June 18, 2020

Comments from Academics, Scientists and Clinicians on: The Application of Systematic Review in TSCA Risk Evaluations.

We are writing on behalf of the undersigned academics, scientists, and clinicians to provide expert comments on the EPA's *Application of Systematic Review in TSCA Risk Evaluations*.

We appreciate the National Academies of Sciences, Engineering, and Medicine (NASEM) is reviewing EPA's guidance document *Application of Systematic Review in TSCA Risk Evaluations* and associated materials (hereafter referred to as the 'TSCA method'). The NASEM provides independent and expert advice into important scientific and evidence use questions and will determine if the TSCA method is 'comprehensive, workable, objective, and transparent'. The TSCA method was developed by EPA as part of the implementation of the Toxic Substances Control Act (TSCA), as amended by the Frank R. Lautenberg Chemical Safety for the 21st Century Act (Lautenberg TSCA). The Act requires that EPA make decisions about chemical risks based on the "best available science" and the "weight of the scientific evidence"[1]. EPA defines "weight of the scientific evidence"; in regulation 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.'' [2]

Our comments highlight that the TSCA method does not follow best practices in systematic review and could result in underestimating risks to environmental chemicals and pollutants. There are currently methods in place that have been reviewed and tested by the NASEM for environmental chemicals that could be implemented by EPA right now to improve the basis of their decision making – our comments outline the six key areas of improvement that are needed to more consistently and fairly evaluate the science and support improved decision making to protect public health.

Systematic review methods originated more than 40 years ago in psychology. In response to empirical evidence demonstrating the need to apply scientific principles both to primary research and to research synthesis methods that inform decision-making in healthcare, the methodology was soon adapted to evaluate the effectiveness of clinical interventions in medicine and related disciplines.[3,4,5] Almost a decade ago, these empirically-proven methods for research synthesis were adapted to environmental health beginning with the development and implementation of the University of California, San Francisco (UCSF) *Navigation Guide*

[1] 15 USC §2625 (h)-(i)

[2] 40 CFR 702.33

[3] Rennie D, Chalmers I. Assessing authority. JAMA. 2009;301(17):1819-21. Epub 2009/05/07. doi: 301/17/1819 [pii]10.1001/jama.2009.559. PubMed PMID: 19417202.

[4] Fox DM. The Convergence of Science and Governance: Research, Health Policy, and American States. Berkeley, CA: University of California Press; 2010

[5] Antman EM, Lau J, Kupelnick B, Mosteller F, Chalmers TC. A comparison of results of meta-analyses of randomized control trials and recommendations of clinical experts. Treatments for myocardial infarction. JAMA. 1992;268(2):240-8. Epub 1992/07/08. PubMed PMID: 1535110.

Systematic Review Method.[6,7,8] This was followed by the publication of the National Toxicology Program's Office of Health Assessment and Translation *OHAT Approach for Systematic Review and Evidence Integration for Health Effects Evaluations,* (hereafter referred to as the OHAT method).[9] Both the Navigation Guide and the OHAT method have been used or recommended by the NASEM[10,11,12] and demonstrated in case studies in the peer-reviewed literature.[13,14,15,16,17, 18,19,20] The World Health Organization and International Labor Organization (WHO/ILO) are using the Navigation Guide to conduct systematic reviews of occupational exposures and disease as part of assessing the global burden of work-related injury and disease

[6] 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. doi:10.1289/ehp.1307175.

[7]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.

[8] 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.

[9] National Toxicology Program Office of Health Assessment and Translation. Handbook for Conducting a Literature-Based Health Assessment Using OHAT Approach for Systematic Review and Evidence Integration. National Institute of Environmental Health Sciences; 2015.

[10] 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

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

[12] 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.

[13]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.

[14]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.

[15] 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

[16]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.

[17] 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.

[18] 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.

[19] 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.

[20] 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.

due to exposure to occupational risk factors.[21] Further, systematic reviews have been adopted by the EPA's Integrated Risk Information System (IRIS) program.[22]

The TSCA method establishes EPA's "rules" for assembling and interpreting the scientific evidence on chemicals in commerce. These "rules" will determine, whether explicitly, implicitly, and/or by default, *what* evidence EPA will consider, and *how* it will evaluate that evidence when it is making decisions about potentially hazardous chemicals in commerce. Exposure to industrial, commercial, and consumer product chemicals is ubiquitous from the time of conception until death. As such, EPA's rules for gathering and interpreting the science that evaluates the relationship between these exposures and adverse health effects are of profound importance to the public and will have even greater impact on the potentially exposed or susceptible sub-populations. Congress explicitly mandated EPA to protect: pregnant women, children, individuals with underlying health conditions, workers, and those with greater exposure and/or greater vulnerability to chemical toxicity and exposure.

With so much at stake, we are deeply concerned by EPA's incomplete TSCA method, which is inconsistent with current, established, best available empirical methods for systematic review. We have detailed below where the method is not consistent with best practices and examples of how the application of EPA's TSCA method has resulted in the exclusion of quality research from EPA's decision-making. Thus, continued use of this method would mean that risks from industrial chemicals and pollutants could be undervalued and underestimated – leaving the public at risk from harmful chemical exposures.

We have commented on the scientific flaws in the TSCA method extensively in previous submissions to EPA on the draft risk evaluations that have already been completed and made

[21] Mandrioli, D., Schlünssen, V., Ádám, B., Cohen, R. A., Colosio, C., Chen, W.et al. (2018). WHO/ILO work-related burden of disease and injury: Protocol for systematic reviews of occupational exposure to dusts and/or fibres and of the effect of occupational exposure to dusts and/or fibres on pneumoconiosis. Environment International, Vol. 119, Pg. 174–185. https://doi.org/10.1016/j.envint.2018.06.005

[22] 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.

publicly available [23,24,25,26,27,28,29,30]and as summarized in a peer-reviewed commentary published in the *American Journal of Public Health*.[31] Further, several of these fundamental systematic review deficiencies in the 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),[32,33] 1,4-dioxane and Cyclic Aliphatic Bromide Cluster (HBCD),[34,35] Bromopropane (1-BP)[36 37] and N-Methylpyrrolidone[38] that echo the comments and recommendations we make here. The SACC has made several comments and critical recommendations necessary to improve the TSCA method which EPA has not addressed in the draft risk evaluations to date; therefore, the scientific flaws in the TSCA method persist.

Based on the most current empirically demonstrated principles of systematic review methods, we provide the NASEM with concrete recommendations and approaches to correct EPA's methodology and inform timely science based, health-protective decision-making around

---

23 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

[24] 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

[25] 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

[26] 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

[27]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

[28] 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

[29] 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

[30] 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

[31] 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

[32] 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

[33] 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

[34] 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

[35] 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

[36] 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

[37] 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

[38] 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

potentially harmful chemicals and achieve the Agency's mission of protecting human health and the environment.

Our comments address the following six main points:

1. **EPA's TSCA method is incomplete and does not follow established methods for systematic review that are based on the best available science.**

   We recommend: EPA should implement a systematic review method that is compatible with empirically based existing methods and aligns with the Institute of Medicine's[39] definition of a systematic review, including but not limited to, using explicit and pre-specified scientific methods for every step of the review. Other methods can be used that have been demonstrated extensively for use in environmental health, and which have been endorsed and utilized by the National Academies of Sciences, Engineering, and Medicine i.e., the National Toxicology Program's *OHAT method* and the *Navigation Guide* developed by the University of California, San Francisco.

2. **EPA's TSCA method fails to provide sufficient information to describe protocol development and EPA has not published a protocol for the first 10 chemicals that have undergone draft risk evaluations.**

   We recommend: EPA should immediately implement the use of pre-established protocols to enhance transparency in the decision-making process and consistency in their draft risk evaluations. EPA should use the published protocols developed for applying the *OHAT method*, and the *Navigation Guide* that can serve as a template to further expedite EPA's TSCA systematic review process.

3. **The literature review step of EPA's TSCA method incorporates select best practices, but also falls short of, or is unclear about, many other best practices for conducting a systematic and transparent literature review.**

   We recommend: EPA should align its framework for conducting a literature review to ensure that it is congruent with all of the Institute of Medicine's best practices and explicitly predefine the eligibility criteria for the included studies before conducting any part of the systematic review process. At this juncture, it is unclear how EPA intends to handle many components of its literature searches. The transparency of the framework would be improved by specifying how EPA is addressing each best practice.

4. **EPA's TSCA method utilizes a quantitative scoring method that is incompatible with the best available science in fundamental ways and can exclude relevant studies from consideration in the risk evaluation:**
   a. **Quantitative scores for assessing the quality of an individual study are arbitrary and not science or evidence-based; the National Academies of Sciences, Engineering, and Medicine recommend against such scoring methods;**
   b. **EPA's scoring method wrongly conflates how well a study is reported with how well**

[39] The Institute of Medicine is now the National Academy of Medicine.

the underlying research was conducted; and

c. EPA's scoring method excludes studies based on one single reporting or methodological limitation.

We recommend: EPA should not use a quantitative scoring method to assess quality in individual studies; it should not conflate study reporting with study quality; and it should not exclude otherwise quality research based on a single reporting or methodological limitation. Rather EPA should employ a scientifically valid method to assess risk of bias of individual studies. We recommend that the approaches of the *OHAT method* or the *Navigation Guide* be used for this step.

5. **EPA's TSCA method does not have a pre-established protocol or methods for evidence integration.**

   We recommend: EPA should immediately implement an evidence integration method that is consistent with best practice in systematic review, such as the *OHAT method* and the *Navigation Guide Systematic* and transparently present how the conclusions are reached in assessing human health hazards for each end point it assesses.

6. **EPA's TSCA method does not consider financial conflicts of interest as a potential source of bias in research.**

   We recommend: EPA should assess study and author funding source as a risk of bias domain for individual studies.

We appreciate the opportunity to provide public input. Please do not hesitate to contact us with any questions regarding these comments. We declare collectively that we have no direct or indirect financial or fiduciary interest in any chemical under consideration in these risk evaluations. The co-signers' institutional affiliations are included for identification purposes only and do not necessarily imply any institutional endorsement or support, unless indicated otherwise.

Sincerely,

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

Tracey J. 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

Lisa Bero PhD
Professor, School of Public Health, School of Medicine (General Internal Medicine)
Chief Scientist, Center for Bioethics and Humanities
The University of Colorado, CU Anschutz Medical Center
Former Co-Chair Cochrane Collaboration
Senior Editor Cochrane Public Health and Health Systems
Senior Editor for Research Integrity Cochrane.
Fellow, Collegium Ramazzini

Courtney Carignan, PhD
Assistant Professor
Michigan State University

Holly Davies, PhD
Senior Toxicologist
Office of Environmental Public Health Sciences
Environmental Public Health Division
Washington State Department of Health

Robert M. Gould, MD
Associate Adjunct Professor
Program on Reproductive Health and the Environment
Department of Obstetrics, Gynecology and Reproductive Sciences
UCSF School of Medicine
and
Past-President
Physicians for Social Responsibility

Alycia Halladay, PhD
Chief Science Officer
Autism Science Foundation
New York, NY

Kim Harley, MPH PhD
Associate Director for Health Effects
Center for Environmental Research and Children's Health (CERCH)
University of California, Berkeley

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

Ronnie Levin
Instructor
Harvard TH Chan School of Public Health

Rainer Lohmann, PhD
Professor of Oceanography

Graduate School of Oceanography
University of Rhode Island

Daniele Mandrioli, MD, PhD
Director, Cesare Maltoni Cancer Research Center
Ramazzini Institute

Rachel Morello-Frosch, PhD, MPH
Professor
University of California, Berkeley School of Public Health
Department of Environmental Science, Policy and Management
University of California, Berkeley

Michele Okoh, JD
Senior Lecturing Fellow of Law
Environmental Law and Policy Clinic
Duke University

Vasantha Padmanabhan, M,S., Ph.D.
Professor Emerita
University of Michigan, Ann Arbor

Christopher J. Portier, PhD
Former Director (retired)
National Center for Environmental Health
Agency for Toxic Substances and Disease Registry

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

Natalie Sampson, PhD, MPH
Associate Professor
University of Michigan,Dearborn

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

Joel Schwartz
Professor of Epidemiology and Biostatistics
Harvard TH Chan School of Public Health

Rachel M. Shaffer, MPH
PhD Candidate

University of Washington, Seattle School of Public Health

Patrice Sutton, MPH
Research Scientist
UCSF Program on Reproductive Health and the Environment
San Francisco, CA

Laura N. Vandenberg, PhD
Associate Professor,
Graduate Program Director
University of Massachusetts – Amherst
School of Public Health & Health Sciences
Department of Environmental Health Sciences

Paul Whaley
Researcher and Consultant
Lancaster Environment Centre, Lancaster University, UK
Evidence Based Toxicology Collaboration Research Fellow
Systematic Reviews Editor, Environment International

Marya Zlatnik, MD, MMS
Professor, Obstetrics, Maternal-Fetal Medicine
Department of Obstetrics, Gynecology and Reproductive Sciences
UCSF School of Medicine

DETAILED POINTS

1. **EPA's TSCA method is incomplete and does not follow established methods for systematic review that are based on the best available science.**

The best available scientific method for a systematic review specifies that all components of a review be established in a publicly available protocol written *prior* to conducting the review to minimize bias and to ensure transparency in decision-making. For example, the Institute of Medicine (IOM), which has 21 standards covering the entire systematic review process that, if adhered to, result in a scientifically valid, transparent, and reproducible systematic review, defines a systematic review as a "scientific investigation that focuses on a specific question and uses explicit, pre-specified scientific methods to identify, select, assess, and summarize the findings of similar but separate studies" (emphasis added).[40] EPA's TSCA method lacks essential systematic review elements, including but not limited to: (1) a protocol for executing a systematic review developed *prior* to conducting the systematic review; (2) the use of a validated tool to assess the risk of bias (study quality) of the included studies in the review (3) an explicit method for evaluating the overall body of each evidence stream, i.e., animal, human, in vitro etc.; and (4) an explicit method for integrating two or more streams of evidence, including defined criteria for the type and level of evidence needed for a decision by

[40] Institute of Medicine. (2011). Finding What Works in Health Care: Standards for Systematic Reviews. Page 1. Washington, DC: The National Academies Press

EPA.

Notably, EPA's TSCA method presents a diagram of a complete systematic review framework in Figure 3-1 (Page 15) and states in footnote 4 on that page that the:

> "Diagram depicts systematic review process to guide the first ten TSCA risk evaluations. It is anticipated that the same basic process will be used to guide future risk evaluations with some potential refinements reflecting efficiencies and other adjustments adopted as EPA/OPPT gains experience in implementing systematic review methods and/or approaches to support risk evaluations within statutory deadlines (e.g., aspects of protocol development would be better defined prior to starting scoping/problem formulation)."[41]

However, EPA's TSCA method then proceeds to describe a flawed systematic review method limited to only the data collection and, to a limited extent, the data evaluation components of a systematic review. Specifically, Figure S-1 below, excerpted from the NASEM 2014 review of the EPA IRIS program's systematic review method,[42] presents all of the components of a systematic review. The red box indicates the parts of a systematic review method that EPA has implemented in the TSCA method.

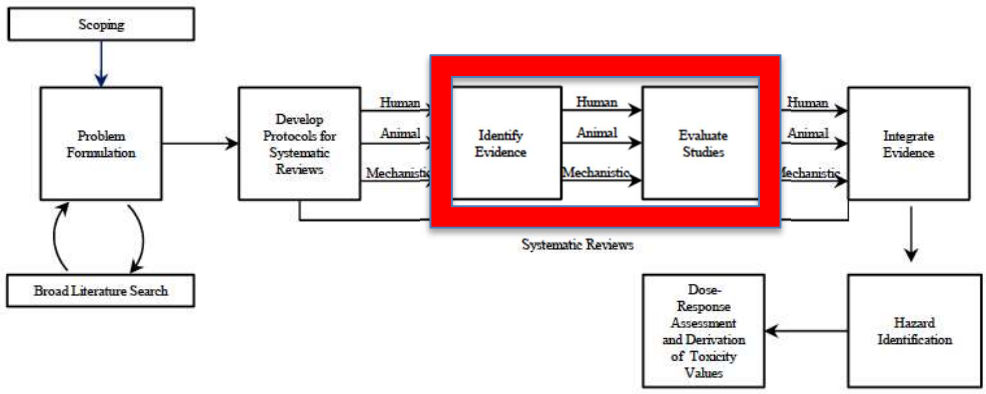


FIGURE S-1 Systematic review in the context of the IRIS process. The committee views public input and peer review as integral parts of the IRIS process, although they are not specifically noted in the figure.

Indicates the steps that EPA has included and implemented in its TSCA method to date. [43]

The EPA TSCA's inadequate approach contradicts best available scientific methods for systematic review, and is incompatible with the regulatory definition of "weight of evidence" in

[41] EPA (2018) Application of Systematic Review in TSCA Risk Evaluations. Page. 15.

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

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

the risk evaluation rule,[44] which specifies a complete method spelled out in a protocol developed *before* conducting the review.

There are additional steps within the method that are not scientifically supported (e.g. EPA's highly quantitative scoring method, which is the main topic of its systematic review framework; see comment #4 below for a detailed critique of the scientific flaws to this approach), and are not completely described a priori (e.g. the method does not have a pre-established protocol or method for evidence integration; see comment #5 below for a detailed critique of the scientific flaws to this approach).

The incomplete nature of EPA's TSCA method is incompatible in a number of additional fundamental ways with best available scientific methods for systematic review (described further in detail below using examples from the draft risk evaluations EPA has completed to date) and with science based methods of systematic review developed, endorsed, and/or advanced by: the NASEM; [45,46,47,48] the IOM;[49] the National Toxicology Program;[50] the Grading of Recommendations Assessment, Development and Evaluation (GRADE) method;[51,52] the international scientific collaboration that developed a framework for the "systematic review and integrated assessment" (SYRINA) of endocrine disrupting chemicals;[53] the SYRCLE systematic review method for animal studies;[54] the Campbell Collaboration's methods;[55] and

[44] EPA's risk evaluation rule (40 CFR 702.33) states: ''Weight of the scientific evidence means 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.''

[45] 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

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

[47] 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.

[48] National Academies of Sciences, Engineering, and Medicine. 2019. Review of DOD's Approach to Deriving an Occupational Exposure Level for Trichloroethylene. Page. 3-4. Washington, DC: The National Academies Press. https://doi.org/10.17226/25610.

[49] Institute of Medicine. Finding What Works in Health Care. Standards for Systematic Review. Washington, D.C.: The National Academies Press.; 2011.

[50] 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; 2015.

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

[52] Guyatt GH, Oxman AD, Vist G, Kunz R, Brozek J, Alonso-Coello P, Montori V, Akl EA, Djulbegovic B, Falck-Ytter Y, Norris SL, Williams JW, Jr., Atkins D, Meerpohl J, Schünemann HJ. GRADE guidelines: 4. Rating the quality of evidence--study limitations (risk of bias). Journal of Clinical Epidemiology. 2011;64(4):407-15. doi: 10.1016/j.jclinepi.2010.07.017.

[53] Vandenberg LN, Ågerstrand M, Beronius A, ., Beausoleild C, Bergman A, Bero LA, Bornehag C, Boyer CS, Cooper GS, Cotgreave I, Gee D, Grandjean P, Guyton KZ, Hass U, Heindel JJ, Jobling S, Kidd KA, Kortenkamp A, Macleod MR, Martin OV, Norinder U, Scheringer M, Thayer KA, Toppari J, Whaley P, Woodrufft TJ, Ruden C. A proposed framework for the systematic review and integrated assessment (SYRINA) of endocrine disrupting chemicals. Environment Health. 2016;In press.

[54] Woodruff TJ, Sutton P. The Navigation Guide Work Group. An Evidence-Based Medicine Methodology To Bridge The Gap Between Clinical And Environmental Health Sciences. Health Affairs. 2011;30(5):931-7. doi: 10.1377/hlthaff.2010.1219; PMCID: 21555477.

[55] Campbell Collaboration. Better evidence for a better world. 2018 [cited 2018 July 29]The Campbell Collaboration promotes positive social and economic change through the production and use of systematic reviews and other evidence synthesis for evidence-based policy and practice.]. Available from: https://campbellcollaboration.org/research-resources/writing-a-campbell-systematic-review.html

the Navigation Guide Systematic Review method developed by a collaboration of scientists led by the University of California San Francisco.[56] Most of these organizations also pre-publish their protocols either online (i.e., the National Toxicology Program)[57] or in PROSPERO (i.e., UCSF). [58]

> **We recommend**: EPA should implement a systematic review method that is compatible with empirically based existing methods and aligns with the Institute of Medicine's definition of a systematic review, including but not limited to, using explicit and pre-specified scientific methods for every step of the review. Other methods can be used that have been demonstrated extensively for use in environmental health, and which have been endorsed and utilized by the National Academies of Sciences, Engineering, and Medicine i.e., the National Toxicology Program's *OHAT method* and the *Navigation Guide Systematic Review Method* developed by the University of California, San Francisco:
>
> - OHAT method: National Toxicology Program Office of Health Assessment and Translation. *Handbook for Conducting a Literature-Based Health Assessment Using OHAT Approach for Systematic Review and Evidence Integration*. National Institute of Environmental Health Sciences; 2015.
>
> - Navigation Guide: 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. doi:10.1289/ehp.1307175

2. **EPA's TSCA method fails to provide sufficient information to describe protocol development and EPA has not published a protocol for the first 10 chemicals that have undergone draft risk evaluations.**

In *Figure 3-1 TSCA Systematic Review Process* (Page 15) EPA presents 'Protocol Development' as the first step.[59] EPA then states that:

> "*Protocol Development* is intended to pre-specify the criteria, approaches and/or methods for data collection, data evaluation and data integration. It is important to plan the systematic review approaches and methods in advance to reduce the risk of introducing bias into the risk evaluation process." And goes on to say that "**EPA had limited ability to develop a protocol document detailing the systematic review approaches and/or methods prior to the initiation of the risk evaluation process for the first ten chemical substances**. **For these reasons, the protocol development is staged in phases while conducting the assessment work.** Additional details on the approach for the

[56] Woodruff TJ, Sutton P. The Navigation Guide Work Group. An Evidence-Based Medicine Methodology To Bridge The Gap Between Clinical And Environmental Health Sciences. Health Affairs. 2011;30(5):931-7. doi: 10.1377/hlthaff.2010.1219; PMCID: 21555477.

[57] National Toxicology Program. Completed Evaluations. Available: https://ntp.niehs.nih.gov/whatwestudy/assessments/noncancer/completed/index.html

[58] PROSPERO International prospective register of systematic reviews https://www.crd.york.ac.uk/prospero/

[59] EPA (2018) Application of Systematic Review in TSCA Risk Evaluations. Page. 15.

> evidence synthesis and integration will be included with the publication of the draft TSCA risk evaluations." [60](emphasis added)

Firstly, this is inconsistent with scientifically-validated approaches to conducting systematic reviews. In its review of the EPA's Integrated Risk Information System (IRIS) program's proposed systematic review methods, the NASEM specified that:

> "Completing the literature search as part of protocol development is inconsistent with current best practices for systematic review, and the IRIS program is encouraged to complete the public-comment process and finalize the protocol before initiating the systematic review." [61]

Secondly, EPA says they are going to have protocols, however, we cannot evaluate them because there is insufficient information provided describing the protocol development in the TSCA method and EPA has not published a protocol for the first 10 chemicals that have undergone draft risk evaluations. 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.[62,63] The goal of the protocol is to ensure that judgements regarding evidence evaluation are made prior to reviewing the evidence so to lower bias (so the evidence does not bias the evaluation of it). However, EPA is assembling and interpreting the evidence at the same time it is applying the rules, which leaves the risk evaluations open to bias.

Lack of time is not a credible rationale for EPA's failure to conduct an evidence-based systematic review for the first 10 TSCA chemicals using pre-established and pre-published protocols. There are multiple well-developed evidence-based, peer-reviewed and validated methods for conducting systematic reviews in environmental health that EPA could readily apply, including the OHAT method[64] and the Navigation Guide Systematic Review Method,

---

[60] EPA (2018) Application of Systematic Review in TSCA Risk Evaluations. Page. 15.

[61] National Academies of Sciences, Engineering, and Medicine. 2018. *Progress Toward Transforming the Integrated Risk Information System (IRIS) Program: A 2018 Evaluation*. Page. 8.Washington, DC: The National Academies Press. https://doi.org/10.17226/25086.

[62] National Research Council. *Review of EPA's Integrated Risk Information System (IRIS) Process*. Washington, D.C.; 2014.

[63] Institute of Medicine. *Finding What Works in Health Care*. Washington, D.C.: National Academies Press; 2011. doi:10.17226/13059.

[64] National Toxicology Program Office of Health Assessment and Translation. (2015). Handbook for Conducting a Literature-Based Health Assessment Using OHAT Approach for Systematic Review and Evidence Integration. National Institute of Environmental Health Sciences; 2015

which has been demonstrated in six case studies. [65,66,67,68,69, 70,71,72] The NASEM has cited both of these systematic review methods as exemplary of the type of methods EPA should use in hazard and risk assessment.[73,74, 75,76] Further, the NASEM utilized both methods in its 2017 assessment of the potential health impacts of endocrine active environmental chemicals.[77] Specifically, in its 2017 review the NASEM found:

> "The two approaches [OHAT and Navigation Guide] are very similar… and they are based on the same established methodology for the conduct of systematic review and evidence assessment (e.g., Cochrane Collaboration, AHRQ Evidence-based Practice Center Program, and GRADE). Both the OHAT and Navigation Guide methods include the key steps recommended by a previous National Academies committee (NRC 2014) for problem formulation, **protocol development**, specifying a study question, developing PECO statement, identifying and selecting the evidence, evaluating the evidence, and integrating the evidence." [78]

---

[65]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.

[66]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.

[67] 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

[68]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.

[69] 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.

[70] 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.

[71] 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.

[72] 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.

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

[74] 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.

[75] 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

[76] National Academies of Sciences, Engineering, and Medicine. 2019. Review of DOD's Approach to Deriving an Occupational Exposure Level for Trichloroethylene. Washington, DC: The National Academies Press. https://doi.org/10.17226/25610.

[77] 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

[78] National Academies of Sciences Engineering, and Medicine. (2017). Application of Systematic Review Methods in an Overall Strategy for Evaluating Low-Dose Toxicity from Endocrine Active Chemicals. Page. 119.Washington, D.C.: The National Academies Press; 2011

To assess the harms in human studies, instead of conducting an entirely new review, the NASEM used the Navigation Guide published systematic review on PBDE flame retardant exposure and IQ and concluded that:

> "To assess the human evidence, the committee critically evaluated the methods of a recent systematic review conducted by Lam et al… Judging that this existing review fulfilled the requirements of a systematic review and that there was **no evidence of risk of bias in the assessment**, the committee used the Lam et al. review as a basis for its own assessment." [79] (emphasis ours)

Protocols developed for applying the OHAT method[80] and the Navigation Guide Systematic Review Method have been published and can serve as a template to further expedite EPA's systematic reviews under TSCA.[81, 82]

Example of how the lack of protocol development influences EPA risk evaluations

Using the Pigment Violet 29 Draft Risk Evaluation as an example, EPA completed the entire systematic review in the absence of a protocol and complete method.[83] How a lack of protocol influenced how EPA evaluated data quality is illustrated by EPA's assessment of the Data Quality ratings for Metric 19 "Blinding of assessors" for animal toxicity studies (studies 1-13, 16-17; see Appendix A).

> "Blinding of assessors" is a risk of bias[84] rating to evaluate if personnel involved in assessing the study animals knew or did not know which animals were assigned to which exposure group (i.e., which animals were in a control or treatment group). This is an important risk of bias domain included in all systematic reviews as there is significant empirical evidence that not blinding assessors can bias the study results.[85,86]

The TSCA method's instructions for what constitutes a serious flaw in the 'Blinding of assessors' (Metric 19) states:

---

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

[80] National Toxicology Program Office of Health Assessment and Translation. (2015). Handbook for Conducting a Literature-Based Health Assessment Using OHAT Approach for Systematic Review and Evidence Integration. National Institute of Environmental Health Sciences; 2015

[81] All Navigation Guide systematic review protocols can be found at: https://prhe.ucsf.edu/navigation-guide The National Toxicology Program's protocol for its systematic review to evaluate the evidence for an association between exposure to PFOA or PFOS and immunotoxicity or immune-related health effects is at: https://ntp.niehs.nih.gov/ntp/ohat/pfoa_pfos/protocol_201506_508.pdf

[82] National Toxicology Program. Completed Evaluations. Available: https://ntp.niehs.nih.gov/whatwestudy/assessments/noncancer/completed/index.html

[83] US. EPA. (2019) Draft Toxic Substances Control Act Risk Evaluations: Colour Index Pigment Violet 29. Available: https://www.regulations.gov/docket?D=EPA-HQ-OPPT-2018-0604

[84] In Application of Systematic Review in TSCA Risk Evaluations, Under 'Table G-3. Data Evaluation Domains and Metrics for Animal Toxicity Studies' Page 175, EPA states, "Items marked with an asterisk (*) are examples of items that can be used to assess internal validity/risk of bias." This includes Metric 19 "Blinding of Assessors."

[85] Bello S, Krogsbøll LT, Gruber J, Zhao ZJ, Fischer D, Hróbjartsson A. Lack of blinding of outcome assessors in animal model experiments implies risk of observer bias. J Clin Epidemiol. 2014;67(9):973–983. doi:10.1016/j.jclinepi.2014.04.008

[86] Hirst JA, Howick J, Aronson JK, et al. The need for randomization in animal trials: an overview of systematic reviews. PLoS One. 2014;9(6):e98856. Published 2014 Jun 6. doi:10.1371/journal.pone.0098856

> "Information in the study report did not report whether assessors were blinded to treatment group for subjective outcomes and suggested that the assessment of subjective outcomes (e.g., functional observational battery, qualitative neurobehavioral endpoints, histopathological re-evaluations) was performed in a biased fashion (e.g., assessors of subjective outcomes were aware of study groups). This is a serious flaw that makes the study unusable."[87]

The above description is appropriate in that it will downgrade studies if it is not reported whether assessors were blinded to treatment and outcomes. In these instructions - 'subjective' refers to outcomes that rely on the experimenter's judgement to 'grade' a particular animal response and the responses can include multiple gradations of possible responses, such as skin irritation, which could be minor, medium, severe or anywhere along this continuum. 'Objective' means there is only one interpretation or measurement of the outcome possible, such as with death, and thus no exercise of judgement is necessary. (Note that we disagree with the rating of 'serious flaw', this is a different issue discussed below).

As shown in Appendix A, all of the animal toxicity studies EPA evaluates in the Pigment Violet 29 Draft Risk Evaluations measure subjective outcomes, and none of them report on blinding, yet EPA's final scores for all the studies is "not rated." According to EPA's own criteria, all but one of the Pigment Violet 29 animal toxicity studies should have been rated "low" or "unacceptable" for blinding of assessors.

In fact, as shown in Appendix A, EPA previously assigned a "medium" or "low" rating to this metric in 60% of these studies in the original 'Data Evaluation Scoring Sheets' in April 2018,[88,] which was subsequently changed to "not rated" in the 'Data Evaluation Scoring Sheets Updated Document' in April, 2019.[89] EPA gives various rationales for its revised scores, ranging from "It is not typically discussed in these studies," to "Blinding is not typically done..." While it is true that many animal studies are not blinded, this does not change the fact that empirical evidence indicates that lack of blinding biases studies, and thus they *should* be blinded—which is why validated risk of bias tools such as the Navigation Guide and OHAT method include this domain.[90, 91]

Table 1 is an example from Appendix A of how EPA previously assigned a "medium" or "low" rating to this metric in the original 'Data Evaluation Scoring Sheets' in April 2018, [92,] which was subsequently changed to "not rated" in the 'Data Evaluation Scoring Sheets Updated Document' in April, 2019.[93]

---

[87] EPA (2018) Application of Systematic Review in TSCA Risk Evaluations. Page. 188

[88] EPA (2018) C.I. Pigment Violet 29 (81-33-4) Systematic Review: Supplemental File for the TSCA Risk Evaluation, Data Evaluation Scoring Sheets

[89] EPA (2019) C.I. Pigment Violet 29 (81-33-4) Systematic Review: Supplemental File for the TSCA Risk Evaluation, Data Evaluation Scoring Sheets, Updated Document, April 2019

[90] Koustas, E., Lam, J., Sutton, P., Johnson, P. I., Atchley, D. S., Sen, S., … Woodruff, T. J. (2014). The Navigation Guide—Evidence-Based Medicine Meets Environmental Health: Systematic Review of Nonhuman Evidence for PFOA Effects on Fetal Growth. *Environmental Health Perspectives*, *122*(10), 1015–1027. https://doi.org/10.1289/ehp.1307177

[91] National Toxicology Program Office of Health Assessment and Translation. *Handbook for Conducting a Literature-Based Health Assessment Using OHAT Approach for Systematic Review and Evidence Integration*. National Institute of Environmental Health Sciences; 2015

[92] EPA (2018) C.I. Pigment Violet 29 (81-33-4) Systematic Review: Supplemental File for the TSCA Risk Evaluation, Data Evaluation Scoring Sheets

[93] EPA (2019) C.I. Pigment Violet 29 (81-33-4) Systematic Review: Supplemental File for the TSCA Risk Evaluation, Data Evaluation Scoring Sheets, Updated Document, April 2019

EPA Metric for "Blinding of assessors" from[1] TSCA Method	Study #9 BASF. 1975. Acute oral toxicity in rats
Blinding of assessors-EPA previous rating (2018)[2]	2
Blinding of assessors- EPA new rating (2019)[3]	NR*
EPA rationale in 2019[3] (Note: no rationales are given for the 2018 ratings)[3]	It is not typically discussed in these studies
Subjective observations in study	"Clinical symptoms of toxicity"

*Not Rated

1. EPA (2018) "Application of Systematic Review in TSCA Risk Evaluations."
2. EPA (2018) C.I. Pigment Violet 29 (81-33-4) Systematic Review: Supplemental File for the TSCA Risk Evaluation, Data Evaluation Scoring Sheets
3. EPA (2019) C.I. Pigment Violet 29 (81-33-4) Systematic Review: Supplemental File for the TSCA Risk Evaluation, Data Evaluation Scoring Sheets, Updated Document, April 2019

Legend

1	High
2	Medium
3	Low
4	Unacceptable
NR	Not Rated

If EPA found some empirical reason why blinding was not relevant to the outcome of these studies, and thus decided to follow criteria that deviated from its own TSCA method, it should have specified this in a pre-established protocol, *prior to rating the studies.* While EPA had a general method (the TSCA method), they did not have a specific protocol for Pigment Violet 29, this would have been the appropriate place to provide additional or alternative instructions about rating this domain. As it stands, without a pre-established protocol, EPA's ratings changes and rationales indicate a lack of scientific expertise at best or intentional changes to bias the evaluation results at worst.

We recommend: EPA should immediately implement the use of pre-established protocols to enhance transparency in the decision-making process and consistency in their draft risk evaluations. EPA should use the published protocols developed for applying the OHAT method and the Navigation Guide that can serve as a template to further expedite EPA's TSCA systematic review process.

3. **The literature review step of EPA's TSCA method incorporates select best practices, but also falls short of, or is unclear about, many other best practices for conducting a systematic and transparent literature review**

Overall, we commend the EPA for its efforts to incorporate many best practices for a comprehensive literature search in its TSCA method. We compared EPA's TSCA method for systematic review to the IOM's best practices for the literature review step of a systematic review (See Chapter 3 and TABLE E-1), [94] which was applied by the NASEM in its review of EPA's IRIS Program methods for systematic review (See Table 4-1).[95]

We found EPA's TSCA Method to be consistent with 12 of IOM's 27 best practices for conducting a literature search (Figure 1 and Appendix B). There are two key features of EPA's framework that are clearly inconsistent with IOM's best practices. EPA fails to:

1. Include or exclude studies based on the protocol's pre-specified criteria, a practice that is critical to avoiding results-based decisions (IOM 3.3.1);[96] and
2. Use two or more members of the review team, working independently, to screen and select studies, which is an essential quality-assurance measure (IOM 3.3.3).[97]

For the remaining 13 IOM best practices, EPA's framework as described in the TSCA method is either unclearly stated (N=7) or the practice is not mentioned at all (N=6). However, based on the literature review methods presented in the first 10 TSCA draft risk evaluations, EPA's appears to have now incorporated five additional practices that are either unclear or not mentioned in EPA's TSCA method as shown in Figure 1, these include:

- Work with a librarian or other information specialist trained in performing systematic reviews to plan the search strategy (IOM 3.1.1);
- Design the search strategy to address each key research question (IOM 3.1.2);
- Search regional bibliographic databases if other databases are unlikely to provide all relevant evidence (IOM 3.1.9);
- Conduct a web search (IOM 3.2.5); and
- Provide a line-by-line description of the search strategy, including the date of search for each database, web browser, etc. (IOM 3.4.1)

Therefore, EPA should immediately update its TSCA Method and make its framework for conducting a literature review transparently congruent with these five additional practices. The transparency of the framework would also be improved by specifying how EPA is addressing each best practice; at this juncture, how EPA intends to specifically handle many components of its literature searches could not readily be identified. For example, the TSCA method is unclear about whether EPA will include papers published in languages other than English. The exclusive reliance on English-language studies may lead to under-representation of the entire body of available evidence, and studies have also suggested that language bias might lead to erroneous conclusions.[98] Furthermore, when considering the inclusion or update of an existing systematic review, studies have found that language-inclusive systematic reviews (including

[94] Institute of Medicine. Finding What Works in Health Care. Standards for Systematic Review. Chapter 3. Washington, D.C.: The National Academies Press; 2011.

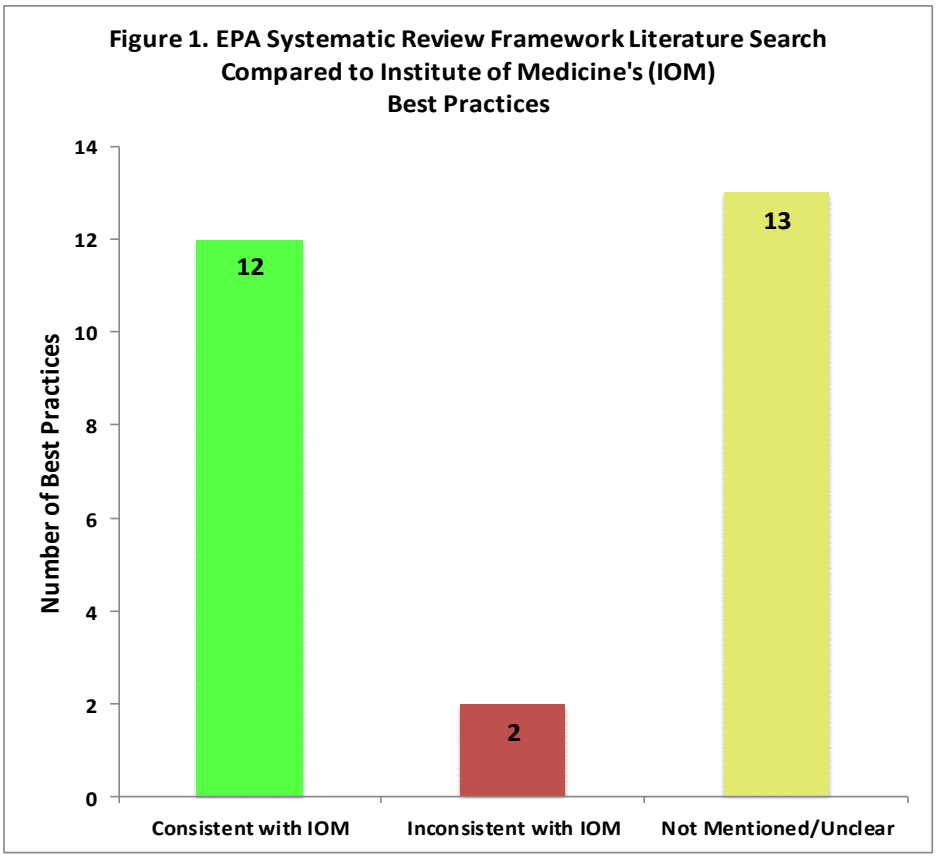
[95] National Research Council. Review of EPA's Integrated Risk Information System (IRIS) Process. Table 4-1, Page 43-45. Washington, DC: National Academies Press; 2014.

[96] See our Comment #2 regarding the TSCA method lack of a pre-defined protocol.

[97] EPA's framework, "Summary of the Title/Abstract Screening Conducted for the First Ten TSCA Risk Evaluations" Pp 24 states that only one screener conducted the screening and categorization of titles and abstracts.

[98] Morrison A, Polisena J, Husereau D, Moulton K, Clark M, Fiander M, Mierzwinski-Urban M, Clifford T, Hutton B, Rabb D. The effect of English-language restriction on systematic review-based meta-analyses: a systematic review of empirical studies. . International journal of technology assessment in health care. 2012;28((2)):138-44.

studies in languages other than English) were of the highest quality, compared with other types of reviews.[99] Online translation tools are readily available to allow screeners to quickly evaluate study abstracts for relevance, and therefore we recommend EPA to incorporate non-English language studies in their screening and not simply exclude in advance these potentially relevant papers.



EPA has failed, however, to adhere to one of the best practices we identified in the TSCA method as being consistent with the IOM's best practices for literature review (as shown in Figure 1 and Appendix B) when conducting the first 10 TSCA draft risk evaluations, that is:

- Document the disposition of each report identified, including reasons for their exclusion if appropriate (IOM 3.4.2).

*Example of how **failing to comply with the IOM standards for conducting a systematic and transparent literature review EPA has threatened the validity of the daft risk evaluations***

[99] Moher D, Pham B, Lawson ML, Klassen TP. The inclusion of reports of randomised trials published in languages other than English in systematic reviews. . Health Technol Assess. 2003;7((41)):1-90.

Below we highlight, using examples from the completed draft risk evaluations for Carbon Tetrachloride and Trichloroethylene, how by failing to comply with two IOM standards, EPA has threatened the validity of the daft risk evaluations. They are:

- *IOM standard 3.3.1 Include or exclude studies based on the protocol's pre-specified criteria* Rationale: On the basis of the study question, inclusion and exclusion criteria for the review would be set a priori, before reviewing the search results (see 3.3.5) so as to avoid results-based decisions. [100]
- *IOM standard 3.4.2 Document the disposition of each report identified, including reasons for their exclusion if appropriate*
  Rationale: The standard supports creation of a flow chart that describes the sequence of events leading to identification of included studies, and it also supports assessment of the sensitivity and precision of the searches a posteriori. [101]

(1)IOM standard 3.3.1 Include or exclude studies based on the protocol's pre-specified criteria

The PECO framework should shape the entire review process, including the search strategy to be used, the study eligibility criteria to be applied, how the data will be extracted from the included studies, the strategy for synthesizing the evidence and how the results will be reported.[102] The IOM states that:

> '**Using prespecified inclusion and exclusion criteria to choose studies is the best way to minimize the risk of researcher biases influencing the ultimate results of the SR** (CRD, 2009; Higgins and Deeks, 2008; Liberati et al., 2009; Silagy et al., 2002)**.** The SR research protocol should make explicit which studies to include or exclude based on the patient population and patient outcomes of interest, the healthcare intervention and comparators, clinical settings (if relevant), and study designs (e.g., randomized vs. observational research) that are appropriate for the research question.' [103] (emphasis ours)

The literature and screening strategy as specifically applied to the Carbon Tetrachloride Draft Risk Evaluation is described in '*Strategy for Conducting Literature Searches for Carbon Tetrachloride (CCL4): Supplemental Document to the TSCA Scope Document'*, which was published in June of 2017.[104] The results of the screening of literature search were published in *'Carbon tetrachloride (CASRN 56-23-5) Bibliography: Supplemental File for the TSCA Scope Document'* (no date is given in this document although the webpage on which this document is made available says 'last updated on June 22, 2017').[105] However, as highlighted by EPA in the

[100] Institute of Medicine. Finding What Works in Health Care. Standards for Systematic Review. Chapter 3, Page. 272. Washington, D.C.: The National Academies Press.; 2011.
[101] Institute of Medicine. Finding What Works in Health Care. Standards for Systematic Review. Chapter 3, Page. 275. Washington, D.C.: The National Academies Press.; 2011.
[102] NTP. (2015). Handbook for conducting a literature-based health assessment using OHAT approach for systematic review and evidence integration. U.S. Dept. of Health and Human Services, National Toxicology Program.
[103] Institute of Medicine. (2011). Finding What Works in Health Care: Standards for Systematic Reviews. Chapter 3, Page. 109. Standards for Finding and Assessing Individual Studies. Washington, DC: The National Academies Press
[104] US EPA. (2017). Strategy for Conducting Literature Searches for Carbon Tetrachloride (CCL4): Supplemental Document to the TSCA Scope Document'. Available: https://www.epa.gov/sites/production/files/2017-06/documents/ccl4_lit_search_strategy_053017_markup_0.pdf
[105] US EPA. (2017). Carbon Tetrachloride(CASRN:56-23-5) Bibliography:
Supplemental File for the TSCA Scope Document. Available: https://www.epa.gov/sites/production/files/2017-06/documents/ccl4_comp_bib_0.pdf

Carbon Tetrachloride Draft Risk Evaluation, for studies determined to be 'on-topic' (or relevant) after title and abstract screening, EPA conducted a full text screening to further exclude references that were not relevant to the risk evaluation:

> "Screening decisions were made based on eligibility criteria documented in the form of the populations, exposures, comparators, and outcomes (PECO) framework or a modified framework…The inclusion and exclusion criteria for full text screening for carbon tetrachloride are available in Appendix F of the Problem Formulation of the Risk Evaluation for Carbon Tetrachloride (U.S. EPA, 2018d.)." [106]

However, the *'Problem Formulation of the Risk Evaluation for Carbon Tetrachloride'* that outlined this PECO framework was published in May 2018, **after** the searches and initial screening had been completed.[107] The fact that the PECO framework was published after the studies had already been identified in the literature search and screened at the title and abstract stage, means that the PECO could be adjusted based on what literature had been included rather than using a more agnostic approach to literature selection. The consequence could be that the eligibility criteria was adjusted to support a pre-expected health hazard conclusion and thus contribute to bias in the evaluation if studies were subsequently excluded or included that supported this pre-expected hazard conclusion. However, it is difficult to judge given the lack of transparency, and leads to less confidence in the conclusions.

While the IOM uses P**I**CO (population, intervention, comparator, outcomes) and not P**E**CO statements as their standards relate to systematic reviews applied in the clinical sciences, the elements of each and the principles underlying them are the same, as they are designed to *'minimize the risk of researcher biases influencing the ultimate results of the SR'.* The critical importance of this is reinforced in *IOM standard 3.3.1 "Include or exclude studies based on the protocol's pre-specified criteria."* [108]

(2) IOM standard 3.4.2 Document the disposition of each report identified, including reasons for their exclusion if appropriate

EPA's method does not adequately cover literature screening, identification and selection. A key element of literature identification is transparency. EPA's process is not transparent and during the development of the risk evaluations using the TSCA method this has resulted in inconsistencies in how studies are identified and the number of studies identified.

The Draft Risk Evaluation for Trichloroethylene is an example of these inconsistencies in the reporting of the included and excluded references.[109] In section 1.5.2 Data Evaluation in the Trichloroethylene Draft Risk Evaluation, EPA states:

---

[106] US EPA. (2020). Carbon Tetrachloride (Methane, Tetrachloro-); Draft Toxic Substances Control Act (TSCA) Risk Evaluation. Page 38-39. Available: EPA Document# EPA-740-R1-8014

[107] US EPA. (2018). Problem Formulation of the Risk Evaluation for Carbon Tetrachloride. Available: https://www.epa.gov/sites/production/files/2018-06/documents/ccl4_problem_formulation_05-31-18.pdf

[108] Institute of Medicine. (2011). Finding What Works in Health Care: Standards for Systematic Reviews. 3. Standards for Finding and Assessing Individual Studies. Washington, DC: The National Academies Press

[109] US EPA. (2020). Trichloroethylene (CASRN 79-01-6); Draft Toxic Substances Control Act (TSCA) Risk Evaluation. Available: https://www.epa.gov/sites/production/files/2020-02/documents/1_draft_risk_evaluation_for_trichloroethylene_tce_public.pdf

> "During the data evaluation stage, the EPA assesses the quality of the methods and reporting of results of the individual studies identified during problem formulation using the evaluation strategies described in Application of Systematic Review in TSCA Risk Evaluations (U.S. EPA, 2018b). The **EPA evaluated the quality of the on-topic TCE study reports identified in** [Trichloroethylene (CASRN 79-01-6) **Bibliography**: Supplemental File for the TSCA Scope Document; (U.S. EPA, 2017i)], **and gave all studies an overall high, medium, low or unacceptable confidence rating during data evaluation**."[110] (emphasis ours)

In the Trichloroethylene Bibliography Supplemental File for the TSCA Scope Document[111] there are 49 pages of 'on topic' study reports for Human Health Hazards, with approximately 25 citations per page, totaling approximately >1200 'on topic' study reports. However, in "Figure 1-9 Literature Flow Diagram for Human Health Hazard" [112] below, EPA states that there are 180 studies that go through Data Evaluation (only 'on topic' studies go through Data Evaluation) leaving over >1000 'on-topic' Trichloroethylene study reports unaccounted for by EPA.

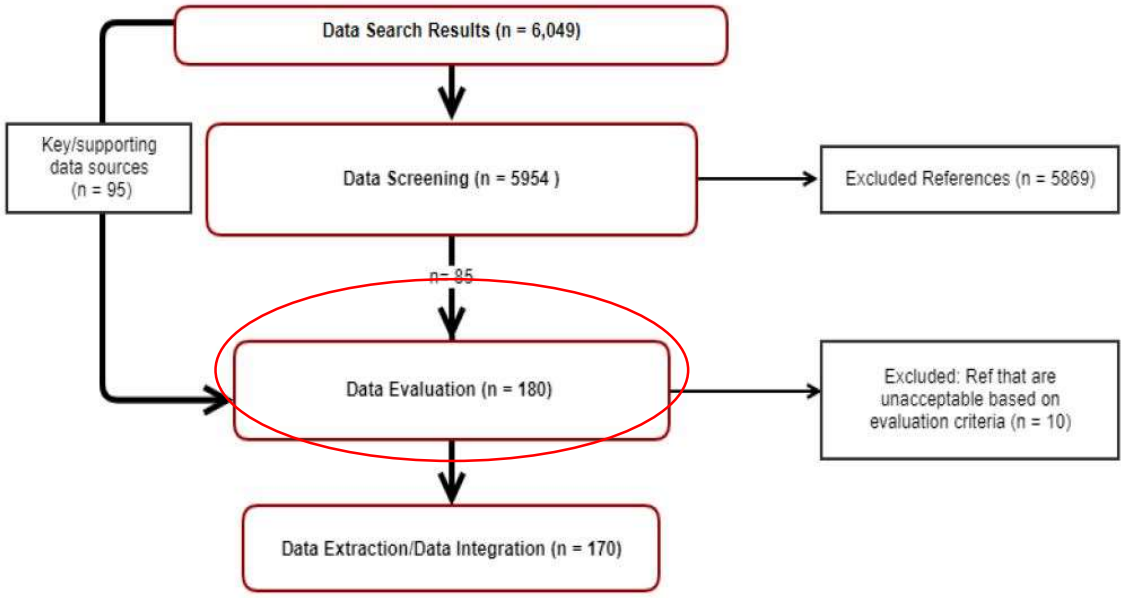


Figure 1-9. Literature Flow Diagram for Human Health Hazard

Of further concern, is the problematic inconsistency between the numbers of studies included in the data evaluation step as recorded in the supplemental files and those shown in the "Figure 1-9 Literature Flow Diagram for Human Health Hazards" above. EPA states in the Trichloroethylene Draft Risk Evaluation that:

---

[110]US EPA. (2020). Trichloroethylene (CASRN 79-01-6); Draft Toxic Substances Control Act (TSCA) Risk Evaluation. Page 66. Available: https://www.epa.gov/sites/production/files/2020-02/documents/1_draft_risk_evaluation_for_trichloroethylene_tce_public.pdf

[111] US EPA. (2017). Trichloroethylene (CASRN 79-01-6) Bibliography: Supplemental File for the TSCA Scope Document; Available: https://www.epa.gov/sites/production/files/2017-06/documents/tce_comp_bib.pdf

[112]US EPA. (2020). Trichloroethylene (CASRN 79-01-6); Draft Toxic Substances Control Act (TSCA) Risk Evaluation. Page 66. Available: https://www.epa.gov/sites/production/files/2020-02/documents/1_draft_risk_evaluation_for_trichloroethylene_tce_public.pdf

"Supplemental files also provide details of the data evaluations including individual metric scores and the overall study score for each data source (Docket: EPA-HQ-OPPT-2019-0500)."[113]

Source in TCE Risk Evaluation	Number that are 'on topic'/go through data evaluation
Systematic Review Supplemental File: Data Quality Evaluation	215
Figure 1-9 Literature Flow Diagram for Human Health Hazard	180
Bibliography	>1,200

Table 2. showing the differences in numbers of on topic/data evaluation studies in different sections of the TCE draft risk evaluation. It is unclear what is the final number of studies or what is included/excluded.

EPA cites in the footnote below this statement 'See Appendix B for the list of all supplemental files.' In Appendix B EPA cites "Systematic Review Supplemental File: Data Quality Evaluation of Human Health Hazard Studies – Animal and Mechanistic Data" [114] and "Systematic Review Supplemental File: Data Quality Evaluation of Human Health Hazard Studies – Epidemiological Data" [115] , which contain **ALL** of the included studies EPA evaluated for Human Health Hazards. However, there are **119** (animal (97) and mechanistic (22)) studies that that go through Data Quality Evaluation as cited in the "Systematic Review Supplemental File: Data Quality Evaluation of Human Health Hazard Studies – Animal and Mechanistic Data" [116] and **96** Epidemiological studies that go through Data Quality Evaluation as cited in the "Systematic Review Supplemental File: Data Quality Evaluation of Human Health Hazard Studies – Epidemiological Data" [117] ,which equals **215** studies that EPA evaluated for quality in assessing Human Health Hazards. Therefore, based on these number **EPA is missing 35 data sources** in 'Figure 1-9' above. Such inconsistencies are concerning and threaten the validity of the draft risk evaluations.

[113]US EPA. (2020). Trichloroethylene (CASRN 79-01-6); Draft Toxic Substances Control Act (TSCA) Risk Evaluation. Page 66. Available: https://www.epa.gov/sites/production/files/2020-02/documents/1_draft_risk_evaluation_for_trichloroethylene_tce_public.pdf

[114]US EPA. (2020). Systematic Review Supplemental File: Data Quality Evaluation of Human Health Hazard Studies – Animal and Mechanistic Data. Available: https://www.epa.gov/sites/production/files/2020-02/documents/14_tce-data_quality_evaluation_of_human_health_hazard_studies_-_animal_and_mechanistic_data.pdf

[115]US EPA. (2020). Systematic Review Supplemental File: Data Quality Evaluation of Human Health Hazard Studies – Epidemiological Data Available: ttps://www.epa.gov/sites/production/files/2020-02/documents/15_tce-data_quality_evaluation_of_human_health_hazard_studies_-_epidemiological_data.pdf

[116]US EPA. (2020). Systematic Review Supplemental File: Data Quality Evaluation of Human Health Hazard Studies – Animal and Mechanistic Data. Available: https://www.epa.gov/sites/production/files/2020-02/documents/14_tce-data_quality_evaluation_of_human_health_hazard_studies_-_animal_and_mechanistic_data.pdf

[117]US EPA. (2020). Systematic Review Supplemental File: Data Quality Evaluation of Human Health Hazard Studies – Epidemiological Data Available: ttps://www.epa.gov/sites/production/files/2020-02/documents/15_tce-data_quality_evaluation_of_human_health_hazard_studies_-_epidemiological_data.pdf

Further, how EPA accounted for the included studies in each step of "Figure 1-9 Literature Flow Diagram for Human Health Hazards" above is inconsistent with the approach used in "Figure 1-8. Literature Flow Diagram for Environmental Hazard" below, despite that these diagrams are only one page apart in the Trichloroethylene Draft Risk Evaluation. 'Figure 1-8' below includes the **appropriate** additional step of reporting the number of studies that are screened at the 'Title/Abstract' stage and the number at the 'Full Text Screening' stage while 'Figure 1-9' does not.[118]

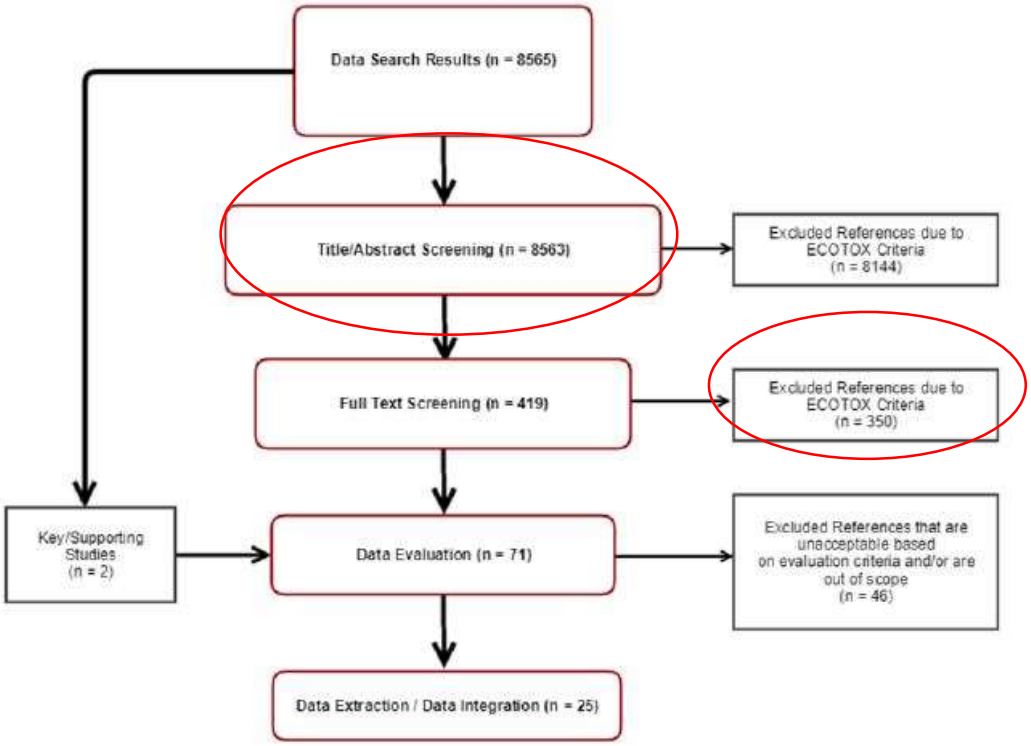


Figure 1-8. Literature Flow Diagram for Environmental Hazard

Further, in a systematic review, studies that make it to 'Full Text Screening' but are excluded thereafter, should only be excluded with an explicit justification. The IOM reports that:

> "In light of the subjective nature of study selection and the large volume of possible citations, the **importance of maintaining a detailed account of study selection cannot be understated**…The SR final report should include a flow chart that shows the number of studies that remain after each stage of the selection process…. **The flow chart documents the number of records identified through electronic databases searched**, whether additional records were identified through other sources, **and the reasons for excluding articles**.

[118]US EPA. (2020). Trichloroethylene (CASRN 79-01-6); Draft Toxic Substances Control Act (TSCA) Risk Evaluation. Page 65. Available: https://www.epa.gov/sites/production/files/2020-02/documents/1_draft_risk_evaluation_for_trichloroethylene_tce_public.pdf

Maintaining a record of excluded as well as selected articles is important." [119](emphasis ours)

The critical importance of stating the rationale for excluding studies throughout the systematic review process is highlighted in *IOM Standard 3.4.2 "Document the disposition of each report identified including reasons for their exclusion if appropriate."*[120] Therefore, as demonstrated in "Figure 1-8. Literature Flow Diagram for Environmental Hazard" above, there are 350 full text references that have been excluded in assessing environmental hazards without a sufficient justification to explain their exclusion. This lack of transparency in documenting the disposition of each report identified, including reasons for their exclusion in the draft risk evaluations EPA has completed to date could lead to bias in that EPA may have excluded studies that are scientifically relevant to the evaluation.

The EPA's SACC has also made comments and recommendations on the literature identification step with recommendations for how this step in the systematic review process should be conducted to comprehensively assess risks as required by law.

The EPA SACC in its Peer Review of 1-BP commented: "The Committee expected all of the quality sources identified in the SR would be used in the DRE and if not, that the general public would be able to follow the rationale as to why they were not used. The Committee generally concluded that it was difficult at best to determine exactly what was done during the SR…..**Committee members expressed that they experienced challenges in trying to follow the actions taken in the SR**, and how the results of the SR were used in the draft risk assessment." [121](emphasis ours)

The EPA SACC in its Peer Review of 1-BP recommended: "Since large percentages of studies are excluded (Section 1.5.1, page 42), the number of items being rejected for each criterion should be summarized to enable readers to determine why studies were excluded." [122]

The EPA SACC in its Peer Review of 1, 4 Dioxane commented: "Committee members did not find the systematic review to be a transparent and objective method for gathering the relevant scientific information, scoring its quality, and integrating the information evaluate."[123]

The EPA SACC in its Peer Review of 1, 4 Dioxane commented: "The Evaluation flow charts suggest a full systematic review was performed, but the text describes a more limited review."[124]

---

[119] Institute of Medicine. (2011). Finding What Works in Health Care: Standards for Systematic Reviews. Chapter 3, Page. 114. Standards for Finding and Assessing Individual Studies. Washington, DC: The National Academies Press

[120] Institute of Medicine. (2011). Finding What Works in Health Care: Standards for Systematic Reviews. Chapter 3, Page. 85. Standards for Finding and Assessing Individual Studies. Washington, DC: The National Academies Press

[121] US EPA. (2019). Peer Review for the United States Environmental Protection Agency (EPA) Draft Risk Evaluation for 1-Bromopropane (1-BP). Page. 22. Available: https://www.regulations.gov/document?D=EPA-HQ-OPPT-2019-0235-0061

[122] US EPA. (2019). Peer Review for the United States Environmental Protection Agency (EPA) Draft Risk Evaluation for 1-Bromopropane (1-BP). Page. 25. Available: https://www.regulations.gov/document?D=EPA-HQ-OPPT-2019-0235-0061

[123] US EPA. (2019). Peer Review for EPA Draft Risk Evaluations for 1,4-Dioxane and Cyclic Aliphatic Bromide Cluster (HBCD). Page.31. Available: https://www.regulations.gov/document?D=EPA-HQ-OPPT-2019-0237-0064

[124] US EPA. (2019). Peer Review for EPA Draft Risk Evaluations for 1,4-Dioxane and Cyclic Aliphatic Bromide Cluster (HBCD). Page. 32 Available: https://www.regulations.gov/document?D=EPA-HQ-OPPT-2019-0237-0064

NAS Recommendation for this Step: "It is crucial that the selection of eligible studies is based on prespecified criteria in a manner that limits potential for bias (IOM Standard 3.3)…Screening studies requires careful judgments and meticulous documentation about eligibility" [125]

We recommend: EPA should align its framework for conducting a literature review to ensure that it is congruent with all of the IOM's best practices and explicitly predefine the eligibility criteria for the included studies before conducting any part of the systematic review process. The transparency of the framework would be improved by specifying how EPA is addressing each best practice.

4. **EPA's TSCA systematic review method utilizes a quantitative scoring method that is incompatible with the best available science in fundamental ways and can exclude relevant studies from consideration in the risk evaluation:**
   - **a. Quantitative scores for assessing the quality of an individual study are arbitrary and not evidence-based; the National Academies of Sciences, Engineering, and Medicine recommend against such scoring methods;**
   - **b. EPA's scoring method wrongly conflates how well a study is reported with how well the underlying research was conducted; and**
   - **c. EPA's scoring method excludes research based on one single reporting or methodological limitation.**

a. Quantitative scores for assessing the quality of an individual study are arbitrary and not evidence-based; the National Academies of Sciences, Engineering, and Medicine recommend against such scoring methods;

EPA's TSCA method employs a quantitative scoring method to assess the quality of individual studies, instead of risk of bias (we discuss this critically important distinction in point #4b. below), assigning, based on its "professional judgment", various weights for quality domains and then summing up the quantitative scores to decide whether a study is of "high", "medium", or "low" quality as follows:[126]

> "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

[125] National Academies of Sciences, Engineering, and Medicine. 2019. Review of DOD's Approach to Deriving an Occupational Exposure Level for Trichloroethylene. Page. 32. Washington, DC: The National Academies Press. https://doi.org/10.17226/25610.

[126] EPA (2018) Application of Systematic Review in TSCA Risk Evaluations. See Appendix A for a more detailed description of the scoring method; how the method will be applied specifically to various streams of evidence, i.e., occupational exposure and release data; animal and in vitro data; epidemiologic studies; etc., is described in subsequent Appendices B-H.

> 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[127] (scores between 1.7 and lower than 2.3 will be assigned an overall quality level of *Medium)*."[128]

This overall 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.[129] (emphasis ours)

Illustrative of the scoring method, in Appendix H "Data Quality Criteria for Epidemiologic Studies," EPA presents how scoring is applied to human studies, explaining:

> "The critical metrics within each domain are those that cover the most important aspects of the domain and are those that more directly evaluate the role of confounding and bias. After pilot testing the evaluation tool, EPA recognized that more attention (or weight) should be given to studies that measure exposure and disease accurately and allow for the consideration of potential confounding factors. Therefore, metrics deemed as critical metrics are those that identify the major biases associated with the domain, evaluate the measurement of exposure and disease, and/or address any potential confounding. … EPA/OPPT assigned a weighting factor that is twice the value of the other metrics within the same domain to each critical metric. Remaining metrics are assigned a weighting factor of 0.5 times the weighting factor assigned to the critical metric(s) in the domain. The sum of the weighting factors for each domain equals one."[130] There is no scientific evidence to support EPA's selection of these "critical metrics" as being more important that other metrics, i.e., why within the "Study Participation" domain "Participant Selection" and "Attrition" are more important than "Comparison Group"; and there are no data supporting EPA's choice of particular numbers for weighting these 'critical metrics' (i.e., some metrics are "twice" as important as the other metrics)."

Overall, there is no scientific justification for EPA to assign these or any other quantitative scoring measures for assessing the quality of an individual study. The implicit assumption in quantitative scoring methods is that we know empirically how much each risk of bias domain contributes to study quality, and that these domains are independent of each other; this is not a scientifically supportable underlying assumption. Research has documented that scoring methods have, at best, unknown validity, may contain invalid items, and that results of a

[127] EPA has incorrectly applied how it will 'round scores up' as "Cut-off values between Medium and Low: 1.67 + 0.67= 2.34, rounded up to 2.3" in fact rounds the score down to 2.3 and not 'up to 2.3' as is stated in The Application of Systematic Review in TSCA Risk Evaluations. While a minor point, it is illustrative of the lack of methodological rigor, care, or consistency found throughout the document.

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

[129] EPA's framework applies quantitative scoring to all types of data; EPA/OPPT "is not applying weighting factors to the general population, consumer, and environmental exposure data types. In practice, it is equivalent to assigning a weighting factor of 1, which statistically assumes that each metric carries an equal amount of weight." (Page. 96).

[130] EPA (2018) Application of Systematic Review in TSCA Risk Evaluations. Page. 225.

quality score are not scientifically meaningful or predictive of the quality of studies.[131,132,133] An examination of the application of quality scores in meta-analysis found that quality-score weighting produced biased effect estimates, with the authors explaining that quality is not a singular dimension that is additive, but that it is possibly non-additive and non-linear.[134,135]

Aggregating across quality criteria to produce a single score is recognized by preeminent systematic review methodologists as problematic and unreliable, because the weights assigned are arbitrary and focus on the quality of reporting rather than the design and conduct of the research.[136,137] As stated by the IOM:

> "Quality scoring systems have not been validated. Studies assessed as excellent quality using one scoring method may be subsequently assessed as lower quality using another scoring method." [138]

The NASEM in its review of the EPA's IRIS program's method for systematic review, strongly supported a methodology that did not incorporate quantitative scoring, stating:

> "…..calculating a score involves choosing a weighting for the subcomponents, and **such scaling generally is nearly impossible to justify** (Juni et al. 1999). Furthermore, a study might be well designed to eliminate bias, but because the study failed to report details in the publication under review, it will receive a low score. Most scoring systems mix criteria that assess risk of bias and reporting. **However, there is no empirical basis for weighting the different criteria in the scores.** Reliability and validity of the scores often are not measured. Furthermore, quality scores have been shown to be invalid for assessing risk of bias in clinical research (Juni et al. 1999). **The current standard in evaluation of clinical research calls for reporting each component of the assessment tool separately and not calculating an overall numeric score** (Higgins and Green 2008)."[139] (emphasis ours)

[131] Herbison P, Hay-Smith J, Gillespie W. Adjustment of meta-analyses on the basis of quality scores should be abandoned. J Clin Epidemiol. 2006;59(12):1249-56. Epub 2006 Sep 11; PMCID: 17098567.

[132] 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. Epub 1999/09/24. doi: joc81641 [pii]. PubMed PMID: 10493204

[133] Stang A. Critical evaluation of the Newcastle-Ottawa scale for the assessment of the quality of nonrandomized studies in meta-analyses. Eur J Epidemiol. 2010;25(9):603-5. Epub 2010 Jul 22. doi: 10.1007/s10654-010-9491-z.; PMCID: 20652370

[134] Herbison P, Hay-Smith J, Gillespie W. Adjustment of meta-analyses on the basis of quality scores should be abandoned. J Clin Epidemiol. 2006;59(12):1249-56. Epub 2006 Sep 11; PMCID: 17098567.

[135] Greenland S, O'Rourke K. On the bias produced by quality scores in meta-analysis, and a hierarchical view of proposed solutions. Biostatistics. 2001;2(4):463-71.

[136] 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

[137] Moher D, Jadad AR, Nichol G, Penman M, Tugwell P, S. W. Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists. Control Clin Trials. 1995;16:62-73.

[138] Institute of Medicine. (2011). Finding What Works in Health Care: Standards for Systematic Reviews. Page 132. Washington, DC: The National Academies Press.

[139] National Research Council. Review of EPA's Integrated Risk Information System (IRIS) Process. Page 69. Washington, DC: National Academies Press; 2014

b. EPA's scoring method wrongly conflates how well a study is reported with how well the underlying research was conducted.

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.[140]

EPA's TSCA method acknowledges that reporting is not the same as an underlying flaw in study methodology, but then proceeds to ignore this distinction by using reporting as a measure of the quality of the underlying research, with consequences that could include excluding relevant scientific data.[141] Although EPA has made "Updates to the Data Quality Criteria for Epidemiological Studies", with its most recent draft risk evaluation for Perchloroethylene,[142] EPA's TSCA method still uses reporting measures in its scoring of the quality of human studies. For example, EPA includes STROBE reporting guidelines into the reasons for scoring studies "low quality" (Metrics 1 and 15) or "unacceptable for use" (Metrics 3, 4, 6, 7).

The authors of the STROBE guidelines specifically note the guidelines <u>are not</u> a measure of the quality of the underlying research, stating:

> "The STROBE Statement is a checklist of items that should be addressed in articles reporting on the 3 main study designs of analytical epidemiology: cohort, case control, and cross-sectional studies. **The intention is solely to provide guidance on how to report observational research well; these recommendations are not prescriptions for designing or conducting studies. Also, while clarity of reporting is a prerequisite to evaluation, the checklist is not an instrument to evaluate the quality of observational research** …Our intention is to explain how to report research well, not how research should be done. We offer a detailed explanation for each checklist item. Each explanation is preceded by an example of what we consider transparent reporting. This does not mean that the study from which the example was taken was uniformly well reported or well done; nor does it mean that its findings were reliable, in the sense that they were later confirmed by others: it only means that this particular item was well reported in that study."[143](emphasis ours)

How completely and clearly a study is reported is not a scientifically valid measure of the

---

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

[141] EPA (2018) Application of Systematic Review in TSCA Risk Evaluations. Page 31.

[142] US EPA. (2020). Risk Evaluation for Perchloroethylene (Ethene, 1,1,2,2-Tetrachloro) CASRN: 127-18-4 Systematic Review Supplemental File: Updates to the Data Quality Criteria for Epidemiological Studies. Available:https://www.epa.gov/sites/production/files/202004/documents/12_pce_updates_to_the_data_quality_criteria_for_epidemiological_studies.pdf

[143] 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.

quality of the underlying research.[144,145,146,147] As GRADE methodologists have succinctly stated, "… just because a safeguard against bias is not reported does not mean it was neglected." [148] Moreover, including many reporting items that are irrelevant to bias in a quality scoring rule (e.g., an indicator of whether power calculations were reported), will disproportionately reduce some of the resulting scores.[149]

The Cochrane Collaboration Handbook for conducting a systematic review clearly distinguishes reporting and bias, the latter which is defined as "a systematic error, or deviation from the truth, in results or inferences".[150] The Cochrane Manual for conducting systematic reviews is explicit about not conflating reporting with bias, stating:

> "Bias may be distinguished from quality. The phrase 'assessment of methodological quality' has been used extensively in the context of systematic review methods to refer to the critical appraisal of included studies. The term suggests an investigation of the extent to which study authors conducted their research to the highest possible standards." [151]

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

1. The key consideration in a Cochrane review is the extent to which results of included studies should be *believed*. Assessing risk of bias targets this question squarely.
2. A study may be performed to the highest possible standards yet still have an important risk of bias. For example, in many situations it is impractical or impossible to blind participants or study personnel to intervention group. It is inappropriately judgemental to describe all such studies as of 'low quality', but that does not mean they are free of 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

[144] 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.

[145] 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

[146] 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

[147] 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.

[148] 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

[149] Greenland S, O'Rourke K. On the bias produced by quality scores in meta-analysis, and a hierarchical view of proposed solutions. Biostatistics. 2001;2(4):463-71.

[150] 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.

[151] 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.

with the CONSORT Statement (Moher 2001d), are unlikely to have direct implications for risk of bias.

4. An emphasis on risk of bias overcomes ambiguity between the quality of reporting and the quality of the underlying research (although does not overcome the problem of having to rely on reports to assess the underlying research)."

Importantly, in the application of EPA's TSCA method, studies can be excluded from EPA's review, based solely on a deficiency in reporting, irrespective of the quality of the underlying research. Research documents that important information is often missing or unclear in published research,[152] as word limits, styles, and other specifications are highly variable, and non-standardized among peer-reviewed journals. As such, efforts to improve reporting are focused on uptake of reporting guidelines by journal editors and researchers.[153,154,155] Improving reporting is needed in academic research, but as stated by the developers of the STROBE guidelines, "We want to provide guidance on how to report observational research well.… **the checklist is not an instrument to evaluate the quality of observational research**." [156]

EPA's method rates as 'unacceptable for use' any human study that does not report even one of four reporting metrics (Metrics 3, 4, 6, 7); as highlighted, reporting metrics are often not included because of historical and present-day deficiencies in how studies are reported in the peer-reviewed literature. This will therefore lead to excluded human studies from consideration that otherwise would provide valuable information to the risk evaluation, leading to biased conclusions and subsequently underestimating health risks - all due to EPA's scoring system. This is not consistent with TSCA mandates to use the "best available science" and "reasonably available information," [157] and contradicts widely accepted empirically based systematic review methodological approaches.

c) EPA's scoring method excludes research based on one single reporting or methodological limitation.

EPA has created an arbitrary list of metrics that make studies "unacceptable for use in the hazard assessment," for each type of evidence stream, i.e., epidemiologic, animal, *in vitro*, etc.. EPA notes that an 'unacceptable score' means that "serious flaws" are noted in the domain metric, specifically

[152] 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.

[153] 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

[154] 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.

[155] 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

[156] 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.

[157] 15 USC §2625(h) and (k)

> "EPA/OPPT plans to use data with an overall quality level of *High, Medium*, or *Low* confidence to quantitatively or qualitatively support the risk evaluations, but does not plan to use data rated as *Unacceptable*. **Studies with any single metric scored as 4 will be automatically assigned an overall quality score of *Unacceptable* and further evaluation of the remaining metrics is not necessary**. An *Unacceptable* score means that serious flaws are noted in the domain metric that consequently make the data unusable (or invalid).[158] (emphasis added)

There is no empirical basis for EPA's selected list of "serious flaws".

For human epidemiologic studies (See Section H.5, Table H-8)[159], EPA lists six domains of study quality with 22 metrics, with varying numbers of metrics per domain: Study Participation (3 metrics); Exposure Characterization (3 metrics); Outcome Assessment (2 metrics); Potential Confounding/Variable Control (3 metrics); Analysis (4 metrics); and Other Considerations for Biomarker Selection and Measurement (7 metrics). Nineteen of the 22 metrics can be scored as a 4 (unacceptable) due to a "serious flaw". A study that has even one 'unacceptable' or "serious flaws" metric is considered to be "unacceptable for use" and is thus excluded from consideration in the risk evaluation (note that EPA has since amended the number of metrics that can be rated as "unacceptable for use" with now only 14 metrics, as shown in the "Updates to the Data Quality Criteria for Epidemiological Studies", in its most recent Draft Risk Evaluation for Perchloroethylene. We show these metrics below in Table 3).[160]

Table 3. EPA's list of 14 metrics that make studies "unacceptable for use in the hazard assessment," shown in "Updates to the Data Quality Criteria for Epidemiological Studies", in its most recent Draft Risk Evaluation for Perchloroethylene

Domain	Metric
Domain 1. Study Participation	Metric 1. Participant selection (selection, performance biases)
	Metric 2. Attrition (missing data/attrition/exclusion, reporting biases)
	Metric 3. Comparison Group (selection, performance biases)
Domain 2. Exposure Characterization	Metric 4. Measurement of Exposure (Detection/measurement/information, performance biases)
	Metric 5. Exposure levels (Detection/measurement/information biases)
	Metric 6. Temporality (Detection/measurement/information biases)

[158] EPA (2018) Application of Systematic Review in TSCA Risk Evaluations. Page. 227.

[159] EPA (2018) Application of Systematic Review in TSCA Risk Evaluations. Page. 231.

[160] US EPA. (2020). Risk Evaluation for Perchloroethylene (Ethene, 1,1,2,2-Tetrachloro) CASRN: 127-18-4 Systematic Review Supplemental File: Updates to the Data Quality Criteria for Epidemiological Studies. Available: https://www.epa.gov/sites/production/files/202004/documents/12_pce_updates_to_the_data_quality_criteria_for_epidemiological_studies.pdf

Domain 3. Outcome Assessment	Metric 7. Outcome measurement or characterization (detection/measurement/information, performance, reporting biases)
Domain 4. Potential Confounding/Variable Control	Metric 9. Covariate Adjustment (confounding)
	Metric 10. Covariate Characterization (measurement/information, confounding biases)
Domain 5. Analysis	Metric 12. Study Design and Methods
	Metric 13. Statistical power (sensitivity)
Domain 6. Other (if applicable) Considerations for Biomarker Selection and Measurement	Metric 16. Use of Biomarker of Exposure (detection/measurement/information biases)
	Metric 17. Effect biomarker (detection/measurement/information biases)
	Metric 20. Sample contamination (detection/measurement/information biases)

There are a number of problems with this approach which include:

- **EPA should not have a single evaluation exclude a study from consideration.**

- The approach is again inconsistent with two previously validated methods used to evaluate the risk of bias in human epidemiological studies recommended by the NASEM, the Navigation Guide[161] and OHAT method. [162] Neither methods recommend the use of excluding a study based on single measure. While the Navigation Guide does not exclude any studies based on the risk of bias assessment, OHAT "*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*."[163] 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 then to demonstrate how conclusions might be affected by the inclusion of high risk of bias studies, performing a sensitivity analysis; present multiple

[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. Environ Health Perspect. 2014;122(10):1007-1014.

[162] National Toxicology Program Office of Health Assessment and Translation. Handbook for Conducting a Literature-Based Health Assessment Using OHAT Approach for Systematic Review and Evidence Integration. National Institute of Environmental Health Sciences; 2015.

[163] 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.

(stratified) analyses; or present every included study and summarize the risk of bias, using structured approaches like GRADE.[164]

- **<u>EPA's list of "serious flaws" are not all equal indicators of study quality</u>**:

- For example, among human observational studies, any one of the 14 metrics listed in "Updates to the Data Quality Criteria for Epidemiological Studies" in the Draft Risk Evaluation for Perchloroethylene[165] can eliminate a study from consideration as EPA considers all of these "flaws" to be of equal importance; as described in detail above (in point #4a.), such weighting is arbitrary and not a science-based method.

- **<u>EPA's list of "serious flaws" are not all related to real flaws in the underlying research</u>**:

- **Reporting** guidelines are wrongly equated with "serious flaws" in study quality as described in detail above (in point #4b). For example, in scoring the quality of human studies, 4 of 14 "serious flaw" metrics (Metrics 3, 4, 6, 7) are STROBE reporting guidelines (STROBE checklist items # 6, 7, 8, 15). **A study would be scored as "unacceptable for use" by EPA based on any one of these STROBE reporting guidelines**. As described above in comment #4b, the STROBE guideline developers explicitly state this is neither the intended nor a scientifically valid use of these guidelines. [166]

- **Analysis** is one of the domains that includes "Statistical Power" (metric 13), which can be rated unacceptable (shown below in Figure 2). Thus, for cohort and cross-sectional studies, a study will be excluded if "The number of participants is inadequate to detect an effect in the exposed population and/or subgroups of the total population" or if the reported statistical power is not high enough (≥ 80%) to detect an effect in the exposure population and/or subgroups of the total population. [167] However, statistical power alone is not a valid measure of study quality and should not be used to exclude studies from consideration. [168]

[164] National Toxicology Program Office of Health Assessment and Translation. Handbook for Conducting a Literature-Based Health Assessment Using OHAT Approach for Systematic Review and Evidence Integration. National Institute of Environmental Health Sciences; 2015.

[165] US EPA. (2020). Risk Evaluation for Perchloroethylene (Ethene, 1,1,2,2-Tetrachloro) CASRN: 127-18-4 Systematic Review Supplemental File: Updates to the Data Quality Criteria for Epidemiological Studies. Available: https://www.epa.gov/sites/production/files/202004/documents/12_pce_updates_to_the_data_quality_criteria_for_epidemiological_studies.pdf

[166] 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.

[167] US EPA. (2020). Risk Evaluation for Trichloroethylene Systematic Review Supplemental File: Updates to the Data Quality Criteria for Epidemiological Studies CASRN: 79-01-6. Available: https://www.epa.gov/sites/production/files/2020-02/documents/16_tce-updates_to_the_data_quality_criteria_for_epidemiological_studies.pdf

[168] A power calculation is an estimate of the size of the study population needed to detect an effect of a given size.

Figure 2. EPA TSCA Method Metric 13 "Statistical Power" (Analysis Domain) Excerpted from Risk Evaluation for Trichloroethylene Systematic Review Supplemental File: Updates to the Data Quality Criteria for Epidemiological Studies CASRN: 79-01-6

Metric 13. Statistical power (sensitivity)	
High (score = 1)	Do not select for this metric
Medium (score = 2)	• ***For cohort and cross-sectional studies:*** The number of participants are adequate to detect an effect in the exposed population and/or subgroups of the total population. **OR** • The paper reported statistical power high enough (≥ 80%) to detect an effect in the exposure population and/or subgroups of the total population. • ***For case-control studies:*** The number of cases and controls are adequate to detect an effect in the exposed population and/or subgroups of the total population. **OR** • The paper reported statistical power was high (≥ 80%) to detect an effect in the exposure population and/or subgroups of the total population.
Low (score =3)	• Do not select for this metric.
Unacceptable (score = 4)	• ***For cohort and cross-sectional studies:*** The number of participants are inadequate to detect an effect in the exposed population and/or subgroups of the total population. • ***For case-control studies:*** The number of cases and controls are inadequate to detect an effect in the exposed population and/or subgroups of the total population.
Not rated/applicable	• *Do not select for this metric*

For this one metric - there are several drawbacks of excluding studies based on statistical power. First, EPA's Metric 13 "Statistical Power" (sensitivity) is not an appropriate measure of bias. For example a small study may be imprecise but that should not be confused with whether it is biased;[169] also a small study can be imprecise but at the same time less biased than a larger study.[170] Individual studies that are "underpowered" (for example, because in the real world the exposed population may not be large enough for statistical purposes even if they are health-impacted) can still be potentially valuable to evidence-based decision-making. Small "underpowered" studies can also be combined in a meta-analysis that increases the statistical power of the body of evidence to reflect the relationship between an exposure and a health

[169] 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.

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

impact. Additionally, "underpowered" studies that find a health effect to be present may be indicative of a larger effect size than anticipated. Thus, omitting such studies would severely bias the conclusions of the review.

Additionally, EPA does not provides a method for how it will determine the "adequacy" of the statistical power of a study on which to base its score and provides no rationale for excluding studies with less than 80% statistical power. According to STROBE guideline developers, … "before a study is conducted power calculations are made with many assumptions that once a study is underway may be upended; further, power calculations are most often not reported."[171]

Example of how EPA's "Statistical Power" metric could result in excluding high quality evidence

Lam et al. published a systematic review *"Developmental PBDE Exposure and IQ/ADHD in Childhood: A Systematic Review and Meta-analysis."*[172] As highlighted above in point #2, this systematic review was used by the NASEM as part of the report "Application of Systematic Review Methods in an Overall Strategy for Evaluating Low-Dose Toxicity from Endocrine Active Chemicals" in which the NASEM stated *"there was* ***no evidence of risk of bias in the [Lam] assessment"***.[173] In the Lam systematic review, there were 15 studies evaluated, and 4 studies, which had in general low risk of bias across seven domains[174], were included in the meta-analysis. None of these 4 studies reported a power calculation, and yet together, these studies found "a 10-fold increase (in other words, times 10) in PBDE exposure associated with a decrement of 3.70 IQ points (95% confidence interval: 0.83, 6.56)."

Thus, the use of the EPA statistical power metric alone in the review process would have excluded highly informative and low risk of bias studies. We demonstrate how the TSCA risk of bias (study quality) method would influence study exclusion in the Lam et al. review in Figure 3 below.[175] As highlighted above, the Lam et al. systematic review, using the best available scientific methods, concluded that there was sufficient evidence supporting an association between developmental PBDE exposure and reduced IQ, a finding that was subsequently reviewed and endorsed by the NASEM ***'no evidence of risk of bias in the assessment'***.[176] Yet as demonstrated in Figure 3. below, EPA's TSCA method would exclude every study from this body of evidence based on the Agency's unvalidated, non-evidence-based criteria for deeming studies 'Unacceptable.'

---

[171] 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.

[172] 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.

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

[174] "High quality" defined as "definitely" or "probably" low or very low risk of bias (Figure 2a in the *Lam et al* paper) based on specific and detailed definitions of risk of bias established before the review was conducted.

[175]

[176] National Academies of Sciences, Engineering, and, Medicine. Application of Systematic Review Methods in an Overall Strategy for Evaluating Low-Dose Toxicity from Endocrine Active Chemicals. Page. 8. Washington, DC: 2017 2017

Figure 3. Individual and overall study ratings of the included studies from *"Developmental PBDE Exposure and IQ/ADHD in Childhood: A Systematic Review and Meta-analysis"* by Lam et al. 2017, using the TSCA method[177]

Further, the exclusion of studies based on one 'Unacceptable' metric is not consistent with the EPA's 2017 framework rules which requires the Agency to consider all relevant science while accounting for "strengths and limitations." [178]

Finally, multiple authoritative review bodies, including the EPA SACC, the NASEM and IOM have concluded that overly quantitative criteria that exclude relevant studies are inappropriate in systematic review methods. Using a scoring method is inappropriate and can exclude relevant evidence. Below are highlights from relevant reports from the EPA SACC, NASEM and IOM.

The EPA SACC Peer Review of 1-BP commented: "Several Committee members discussed in depth that it was not appropriate to determine an "unacceptable" rating during data quality evaluation based solely on one criterion."[179]

The EPA SACC Peer Review of 1, 4 Dioxane recommended: "Do not be overly stringent and exclude studies based on a single criterion."[180]

The EPA SACC Peer Review of 1, 4 Dioxane recommended: "Follow best practices in the field and simplify the data quality criteria." [181]

[177] Eick S, Goin DE, Chartres N, Lam J; Woodruff TW. (2020) Assessing Risk of Bias in Human Epidemiologic Studies Using Three Tools: Different Conclusions from Different Tools. *Systematic Reviews.* Under Review.
[178] 40 CFR 702 Pg. 33733
[179] US EPA. (2019). Peer Review for the United States Environmental Protection Agency (EPA) Draft Risk Evaluation for 1-Bromopropane (1-BP). Page. 21. Available: https://www.regulations.gov/document?D=EPA-HQ-OPPT-2019-0235-0061
[180] US EPA. (2019). Peer Review for EPA Draft Risk Evaluations for 1,4-Dioxane and Cyclic Aliphatic Bromide Cluster (HBCD). Page. 38. Available: https://www.regulations.gov/document?D=EPA-HQ-OPPT-2019-0237-0064
[181] US EPA. (2019). Peer Review for EPA Draft Risk Evaluations for 1,4-Dioxane and Cyclic Aliphatic Bromide Cluster (HBCD). Page. 38. Available: https://www.regulations.gov/document?D=EPA-HQ-OPPT-2019-0237-0064

NAS Recommendation for this Step: "Most significantly, the quantitative scores are contrary to standard systematic review practices, as numerical scores falsely imply a relationship between scores and effect or association, along with several other critical limitations"[182]

NAS Recommendation for this Step: "The committee recommends that DOD abandon the use of this study applicability tool in favor of established tools to assess risk of bias of animal and human studies. For example, one option could be the approach developed by the National Toxicology Program's Office of Health Assessment and Translation."[183]

The IOM Recommendation for this step: "Quality scoring systems have not been validated. Studies assessed as excellent quality using one scoring method may be subsequently assessed as lower quality using another scoring method. Moreover, with an emphasis on risk of bias, the SR more appropriately assesses the quality of study design and conduct rather than the quality of reporting. The committee chose the term "risk of bias" to describe the focus of the assessment of individual studies and the term "quality" to describe the focus of the assessment of a body of evidence." [184]

We recommend: EPA should not use a quantitative scoring method to assess quality in individual studies; it should not conflate study reporting with study quality; and it should not exclude otherwise quality research based on a single reporting or methodological limitation. Rather EPA should employ a scientifically valid method to assess risk of bias of individual studies.

5. **EPA's TSCA systematic review method does not have a pre-established protocol or methods for evidence integration.**

EPA has not fully described how they will conduct their evidence integration in a transparent manner. The 2 following paragraphs set for the outline of EPA's approach:

> "In other words, it will involve assembling the relevant data and evaluating the data for quality and relevance, followed by synthesis and integration of the evidence to support conclusions (U.S. EPA, 2016). The significant issues, strengths, and limitations of the data and the uncertainties that require consideration will be presented, and the major points of interpretation will be highlighted. Professional judgment will be used at every step of the process and will be applied transparently, clearly documented, and to the extent possible, follow principles and procedures that are articulated prior to conducting the assessment (U.S. EPA, 2016)" [185]
>
> "The last step of the systematic review process is the summary of findings in which the evidence is summarized, the approaches or methods used to weigh

[182] National Research Council. (2014). Review of EPA's Integrated Risk Information System (IRIS) Process. Page 69. Washington, DC: The National Academies Press; 2014.
[183] National Academies of Sciences, Engineering, and Medicine. 2019. Review of DOD's Approach to Deriving an Occupational Exposure Level for Trichloroethylene. Page. 4. Washington, DC: The National Academies Press. https://doi.org/10.17226/25610.
[184] Institute of Medicine. (2011). Finding What Works in Health Care: Standards for Systematic Reviews. Page 132. Washington, DC: The National Academies Press
[185] EPA (2018) Application of Systematic Review in TSCA Risk Evaluations. Page. 27.

> the evidence are discussed, and the basis for the conclusion(s), recommendation(s), and any uncertainties are fully described. This step occurs in each of the components of the risk assessment (i.e., exposure assessment and hazard assessment) and is summarized in the risk characterization section of the TSCA risk evaluation." [186]

There are numerous challenges and problems with this step in the TSCA method. First, EPA's TSCA regulation governing procedures for chemical risk evaluations requires that it use a systematic review method to "integrate evidence."[187] However, the current language does not pre define how the TSCA method will address this step. In particular EPA writes:

> "Data integration activities for the first ten TSCA risk evaluation are anticipated to occur after the TSCA Problem Formulation documents are released (Figure 1-1). EPA/OPPT will provide further details about the data integration strategy along with the publication of the draft TSCA risk evaluations."[188]

A predefined method should be included in the protocol before EPA conducts their review. Consequently, without definition of how overall quality of evidence will be evaluated within each stream and how the evidence streams will be integrated, this leaves the assessment open to bias.

Using the Carbon Tetrachloride Draft Risk Evaluation as an example, EPA fails to clearly define how the quality of the body of evidence has been evaluated for each evidence stream and it failed to pre-specify the method for integrating two or more streams of evidence in formulating the final conclusions. [189] The Carbon Tetrachloride Draft Risk Evaluation. As shown below *in 'Figure 3-1 Hazard Identification and Dose Response Process'* of the Carbon Tetrachloride Draft Risk Evaluation, EPA conflates data quality evaluation and evidence integration in the *'Human Health Hazard Assessment'* and does not clearly outline how these two critically important steps were completed. [190]

[186] EPA (2018) Application of Systematic Review in TSCA Risk Evaluations. Page. 27.
[187] 40 CFR 702.33
[188] EPA (2018) Application of Systematic Review in TSCA Risk Evaluations. Page. 27.
[189] US EPA. (2020). Carbon Tetrachloride (Methane, Tetrachloro-); Draft Toxic Substances Control Act (TSCA) Risk Evaluation. Available: https://www.epa.gov/sites/production/files/2020-01/documents/1_ccl4_draft_risk_evaluation_draft_public_updated.pdf
[190] US EPA. (2020). Carbon Tetrachloride (Methane, Tetrachloro-); Draft Toxic Substances Control Act (TSCA) Risk Evaluation. Page. 98. Available: https://www.epa.gov/sites/production/files/2020-01/documents/1_ccl4_draft_risk_evaluation_draft_public_updated.pdf

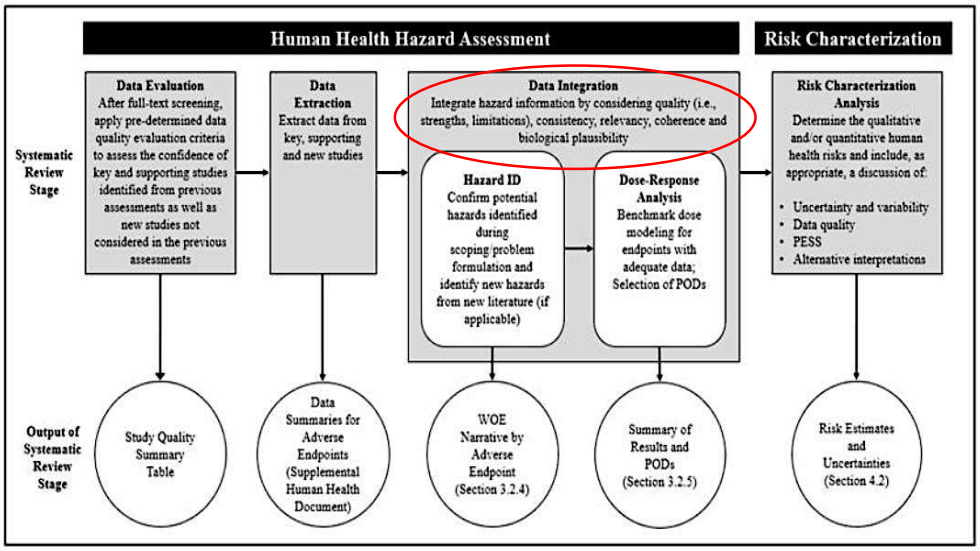


Figure 3-1. Hazard Identification and Dose-Response Process

In section *'3.2.4 Weight of Scientific Evidence*' of the Carbon Tetrachloride Draft Risk Evaluation EPA goes onto describe how it conflates both an evaluation of the quality of the body of evidence and the evidence integration steps during the 'weight of the scientific evidence' process:

> "The following sections describe the weight of the scientific evidence for both non-cancer and cancer hazard endpoints. Factors considered in weighing the scientific evidence included **consistency and coherence among human and animal studies, quality of the studies (such as whether studies exhibited design flaws that made them unacceptable) and biological plausibility**. **Relevance of data was considered primarily during the screening process but may also have been considered when weighing the evidence**." [191] (emphasis ours)

It is critical to clearly define how these <u>two separate steps</u> are to be conducted. The body of evidence for each stream of evidence must first clearly be assessed, integrating the findings from the internal validity (risk of bias) with considerations of external validity. Systematically rating the quality of individual studies and the quality of the overall body of evidence for each evidence stream based on pre-established and clearly stated criteria allows for greater transparency and consistency in the systematic review process. As the criteria to assess the internal validity of the body of evidence, such as the risk of bias, and the external validity, such as directness, is different for each evidence stream, it is only appropriate to first asses these evidence streams separately. Once the body of evidence has been assessed for each stream and a conclusion on the confidence in the body of evidence has been reached, the evidence streams can then be integrated.

[191] US EPA. (2020). Carbon Tetrachloride (Methane, Tetrachloro-); Draft Toxic Substances Control Act (TSCA) Risk Evaluation. Page. 120. Available: https://www.epa.gov/sites/production/files/2020-01/documents/1_ccl4_draft_risk_evaluation_draft_public_updated.pdf

Further, EPA states that they will continue to consider relevance of the data in the 'weighing the evidence' step – indicating that they could exclude (or include) studies even after they have gone through exclusion/inclusion and study quality evaluation. This could be interpreted that if there are studies that are not consistent with the finding (or support a finding) that EPA would like to achieve that they could apply a reason, which has not been pre-defined, to alter the list of studies they are including or excluding.

The incomplete nature of EPA's TSCA method for evidence integration is inconsistent with evidence-based methods of systematic review developed, endorsed, and/or advanced by the: NASEM;[192] the IOM;[193] the OHAT method;[194] the GRADE method;[195] SYRINA;[196] the SYRCLE method;[197] the Campbell Collaboration's methods;[198] and the Navigation Guide.[199] In each of these methods, there is first an overall evaluation in the confidence of the body of evidence using explicit, predefined criteria integrating the findings from the internal validity (risk of bias) with considerations of external validity. Some methods are designed for human evidence, and primarily RCTs (e.g. GRADE and IOM). While other methods have extended this approach to include evaluation of each stream of human and nonhuman evidence (e.g. OHAT method, Navigation Guide, SYRINA). What is universal among these systematic methods is that there are explicit, predefined criteria that are described in the protocol before the assessment has been conducted.

Note that the methods are in general consistent in that confidence in the body of evidence is based on integrating the findings from the internal validity (risk of bias) with considerations of external validity. And the approach is predefined in the protocol. Note that the overall evaluation in the confidence in the body of evidence is based on considerations from Bradford Hill as shown below.[200]

---

[192] NAS. (2017). Application of Systematic Review Methods in an Overall Stretegy for Evaluating Low-Dose Toxicity from Endocrine Active Chemicals. Washington, D.C.: The National Academies Press.; 2011

[193] Institute of Medicine. Finding What Works in Health Care. Standards for Systematic Review. Washington, D.C.: The National Academies Press.; 2011

[194] National Toxicology Program Office of Health Assessment and Translation. *Handbook for Conducting a Literature-Based Health Assessment Using OHAT Approach for Systematic Review and Evidence Integration*. National Institute of Environmental Health Sciences; 2015

[195] 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

Guyatt GH, Oxman AD, Vist G, Kunz R, Brozek J, Alonso-Coello P, Montori V, Akl EA, Djulbegovic B, Falck-Ytter Y, Norris SL, Williams JW, Jr., Atkins D, Meerpohl J, Schünemann HJ. GRADE guidelines: 4. Rating the quality of evidence--study limitations (risk of bias). Journal of Clinical Epidemiology. 2011;64(4):407-15. doi: 10.1016/j.jclinepi.2010.07.017

[196] Vandenberg, L. N., Ågerstrand, M., Beronius, A., Beausoleil, C., Bergman, Å., Bero, L. A., Rudén, C. (2016). A proposed framework for the systematic review and integrated assessment (SYRINA) of endocrine disrupting chemicals. *Environmental Health*, *15*(1), 74. https://doi.org/10.1186/s12940-016-0156-6

[197] Hooijmans CR, Rovers MM, de Vries RB, Leenaars M, Ritskes-Hoitinga M, Langendam MW. SYRCLE's risk of bias tool for animal studies. BMC medical research methodology. 2014;14:43. Epub 2014/03/29. doi: 10.1186/1471-2288-14-43. PubMed PMID: 24667063.

[198] Campbell Collaboration. Better evidence for a better world. 2018 [cited 2018 July 29]The Campbell Collaboration promotes positive social and economic change through the production and use of systematic reviews and other evidence synthesis for evidence-based policy and practice.]. Available from: https://campbellcollaboration.org/research-resources/writing-a-campbell-systematic-review.html

[199] Woodruff TJ, Sutton P, The Navigation Guide Work Group. An Evidence-Based Medicine Methodology To Bridge The Gap Between Clinical And Environmental Health Sciences. Health Affairs. 2011;30(5):931-7. doi: 10.1377/hlthaff.2010.1219; PMCID: 21555477

[200] Descatha, A., G. Sembajwe, F. Pega, Y. Ujita, M. Baer, F. Boccuni, C. Di Tecco, C. Duret, B. A. Evanoff, D. Gagliardi, L. Godderis, S.-K. Kang, B. J. Kim, J. Li, L. L. Magnusson Hanson, A. Marinaccio, A. Ozguler, D. Pachito, J. Pell, F. Pico, M. Ronchetti, Y. Roquelaure, R. Rugulies, M. Schouteden, J. Siegrist, A. Tsutsumi and S. Iavicoli (2020). "The effect of exposure to long working hours on stroke: A

Table 3.- Bradford Hill considerations and their relationship to GRADE and the Navigation Guide for evaluating the overall quality of the evidence for human observational studies.

Bradford Hill	GRADE	Navigation Guide
Strength	Strength of association and imprecision in effect estimate	Strength of association and imprecision in effect estimate
Consistency	Consistency across studies, i.e., across different situations (different researchers)	Consistency across studies, i.e., across different situations (different researchers)
Temporality	Study design, properly designed and conducted observational studies	Study design, properly designed and conducted observational studies
Biological Gradient	Dose response gradient	Dose response gradient
Specificity	Indirectness	Indirectness
Coherence	Indirectness	Indirectness
Experiment	Study design, properly designed and conducted observational studies	Study design, properly designed and conducted observational studies
Analogy	Existing association for critical outcomes leads to not downgrading the quality, indirectness	Existing association for critical outcomes leads to not downgrading the quality, indirectness. Evaluating the overall strength of body of human evidence allows consideration of other compelling attributes of the data that may influence certainty.

Table adapted from (Schunemann et al. 2011).

Figure 4. below shows an example from the OHAT method of how these criteria are applied consistently and transparently to each evidence stream. The OHAT method has been used by the NASEM in two systematic reviews. In general, the evidence integration consists of steps that are explicit in translating the overall rating into a conclusion on the level of evidence for a health effect (see example from the OHAT method Figure 5.); and then using this integration to formulate a hazard identification conclusion. Human and animal evidence when available should be integrated, and some approaches can use mechanistic data to help inform the final conclusions (see example from OHAT method Figure 6.).[201]

EPA does not rate its confidence in the body of evidence in any of the draft risk evaluations it has completed to date, nor does it follow a proper evidence integration protocol to come to its final conclusion. Thus, EPA's method is inconsistent with best practices in the field. We therefore strongly recommend that EPA use these validated, peer review methods that are

systematic review and meta-analysis from the WHO/ILO Joint Estimates of the Work-related Burden of Disease and Injury." Environment International 142: 105746.

[201] Vandenberg, L. N., Ågerstrand, M., Beronius, A., Beausoleil, C., Bergman, Å., Bero, L. A., Rudén, C. (2016). A proposed framework for the systematic review and integrated assessment (SYRINA) of endocrine disrupting chemicals. *Environmental Health*, *15*(1), 74. https://doi.org/10.1186/s12940-016-0156-6

consistent with best practice for the evidence integration step in all risk evaluations it conducts.

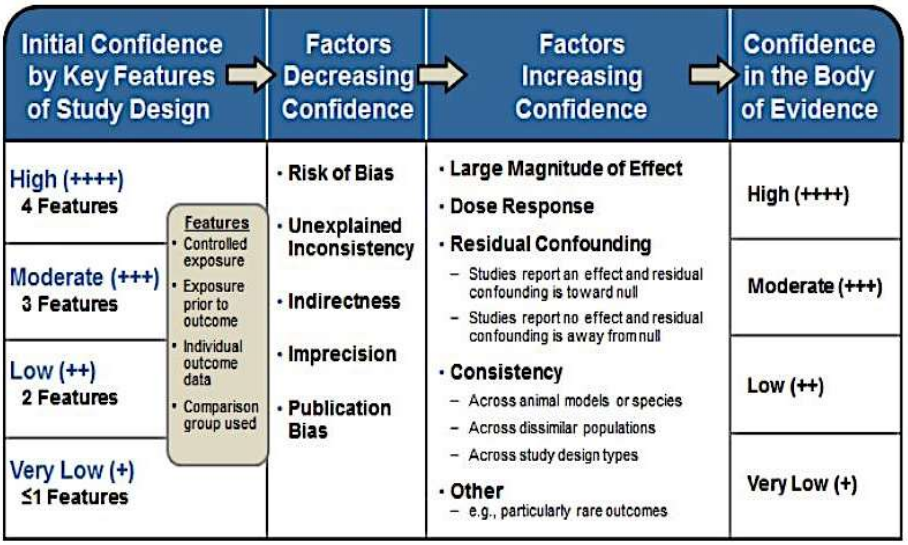


Figure 4: OHAT's method for rating the confidence in the body of evidence. This step is missing from every Draft Risk Evaluation. [202]

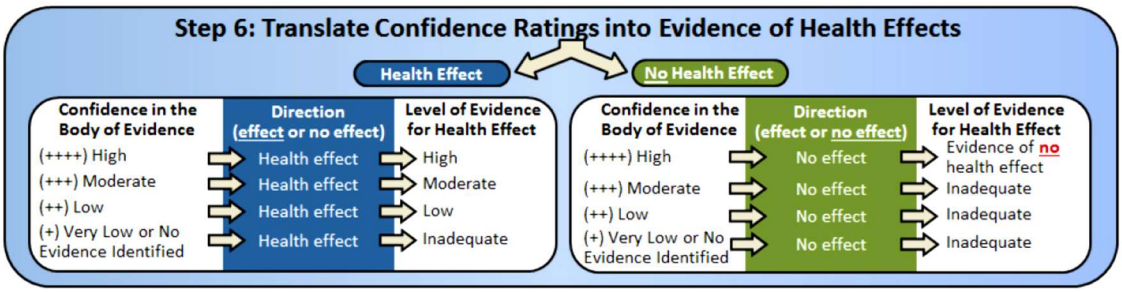


Figure 5: OHAT's method to translate confidence in the body of evidence to come to a conclusion on the level of evidence for a health effect. [203] This step is missing from every Draft Risk Evaluation. [204]

[202] National Toxicology Program Office of Health Assessment and Translation. (2015). Handbook for Conducting a Literature-Based Health Assessment Using OHAT Approach for Systematic Review and Evidence Integration. National Institute of Environmental Health Sciences; 2015. Pg. 48.

[203] National Toxicology Program Office of Health Assessment and Translation. (2015). *Handbook for Conducting a Literature-Based Health Assessment Using OHAT Approach for Systematic Review and Evidence Integration*. National Institute of Environmental Health Sciences; 2015. Pg. 64

[204] National Toxicology Program Office of Health Assessment and Translation. (2015). Handbook for Conducting a Literature-Based Health Assessment Using OHAT Approach for Systematic Review and Evidence Integration. National Institute of Environmental Health Sciences; 2015. Pg. 64.

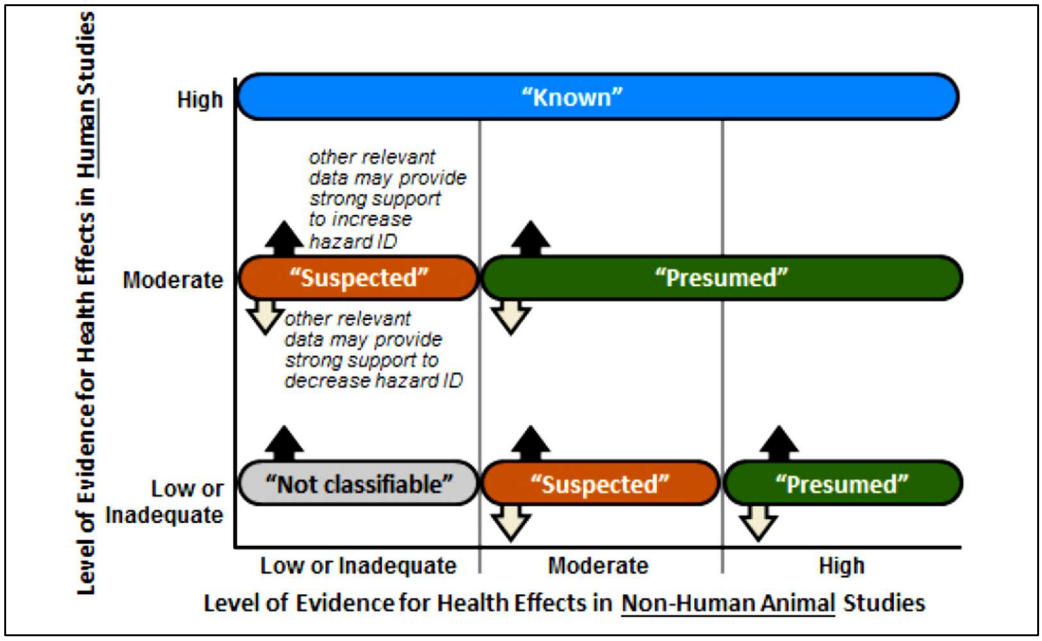


Figure 6: OHAT's process to translate the level of evidence for a health effect into a hazard identification conclusion. This step is missing from every Draft Risk Evaluation. [205]

Finally, multiple authoritative review bodies, including the EPA SACC and the NASEM, have recommended that EPA provide more transparency around the evidence integration and final steps in the hazard identification.

The EPA SACC Peer Review of 1-BP commented: "Improve the clarity of data integration. Multiple times papers that had been identified for data extraction and integration were not used with no explanation as to why." [206]

The EPA SACC Peer Review of PV29 commented: "Regarding data integration, the Committee discussed the benefits of including a more thorough and inclusive data integration discussion in the TSCA SR for PV29… there is a need in the Evaluation for a thorough description and outline for how all evidence and data are integrated into a final weight of evidence conclusion. This was not transparent from reading the documents provided." [207]

NAS Recommendation for this Step: "EPA should continue to improve its evidence-integration process incrementally and enhance the transparency of its process. If EPA does move to a structured evidence-integration process, it should combine resources with NTP to leverage the intellectual resources and scientific experience in both organizations" [208]

---

[205] National Toxicology Program Office of Health Assessment and Translation. (2015). Handbook for Conducting a Literature-Based Health Assessment Using OHAT Approach for Systematic Review and Evidence Integration. National Institute of Environmental Health Sciences; 2015. Page. 67.

[206] US EPA. (2019). Peer Review for the United States Environmental Protection Agency (EPA) Draft Risk Evaluation for 1-Bromopropane (1-BP). Page. 25.Available: https://www.regulations.gov/document?D=EPA-HQ-OPPT-2019-0235-0061

[207] US EPA. (2019). Peer Review for EPA Draft Risk Evaluation of C.I. Pigment Violet 29 (PV29). Page 27. Available: https://www.regulations.gov/docket?D=EPA-HQ-OPPT-2018-0604 D=EPA-HQ-OPPT-2018-0604

[208] National Research Council. (2014). Review of EPA's Integrated Risk Information System (IRIS) Process. Page 105. Washington, DC: The National Academies Press; 2014.

We recommend: EPA should immediately implement an evidence integration method that is consistent with best practice in systematic review, such as the OHAT method and the Navigation Guide and transparently present how the conclusions are reached in assessing human health hazards for each end point it assesses.

6. EPA's TSCA systematic review method does not consider financial conflicts of interest as a potential source of bias in research.

As observed by the Deputy Editor (West) of JAMA in 2010, "the biggest threat to [scientific] integrity [is] financial conflicts of interest."[209] Yet EPA's TSCA method is silent on how it will take into account this empirically documented influence on the results of scientific research. Underscoring this EPA TSCA method deficiency is the fact that recent studies empirically document that industry sponsorship produces research that is favorable to the sponsor.[210,211] The influence of financial ties on research can be traced to a variety of types of biases, and this conflict of interest needs to be distinguished from non-financial interests in the research, which can also affect research.[212]

The fact that funding source needs to be accounted for in some manner is empirically supported and not a subject of scientific debate; what scientists differ on is *how* to best address funding as a potential source of bias;[213,214] for example, whether funding source is assessed as a specific risk of bias domain[215] or considered at multiple points in the evaluation.[216,217] For example, funding source is recommended as a factor to consider when evaluating risk of bias of individual studies for selective reporting, and then again for evaluating the body of evidence for publication bias,[218]and/or to be considered as a potential factor to explain apparent inconsistency within a body of evidence. [219]

---

[209] Rennie D. Integrity in scientific publishing. Health Serv Res. 2010;45(3):885-96. Epub 2010/03/27. doi: HESR1088 [pii] 10.1111/j.1475-6773.2010.01088.x. PubMed PMID: 20337732

[210] Lundh A, Lexchin J, Mintzes B, Schroll JB, Bero L. Industry sponsorship and research outcome. Cochrane Database Syst Rev. 2017(2:MR000033.). doi: 10.1002/14651858.MR000033.pub3.; PMCID: 28207928.

[211] White J, Bero LA. Corporate manipulation of research: Strategies are similar across five industries. Stanford Law & Policy Review. 2010;21((1)):105-34.

[212] Bero L. Addressing Bias and Conflict of Interest Among Biomedical Researchers. JAMA. 2017;317(17):1723-4. doi: 10.1001/jama.2017.3854; PMCID: 28464166

[213] Bero L. Why the Cochrane risk of bias tool should include funding source as a standard item [editorial]. Cochrane Database Syst Rev. 2013;12:ED000075.

[214] Sterne JA. Why the Cochrane risk of bias tool should not include funding source as a standard item. Cochrane Database Syst Rev. 2013;(12)(:ED000076.). doi: 10.1002/14651858.ED000076.; PMCID: PMID: 24575440

[215] Bero L. Why the Cochrane risk of bias tool should include funding source as a standard item [editorial]. Cochrane Database Syst Rev. 2013;12:ED000075

[216] 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; 2015.

[217] Sterne JA. Why the Cochrane risk of bias tool should not include funding source as a standard item. Cochrane Database Syst Rev. 2013;(12)(:ED000076.). doi: 10.1002/14651858.ED000076.; PMCID: PMID: 24575440

[218] 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 in Systematic Reviews of Health Care Interventions. 2012 AHRQ Publication No. 12-EHC047-EF.

[219] 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; 2015.

A 2017 Cochrane systematic review of industry sponsorship and research outcomes concluded that "industry sponsorship should be treated as bias-inducing and industry bias should be treated as a separate domain."[220] The NASEM in its review of the EPA IRIS program's systematic review method found that "Funding sources should be considered in the risk-of-bias assessment conducted for systematic reviews that are part of an IRIS assessment."[221]

Notably, EPA's exclusion of consideration of funding source and other potential conflicts of interests is also internally inconsistent with its own improper reliance on STROBE guidelines as quality measures: STROBE guidelines item #22 specified that "the source of funding and the role of funders, could be addressed in an appendix or in the methods section of the article."[222]

Importantly, including funding as a risk of bias as a domain **does not** mean excluding industry sponsored studies from EPA's hazard and risk assessment; it only means documenting funding as one of many domains of potential bias and evaluating its impact on the overall quality of the body of evidence.

NAS Recommendation for this Step: Funding sources should be considered in the risk-of-bias assessment conducted for systematic reviews that are part of an IRIS assessment. [223]

We recommend: EPA should assess study funding source and author financial conflicts of interests as a risk of bias domain for individual studies.

[220] Lundh A, Lexchin J, Mintzes B, Schroll JB, Bero L. Industry sponsorship and research outcome. Cochrane Database Syst Rev. 2017(2:MR000033.). doi: 10.1002/14651858.MR000033.pub3.; PMCID: 28207928.

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

[222] 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.

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

Appendix A. Detailed Ratings for EPA Metric “Blinding of Assessors”

EPA Metric from TSCA Method[1]	Study #9 BASF. 1975. Acute oral toxicity in rats	Study #10 BASF. 1978. Acute oral toxicity in rats	Study #11 Rupprich. 1984. Acute toxicity in Wistar rats	Study #5 BASF. 1975. Acute inhalation in rats	Study #6 BASF. 1978. Acute inhalation in rats	Study #7 BASF. 1975. Acute IP toxicity in mice	Study #8 BASF 1978. Acute IP toxicity in mice	Study #17 Stark. 2013. Repro/dev Toxicity in wistar rats	Study #12 BASF. 1975. Skin irritation study XXV/454	Study #13 BASF. 1978. Skin Irritation study 77/360	Study #1 BASF. 1975. Eye Irritation study	Study #3 Rupprich. 1984. Perylimid - acute dermal irritant	Study #2 BASF. 1978. Eye irritation study.	Study #4 Rupprich 1984. Acute irritant - rabbit eye. 840229.	Study #16 Johnson 1999. Local Lymph node assay
Blinding of assessors- EPA previous rating (2018)[2]	2	2	2	2	2	2	2	N/A	3	3	N/A	N/A	N/A	N/A	N/A
Blinding of assessors- EPA new rating (2019)[3]	NR*	NR*	NR*	NR*	NR*	NR*	NR*	NR	NR*	NR*	NR	NR	NR	NR	NR
EPA rationale in 2019[3] (Note: no rationales are given for the 2018 ratings)	It is not typically discussed in these studies	It is not typically discussed in these studies	It is not typically discussed in these studies	Blinding is not typically done for acute inhalation studies that are assessing mortality, clinical signs (e.g., irritation) and gross pathology.	Blinding is not typically done for acute inhalation studies that are assessing mortality, clinical signs (e.g., irritation) and gross pathology.	It is not typically discussed in these studies	It is not typically discussed in these studies	Initial histopathology review was the only subjective assessment conducted, and this metric is not applicable.	It is not typically discussed in these studies. Note that the grading of dermal responses is subjective. Training in observing the dermal responses and translating them to a score promotes harmonization of subjective results.	It is not typically done. Note that the grading of dermal responses is subjective. Training in observing the dermal responses and translating them to a score promotes harmonization of subjective results.	It is not discussed in these studies. Note that the grading of ocular responses is subjective. Training in observing the ocular responses and translating them to a score promotes harmonization of subjective results.	It is not typically discussed in these studies. Note that the grading of dermal responses is subjective. Training in observing the dermal responses and translating them to a score promotes harmonization of subjective results.	It is not discussed in these studies. Note that the grading of ocular responses is subjective. Training in observing the ocular responses and translating them to a score promotes harmonization of subjective results.	No subjective outcomes were assessed.	It is not typically discussed in these studies
Subjective observations in study	"Clinical symptoms of toxicity"	"Clinical symptoms of toxicity"	"Clinical toxic reactions/ symptoms of being poisoned"	Clinical signs such as irritation are subjective	Clinical signs such as irritation are subjective	"Clinical symptoms of toxicity"	"Clinical symptoms of toxicity"		EPA acknowledges that grading of dermal responses is subjective	EPA acknowledges that grading of dermal responses is subjective	EPA acknowledges that grading of ocular responses is subjective	EPA acknowledges that grading of dermal responses is subjective	EPA acknowledges that grading of ocular responses is subjective	As acknowledged by EPA in the other eye irritation studies, grading of ocular responses is subjective.	"Clinical observation s- signs of systemic toxicity"

Legend	
1	High
2	Medium
3	Low
4	Unacceptable
NR	Not Rated

References:

1 EPA (2018) "Application of Systematic Review in TSCA Risk Evaluations."

2 EPA (2018) C.I. Pigment Violet 29 (81-33-4) Systematic Review: Supplemental File for the TSCA Risk Evaluation, Data Evaluation Scoring Sheets

3 EPA (2019) C.I. Pigment Violet 29 (81-33-4) Systematic Review: Supplemental File for the TSCA Risk Evaluation, Data Evaluation Scoring Sheets, Updated Document, April 20

Appendix B: Comparison of IOM literature review best practices with EPA systematic review framework

IOM Standard (IOM 2011) and Rationale as cited in 2014 National Academy Review of the IRIS program (pp 43-55)	EPA Systematic Review Framework	Consistent with IOM	Inconsistent with IOM	Not Mentioned	Unclear	Applied to first 10 chemicals	Not Applied in first 10 chemicals
3.1 Conduct a comprehensive systematic search for evidence							
3.1.1 Work with a librarian or other information specialist trained in performing systematic reviews to plan the search strategy (p. 266). Rationale: As with other aspects of research, specific skills and training are required to navigate a wide range of bibliographic databases and electronic information sources.	**Not mentioned in the EPA Systematic Review Framework; based on first 10 chemicals EPA Systematic Review Framework consistent with this best practice.** Table 3-2 page 29 provides web links to the Strategy for Conducting Literature Searches and Bibliography documents published in June 2017 along with each of the first ten TSCA Scope documents. Within these documents it states that a professional librarian developed the search.			1		1	
3.1.2 Design the search strategy to address each key research question (p. 266). Rationale: The goal of the search strategy is to maximize both sensitivity (the proportion of all eligible articles that are correctly identified) and precision (the proportion of all articles identified by the search that are eligible). With multiple research questions, a single search strategy is unlikely to cover all questions posed with any precision.	**Unclear in the EPA Systematic Review Framework; based on first 10 chemicals EPA Systematic Review Framework consistent with this best practice.** Table 3-2 page 29 provides web links to the Strategy for Conducting Literature Searches and Bibliography documents published in June 2017 along with each of the first ten TSCA Scope documents. Within these documents multiple search strategies are presented.				1	1	
3.1.3 Use an independent librarian or other information specialist to peer review the search strategy (p. 267). Rationale: This part of the evidence review requires peer review like any other part. Given the specialized skills required, a person with similar skills would be expected to serve as peer reviewer.	**Not mentioned in the EPA Systematic Review Framework;**			1			
3.1.4 Search bibliographic databases (p. 267). Rationale: A single database is typically not sufficient to cover all publications (journals, books, monographs, government reports, and others) for clinical research. Databases for reports published in languages other than English and for the gray literature could also be searched.	**EPA Systematic Review Framework is consistent with this best practice.** pp 21-22 EPA SR Framework - "EPA/OPPT designed its initial data search to be broad enough to capture a comprehensive set of sources containing data/information potentially relevant to the risk evaluation process. Generally, the search was conducted on a wide range of data/information sources, includingbut not limited to peer-reviewed and grey literature8. When available, EPA/OPPT relied on the search strategies from recent assessments (e.g., EPA Integrated Risk Information System (IRIS) assessments) as a starting point to identify relevant references and supplemented these searches to identify relevant information published after the end date of the previous search to capture more recent literature." "Following the initial search of data for the first ten risk evaluations, EPA/OPPT searched for data submitted to EPA under TSCA sections 4, 5, 8(e), and 8(d), as well as for your information (FYI) submissions, to find additional data relevant to human health and environmental hazard, exposure, fate, engineering, physical-chemical properties, and TSCA conditions of use."	1					
3.1.5 Search citation indexes (p. 267). Rationale: Citation indexes are a good way to ensure that eligible reports were not missed.	**EPA Systematic Review Framework is consistent with this best practice.** EPA is searching Web of Science, a citation index, which searches Science, Social Science, and Arts & Humanities citation indexes	1					

3.1.6 Search literature cited by eligible studies (p. 268). Rationale: The literature cited by eligible studies (for example, references provided in a journal article or thesis) is a good way to ensure eligible reports were not missed.	**EPA Systematic Review Framework is consistent with this best practice.** EPA/OPPT identified additional environmental fate and exposure references that were not captured in the initial categorization of the on-topic references for the first ten risk evaluations published on June 22, 2017. Specifically, assessors identified references by checking the list of references of data sources frequently used to support EPA/OPPT's risk assessments (e.g., previous assessments cited in Table 1-1 of the TSCA Scope documents). This method, called backward reference searching (or snowballing), was not part of the initial literature search strategy. The inclusion of these additional on-topic references is not expected to change the information presented in the TSCA Scope and Problem Formulation documents. Also, EPA/OPPT anticipates targeted supplemental searches during the analysis phase (e.g., to locate specific information for exposure modeling). Backward reference searching will be included in the literature search strategy for supplemental searches.	1					
3.1.7 Update the search at intervals appropriate to the pace of generation of new information for the research question being addressed (p. 268). Rationale: Given that new articles and reports are being generated in an ongoing manner, searches would be updated regularly to reflect new information relevant to the topic.	**Not mentioned in the EPA Systematic Review Framework;**			1			
3.1.8 Search subject specific databases if other databases are unlikely to provide all relevant evidence (p. 268). Rationale: If other databases are unlikely to be comprehensive, search a variety of other sources to cover the missing areas.	**EPA Systematic Review Framework is consistent with this best practice.** The databases searched are not named in the EPA Systematic Review Framework. However, Table 3-2 page 29 provides web links to the Strategy for Conducting Literature Searches and Bibliography documents published in June 2017 along with each of the first ten TSCA Scope documents. Within these documents subject specific databases are searched. pp 21-22 "EPA/OPPT designed its initial data search to be broad enough to capture a comprehensive set of sources containing data/information potentially relevant to the risk evaluation process. Generally, the search was conducted on a wide range of data/information sources, including but not limited to peer-reviewed and grey literature8. When available, EPA/OPPT relied on the search strategies from recent assessments (e.g., EPA Integrated Risk Information System (IRIS) assessments) as a starting point to identify relevant references and supplemented these searches to identify relevant information published after the end date of the previous search to capture more recent literature. For human health hazards, the literature search strategy was designed to identify relevant data/information in favor (e.g., positive study) or against (e.g., negative study) a given hypothesis within the context of the assessment question(s) being evaluated in the risk evaluation." "Following the initial search of data for the first ten risk evaluations, EPA/OPPT searched for data submitted to EPA under TSCA sections 4, 5, 8(e), and 8(d), as well as for your information (FYI) submissions, to find additional data relevant to human health and environmental hazard, exposure, fate, engineering, physical-chemical properties, and TSCA conditions of use. Searches were conducted of CBI and non-CBI databases followed by a duplicate identification step. Many of the non-CBI data submissions were captured in the initial search published on June 22, 2017, but some were found and added to the pool of new references to undergo data screening."	1					
3.1.9 Search regional bibliographic databases if other databases are unlikely to provide all relevant evidence (p. 269). Rationale: Many countries have their own databases and either because of language or other regional factors the reports are not necessarily also present in US-based databases	**Unclear in the EPA Systematic Review Framework; based on first 10 chemicals EPA Systematic Review Framework consistent with this best practice in that state databases are searched.**				1	1	
		4	**0**	**3**	**2**	**3**	**0**
3.2 Take action to address potentially biased reporting of research results							
3.2.1 Search gray literature databases, clinical trial registries, and other sources of unpublished information about studies (p. 269). Rationale: Negative or null results, or undesirable results, might be published in difficult to access sources.	**EPA Systematic Review Framework is consistent with this best practice**. p 21-22 "Generally, the search was conducted on a wide range of data/information sources, including but not limited to peer-reviewed and grey literature"	1					

3.2.2 Invite researchers to clarify information about study eligibility, study characteristics, and risk of bias (p. 269). Rationale: Rather than classify identified studies as missing critical information, it is preferable to ask the investigators directly for the information.	**EPA Systematic Review Framework is consistent with this best practice.** Page 26 "When applicable and feasible, EPA/OPPT will reach out to the authors of the data/information source to obtain raw data or missing elements that would be important to support the data evaluation and data integration steps. In such cases, the request(s) for additional data/information, number of contact attempts, and responses from the authors will be documented."*	1					
3.2.3 Invite all study sponsors and researchers to submit unpublished data, including unreported outcomes, for possible inclusion in the systematic review (p. 270). Rationale: So as to include all relevant studies and data in the review, ask sponsors and researchers for information about unpublished studies or data.	**Not mentioned in the EPA Systematic Review Framework;**			1			
3.2.4 Hand search selected journals and conference abstracts (p. 270). Rationale: Hand searching of sources most likely provides relevant up-to-date information and contributes to the likelihood of comprehensive identification of eligible studies.	**Not mentioned in the EPA Systematic Review Framework;**			1			
3.2.5 Conduct a web search (p. 271). Rationale: Web searches, even when broad and relatively untargeted, can contribute to the likelihood that all eligible studies have been identified.	**Unclear in the EPA Systematic Review Framework; based on first 10 chemicals EPA Systematic Review Framework consistent with this best practice.**				1	1	
3.2.6 Search for studies reported in languages other than English if appropriate (p. 271). Rationale: There is limited evidence that negative, null, or undesirable findings might be published in languages other than English.	**Not mentioned in EPA Systematic Review Framework; unlcear in first 10 chemicals EPA Systematic Review, for example ecotox on methylene chloride excludes non english papers**			1			
		2	**0**	**3**	**1**	**1**	**0**
3.3 Screen and select studies							
3.3.1 Include or exclude studies based on the protocol's pre-specified criteria (p. 272). Rationale: On the basis of the study question, inclusion and exclusion criteria for the review would be set a priori, before reviewing the search results (see 3.3.5) so as to avoid results-based decisions.	**EPA Systematic Review Framework is inconsistent with this best practice; no pre-specified protocols developed for the first 10 chemicals; criteria listed in chemical specific strategies for conducting literature searches lack specificity needed to rapidly and transparently screen relevant papers.** Figure 3-1 includes protocol development as a first step. However, Table 3-1 begins with the data search phase of EPA's systematic review method. On page 19 EPA states, "The timeframe for development of the TSCA Scope documents has been very compressed. ... EPA had limited ability to develop a protocol document detailing the systematic review approaches and/or methods prior to the initiation of the risk evaluation process for the first ten chemical substances. For these reasons, the protocol development is staged in phases while conducting the assessment work." EPA's application of inclusion/exclusion criteria for the first 10 chemicals (based on asbestos and methlyene chloride) only generally lists inclusion and exclusion criteria. Methlylene chloride: page 80 INCLUDE: Studies evaluating human health effects resulting from exposure to the chemical. Includes epidemiology studies (measure an adverse outcome in an exposed population), experimental studies (e.g. individuals exposed to chemical in a controlled study) and case studies (e.g. individual case report on accidental exposure to chemical) Acute, subchronic, and chronic exposures **Also choose applicable health effect tags in next section "Methylene Chloride (DCM) Health Effect Tags" EXCLUDE: Occupational studies that do not specify specific solvent exposure page 83 asbestos - INCLUDE: Studies evaluating human health effects resulting from exposure to the chemical. Includes epidemiology studies (measure an adverse outcome in an exposed population), experimental studies (e.g. individuals exposed to chemical in a controlled study) and case studies (e.g.individual case report on accidental exposure to chemical) Acute, subchronic, and chronic exposures **Also choose applicable health effect tags in next section "asbestos Health Effect Tags"		1				

3.3.2 Use observational studies in addition to randomized controlled trials to evaluate harms of interventions (p. 272). Rationale: Predetermine study designs that will be eligible for each study question.	**EPA Systematic Review Framework is consistent with this best practice.**	**1**					
3.3.3 Use two or more members of the review team, working independently, to screen and select studies (p. 273). Rationale: Because reporting is often not clear or logically placed, having two independent reviewers is a quality-assurance approach.	**EPA Systematic Review Framework is not consistent with this best practice.** Based on first 10 chemicals EPA Systematic Review Framework one reviewer was used for title and abstract screening. Section 3.2.2.1 Title and abstract screening - page 23. "Each article is generally screened by two independent reviewers using specialized web-based software (i.e., DistillerSR)9. Screeners are assigned batches of references after conducing pilot testing. Screening forms are typically used to facilitate the screening process by asking a series of questions based on pre-determined inclusion and exclusion criteria. The screeners resolve conflicts by consensus, or consultation with an independent individual(s)." p. 24 "3.2.2.1.1 Summary of the Title/Abstract Screening Conducted for the First Ten TSCA Risk Evaluations **One screener (11)** conducted the screening and categorization of titles and abstracts. Relevant studies were identified according to inclusion and exclusion criteria as described in the Strategy for Conducting Literature Searches documents (Table 3-2)." (11) "Systematic review guidelines typically recommend at least two screeners to review each article to minimize bias. EPA had less than 6 months to conduct data collection and screening activities for 10 chemical substances; thus, one screener was used for the title/abstract screening to meet the statutory deadline in June 2017. However, full text screening generally used two independent screeners (see Section 3.2.2.2)."		1				
3.3.4 Train screeners using written documentation; test and retest screeners to improve accuracy and consistency (p. 273). Rationale: Training and documentation are standard quality-assurance approaches.	**EPA Systematic Review Framework is consistent with this best practice.** Table 3-1 states that EPA will train screeners in the data title abstract and full text screening, i.e., EPA states it will: "Conduct pilot study to test the criteria for title/abstract screening and tagging and conflict resolution strategy"; "Develop pilot plan to test criteria for the title/abstract screening and tagging." "Conduct pilot study to test the criteria for title/abstract screening and tagging and conflict resolution strategy.Unless major changes are made, piloting may only need to be conducted once and not after each update." and " Refine the screening and tagging criteria before application."	1					
3.3.5 Use one of two strategies to select studies: 1) read all full-text articles identified in the search or 2) screen titles and abstracts of all articles and then read the full-text of articles identified in initial screening (p. 273). Rationale: Data are not clear, even for clinical intervention questions, regarding which method is best, although 2) appears to be more common.	**EPA Systematic Review Framework is unclear on this best practice.**				**1**		
3.3.6 Taking account of the risk of bias, consider using observational studies to address gaps in the evidence from randomized clinical trials on the benefits of interventions (p. 274). Rationale: Rather than exclude evidence where it is sparse, it might be necessary to use data from studies using design more susceptible to bias than a preferred design.	**EPA Systematic Review Framework is consistent with this best practice. Human observational studies included in search strategy.**	**1**					
		3	**2**	**0**	**1**	**0**	**0**

3.4 Document the search							
3.4.1 Provide a line-by-line description of the search strategy, including the date of search for each database, web browser, etc. (p. 274) Rationale: Appropriate documentation of the search processes ensures transparency of the methods used in the review, and appropriate peer review by information specialists.	**EPA Systematic Review Framework is unclear on this best practice; based on first 10 chemicals EPA Systematic Review Framework consistent with this best practice.**				1	1	
3.4.2 Document the disposition of each report identified, including reasons for their exclusion if appropriate (p. 275). Rationale: The standard supports creation of a flow chart that describes the sequence of events leading to identification of included studies, and it also supports assessment of the sensitivity and precision of the searches a posteriori.	**EPA Systematic Review Framework is consistent with this best practice.** **Page 25** EPA states "Each article was generally screened by two independent reviewers using specialized web-based software (i.e., DistillerSR)13. Screeners were assigned batches of references after conducing pilot testing. Screening forms facilitated the reference review process by asking a series of questions based on pre-determined eligibility criteria. **DistillerSR was used to manage the work flow of the screening process and document the eligibility decisions for each reference**. The screeners resolved conflicts by consensus, or consultation with an independent individual(s). Footnote 9 page 23 also states "In addition to using DistillerSR, EPA/OPPT is exploring automation and machine learning tools for data screening and prioritization activities (e.g., SWIFT-Review, SWIFT-Active Screener, Dragon, DocTER). SWIFT is an acronym for “Sciome Workbench for Interactive Computer-Facilitated Text-mining”. Flow Diagrams throught the first 10 draft risk evaluations fail to offer an explicit reason for why studies that	1					1
		1	0	0	1	1	1
3.5 Manage data collection							
3.5.1 At a minimum, use two or more researchers, working independently, to extract quantitative or other critical data from each study. For other types of data, one individual could extract the data while the second individual independently checks for accuracy and completeness. Establish a fair procedure for resolving discrepancies—do not simply give final decisionmaking power to the senior reviewer (p. 275). Rationale: Because reporting is often not clear or logically placed, having two independent reviewers is a quality-assurance approach. The evidence supporting two independent data extractors is limited and so some reviewers prefer that one person extracts and the other verifies, a time- saving approach. Discrepancies would be decided by discussion so that each person’s viewpoint is heard.	**EPA Systematic Review Framework is unclear on this best practice.** Table 3-1 states only to "Specify number and expertise of reviewers involved in the data extraction process." It does not specify that at a minimum two or more researchers working independently, will extract quantitative or other critical data from each stud.y				1		
3.5.2 Link publications from the same study to avoid including data from the same study more than once (p. 276). Rationale: There are numerous examples in the literature where two articles reporting the same study are thought to represent two separate studies.	**EPA Systematic Review Framework is unclear on this best practice.**				1		

3.5.3 Use standard data extraction forms developed for the specific systematic review (p. 276). Rationale: Standardized data forms are broadly applied quality assurance approaches.	**EPA Systematic Review Framework is consistent with this best practice.** Table 3-1 states that EPA will " Extract data/information using pre-defined templates." page 25 EPA/OPPT will use various extraction tools to meet the needs of each chemical assessment. These may include specialized web-based software (e.g., DistillerSR, HAWC14).footnote 14 states: EPA/OPPT is exploring HAWC for extracting data supporting TSCA risk evaluations. HAWC stands for Health Assessment Workspace Collaborative.	**1**					
3.5.4 Pilot-test the data extraction forms and process (p. 276). Rationale: Pre-testing of the data collection forms and processes are broadly applied quality assurance approaches.	**EPA Systematic Review Framework is consistent with this best practice.** Table 3-1 states that EPA will "Conduct pilot study to test the extraction process and conflict resolution strategy. Unless major changes are made, piloting may only need to be conducted once and not after each update.;	**1**					
*In case raw data are not available, this should not be considered a reason for exclusion of a study.		2	0	0	2	0	0
		Consistent with IOM	Inconsistent with IOM	Not Mentioned	Unclear	Not mentioned/unclear but apparently applied to First 10 TSCA chemicals	Not Applied in first 10 chemicals
	TOTALS	12	2	6	7	5	1