appropriate. The challenge, in many cases, is to distinguish a deficit in reporting from a problem in the underlying study methods.[143]

EPA's characterization of this issue does not take into account the NASEM statement in *The Use of Systematic Review in EPA's Toxic Substances Control Act Risk Evaluations (2021)* 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**[144] (emphasis added).

The *Cochrane Handbook* draws a distinction between assessment of methodological quality (including reporting 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.

[140] 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

[141] 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.

[142] 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

[143] US EPA. (2021). Draft systematic review protocol supporting TSCA risk evaluations for chemical substances version 1.0: A generic TSCA systematic review protocol with chemical-specific methodologies, pp. 69. https://www.epa.gov/system/files/documents/2021-12/draft-systematic-review-protocol-supporting-tsca-risk-evaluations-for-chemical-substances_0.pdf

[144] 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.

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).

In *The Use of Systematic Review in EPA's Toxic Substances Control Act Risk Evaluations (2021)* the NASEM stated that:

> Many markers of a high-quality study (e.g., whether a study's investigator has performed a sample size calculation and whether the study is reported adequately or has received appropriate ethical approvals) are unlikely to have any direct implication for the potential for a study to be affected by bias.[145]
>
> **Statistical power and statistical significance are not markers of risk of bias or quality**. **Statistical significance is not a measure of association or strength of association and should not be used to evaluate studies**. In fact, combining multiple small, low-powered but similar studies in a synthesis is one of the potential benefits of systematic review [146] (emphasis added).

However, despite these very explicit NASEM statements about the inappropriateness of these metrics being included in the study quality evaluations, EPA continues to use statistical power (sensitivity) as a study quality metric (***Table_Apx R-7. Evaluation Criteria for Epidemiological Studies***, Metric 13). Additionally, the NASEM review of the *ORD Staff Handbook for Developing IRIS Assessments* expressed similar concerns regarding non-standard considerations for assessing study quality:

> While standard methods for systematic review would consider only risk of bias when evaluating individual studies (Higgins and Thomas, 2019), study evaluation in the handbook is intentionally broadened to include the additional study evaluation domains of "sensitivity" and "reporting quality." **It is not standard practice to include the concepts of sensitivity or reporting quality as part of the evaluation of individual studies included in systematic reviews of human research**, although these domains are sometimes part of study evaluation for systematic reviews of animal studies (SYRCLE [Hooijmans et al., 2018).[147] (emphasis added)

Finally, in *The Use of Systematic Review in EPA's Toxic Substances Control Act Risk Evaluations (2021)* the NASEM recommended that EPA "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**" [148] (emphasis added).

e. Continues to ignore evaluation of funding bias.

[145] 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.

[146] 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.

[147] National Academies of Sciences, Engineering, and Medicine 2021. Review of U.S. EPA's ORD Staff Handbook for Developing IRIS Assessments: 2020 Version. Washington, DC: The National Academies Press. pp 47 https://doi.org/10.17226/26289.

[148] 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.

The NAS 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"[149] and in its recent review of the *ORD Staff Handbook for Developing IRIS Assessments*, the NASEM stated that "The handbook should describe how to detect and assess the effect of funding bias on the confidence of study ratings from evidence evaluation or effect estimates from synthesis."[150]

The NASEM 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.[151,152] 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. 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.[153,154,155,156]

The Navigation Guide assesses financial conflicts of interest as a separate domain within its risk of bias evaluation. [157] 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 assessment of the selective reporting of results, publication bias, and in assessing inconsistency in a body of evidence. [158,159,160] 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.[161]

---

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

[150] National Academies of Sciences, Engineering, and Medicine 2021. Review of U.S. EPA's ORD Staff Handbook for Developing IRIS Assessments: 2020 Version. Washington, DC: The National Academies Press. pp 4 https://doi.org/10.17226/26289.

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

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

[153] 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.

[154] 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

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

[156] 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

[157] 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.

[158] 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.

[159] 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

[160] 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.

[161] 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.

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 NASEM 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."[162]

f. Continues to be infeasible to implement.

In the recent consensus report *The Use of Systematic Review in EPA's Toxic Substances Control Act Risk Evaluations (2021)* the NASEM highlighted that:

> …completing the detailed evaluations of each study that may be included with risk evaluation is time consuming. In a study comparing the risk-of-bias assessments for epidemiologic studies from OHAT, the IRIS Program, and TSCA, the authors found that the TSCA evaluation tool took the most time to complete with a mean of 40 minutes per study, compared to 32 minutes (IRIS) and 20 minutes (OHAT) (Eick et al. 2020). While a mean increase of 8 minutes of review time per study may not seem that laborious, it is potentially severely burdensome for reviews with many studies." [163]

As EPA still has 22 metrics for the Evaluation Criteria for Epidemiological Studies, the feasibility of conducting study quality evaluations in TSCA risk evaluations is therefore still a great concern. Once again, this this further supports the use of an established risk of bias approach such as the NTP OHAT method, as recommend by the NASEM to EPA "Use established tools for assessing risk of bias and study quality such as those developed for use by OHAT or the Navigation Guide." [164]

We recommend:

- **Using existing, validated methods for assessing study quality.**
- **Considering all relevant studies in evidence synthesis and evidence integration, with no exclusion of studies based on study quality ratings.**
- **Following NASEM recommendations regarding appropriate considerations for study quality evaluation.**

5. The *2021 Draft TSCA Method's* approach to evidence integration is unclear and inconsistent.

a. The *2021 Draft TSCA Method* fails to draw a clear distinction between evidence synthesis and evidence integration.

The *2021 Draft TSCA Method l* uses the term "integration" to describe two separate steps in the process of deriving conclusions from the evidence: "The integration of separate bodies of evidence (*i.e.,* human, animal, and mechanistic evidence) described in this section directly informs the integration across all

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

[163] 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.

[164] 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.

evidence to draw an overall judgment for each of the assessed human health effects."[165]Integration is first described as evaluation of separate bodies of human, animal and mechanistic evidence, then it is described as an overall judgment across the bodies of evidence.

This usage is contrary to the recommendation regarding terminology in the NASEM 2021 review of the TSCA systematic review method: "Separate evidence synthesis from evidence integration. Evidence synthesis deals with more homogeneous data within a single stream, and evidence integration deals with more heterogeneous data from multiple streams."[166]

EPA could improve the clarity of its process by adopting the NASEM recommendation to use "synthesis" for drawing conclusions separately for each evidence stream (*i.e.,* human, animal, and mechanistic evidence) and use "integration" for drawing conclusions considering all evidence streams in combination.

b. The TSCA approach to within-stream evidence synthesis needs improvement.

Table 7-13 presents several considerations for within-stream evidence synthesis (as discussed above, the *2021 Draft TSCA Method* uses the term "integration" for this step), including quality of the database, consistency, magnitude and precision, biological gradient/dose-response, and biological plausibility and relevance to humans. A similar table is found in the draft *IRIS Handbook* (***Table 11-2***), but the TSCA version includes some important differences that significantly detract from the validity of the approach.

Regarding quality of the database, the table includes among the "factors that decrease strength" this statement: "a very limited database (*e.g.,* only one relevant study available) would also decrease strength for a given observed outcome." This is inappropriate and conflicts with the draft *IRIS Handbook* ***Table 11-2***, which states that "A health effect evidence base of a single or a few studies does not, on its own, decrease evidence strength." In addition, EPA's 2005 *Guidelines for Carcinogen Risk Assessment* identified instances in which a single study is sufficient to support a conclusion of "Likely to be Carcinogenic to Humans."[167]

Regarding magnitude and precision, among factors that increase strength is a statement that "biological significance should be considered in addition to **merely** statistical significance" (emphasis added). The IRIS version of the table, in contrast, places greater emphasis on biological significance relative to statistical significance: "Precise results from individual studies or across the set of studies increases strength, noting that biological significance is prioritized over statistical significance." Lack of statistical significance may be due to a small sample size and is not necessarily an indication of either lack of quality of absence of association. Leading statisticians have identified numerous problems with classifying studies as either statistically significant or not statistically significant and emphasize more in-depth consideration of confidence intervals. They have also observed that, regarding the standard use

[165] US EPA. (2021). Draft systematic review protocol supporting TSCA risk evaluations for chemical substances version 1.0: A generic TSCA systematic review protocol with chemical-specific methodologies, pp. 111. https://www.epa.gov/system/files/documents/2021-12/draft-systematic-review-protocol-supporting-tsca-risk-evaluations-for-chemical-substances_0.pdf

[166] 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 45. https://doi.org/10.17226/25952.

[167] US EPA. (2005). Guidelines for carcinogen risk assessment, pp. 2-55. https://www.epa.gov/sites/default/files/2013-09/documents/cancer_guidelines_final_3-25-05.pdf

of 95 percent confidence intervals, "the default 95% used to compute intervals is itself an arbitrary convention."[168]
Regarding biological gradient/dose-response, it is unclear why EPA departs from the GRADE approach, in which the dose-response consideration is applied only to "increase strength" and not to "decrease strength."

Regarding biological plausibility and relevance to humans, the NASEM review of the draft *IRIS Handbook* commented that this consideration should not be included at the within-stream evidence synthesis stage, but should only be included at the evidence integration stage:

> Overall, the applications of "biological plausibility" in the handbook appear to either (1) address considerations already covered elsewhere, such as consistency; (2) be more appropriate to consider during evidence integration; or (3) involve comparison with data on mechanistic effects. The latter two are more appropriate to incorporate during the evidence integration step.[169]

Regardless of where the consideration of biological plausibility is included, it should be amended with a statement that lack of knowledge re mechanism neither increases nor decreases strength. The draft *IRIS Handbook* table makes this point as follows:

> Mechanistic understanding is not a prerequisite for drawing a conclusion that a chemical causes a given health effect; thus, an absence of knowledge should not be used a basis for decreasing strength (NTP, 2015; NRC, 2014).[170]

> The *2021 Draft TSCA Method* includes a similar statement in the text: "However, mechanistic understanding is not a prerequisite for drawing a conclusion that a chemical causes a given health effect (NTP, 2015a; NRC, 2014)."[171]

This point should be incorporated in ***Table 7-13***, with the additional phrase found in the IRIS version that "thus, an absence of knowledge [regarding mechanisms] should not be used a basis for decreasing strength."

c. EPA's example using six human neurotoxicity studies reaches a poorly-justified conclusion that does not appropriately apply the identified considerations for evidence synthesis.

EPA presents an example of within-stream evidence synthesis, based on the considerations in ***Table 7-13***, for a set of six human neurotoxicity studies.[172] The example concludes that the evidence is "indeterminate," but this conclusion is not supported by the described evidence and relies excessively on consideration of statistical significance. The example describes "high quality yet small cohort studies"

[168] Amrhein, V, Greenland, S, McShane, B. "Scientists rise up against statistical significance." Nature. 20 March 2019. https://www.nature.com/articles/d41586-019-00857-9
[169] National Academies of Sciences, Engineering, and Medicine. 2022. Review of U.S. EPA's ORD Staff Handbook for Developing IRIS Assessments: 2020 Version, pp. 67. Washington, DC: The National Academies Press. https://doi.org/10.17226/26289.
[170] US EPA. (2021). ORD staff handbook for developing IRIS assessments, pp. 11-12 https://cfpub.epa.gov/ncea/iris_drafts/recordisplay.cfm?deid=350086
[171] US EPA. (2021). Draft systematic review protocol supporting TSCA risk evaluations for chemical substances version 1.0: A generic TSCA systematic review protocol with chemical-specific methodologies, pp. 69. https://www.epa.gov/system/files/documents/2021-12/draft-systematic-review-protocol-supporting-tsca-risk-evaluations-for-chemical-substances_0.pdf
[172] US EPA. (2021). Draft systematic review protocol supporting TSCA risk evaluations for chemical substances version 1.0: A generic TSCA systematic review protocol with chemical-specific methodologies, pp. 119-120. https://www.epa.gov/system/files/documents/2021-12/draft-systematic-review-protocol-supporting-tsca-risk-evaluations-for-chemical-substances_0.pdf

with findings of elevated relative risks that are dose-responsive but not statistically significant. Medium and low-quality cross-sectional studies "did not report any neurotoxicological outcomes, however neurotoxicity was not the focus of those studies." The cross-sectional studies appear to be irrelevant to the question of neurotoxicity, so it is unclear why they are even mentioned.

EPA's summary description of the evidence for this example is inappropriate, as it reports point estimates of odds ratios and emphasizes determinations regarding statistical significance without any presentation of confidence intervals. Confidence intervals are critical to appropriate interpretation of results.[173]

EPA's evaluation of the evidence is confusing because it does not follow clearly from the set of considerations outlined in ***Table 7-13***. Two bullet points assign judgments of "increased strength" and three bullet points judge "decreased strength," but it is unclear, for example, which of these judgments pertains to the consideration "quality of the database."

Among these bullet points representing EPA's evaluation of the evidence is a judgment of "decreased strength" for "low coherence across studies with four of six studies not identifying any association (although three studies did not focus on neurotoxicity) and the other two reporting different outcomes." It must first be noted that the appropriate consideration identified in ***Table 7-13*** is "consistency," not "coherence." In addition, ***Table 7-13*** states that "decreased strength" is applied only to a situation in which a body of evidence consists of unexplained "conflicting evidence." In this example, the only studies that lack evidence of neurotoxicity are the cross-sectional studies that did not focus on neurotoxicity. The fact that all studies designed to consider neurotoxicity found evidence of relevant outcomes is a strong indicator of consistency in the human evidence stream; the summary statement asserting "four of six studies not identifying any association" does not correspond to the description of the evidence and seems to disregard all studies lacking statistical significance. Further, any variation in neurotoxic outcomes reported (which is not at all apparent from the description of the studies) does not constitute a lack of consistency.

EPA's analysis also assigns "decreased strength" based on "inconsistent dose-response across exposure duration groups in the case-control study." The dose-response relationships in the two cohort studies seem to have been disregarded by EPA in drawing this judgment. Further, as noted above, the GRADE approach to evidence synthesis only uses the dose-response consideration to increase strength, and absence of a dose-response pattern does not decrease strength.

EPA's analysis also assigns "decreased strength" due to "no statistically-significant association with any neurological outcomes in two high-quality studies (despite a dose-responsive increase in incidence)," which overemphasizes statistical significance, particularly when it is explained that these studies are "high quality yet small." Further, EPA's overall conclusion is that a judgment of "indeterminate" likely applies because "only one of five studies demonstrated a statistically-significant association," which ultimately discounts the multiple indications of neurotoxic effects across the relevant studies.

Based on the description of the body of evidence, a more appropriate judgment would be "moderate" evidence of neurotoxicity. EPA's assessment of the evidence could also be informed by conducting a meta-analysis of the cohort and case-control studies. These significant errors in evidence synthesis raise

[173] Amrhein, V, Greenland, S, McShane, B. "Scientists rise up against statistical significance." Nature. 20 March 2019. https://www.nature.com/articles/d41586-019-00857-9

concerns about how EPA is applying the approach to evidence synthesis outlined in this document to its ongoing TSCA risk evaluations.

d. The "Evidence Profile Figure Template" (*Table 7-15*) introduces new considerations for within-stream evidence synthesis that are inconsistent with the earlier discussion of within-stream considerations.

Table 7-15 presents the format for integrating evidence across evidence streams. The intention appears to be that the table recaps the synthesis for each evidence stream separately and then uses those recaps as the basis for evidence integration across evidence streams. However, the table inserts considerations for within-stream evidence synthesis that were not presented in ***Table 7-13***.

Columns 2 and 3 of ***Table 7-15*** provide "Factors that Increase Strength" and "Factors that Decrease Strength" for each evidence stream. These columns include factors that were not presented in Table 7-13. In ***Table 7-15*** the factors that increase strength include both "consistency" and "coherence of observed effects;" the latter consideration is not mentioned in ***Table 7-13***, and the text on evidence integration indicates that "coherence" is a consideration for integrating across evidence streams. "Imprecision" is listed in ***Table 7-15*** as a consideration that decreases strength, but there is no corresponding entry for "precision" as a consideration that increases strength.

In some cases, different terminology is used, which could introduce confusion. ***Table 7-15*** identifies "effect size" as a consideration, replacing the term "magnitude" found in ***Table 7-13***. Similarly, ***Table 7-15*** uses "mechanistic evidence providing plausibility" replacing the term "biological plausibility and relevance to humans" (as noted above, this consideration does not belong in the within-stream synthesis step but rather is part of integrating evidence across streams). For the mechanistic evidence stream alone, ***Table 7-15*** introduces new phrasing that was not previously described for the within-stream synthesis: in addition to "Consistency and/or Replication" there are separate, new entries for "Results are consistent with related apical endpoints" and "Mechanistic data is supported by available AOP or signaling pathway information." Consistency with apical endpoints seems more like a consideration for integrating across evidence streams. Regardless, these new entries should not be presented as elements of within-stream synthesis for the first time in ***Table 7-15.***

e. EPA has imposed an inappropriate limitation on evidence conclusions that can be drawn from mechanistic evidence.

In ***Table 7-14***, EPA states that robust mechanistic evidence alone (without human or animal evidence) of a health outcome can support no stronger conclusion than "evidence suggests:"

> In the absence of informative conventional studies in humans or in animals (*i.e., indeterminate* evidence in both, mechanistic evidence, including information obtained outside the systematic review process (*e.g.,* read-across, QSAR, see Section 7.5.1.4), could also be used to conclude *"evidence suggests"* if the mechanistic evidence is sufficient to highlight potential human toxicity (*i.e., moderate* or *robust* evidence).

A separate entry in the table indicates that at least some evidence from animals or humans is necessary alongside robust mechanistic evidence to conclude "evidence indicates likely." This is an inappropriate limitation on the conclusions that can be drawn from mechanistic evidence. While it may be unusual for

mechanistic evidence alone to be "robust," if such an instance is found in a TSCA risk evaluation that finding should be sufficient for a judgment of "evidence indicates likely." For example, there may be a well-established adverse outcome pathway demonstrating that a chemical with certain mechanistic activity will very likely lead to a specified adverse outcome. A chemical with robust evidence of such mechanistic activity could support an "evidence indicates likely" conclusion even without human or animal evidence of the same outcome. This would be consistent with treatment of strong mechanistic evidence in the OHAT Handbook. Given the extensive current efforts to develop "new approach methods" (NAMs) suitable for risk assessment, it is surprising that EPA would impose this limitation on their use under TSCA. Limiting the conclusions that can be drawn from NAMs in TSCA risk evaluations will be a barrier to use of new mechanistic findings.

f. The concept presented in the evidence integration chapter of mechanistic data as a third stream of evidence, alongside human and animal evidence, is unclear and not carried out consistently in other sections of the document.

Incorporation of evidence integration is an important improvement in the *2021 Draft TSCA Method* compared with the *2018 TSCA Method*. The *2021 Draft TSCA Method* indicates that the approach to evidence integration for human health hazards is modelled after the approach of the IRIS program, with "minor deviations,"[174]and further notes that it incorporates a "more formalized evidence integration process for mechanistic data in this TSCA SR Protocol compared to the IRIS Handbook." Specifically, the *2021 Draft TSCA Method* identifies mechanistic evidence as a separate stream of evidence, independent of human and animal evidence. What this means in practice for how TSCA systematic reviews are conducted, however, is unclear. If mechanistic evidence is a separate evidence stream, then its treatment throughout the systematic review process should be parallel to the treatment of the human and animal evidence streams. However, the TSCA PECO statements do not include explicit components necessary to identify mechanistic evidence as a separate evidence stream, except with statements that such evidence is to be tagged as "potentially relevant supplemental material."

Other statements indicate that the studies tagged as supplemental-mechanistic may or may not go through the steps of study evaluation and data extraction: "these references are included as supplemental references during the screening phase, a subset is evaluated as needed for any given chemical"[175] and "may be reviewed, evaluated for data quality, and incorporated into risk evaluations as needed for each chemical assessment."[176] This is quite different from the human and animal evidence streams, in which all relevant studies are evaluated for data quality and considered for evidence integration.

There does not appear to be any further explanation of when and how decisions are made to examine, evaluate, and extract data from mechanistic studies, and how these steps are affected by considering mechanistic data as a separate, independent evidence stream. The only relevant statement does not provide any clarity:

[174] US EPA. (2021). Draft systematic review protocol supporting TSCA risk evaluations for chemical substances version 1.0: A generic TSCA systematic review protocol with chemical-specific methodologies, pp. 111. https://www.epa.gov/system/files/documents/2021-12/draft-systematic-review-protocol-supporting-tsca-risk-evaluations-for-chemical-substances_0.pdf

[175] US EPA. (2021). Draft systematic review protocol supporting TSCA risk evaluations for chemical substances version 1.0: A generic TSCA systematic review protocol with chemical-specific methodologies, pp. 112. https://www.epa.gov/system/files/documents/2021-12/draft-systematic-review-protocol-supporting-tsca-risk-evaluations-for-chemical-substances_0.pdf

[176] US EPA. (2021). Draft systematic review protocol supporting TSCA risk evaluations for chemical substances version 1.0: A generic TSCA systematic review protocol with chemical-specific methodologies, pp. 345. https://www.epa.gov/system/files/documents/2021-12/draft-systematic-review-protocol-supporting-tsca-risk-evaluations-for-chemical-substances_0.pdf

> EPA plans to prioritize the evaluation of mechanistic evidence instead of evaluating all of the identified evidence upfront. This approach has the advantage of conducting a focused review of those mechanistic studies that are most relevant to the hazards under evaluation. The prioritization approach is generally initiated during the data screening step. For example, the hazard PECOs consider the mechanistic evidence as supplemental information during full-text screening. The assessor can eventually mine the supplemental information when specific questions or hypotheses arise related to the chemical's MOA/AOP.[177]

It is unclear why the data screening step would be where decisions are made regarding review of mechanistic evidence (this would not be where "specific questions or hypotheses arise related to the chemical's MOA/AOP"), and no further explanation is provided. The *2021 Draft TSCA Method* should be revised to incorporate an explicit step that indicates how assessment teams are to determine what mechanistic evidence to review and how that evidence is selected from the larger body of studies tagged as supplemental-mechanistic.

The *2021 Draft TSCA Method* also imposes an overly narrow conception of how mechanistic evidence and NAMs may contribute to a risk evaluation, by using mechanistic research only for hazard identification and dose-response assessment of phenotypic or apical outcomes: "Mechanistic evidence is also considered in targeted analyses conducted prior to, during, and after the integration of phenotypic human and animal evidence."[178] This excludes use of novel mechanistic findings from emerging NAMs that can inform how a chemical may interact within a specific cell system long before epidemiological or whole-animal studies can show the same. EPA should instead incorporate consideration of mechanistic events that do not yet have a known relationship to a phenotypic outcome. This broader concept should also include advanced predictive modeling such as machine learning or artificial intelligence to understand effects of chemical exposure, based on modern "big data" statistical methods applied to data being collected from high throughput studies and other NAMs. All mechanistic data used in a risk evaluation should be evaluated for risk of bias.

The *2021 Draft TSCA Method* indicates that, as a separate evidence stream, mechanistic evidence will be assigned a descriptor of robust, moderate, slight, or indeterminate for each applicable health outcome. While there may be some mechanistic evidence that can stand on its own, the document describes the primary use of mechanistic evidence as informing interpretation of animal and human evidence:

> …mechanistic data evaluations consider the support for and involvement of specific events or sets of events within the context of a broader research question (*e.g.,* support for a hypothesized mechanism, consistency with known biological processes), rather than evaluations of individual apical endpoints considered in relative isolation…while mechanistic data is

[177] US EPA. (2021). Draft systematic review protocol supporting TSCA risk evaluations for chemical substances version 1.0: A generic TSCA systematic review protocol with chemical-specific methodologies, pp. 589. https://www.epa.gov/system/files/documents/2021-12/draft-systematic-review-protocol-supporting-tsca-risk-evaluations-for-chemical-substances_0.pdf

[178] US EPA. (2021). Draft systematic review protocol supporting TSCA risk evaluations for chemical substances version 1.0: A generic TSCA systematic review protocol with chemical-specific methodologies, pp. 112. https://www.epa.gov/system/files/documents/2021-12/draft-systematic-review-protocol-supporting-tsca-risk-evaluations-for-chemical-substances_0.pdf

evaluated on its own, the mechanistic evidence is most useful in demonstrating the relevance and reliability of apical outcome findings in animal and human studies.[179]

It is unclear how a descriptor such as robust or slight is to be selected for mechanistic evidence that is considered for the purpose of evaluating the relevance and reliability of findings from other evidence streams. Does "robust" mechanistic evidence mean strong evidence of a particular health outcome, or does it mean strong evidence for a particular mechanistic hypothesis that may or may not support a hazard conclusion?

The *2021 Draft TSCA Method* should provide more information on how EPA will develop, and how readers should interpret, descriptors such as "robust" for mechanistic evidence that is not considered in isolation, which contrasts to the assignment of descriptors for animal and human evidence streams that are each synthesized independently.

Approaches to use of mechanistic information in systematic review are still in development. The NASEM review of the draft *IRIS Handbook* provided numerous suggestions for clarifying the role of mechanistic information, and the TSCA program should carefully review these suggestions.

g. The relationship of the section on characterizing key sources of uncertainty to the rest of the evidence integration chapter is unclear.

The evidence integration chapter concludes with section 7.5.2.4, "Characterization of Strengths, Limitations, Assumptions, and Key Sources of Uncertainty in the Human Health Hazard Assessments." Any risk assessment document should of course characterize strengths, limitations, assumptions, and uncertainties to provide context to its conclusions and results. In this case, however, this section seems to have been drafted without reference to the rest of the evidence integration chapter. It highlights issues such as consistency of evidence and relevance of animal evidence to humans that are integral to the evidence integration process that has already been outlined, and the important points have been covered in section 7.5.2.3, "Overall Weight of the Scientific Evidence Judgments." In addition, some of the text of this chapter is most relevant to characterizing the results of dose-response assessment or risk characterization, which are not subjects of this document.

Some of the phrasing in this section also seems unbalanced and pointed to a particular conclusion. For example, there is no reason for the inclusion of the word "weak" in this question: "Is there weak epidemiological or mechanistic support for the human relevance or adversity of an apical endpoint observed in animal studies?" Further, as noted above, mechanistic evidence is not necessary to demonstrate a hazard; and additionally, for many chemical/hazard combinations there will be no epidemiological evidence and such evidence is not required to conclude that a hazard exists. Absence of epidemiological or mechanistic evidence should not be characterized as an uncertainty.

Another uncertainty identified in this section is "Are the exposure conditions associated with adverse health outcomes unlikely to be relevant to human exposure scenarios, based on the conditions of use for that chemical?" It may be unwarranted to draw conclusions of relevance in a hazard assessment

[179] US EPA. (2021). Draft systematic review protocol supporting TSCA risk evaluations for chemical substances version 1.0: A generic TSCA systematic review protocol with chemical-specific methodologies, pp. 112. https://www.epa.gov/system/files/documents/2021-12/draft-systematic-review-protocol-supporting-tsca-risk-evaluations-for-chemical-substances_0.pdf

based on exposure scenarios until the exposure assessment is finalized; such questions are better considered in risk characterization, which is not addressed in this document.

h. The *2021 Draft TSCA Method* provides insufficient guidance on toxicokinetic modeling.

The *2021 Draft TSCA Method* does not include any instructions on incorporating PBTK methods. Instead on page 590, EPA makes it clear they are triaging understanding PBPK modeling until they have more information: "Although EPA is not including an evaluation strategy for PBPK models in this document, when necessary, it plans to document the model evaluation process based on the Quality Assurance Project Plan for PBPK models (U.S. EPA. 2018e)."[180]

Understanding the limitations of the current literature is important, but we strongly encourage EPA to consider developing criteria for use of kinetic modeling in TSCA risk evaluations Such criteria need to be developed in advance of consideration of PBPK modeling in a particular risk evaluation, so that development of the criteria is not biased by chemical-specific or model-specific considerations that could affect the outcome of the risk evaluation. Currently, there is limited PBPK data available, and EPA should also consider creating guidance on how to address data gaps on chemical toxicokinetics. Improved understanding of how external chemical concentrations relate to internal chemical concentrations could contribute additional information supporting the application of NAMs in risk evaluation.

The NASEM review of the draft *IRIS Handbook*[181] recommends more clear and concise definitions of TK modeling as well as creating PECO statements for each outcome of interest:

> **Recommendation 3.9:** The handbook should describe how the IRIS assessment plan and IRIS assessment protocol can identify the potential roles of mechanistic and TK data, including if they are to be units of analysis for systematic review, synthesis, and strength of evidence judgments. At a minimum, all endpoints that may be used for toxicity values, including so-called "precursor" endpoints that might be viewed as "mechanistic," should require separate PECO statements; however, application of systematic review methods to other mechanistic endpoints, such as mutagenicity, may depend on the needs of the assessment. The key mechanistic and TK questions should be identified to the extent possible in the IRIS assessment plan and IRIS assessment protocol documents.
>
> **Recommendation 3.10:** When available, KCs should be used to search for and organize mechanistic data, identify data gaps, and evaluate biological plausibility. Those uses should be reflected in the IRIS assessment plan and IRIS assessment protocol.[182]

We concur with NASEM recommendations that key issues regarding mechanisms and toxicokinetics require thorough documentation in assessment-specific protocols and peer review.

---

[180] US EPA. (2021). Draft systematic review protocol supporting TSCA risk evaluations for chemical substances version 1.0: A generic TSCA systematic review protocol with chemical-specific methodologies, pp. 590. https://www.epa.gov/system/files/documents/2021-12/draft-systematic-review-protocol-supporting-tsca-risk-evaluations-for-chemical-substances_0.pdf

[181] National Academies of Sciences, Engineering, and Medicine 2021. Review of U.S.EPA's ORD Staff Handbook for Developing IRIS Assessments: 2020 Version.Washington, DC: The National Academies Press. https://doi.org/10.17226/26289.

[182] US EPA. (2021). Draft systematic review protocol supporting TSCA risk evaluations for chemical substances version 1.0: A generic TSCA systematic review protocol with chemical-specific methodologies, pp. 126. https://www.epa.gov/system/files/documents/2021-12/draft-systematic-review-protocol-supporting-tsca-risk-evaluations-for-chemical-substances_0.pdf

We recommend:

- **Identifying evidence synthesis and evidence integration as two distinct steps in the systematic review process, with synthesis conducted for each evidence stream separately, and integration combining the findings across evidence streams.**
- **Clarifying considerations that are applied in conducting within-stream evidence synthesis and applying a consistent set of considerations throughout the process.**
- **Clarifying the role of mechanistic evidence in evidence integration.**

6. Dose-response assessment should be conducted for all identified and suspected health hazards.

The *2021 Draft TSCA Method* provides this statement regarding use of evidence integration judgments in selection of effects for dose-response assessment:

> When "evidence suggests but is not sufficient to conclude…" a non-cancer effect or there is "suggestive evidence of carcinogenicity" to humans, in most instances EPA generally would not conduct a dose-response assessment or derive a cancer value except when the evidence includes a high-quality study and quantitative analyses may be useful for some purposes. For example, quantitative results for endpoints in this category may help for providing a sense of the magnitude and uncertainty of potential risks, ranking potential hazards, or setting research priorities (U.S. EPA, 2005b). It is critical to transparently convey the large uncertainty in any such estimates. [183]

EPA should conduct dose-response assessment for all outcomes with an "evidence suggests" finding. This hazard descriptor captures only one dimension of risk – the likelihood that the effect in question results from exposure to the chemical being assessed. The other dimensions are equally important for decision-making: potency of the chemical (dose-response); the severity of the effect; the magnitude of exposures; and the number of people exposed. An effect judged to have suggestive evidence may still potentially pose significant risks when the other dimensions are considered; and could even prove to be more important than an effect with stronger evidence (e.g., a severe outcome (such as cardiovascular disease) with suggestive evidence may be of greater concern than a mild outcome (such as one day of mild respiratory symptoms) with a conclusion of "evidence demonstrates likely"). Dose-response assessment is necessary to ensure that all dimensions of the potential risk are considered in making the TSCA unreasonable risk determination.

We recommend:

- **Conducting dose-response assessment for all outcomes with conclusions of "evidence demonstrates," "evidence indicates likely," and "evidence suggests but is not sufficient to conclude."**

[183] US EPA. (2021). Draft systematic review protocol supporting TSCA risk evaluations for chemical substances version 1.0: A generic TSCA systematic review protocol with chemical-specific methodologies, pp. 127. https://www.epa.gov/system/files/documents/2021-12/draft-systematic-review-protocol-supporting-tsca-risk-evaluations-for-chemical-substances_0.pdf