May 17, 2019

# Comments from Academics, Scientists and Clinicians on Materials Supporting the Colour Index (C. I.) Pigment Violet 29 Risk Evaluation

Submitted online via *Regulations.gov* to docket EPA-HQ-OPPT-2018-0604

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

We appreciate the opportunity to provide written comments on the updated materials for Pigment Violet 29,[1] issued under EPA's Toxic Substances Control Act (TSCA), amended by the Frank R. Lautenberg Chemical Safety for the 21st Century Act ("amended TSCA"). The law requires EPA to make decisions about chemical risks based on the "best available science" and "weight of the scientific evidence,"[2] which EPA regulation defined 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.''[3]

We previously commented in January 2019 that EPA does not have adequate information to conclude that Pigment Violet 29 does not pose an unreasonable risk.[4] Since then, EPA released more information about the 24 studies underlying the Pigment Violet 29 risk assessment and revised its evaluations of these studies. However, this new information does not change our original conclusion; rather, it strengthens our assertions that EPA's data is insufficient for risk assessment because of quality deficiencies and critical data gaps.

The new information also sheds further light on the specific ways that EPA's systematic review method developed under TSCA (hereafter referred to as the "TSCA method")[5] fails to accurately evaluate the evidence. We commented on the scientific flaws in the TSCA method previously; these comments are provided in Appendix A and summarized in a recent peer-reviewed commentary published in the

[1] 84 FR 16011

[2] 15 USC §2625 (h)-(i)

[3] 40 CFR 702.33

[4] UCSF PRHE, et al. (2019) Comments from Academics, Scientists and Clinicians on the Draft Risk Evaluation for C. I. Pigment Violet 29. Available: https://prhe.ucsf.edu/sites/g/files/tkssra341/f/wysiwyg/2019%2001%2014_PV%2029%20Risk%20Eval_UCSF%20PRHE_comments_EPA.pdf

[5] EPA (2018) Application of Systematic Review in TSCA Risk Evaluations.

*American Journal of Public Health*.[6] The Pigment Violet 29 evaluation clearly demonstrates the TSCA method's fundamental deficiencies.

Our comments address the following main issues:

1. **EPA should use a peer-reviewed, validated systematic review method for chemical evaluations instead of "Application of systematic review in TSCA risk evaluations."**
2. **The Pigment Violet 29 evaluation does not use a pre-established protocol as required by EPA regulation under TSCA.**
3. **The TSCA method is not applied consistently to evaluate studies as required by EPA regulation under TSCA.**
4. **The TSCA method does not have a pre-established protocol or methods for evidence integration as required by EPA regulation under TSCA.**
5. **The TSCA method and the Pigment Violet 29 evaluation do not follow best scientific practices for systematic reviews.**
6. **A comparison to a tool with validated domains for assessing risk of bias finds that EPA's TSCA method does not accurately evaluate study quality, and the overall evidence base of Pigment Violet 29 toxicity studies is probably high risk of bias.**
7. **EPA's conclusion that Pigment Violet 29 presents a low hazard to human health by all routes of exposure is not supported by the evidence.**

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

Sincerely,

Hanna Vesterinen, PhD
Research Consultant to UCSF

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

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

[6] 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. American Journal of Public Health. *In Press.*

Steven G. Gilbert PhD, DABT
Director, Institute of Neurotoxicology and Neurological Disorders
Department of Environmental and Occupational Health Sciences
University of Washington, Seattle School of Public Health
University of Washington, Seattle

Paolo Rinaudo, MD PhD
Professor, Department of Obstetrics, Gynecology and Reproductive Sciences
University of California, San Francisco

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

Robert M. Gould, MD
Associate Adjunct Professor, UCSF Program on Reproductive Health and the Environment
Department of Obstetrics, Gynecology and Reproductive Sciences
University of California, San Francisco
President, Physicians for Social Responsibility, San Francisco Bay Area Chapter

Victoria Leonard, RN, NP, PhD
Staff Specialist
Department of Occupational Environmental Medicine
University of California, San Francisco

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

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

Jyotsna Jagai, MS, MPH, PhD
Research Assistant Professor
University of Illinois at Chicago
Chicago, IL

Ondine von Ehrenstein, PhD, MPH, MSc
Associate Professor, University of California, Los Angeles
Los Angeles, CA

## DETAILED COMMENTS

### 1. EPA should use a peer-reviewed, validated systematic review method for evaluating study quality instead of "Application of systematic review in TSCA risk evaluations."

Our previous comments provided detailed evidence on the scientific shortcomings of the TSCA method (Appendix A). Briefly, one of the major problems is the TSCA method's inappropriate 'scoring' scheme for rating quality of studies that assigns numerical scores to various study components and then calculates an overall "quality score." The implicit assumption in such quantitative scoring methods is that we understand how much each factor used to evaluate study quality contributes to the overall quality, and that these factors are independent of each other. This is not a scientifically supportable underlying assumption, as researchers have documented that such scoring methods have unknown validity, may contain invalid items, and that results of a quality score are not predictive of the quality of studies.[7] An examination of the application of quality scores in meta-analysis found that quality-score weighting produced biased effect estimates because quality is not a singular dimension that is additive, but may be non-additive and non-linear.[8] A relevant metaphor is the saying "the whole is greater than the sum of the parts," which captures the idea that quantitative measures cannot accurately reflect some qualities. The National Academies of Sciences (NAS) recommended against use of scoring systems, concluding that "… there is no empirical basis for weighting the different criteria in the scores…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."[9]

Instead of the unscientific TSCA method, we recommend that EPA adopt and implement one of the three existing empirically-based systematic review methodologies below. Having been peer-reviewed, validated, demonstrated in case studies and recommended for chemical evaluations by the NAS,[10] these are the best available science for systematic review:

- 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.
- OHAT: 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
- Integrated Risk Information System (IRIS): 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. https://doi.org/10.17226/25086.

---

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

[8] 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):463-471. doi:10.1093/biostatistics/2.4.463.

[9] National Research Council. *Review of EPA's Integrated Risk Information System (IRIS) Process*. Washington, D.C.; 2014. Pg. 69.

[10] The National Academies of Sciences. *Application of Systematic Review Methods in an Overall Strategy for Evaluating Low-Dose Toxicity from Endocrine Active Chemicals*. Washington, D.C.: National Academies Press; 2017. doi:10.17226/24758.

While the scoring system in the TSCA method is not empirically based and should not be used, we nonetheless provide analysis in our comments below of how it has been applied in the Pigment Violet 29 evaluation to demonstrate its shortcomings.

**2. The Pigment Violet 29 evaluation does not use a pre-established protocol as required by EPA regulation under TSCA.**

EPA has not created a protocol for the Pigment Violet 29 systematic review. 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 best practice by all established systematic review methods.[11,12,13] Further, a "pre-established protocol" is required by EPA's regulation under TSCA. [14]

Thus, EPA's approach of conducting the Pigment Violet 29 review without a pre-established protocol is in clear violation of scientifically validated approaches to conducting systematic reviews. In its review of the EPA IRIS program's proposed systematic review methods, the NAS 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."[15] In the case of the Pigment Violet 29 risk assessment, EPA not only completed the literature search without a complete protocol, it completed the entire systematic review in the absence of a protocol and complete method. It is blatantly biased to write the rules of evidence assembly and interpretation at the same time one is applying the rules, and as such, EPA's review of Pigment Violet 29 cannot be validly referred to as a science-based systematic review.

A case in point are EPA's ratings for the metric "blinding of assessors" for animal toxicity studies (studies 1-13, 16-17; see Appendix B), which show serious deviations from EPA's own criteria. "Blinding of assessors" refers to the importance of ensuring that personnel involved in assessing the study animals did not know which animals were assigned to which group (i.e., which animals were in a control or treatment group), as there is significant empirical evidence that not blinding assessors can bias their evaluation.

The TSCA method's description for a serious flaw in the data source for this metric states:

> "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-

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

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

[13] Higgins J, Green S. Chapter 2: Preparing a Cochrane review. In: *Cochrane Handbook for Systematic Reviews of Interventions*. 5.1. The Cochrane Collaboration; 2011.

[14] 40 CFR 702.33

[15] 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. https://doi.org/10.17226/25086. Pg. 8

evaluations) was performed in a biased fashion (e.g., assessors of subjective outcomes were aware of study groups)."[16]

The above description appropriately indicates that blinding of assessors is critical if the study in question measured subjective outcomes. Note in this context, 'subjective' refers to outcomes with multiple gradations of possible responses, such as skin irritation, which could be minor, medium, severe or anywhere along this continuum—it is subjective because it relies on the assessor's judgement to categorize a particular animal's response. 'Objective' means there is only one interpretation of the outcome possible, such as with death, and thus no exercise of judgement is necessary.

As shown in Appendix B, all of the animal toxicity studies EPA relies on the Pigment Violet 29 assessment actually do 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, EPA previously assigned a "medium" or "low" rating to this metric in 60% of these studies, which was subsequently changed to "not rated." [17,18] 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 include this domain.[19, 20]

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 TSCA systematic review method, it should have specified this in a pre-established protocol, *prior to rating the studies.* 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 that bias the evaluation results at worst.

**3. The TSCA method is not applied consistently to evaluate studies as required by EPA regulation under TSCA.**

EPA's TSCA regulation requires that the systematic review method be applied "consistently" to each evidence stream.[21] We have serious concerns about the consistency of applying the TSCA method to evaluate study quality. We were surprised to identify that across all 15 animal toxicity studies EPA relies on for its draft risk evaluation, there were 124 changes in the rating of metrics of a possible 360 total between the initially released systematic review documents and the updated systematic review

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

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

[18] 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

[19] 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

[20] 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

[21] 40 CFR 702.33

documents. This means that 34% of the ratings for the metrics were changed, and two studies alone had 75% or greater of their 24 metrics ratings altered (see Appendix C). Moreover, 18 of these quality ratings were changed but not reported in the updated ratings document (Appendix C). Concerningly, in the initial systematic review, the default rating for a number of domains appeared to be "high" with no evidence to support such a rating, and were subsequently changed to "low" or "unacceptable" upon re-evaluation.[22, 23] It is difficult to understand how, if the criteria for the metric are clear, assessors could assign such vastly different ratings.

Another example is that different ratings are being assigned to the same metric using the same rationale. For example, metric 14 - "test animal characteristics" is supposed to rate the reporting of the test animal species, and/ or appropriateness of the test animal species for the outcome in question.[24] For study #9, which reported the below on the test animals, EPA gave a final rating of 2 (medium) stating that "Health status and age at initiation were not reported."[25]

- Test animals: Sprague-Dawley rats, male and female
- Average weights at study initiation: males: 250 g, females: 160 g[26]

However, for study #5, which reported the below almost identical information on test animals, EPA gave a final rating of 3 (low), stating that "Study provided minimal information on the test animal characteristics (e.g., strain, health status, age)."[27]

- Test animals: Rats, male and female
- Average weights at study initiation: 182 g[28]

Neither study reported on health status or age at initiation, yet the first was rated "2" and the second "3." This is an inconsistent application of the method and speaks to the lack of clarity in EPA's criteria as its TSCA method has never been applied, validated, peer-reviewed or used by scientists.

Another example of inconsistency is in study 17; EPA rated 18 metrics as "high," even though the authors provide no rationale for their ratings other than "This metric met the criteria for high confidence as expected for this type of study." This rationale is not adequate because it is unclear *why* the rater considered the metric to have met the criteria. In fact, we found discrepancies in our evaluation of this study. For example, for question 8, "Consistency of exposure administration," EPA rated this domain as "high" per its TSCA method which states:

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

[23] 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

[24] EPA (2018) Application of Systematic Review in TSCA Risk Evaluations. Pg. 187

[25] 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. Pg. 15

[26] EPA (2019) BASF Summary of Toxicological Investigations with CAS 81-33-4.

[27] 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. Pg. 21

[28] EPA (2019) Study #s 1, 2, 5-10, 12, 13: Toxicological investigation summaries, non-confidential. Pg. 2

> "Were exposures administered consistently across study groups (e.g., same exposure frequency; same time of day; consistent gavage volumes or diet compositions in oral studies; consistent chamber designs, animals/chamber, and comparable particle size characteristics in inhalation studies; consistent application methods and volumes in dermal studies)?"[29]

However, we did not identify any information in the report regarding the timing of the dose across study groups, which would appear to warrant a lower rating than "high," according to the TSCA method.

In addition, for question 11 "Number of exposure groups and dose spacing," the report states that the doses were selected at the request of the sponsor. The TSCA method criteria for this metric states:

> "Were the number of exposure groups and dose/concentration spacing justified by study authors (e.g., based on range-finding studies) and adequate to address the purpose of the study (e.g., to evaluate dose-response relationships, identify points of departure, inform MOA/AOP, etc.)?"[30]

Having the doses specified by the sponsor does not appear to meet this criterion, but yet again, EPA rated this study "high" quality for this domain.

Overall, these problems are further evidence of the fundamental flaws in the TSCA method. They could also reflect other biases that are imposed on the evaluation of the studies to achieve a desired outcome, rather than following the science; this further highlights the importance of having pre-specified protocol for the systematic review.

**4. The TSCA method does not have a pre-established protocol or methods for evidence integration as required by EPA regulation under TSCA.**

EPA's TSCA regulation governing procedures for chemical risk evaluations requires that it use a systematic review method to "integrate evidence,"[31] but EPA's TSCA method does not address this step, nor does EPA's Pigment Violet 29 risk evaluation.

The ad hoc and incomplete nature of EPA's TSCA method is incompatible with science-based methods of systematic review developed, endorsed, and/or advanced by the: National Academy of Sciences;[32] the Institute of Medicine;[33] the National Toxicology Program;[34] the Cochrane Collaboration;[35] the Grading of

[29] EPA (2018) Application of Systematic Review in TSCA Risk Evaluations. Pg. 193
[30] Id. pg. 196
[31] 40 CFR 702.33
[32] 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
[33] Institute of Medicine. Finding What Works in Health Care. Standards for Systematic Review. Washington, D.C.: The National Academies Press.; 2011
[34] 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
[35] 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.

Recommendations Assessment, Development and Evaluation (GRADE) method;[36] the international scientific collaboration that developed a framework for the "systematic review and integrated assessment" (SYRINA) of endocrine disrupting chemicals;[37] the SYRCLE systematic review method for animal studies;[38] the Campbell Collaboration's methods;[39] and the Navigation Guide systematic review method developed by a collaboration of scientists led by the University of California, San Francisco.[40]

Evidence integration consists of, at minimum, qualitatively rating the confidence in the overall body of evidence for a specific outcome, translating that confidence rating into a conclusion on the level of evidence for a health effect, and then developing a hazard identification conclusion. Where available, animal and human evidence would be integrated, and mechanistic data would be used to inform the final conclusion. Examples from the OHAT method of the translation and hazard identification steps are below.

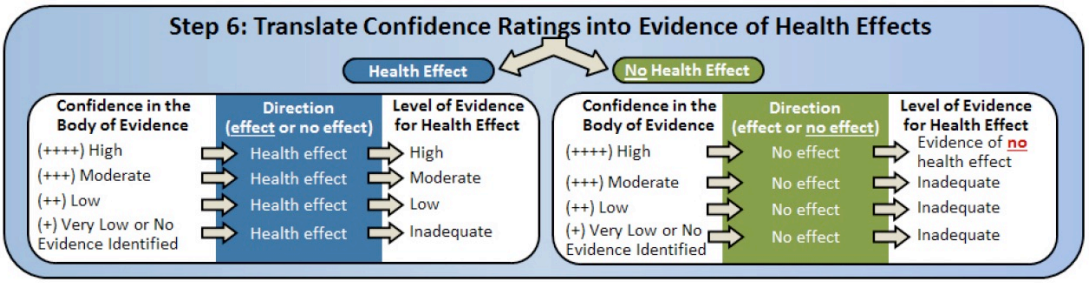



*Figure 1: OHAT's process to translate confidence in the body of evidence to come to a conclusion on the level of evidence for a health effect.[41] This step is missing from the TSCA method.*

[36] 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

[37] 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

[38] 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.

[39] 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

[40] 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

[41] 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. Pg. 64

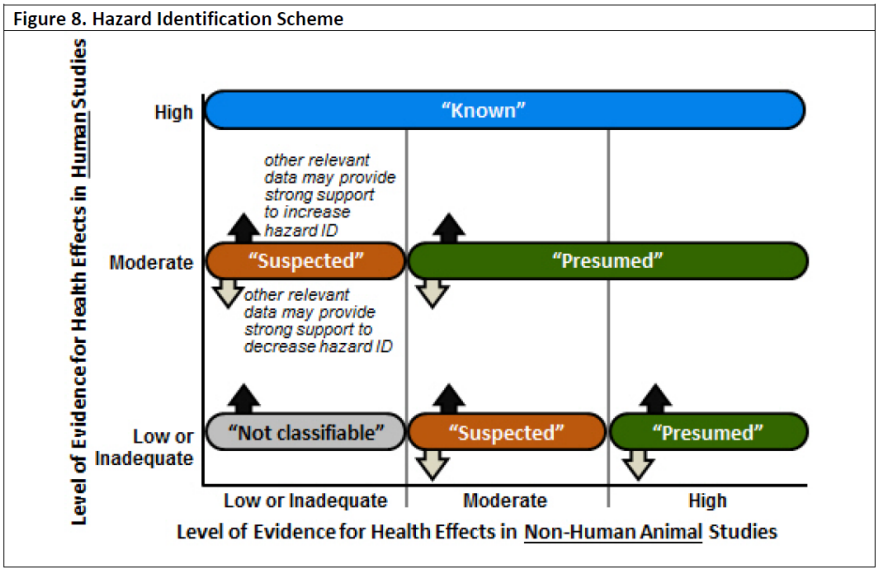



*Figure 2: OHAT's process to translate the level of evidence for a health effect into a hazard identification conclusion.*[42] *This step is missing from the TSCA method.*

EPA does not rate the confidence in the body of evidence on Pigment Violet 29, nor does it follow a proper evidence integration protocol to come to its final conclusion that Pigment Violet 29 does not pose an unreasonable risk. Therefore, it is unclear how EPA translated the available evidence into its final conclusion.

**5. The TSCA method and the Pigment Violet 29 evaluation do not follow best scientific practices for systematic reviews.**

All established systematic review methods require two independent reviewers to rate the risk of bias of individual studies, with discrepancies to be resolved by a third reviewer. In contrast to this, the TSCA method specifically states that only one reviewer may be used.[43] EPA does not state how many reviewers were involved in the data quality evaluation for Pigment Violet 29. We presume that in 2018 there was only one reviewer because it seems unlikely that the myriad errors in scoring results across many domains would be the result of two individual reviewers.[44] However, we see no evidence that there was more than one reviewer involved in the re-evaluation in 2019.[45] If there was more than one reviewer involved, we recommend stating this; and if not, we would urge the EPA to assign at least one additional reviewer in future reviews to avoid repeating a similar scenario.

[42] Id. pg. 67

[43] EPA (2018) Application of Systematic Review in TSCA Risk Evaluations. Pg. 26 "Ideally, each data/information source will be screened by two reviewers but one reviewer may be used."

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

[45] 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

**6. A comparison to a tool with validated domains for assessing risk of bias finds that EPA's TSCA method does not accurately evaluate study quality, and the overall evidence base of toxicity studies is low quality.**

We have conducted an initial evaluation of risk of bias for the 15 Pigment Violet 29 animal toxicity studies by applying the Navigation Guide risk of bias tool (see Figure 3 below). This evaluation indicates that the overall quality of the body of evidence is poor.

The Navigation Guide is a systematic review method developed by an international, interdisciplinary collaboration of clinical and environmental health scientists from academia, government and NGOs.[46] It has been applied in multiple case studies and the National Academy of Sciences has evaluated and recommended both the Navigation Guide and OHAT methods for systematic reviews, noting that:

> "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."[47]

Currently, the World Health Organization is using Navigation Guide methodology to conduct an analysis of the global burden of work-related injury and disease.[48]

Risk of bias means characteristics of a study that can introduce a systematic error in the magnitude or direction of the results of the study, thus 'biasing' the results away from the true result.[49] 'Blinding of assessors' is an example of a characteristic that can bias study results, discussed in point 2 above. If assessors are not blinded to treatment groups, they may expect to see differences in the animals treated with the chemical—and evidence shows that this can bias their judgement.

The Navigation Guide risk of bias assessment for animal toxicity studies was developed based on the Cochrane Collaboration's Tool for Assessing Risk of Bias,[50] which includes domains that have been empirically demonstrated to have a material effect on study outcomes, outlined in the table below.

---

[46] 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

[47] The National Academies of Sciences. *Application of Systematic Review Methods in an Overall Strategy for Evaluating Low-Dose Toxicity from Endocrine Active Chemicals*. Washington, D.C.: National Academies Press; 2017. doi:10.17226/24758. Pg. 119

[48] Mandrioli, D., Schlünssen, V., Ádám, B., Cohen, R. A., Colosio, C., Chen, W., … Scheepers, P. T. J. (2018, October 1). 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, pp. 174–185. https://doi.org/10.1016/j.envint.2018.06.005

[49] Higgins JPT, Altman DJ, Sterne JAC, eds. 2011. Chapter 8: Assessing risk of bias in included studies. In: Cochrane Handbook for Systematic Reviews of Interventions, Version 5.1.0 (Higgins JPT, Green S, eds).

[50] Id.

| Domain | Criteria for low risk of bias rating |
|---|---|
| Sequence generation | Study authors reported the use of a random component in the sequence generation process. |
| Allocation concealment | Study authors reported that study personnel could not foresee which animals were allocated to the various experimental groups. |
| Blinding | Study authors reported that personnel and outcome assessors were adequately prevented from knowledge of the allocated exposures during the study. |
| Incomplete outcome data | Study authors reported when and why participants left the study. |
| Selective reporting | The study's prespecified outcomes that are of interest in the review were reported in a prespecified way. |
| Conflict of interest | The study was free of support from a company, study author, or other entity having a financial interest in the exposures of interest in the review. |
| Other bias | Study appears to be free of other sources of bias. |

*Table 1: Navigation Guide tool for assessing risk of bias of animal toxicity studies contains seven domains empirically shown to affect bias.* [51]

Figure 3 shows the results of our initial evaluation of the Pigment Violet 29 animal toxicity studies with the Navigation Guide risk of bias tool, finding that the overall evidence base appears to be of poor quality.

[51] 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

| | | Navigation Guide Risk of Bias Domains | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Study Number** | **Study Name** | **Sequence generation** | **Allocation concealment** | **Blinding** | **Incomplete outcome data** | **Selective outcome reporting** | **Conflict of interest** | **Other bias** |
| #9 | BASF 1975. XXV/454. Acute oral | | | | | | | |
| #10 | BASF 1978. 77/360. Acute oral toxicity | | | | | | | |
| #5 | BASF 1975. XXC/454. Acute Inhalation | | | | | | | |
| #6 | BASF 1978. 77/360. Acute inhalation toxicity in rats | | | | | | | |
| #7 | BASF 1975. XXV/454. Acute intraperitoneal toxicity in mice | | | | | | | |
| #8 | BASF 1978. 77/360. Acute intraperitoneal toxicity in mice | | | | | | | |
| #17 | Stark 2013. Repro/ dev. Tox screening in Wistar rats (oral gavage) | | | | | | | |
| #1 | BASF 1975. XXV/454. Eye Irritation Study | | | | | | | |
| #2 | BASF 1978. Eye irritation study. | | | | | | | |
| #11 | Rupprich & Weigand 1984. Acute oral toxicity in female wistar rat | | | | | | | |
| #12 | BASF 1975 XV/454. Skin Irritation in 2 rabbits | | | | | | | |
| #13 | BASF 1978 77/360. Skin Irritation in 3 rabbits | | | | | | | |
| #3 | Rupprich & Weigand 1984. Skin irritation in rabbits | | | | | | | |
| #4 | Rupprich & Weigand 1984. Eye irritation in rabbits | | | | | | | |

| Legend | |
|---|---|
| | Low Risk |
| | Probably Low |
| | Probably High |
| | High Risk |
| | Not Applicable |

*Figure 3: Initial evaluation of Pigment Violet 29 animal toxicity studies with the Navigation Guide risk of bias assessment, finding the evidence base is overall probably high risk of bias.*

For the majority of studies, we only had a half-page report upon which to base our ratings; without information to the contrary, we can only assume this is the same information the EPA had to base their ratings. It is not clear who prepared the reports or if they accurately reflect the study data and findings. Moreover, these studies were subject to a potential conflict of interest, often with no control group, and with no indication that study authors randomized or blinded their assessment of outcome.

As we described in detail in our previous comments (Appendix A), the TSCA method's scoring system wrongly conflates reporting in a study with risk of bias in a study, with low scores given for less reporting and high scores given for more reporting. Briefly, 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. Numerous studies document that how completely and clearly a study is reported is not a scientifically valid measure of the quality of the underlying research.[52, 53, 54, 55] Yet, EPA's metrics for scoring study quality explicitly encompass reporting (see Appendix C).

In comparison to a validated risk of bias tool like Navigation Guide, the metrics evaluated by the TSCA method include many elements that are not related to bias (i.e., reporting), and as such, "high" ratings under the TSCA system are not a meaningful reflection of study quality. The TSCA method rates most of the Pigment Violet 29 animal toxicity studies as of "Medium" or "High" quality (see Appendix C), in contrast to the Navigation Guide assessment where the studies generally appear to be "probably high risk of bias."

Based on validated risk of bias criteria, the evidence base of toxicity studies for Pigment Violet 29 is low quality.

**7. EPA's conclusion that Pigment Violet 29 presents a low hazard to human health by all routes of exposure is not supported by the evidence.**

EPA's draft conclusion that Pigment Violet 29 presents a low hazard to human health[56] is not supported by the evidence for a number of reasons. We commented in detail previously about the lack of adequate information to draw conclusions about cancer, reproductive and developmental toxicity hazards.[57] Below, we detail additional concerns raised by the new information EPA has released.

---

[52] 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.

[53] 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

[54] 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.

[55] 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.

[56] US EPA (2018) Draft Risk Evaluation for C.I. Pigment Violet 29 (Anthra[2,1,9-def:6,5,10-d'e'f']diisoquinoline-1,3,8,10(2H,9H)-tetrone) Pg. 25

[57] UCSF PRHE, et al. (2019) Comments from Academics, Scientists and Clinicians on the Draft Risk Evaluation for C. I. Pigment Violet 29. Available: https://prhe.ucsf.edu/sites/g/files/tkssra341/f/wysiwyg/2019%2001%2014_PV%2029%20Risk%20Eval_UCSF%20PRHE_comments_EPA.pdf

Firstly, EPA does not have a reliable inhalation toxicity study even though inhalation is expected to be a main exposure pathway for workers.[58] Studies 5 and 6 are the only two inhalation studies; both were conducted over four decades ago and EPA has newly downgraded these from high quality to unacceptable; these were also generally rated as "probably high risk of bias" in our analysis (Figure 3).

Whether the studies are rated as unacceptable or overall low quality, in either case this results in a poor body of evidence to assess inhalation toxicity. Additionally, there are concerns about the quality of the evidence on toxicity via the intraperitoneal and oral routes—which is far from the robust evidence needed to conclude that Pigment Violet 29 is not toxic, as detailed in our previous comments.[59]

Secondly, these conclusions were not derived from peer-reviewed studies that were free of bias. The majority of studies were industry-sponsored, which is a proven source of bias.[60, 61] Indeed, because it is a documented source of bias, the NAS has recommended that "Funding sources should be considered in the risk-of-bias assessment conducted for systematic reviews…"[62] Moreover, only one study reported randomizing to group, none reported being blinded and the majority did not report a control group at all, let alone a suitable concurrent control.

Additionally, we have concerns over the sample sizes used in these studies. EPA has specifically addressed the issue of sample size in question 15 for animal studies ("Numbers per group").[63] In the 9 animal studies we assessed, the maximum group size was 10 (study 17). The others had a typical group size of 5 per sex, and one had as few as 3 animals in the control group; however, we also noted several studies did not report a control group at all. With such small group sizes these studies simply lack the power to draw robust conclusions.

Lastly, the majority of Pigment Violet 29 animal toxicity studies are many decades old. In general, these studies do not use the most current scientific methods, instruments and measures to assess the most sensitive and relevant health endpoints that have been identified by modern biology. Further, specific definitions and guidelines have evolved based on the science—for example, for studies #3 and #4, "irritation" is defined as lasting more than 24 hours. But newer guidelines specify that irritation is reversible damage observed after application of a chemical, with application time up to four hours.[64] By this definition, the animals clearly suffered irritation (erythema, edema), calling into question Pigment Violet 29's classification as a non-irritant.

---

[58] US EPA (2018) Draft Risk Evaluation for C.I. Pigment Violet 29 (Anthra[2,1,9-def:6,5,10-d'e'f']diisoquinoline-1,3,8,10(2H,9H)-tetrone) Pg. 22

[59] UCSF PRHE, et al. (2019) Comments from Academics, Scientists and Clinicians on the Draft Risk Evaluation for C. I. Pigment Violet 29.

[60] 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.

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

[62] National Research Council. *Review of EPA's Integrated Risk Information System (IRIS) Process*. Washington, D.C.; 2014. Pg. 79

[63] EPA (2018) Application of Systematic Review in TSCA Risk Evaluations. Pg. 198

[64] OECD (2015) OECD Guideline for Testing of Chemicals 404: Acute Dermal Irritation/ Corrosion. Pg. 8. Available: https://read.oecd-ilibrary.org/environment/test-no-404-acute-dermal-irritation-corrosion_9789264242678-en#page1

Recently the Belgian Competent Authority released a new evaluation of Pigment Violet 29 as part of the Community Rolling Action Plan under REACH.[65] The review finds that current data indicate a concern for potential persistence and bioaccumulation, and indicates that additional data on toxicity, fate, environmental behavior, and physical-chemical properties is needed.

Because the data on Pigment Violet 29 do not constitute adequate information for risk assessment, EPA should request needed data using TSCA authorities and complete a new risk assessment using an established, validated systematic review method once it has additional data.

[65] Belgian Competent Authority (Be CA) (2019) Justification Document for the Selection of a CoRAP Substance. Group Name: Diisoquinoline tetrones. Available: https://echa.europa.eu/documents/10162/387374b8-62fa-c857-e60f-65e1cd9fd821

## Appendices:

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

**Appendix B:** Detailed Ratings for EPA Metric “Blinding of Assessors”

**Appendix C:** Analysis of EPA Ratings Changes

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

August 16, 2018

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

Submitted online via *Regulations.gov* to docket EPA-HQ-OPPT-2018-0210

These comments are submitted on behalf of the undersigned academic, scientists, and clinicians. 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.

We appreciate the opportunity to provide written comments on the Application of Systematic Review in TSCA Risk Evaluations,[a] pursuant to the Toxic Substances Control Act (TSCA), as amended by the Frank R. Lautenberg Chemical Safety of the 21st Century Act (Lautenberg TSCA). TSCA requires that EPA make decisions about chemical risks based on the "best available science" and the "weight of the scientific evidence"[b] which EPA defined 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.''[c]

Systematic review methods originated more than 40 years ago in psychology. The methodology was soon adapted to evaluating the effectiveness of clinical interventions in medicine and related disciplines in response to empirical evidence demonstrating the need to apply scientific principles not only to primary research, but also to research synthesis methods that inform decision-making in healthcare (1-3). Almost a decade ago, these empirically-proven methods for research synthesis were adapted to environmental health (4, 5). To date, science-based methods for systematic review in environmental health have been demonstrated in case studies in the peer-reviewed literature (6-13), and adopted by the National Toxicology Program (14) and the U.S. EPA's Integrated Risk Information System (IRIS) program (15).

EPA's systematic review framework under TSCA 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 general 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.

[a] 83 FR 26998, June 11, 2018
[b] 15 USC §2625 (h)-(i)
[c] 40 CFR 704.33

With so much at stake, we are deeply concerned by EPA's ad hoc and incomplete TSCA systematic review framework, which is inconsistent with current, established, best available empirical methods for systematic review. Moreover, as we detail below, the application of EPA's TSCA framework would likely result in the exclusion of quality research from EPA's decision-making. Accordingly, the TSCA systematic review method does not meet the mandate of the law to use the "best available science." [d]

Based on the most current empirically demonstrated principles of systematic review methods, we provide EPA with concrete recommendations and approaches to correct its methodology and inform timely science-based decision-making to achieve the Agency's mission of protecting the public from harmful chemicals.

**Our comments address the following six main points:**

1. **EPA's TSCA systematic review framework is ad hoc, 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[e] definition of a systematic review, including but not limited to, using explicit and pre-specified scientific methods for every step of the review. EPA should consider methods demonstrated for use in environmental health, and which have been endorsed and utilized by the National Academy of Sciences, i.e., the National Toxicology's Office of Heath Assessment and Translation systematic review method, and the Navigation Guide Systematic Review Method. EPA's TSCA systematic review framework should be peer-reviewed by qualified external experts in the field.

2. **EPA's TSCA systematic review framework utilizes a quantitative scoring method that is incompatible with the best available science in fundamental ways:**

   a. **Quantitative scores for assessing the quality of an individual study are arbitrary and not science-based; the Cochrane Collaboration and National Academy of Sciences 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.**

   **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.

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

[d] 15 USC §2625 (h)

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

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

4. **The literature review step of EPA's TSCA systematic review framework 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 make its framework for conducting a literature review congruent with all of the Institute of Medicine's best practices and explicitly include rules for when the list of relevant studies will be considered final.

5. **EPA's TSCA systematic review framework correctly recognizes that mechanistic data are not required for a hazard assessment, but EPA is not clear that these data, if available, can only be used to increase, and not to decrease, confidence in a body of evidence.**

   **We recommend:** EPA should be explicit that mechanistic data can only be used to upgrade a hazard classification, or increase the confidence of a finding made based on evaluation of animal and human data, and that these data will not be used to decrease confidence in a body of evidence.

6. **EPA's TSCA systematic review framework is not independent of the regulatory end user of the review.**

   **We recommend**: EPA's TSCA systematic reviews should be produced independently of the regulatory end user of the review.

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

Sincerely,

Veena Singla, PhD
Associate Director, Science and Policy, Program on Reproductive Health and the Environment
University of California, San Francisco

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

Tracey Woodruff, PhD, MPH
Director, Program on Reproductive Health and the Environment
University of California, San Francisco

Juleen Lam, PhD, MHS, MS
Assistant Professor, Department of Health Sciences
California State University, East Bay

Patricia D. Koman, PhD, MPP
President and Senior Health Scientist
Green Barn Research Associates*
Ann Arbor, Michigan

Lisa Bero, PhD
Chair of Medicines Use and Health Outcomes, Charles Perkins Centre
The University of Sydney

Liz Borkowski, MPH
Senior Research Scientist, Milken Institute School of Public Health
George Washington University

Sheila Brear, BDS
Associate Dean, Academic Affairs, School of Dentistry
University of California, San Francisco

Adelita G. Cantu, PhD, RN
Associate Professor
Alliance of Nurses for Healthy Environments

Courtney Carignan, PhD
Assistant Professor
Michigan State University

Daniel M. Fox, PhD
President Emeritus
Milbank Memorial Fund

Danielle Fries, MPH
Science Associate, Program on Reproductive Health and the Environment
University of California, San Francisco

Mary Gant, MS
Retired Policy Analyst
National Institute of Environmental Health Sciences

Steven G. Gilbert, PhD, DABT
Affiliate Professor
University of Washington

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

Maeve Howett, PhD, APRN, CPNP, IBCLC, CNE
Clinical Professor and Assistant Dean

University of Massachusetts Amherst

Jyotsna Jagai, MS, MPH, PhD
Research Assistant Professor, School of Public Health
University of Illinois at Chicago

Paula I. Johnson, PhD, MPH
Research Scientist, Safe Cosmetics Program
California Department of Public Health

Jean-Marie Kauth, PhD, MPH
Professor
Benedictine University

Carol Kwiatkowski, PhD
Executive Director
The Endocrine Disruption Exchange*

Joseph Laakso, PhD
Director, Science Policy
Endocrine Society*

Gail Lee, RD, REHS Hem
Sustainability Director
University of California, San Francisco

Michael J. Martin, MD, MPH, MBA
Associate Clinical Professor
University of California, San Francisco

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

Katherine Pelch, PhD
Senior Scientist
The Endocrine Disruption Exchange

Janet Pelrman, MD, MPH
Physician
Stanford Children's Hospital

Jeanne Rizzo, RN
President & CEO
Breast Cancer Prevention Partners

Ted Schettler, MD, MPH
Science Director

Science and Environmental Health Network

Rachel M. Shaffer, MPH
PhD Candidate, School of Public Health
University of Washington, Seattle

Laura N. Vandenberg, PhD
Associate Professor
University of Massachusetts, Amherst

Ellen M. Wells, PhD
Assistant Professor of Environmental & Occupational Health
Purdue University School of Health Sciences

Nsedu Obot Witherspoon, MPH
Executive Director
Children's Environmental Health Network

Marya Zlatnik, MD, MMS
Professor, Department of Obstetrics, Gynecology & Reproductive Sciences
University of California, San Francisco

*indicates organizational support

DETAILED COMMENTS

**1. EPA's TSCA systematic review framework is ad hoc, 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 (SR) specifies that all components of a review be established in a publically available protocol written *prior* to conducting the review to minimize bias and to ensure transparency in decision-making. For example, the Institute of Medicine defines a systematic review as a "scientific investigation that focuses on a specific question and uses explicit, prespecified scientific methods to identify, select, assess, and summarize the findings of similar but separate studies" (emphasis added) (16)(p.1). A fatal flaw in EPA's SR framework is that it lacks essential SR elements, including but not limited to: (1) a protocol for executing a SR developed *prior* to conducting the SR; (2) an explicit method for evaluating the overall body of each evidence stream, i.e., animal, human, etc.; and (3) 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 EPA.

Notably, EPA's TSCA SR Framework presents a diagram of a complete SR 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).

However, EPA's TSCA SR Framework then proceeds to describe an ad hoc and highly flawed method limited to only the data collection and, to a limited extent, the data evaluation components of a SR. Specifically, Figure S-1 below, excerpted from the National Academy of Sciences 2014 review of the EPA IRIS program's systematic review method (17), presents all of the components of a science-based SR. The red box indicates the parts of a SR method that EPA has included in its proposed framework.

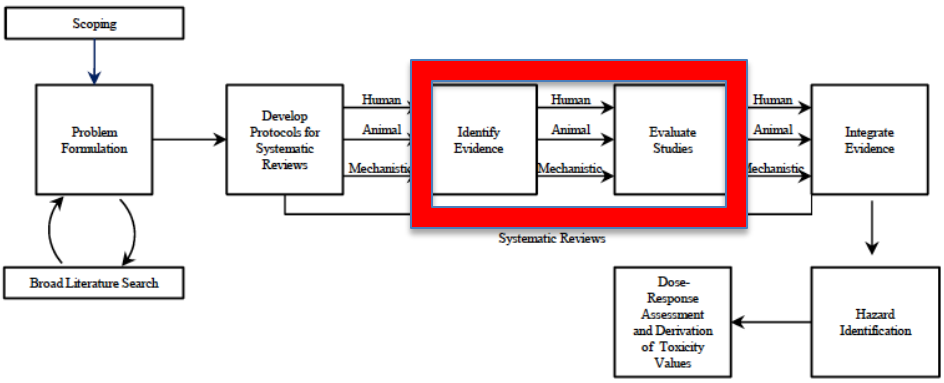



**FIGURE S-1** Systematic review in the context of the IRIS process. The committee views public input and peer review as integral parts of the IRIS process, although they are not specifically noted in the figure.

EPA's piecemeal approach is not only in direct contradiction with the best available scientific methods for SR, but also incompatible with the regulatory definition of[f] "weight of evidence" in the risk evaluation rule, which specifies a complete method spelled out in a protocol developed *before* conducting the review. Therefore, the TSCA systematic review method violates both TSCA statute and regulation. [g]

EPA explicitly states that it is proceeding with its first ten risk assessments in the absence of a pre-defined protocol and a complete method for systematic review. Specifically, EPA's SR Framework states:

> (p. 9) … the purpose of the document is internal guidance that … sets out general principles to guide EPA's application of systematic review in the risk evaluation process for the first ten chemicals … **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" (emphasis added).** Additional details on the approach for the evidence synthesis and integration will be included with the publication of the draft TSCA risk evaluations.

In effect, EPA is saying it does not have time to comply with its regulatory requirement to conduct a science-based systematic review, and will not actually develop its protocol until it completes the first ten systematic reviews.

First, this approach is in clear violation with scientifically-validated approaches to conducting systematic reviews. In its review of the EPA's Integrated Risk Information System (IRIS) program's proposed SR methods, the National Academy of Sciences specified that, "Completing the literature search as part of

[f] EPA's risk evaluation rule (40 CFR 704.33) 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.''

[g] 15 USC §2625 (h)-(i) and 40 CFR 704.33

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" (15)(Pg. 8). In the case of TSCA risk assessments, EPA is not only completing the literature search as part of protocol development, it is completing the entire systematic review in the absence of a protocol and complete method. It is blatantly biased to write the rules of evidence assembly and interpretation at the same time one is applying the rules, and as such, this method cannot be validly referred to as a science-based systematic review.

Second, a lack of time is not a credible rationale for EPA's failure to conduct a science-based systematic review for the first ten TSCA chemicals. There are multiple well-developed, science-based, peer-reviewed and validated methods for conducting systematic reviews in environmental health that EPA could readily apply, including the SR method and handbook developed by the Office of Health Assessment and Translation at the National Toxicology Program (14), and the Navigation Guide Systematic Review Method, which has been demonstrated in six case studies (6-13). The National Academy of Sciences cited both of these SR methods as exemplary of the type of methods EPA should use in hazard and risk assessment (17, 18). Further, the National Academy of Sciences utilized both methods in its 2017 assessment of the potential health impacts of endocrine active environmental chemicals (19). Specifically, in its 2017 review the National Academy of Sciences 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" (19)(page 119).

Protocols developed for applying the Navigation Guide and the OHAT method have been published and can serve as a template to further expedite EPA's TSCA reviews.[h]

Furthermore, the language of EPA's systematic review framework is confusing, contradictory, and poorly and incorrectly referenced with little science or policy foundation. This suggests the authors of EPA's TSCA Systematic Review Framework lack sufficient understanding of the scientific process integral to this work. A particularly egregious example is EPA's stated understanding of EPA's TSCA statutory science standards:

> (Pg. 26) EPA/OPPT is required by TSCA to use the weight of the scientific evidence in TSCA risk evaluations. Application of weight of evidence analysis is an integrative and interpretive process that considers both 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.

This directly contradicts EPA's own published rule which defines what a systematic review is (see

[h] 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

footnote “e”, above) and such an understanding completely subverts the purpose of a systematic review which is to explicitly avoid a simplistic analysis that would led to erroneous conclusions along the lines of stating that, for instance, “five studies are in favor (positive) and ten are against (negative) and therefore the weight is … ”

Another bewildering statement by EPA concerns its highly quantitative scoring method, which is the main topic of its systematic review framework (see comment #2, below). EPA adds a caveat to the scoring method that says quantitative scoring is actually a qualitative method, and further: “The [scoring] system is not intended to imply precision and/or accuracy of the scoring results” (Pg. 35).

The ad hoc and incomplete nature of EPA’s systematic review framework is incompatible in many additional fundamental ways, described further in detail below, with science based methods of systematic review developed, endorsed, and/or advanced by the: National Academy of Sciences (17-19); the Institute of Medicine (16); the National Toxicology Program (14); the Cochrane Collaboration (20); the Grading of Recommendations Assessment, Development and Evaluation (GRADE) method (21, 22); the international scientific collaboration that developed a framework for the “systematic review and integrated assessment” (SYRINA) of endocrine disrupting chemicals (23); the SYRCLE systematic review method for animal studies (24); the Campbell Collaboration’s methods (25); and the Navigation Guide systematic review method developed by a collaboration of scientists led by the University of California San Francisco (4). Most of these organizations also pre-publish their protocols either online (i.e., the National Toxicology Program) or in PROSPERO[i] (i.e., UCSF).

**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. EPA should consider methods demonstrated for use in environmental health, and which have been endorsed and utilized by the National Academy of Sciences, i.e., the National Toxicology’s Office of Heath Assessment and Translation systematic review method, and the Navigation Guide Systematic Review Method. EPA’s TSCA systematic review framework should be peer-reviewed by qualified external experts in the field.

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

**2. EPA's TSCA systematic review framework utilizes a quantitative scoring method that is incompatible with the best available science in fundamental ways:**

- **a. Quantitative scores for assessing the quality of an individual study are arbitrary and not science-based; the Cochrane Collaboration and National Academy of Sciences 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 detailed explanation of each of these scientific shortcomings is provided below.

**(a) Quantitative scores for assessing the quality of an individual study are arbitrary and not science-based.**

EPA's SR framework employs a quantitative scoring method to assess the quality of individual studies, 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:[j]

> (Pg. 33) A numerical scoring method is used to convert the confidence level for each metric into the overall quality level for the data/information source. The overall study score is equated to an overall quality level (*High*, *Medium*, or *Low*) using the level definitions and scoring scale shown in Table A-1. The scoring scale was obtained by calculating the difference between the highest possible score of 3 and the lowest possible score of 1 (i.e., 3-1= 2) and dividing into three equal parts (2 ÷ 3 = 0.67). This results in a range of approximately 0.7 for each overall data quality level, which was used to estimate the transition points (cut-off values) in the scale between *High* and *Medium* scores, and *Medium* and *Low* scores. These transition points between the ranges of 1 and 3 were calculated as follows: Cut-off values between *High* and *Medium*: 1 + 0.67= 1.67, rounded up to 1.7 (scores lower than 1.7 will be assigned an overall quality level of *High*) Cut-off values between *Medium* and *Low*: 1.67 + 0.67= 2.34, rounded up to 2.3 (scores between 1.7 and lower than 2.3 will be assigned an overall quality level of *Medium)*

<u>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</u>.[k]

Illustrative of the scoring method, in Appendix H "Data Quality Criteria for Epidemiologic Studies," (page

[j] 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.

[k] 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." (Pg. 96).

225) EPA presents how scoring is further 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.

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 "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 quality score are not scientifically meaningful or predictive of the quality of studies (26-28). 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 (29).

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 (21, 30). Scoring is not utilized by empirically based systematic review methodologies, such as the Cochrane Collaboration or GRADE (21, 31). As stated by the Institute of Medicine, "… systematic review teams have moved away from scoring systems to assess the quality of individual studies toward a focus on the components of quality and risk of bias" (16).

The Cochrane Collaboration, founded in 1993, is an international non-profit and independent organization that produces and disseminates systematic reviews of healthcare interventions and is a key locus of the world's most authoritative expertise on systematic review methods. Cochrane's methodology states: "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** **(emphasis added**)"(31).

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

> … Cochrane discourages using a numerical scale because 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) (17)(Pg. 69).

**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.[l]

EPA's SR Framework uses reporting measures in its scoring of the quality of human studies, including incorporating reporting guidelines into the reasons for scoring studies "low quality" (Metrics 1 and 15) or "unacceptable for use" (Metrics 2, 3, 4, 6, 7). EPA's SR Framework acknowledges that reporting is not the same as an underlying flaw in study methodology (Pg. 31), but then proceeds to ignore this distinction by using reporting as a measure of the quality of the underlying research. EPA's SR Framework not only does not "untangle" reporting from quality, it specifically conflates the two by using metrics in the STROBE reporting guidelines to score individual studies. The authors of the STROBE guidelines specifically note the guidelines are not 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 (emphasis added). … 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."(32)

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

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

underlying research (20, 21, 33, 34). As GRADE methodologists have succinctly stated, "… just because a safeguard against bias is not reported does not mean it was neglected"(21). 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 (29).

The Cochrane Collaboration Handbook for conducting a SR clearly distinguishes reporting and bias, the latter which is defined as "a systematic error, or deviation from the truth, in results or inferences" (20). 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. 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 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 SR Framework, studies can be scored as "low quality," and even 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 (35), 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 (32, 36, 37). 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."

Given the historical and present-day deficiencies in how studies are reported in the peer-reviewed literature, and because EPA's scoring system rates as 'unacceptable for use' any human study that does not report even one of five reporting metrics, EPA's proposal could reasonably be expected to lead to the exclusion from EPA's consideration much of the existing body of knowledge on the impact of

environmental chemicals on human health, and is inconsistent with TSCA mandates to use the "best available science" and "reasonably available information." [m] Applying flawed exclusion criteria that directly contradicts widely accepted empirically based SR methodological approaches will almost certainly result in flawed conclusions and threaten the protection of the public's health.

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

In the "fatal flaw" component of EPA's SR Framework's scoring system, for each type of evidence stream, i.e., epidemiologic, animal, *in vitro*, etc., EPA created an arbitrary list of metrics that make studies "unacceptable for use in the hazard assessment," stating:

> 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 (emphasis added). An *Unacceptable* score means that serious flaws are noted in the domain metric that consequently make the data unusable (or invalid) (Pg. 227).

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

Illustrative of this "fatal flaw" aspect of EPA' scoring system, for human epidemiologic studies (See Section H.5, Table H-8 (page 231), EPA lists six domains of study quality, i.e., study participation; exposure characterization; outcome assessment; potential confounding/variable control; analysis**;** and other considerations for biomarker selection and measurement, and 19 metrics to assess the six domains. A study that has even one of the 19 "serious flaws" metrics is considered to be "unacceptable for use."

The underlying assumptions of EPA's "serious flaws" metrics are not science-based because:

- **EPA's list of "serious flaws" are not all equal indicators of study quality**:
  For example, among human observational studies, any one of the list of 19 metrics can eliminate a study from consideration as EPA considers all of these "flaws" to be of equal import; as described in detail above, such weighting is arbitrary and not a science-based method.

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

  - **Reporting** guidelines are wrongly equated with "serious flaws" in study quality.
    For example, in scoring the quality of human studies, 5 of 19 "serious flaw" metrics (Table H-8) are STROBE reporting guidelines (STROBE checklist items # 6,7,8,13,15). 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 #2a, the STROBE guideline developers explicitly state this is neither the intended nor a scientifically valid use of these guidelines. (32)

[m] 15 USC §2625(h) and (k)

- **Analysis**[n] is equated with a "serious flaw" in study quality, but statistical power[o] alone is not a valid measure of study quality. For example, EPA's framework excludes human studies that do not meet EPA's criteria for "high" in the analysis domain. EPA does not state how it will calculate whether a study is "adequately" powered. According to EPA's framework, to be included in an EPA review, a study must meet the "high" criteria in EPA's "Metric 13, Statistical power (sensitivity, reporting bias)" as presented in the box below. Studies that are not "high" quality for this metric would be designated as "unacceptable for use" by EPA:

EPA Metric 13. Excerpted from Table H-9 (page 243)

| **Metric 13. Statistical power (sensitivity, reporting bias)**<br>**Instructions: To meet criteria for confidence ratings for metrics where 'AND' is included, studies must address both of the conditions where "AND" is stipulated. To meet criteria for confidence ratings for metrics where 'OR' is included studies must address at least one of the conditions stipulated.** | |
|---|---|
| High<br>(score = 1) | ***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.<br>**OR**<br>The paper reported statistical power high enough (≥ 80%) to detect an effect in the exposure population and/or subgroups of the total population.<br><br>***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.<br>**OR**<br>The paper reported statistical power was high (≥ 80%) to detect an effect in the exposure population and/or subgroups of the total population. |
| Medium<br>(score = 2) | • Do not select for this metric. |
| Low<br>(score = 3) | • Do not select for this metric. |
| Unacceptable<br>(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.<br>• ***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. |

[n] See Table H-8 "Serious Flaws that Would Make Epidemiological Studies Unacceptable for Use in the Hazard Assessment" under the "analysis domain" "statistical power/sensitivity" metric (page 233) " in conjunction with Table H-9 "Evaluation Criteria for Epidemiologic Studies, Metric 13 "statistical power (sensitivity, reporting bias) (page 243).

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

First and foremost, EPA provides no 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" (32).

EPA's Metric 13 statistical power/sensitivity also appears to confuse bias with imprecision. 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 science-based decision-making. For example a small study may be imprecise but that should not be confused with whether it is biased (20); a small study can be imprecise but at the same time less biased than a larger study (17). 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 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.

Illustrative of how EPA's "analysis" metric could result in excluding high quality research that can inform science-based decision-making by EPA, in a 2017 systematic review by Lam et al. *"Developmental PBDE Exposure and IQ/ADHD in Childhood: A Systematic Review and Meta-analysis,"*(12) none of the 4 high-quality[p] studies included in the meta-analysis 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)." It is also notable that one of the studies in the meta-analysis, Herbstman et al. 2010, (38) was assessed by the review authors to be "probably high risk of bias" for "Incomplete Outcome Data."[q] As such, this otherwise high quality study, i.e., all of the other domains were "definitely" or "probably" low risk of bias, would meet EPA's criteria for "unacceptable for use" based on STROBE reporting guideline #15, "Report numbers of outcome events or summary measures over time".[r]

In short, the *Lam et al* systematic review, using the best available scientific methods, found that a ubiquitous environmental contaminant is impacting human intelligence, a finding that was subsequently reviewed and endorsed by the National Academy of Sciences (19). Yet EPA's SR review framework would exclude crucial pieces of this body of evidence based on the Agency's inaccurate, non-science-based criteria for deeming studies 'unacceptable.' This is contrary to TSCA's mandate to use the best available science. [s]

- **"Level of exposure"** is equated with a "serious flaw".

[p] "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.

[q] The authors of the systematic review rated the Herbstman 2010 study "probably high risk of bias" for "incomplete outcome data" based on the following rationale: "Concerns regarding missing outcome data at each follow-up time on almost half the cohort of 210 with cord blood PBDE measurements; no argument is presented that would invalidate the possibility of a selection bias (i.e., likelihood that outcome data is missing is related both to outcome status and exposure)."

[r] See Table H-8 "Serious Flaws that Would Make Epidemiological Studies Unacceptable for Use in the Hazard Assessment" under the "outcome assessment domain" "Outcome measurement or characterization" metric (page 232) which specified STROBE guideline #15 to assess this metric.

[s] 15 USC §2625 (h)

EPA's "exposure characterization" domain for human studies includes the level of exposure as a fatal flaw, stating: "For all study types: The **levels** of exposure are not sufficient or adequate (as defined above)[t] to detect an effect of exposure (Cooper et al., 2016)." Unlike human experimental studies, which are largely precluded for ethical reasons, human observational studies can only be based on what exposures actually occur in the real world. EPA offers no explanation of how one could know whether the levels would be "sufficient or adequate" enough to detect an effect. Given the vagaries of this metric, it could be reasonably anticipated that it would permit EPA to arbitrarily exclude quality research from its decision-making.

**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.

[t] EPA "as defined above" is unclear, presumably "as defined above" refers to the definition of the domain in Table H-2 page 223, "Evaluation of exposure assessment methodology that includes consideration of methodological quality, sensitivity, and validation of the methods used, degree of variation in participants, and an established time order between exposure and outcome."

**3. EPA's TSCA systematic review framework 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" (39). Yet EPA's systematic review framework is silent on how it will take into account this empirically documented influence on the results of scientific research. Underscoring this EPA SR framework deficiency is the fact that recent studies empirically document that industry sponsorship produces research that is favorable to the sponsor (40, 41). 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 (42).

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 (43, 44); for example, whether funding source is assessed as a specific risk of bias domain (43) or considered at multiple points in the evaluation (20, 44). 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, (45) and/or to be considered as a potential factor to explain apparent inconsistency within a body of evidence (14).

A 2017 Cochrane systematic review of industry sponsorship and research outcome concluded … "industry sponsorship should be treated as bias-inducing and industry bias should be treated as a separate domain" (40). The National Academy of Sciences in its review of the EPA IRIS program's SR 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 (17)(p 79).

Notably, EPA's exclusion of consideration of funding source and other potential conflicts of interests is also internally inconsistent with EPA's 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" (32).

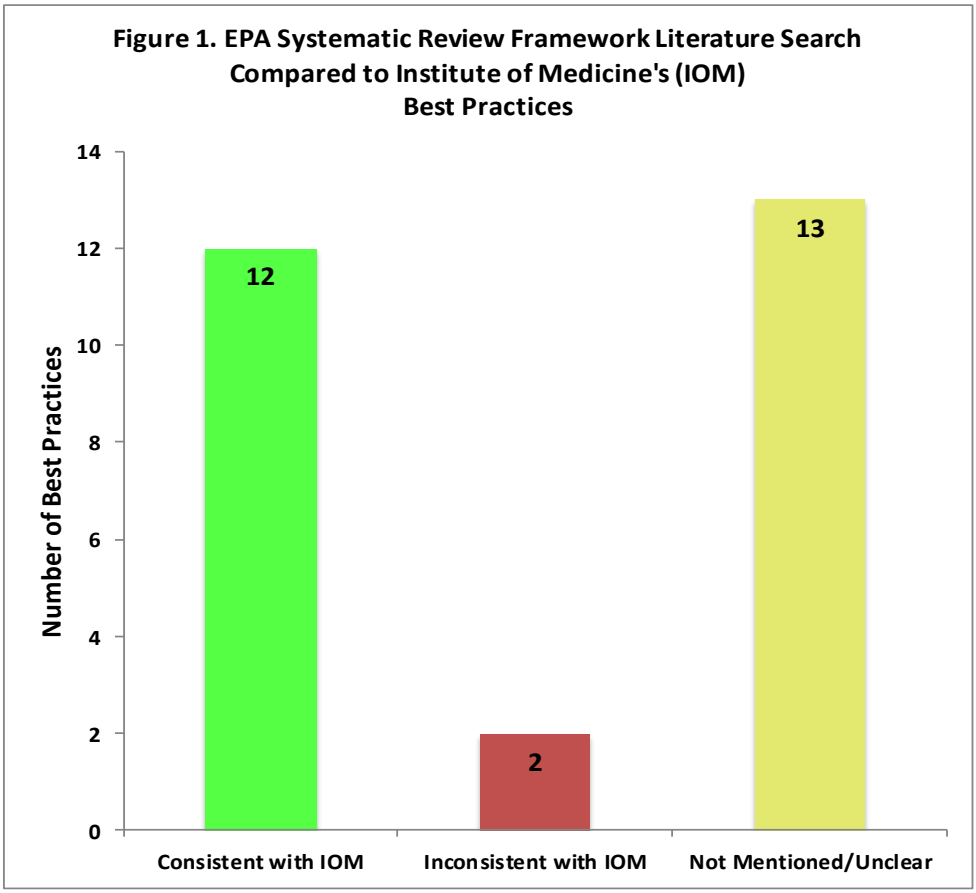
Importantly, including funding as a risk of bias as a domain does not mean excluding industry sponsored studies from EPA's hazard and risk assessment; it only means documenting funding as one of many domains of potential bias and evaluating its impact on the overall quality of the body of evidence.

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

**4. The literature review step of EPA's TSCA systematic review framework 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 systematic review framework. We compared EPA's framework for systematic review to the Institute of Medicine's (IOM's) best practices for the literature review step of a systematic review (16)(See IOM 2011 Chapter 3. and TABLE E-1), which was applied by the National Academy of Sciences in its review of EPA's IRIS Program methods for systematic review (17)(See Table 4-1 Pp. 43-55).

We found EPA's framework to be consistent with 12 of IOM's 27 best practices for conducting a literature search (Figure 1 and Appendix 1). There are two key features of EPA's framework that are clearly inconsistent with IOM's best practices. EPA fails: (1) to include or exclude studies based on the protocol's pre-specified criteria, a practice that is critical to avoiding results-based decisions;[u] and (2) to use two or more members of the review team, working independently, to screen and select studies, which is an essential quality-assurance measure.[v]




[u] See our Comment #1 regarding the EPA framework's lack of a pre-defined protocol.

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

For the remaining 13 IOM best practices, EPA's framework 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 Ten TSCA Risk Evaluations, EPA's framework appears to have incorporated six additional best practices that are either unclear or not mentioned in EPA's SR framework: (1) work with a librarian or other information specialist trained in performing systematic reviews to plan the search strategy (IOM 3.1.1); (2) Design the search strategy to address each key research question (IOM 3.1.2); (3) Search regional bibliographic databases if other databases are unlikely to provide all relevant evidence (IOM 3.1.9); (4) Conduct a web search (IOM 3.2.5); and (5) 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).

EPA should make its framework for conducting a literature review transparently congruent with all of IOM's best practices. This includes addressing two critical inconsistencies: (1) include or exclude studies based on the protocol's pre-specified criteria to prevent results-based decisions; and (2) Use two or more members of the review team, working independently, to screen and select studies, to ensure quality assurance. The transparency of the framework would 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 framework 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 (46). Furthermore, when considering the inclusion or update of an existing systematic review, studies have found that language-inclusive systematic reviews (including studies in languages other than English) were of the highest quality, compared with other types of reviews (47). 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.

Additionally, EPA's framework should explicitly include rules for determining when the list of relevant studies will be considered final i.e., "stopping rules." Newer scientific studies will inevitably continue to appear in scientific journals and it will be impossible to continually attempt to include all these studies in a chemical assessment. To meet the deadlines as mandated by the Lautenberg Amendments, EPA should state clear stopping rules in the form of deadlines or criteria for when the body of included relevant studies will be finalized for the purposes of the chemicals assessment. We also strongly encourage EPA in its stated exploration of automation and machine learning tools,[w] which can help speed the production of EPA's systematic reviews.

**We recommend:** EPA should make its framework for conducting a literature review congruent with all of the Institute of Medicine's best practices, and explicitly include rules for when the list of relevant studies will be considered final.

[w] Footnote 9 page 23 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".

**5. EPA's TSCA systematic review framework correctly recognizes that mechanistic data are not required for a hazard assessment, but EPA is not clear that these data, if available, can only be used to increase, and not to decrease, confidence in a body of evidence.**

EPA's TSCA framework (page 172) states that EPA will use the evaluation strategies for animal and *in vitro* toxicity data to assess the quality of mechanistic and pharmacokinetic data supporting the model, and may tailor its criteria further to evaluate new approach methodologies (NAMs). We agree with EPA that mechanistic data need to be evaluated in a manner comparable to how other streams of evidence are evaluated. Data generated by alternative test methods (such as high-throughput screening methods) are not different than any other type of *in vitro* or cell-based assay data that would be considered in a systematic review. These kinds of assays provide mechanistic data. However, in this case, as described in comment # 2 above, EPA's use of its evaluation strategies for animal and *in vitro* toxicity data would entail using a quantitative scoring method that is incompatible with the best available science in fundamental ways. EPA should employ a scientifically valid method to assess risk of bias of individual studies in *all* streams of evidence, including mechanistic data.

EPA's framework (page 172) states, "the availability of a fully elucidated mode of action (MOA) or adverse outcome pathway (AOP) is not required to conduct the human health hazard assessment for a given chemical (emphasis added)." We strongly agree with EPA that mechanistic data are not needed for a hazard assessment. In addition, EPA's framework should be explicit that mechanistic data are only used to increase confidence in a hazard assessment, and never to decrease confidence.

The National Academy of Sciences explicitly considered how mechanistic data could be utilized in a systematic review for evidence integration (19). The committee came to two conclusions. First, the same protocol for evaluating relevance and study quality must be used with mechanistic data as for any other study. For example, in the report's case study on phthalates, the committee was not able to integrate results from high-throughput assays because the cell lines used were of unknown relevance to the *in vivo* mechanism of phthalate toxicity (19)(pg.78). Second, the foundation of the hazard classification in a systematic review is the animal and human data, with the mechanistic data playing a supporting role. If mechanistic data is relevant, it can be used to upgrade a hazard classification, or increase the confidence of a finding made based on evaluation of animal and human data. A hazard classification is never made based on high-throughput or other kinds of mechanistic data alone (19)(Pp. 158-9).

**We recommend:** EPA should be explicit that mechanistic data can only be used to upgrade a hazard classification, or increase the confidence of a finding made based on evaluation of animal and human data, and that these data will not be used to decrease confidence in a body of evidence.

**6. EPA's TSCA systematic review framework is not independent of the regulatory end user of the review.**

EPA's TSCA systematic review/risk assessment process is not independent of the TSCA risk management process, a conflict that is incompatible with best scientific methods. EPA's SR framework was developed and is being implemented by the Office of Chemical Safety and Pollution Prevention (OCSPP), which is also responsible for regulating the environmental exposures under TSCA review. In contrast, other EPA chemical assessment programs such as the IRIS program are intentionally placed in a non-regulatory research arm (the Office of Research and Development), to create separation from the Agency's program office responsible for regulatory decisions. This separation supports IRIS's ability to develop impartial chemical toxicity information independent of its ultimate use by EPA's program and regional office in risk assessment and risk management decisions. The National Academy of Sciences supported this in its 2018 report, stating that: "Current best practices [for systematic reviews in other medical disciplines] recommended by the Institute of Medicine (IOM 2011) suggest that the IRIS teams involved in the systematic-review process **should be independent of those involved in regulatory decision-making** who use the products of the systematic-review teams **(emphasis added)**" (15). This same principle should also be implemented across the Agency and specifically for TSCA assessments.

**We recommend**: EPA's systematic reviews should be produced independently of the regulatory end user of the review.

**REFERENCES**


1. 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.
2. Fox DM. The Convergence of Science and Governance: Research, Health Policy, and American States. Berkeley, CA: University of California Press; 2010.
3. 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.
4. 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.
5. Woodruff TJ, Sutton P. The Navigation Guide sytematic review methodology: a rigorous and transparent method for translating environmental health science into better health outcomes. Environmenal Health Perspectives. 2014;122(10):A283.
6. 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.
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. 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.
10. 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.
11. 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.

12. 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.
13. 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. 2018.
14. 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.
15. National Academies of Sciences, Engineering, and, Medicine. Progress Toward Transforming the Integrated Risk Information System (IRIS) Program: A 2018 Evaluation. Washington, D.C.: The National Academies Press; 2018.
16. Institute of Medicine. Finding What Works in Health Care. Standards for Systematic Review. Washington, D.C.: The National Academies Press.; 2011.
17. National Research Council. Review of EPA's Integrated Risk Information System (IRIS) Process. Washington, DC: National Academies Press; 2014.
18. National Research Council. Review of the Environmental Protection Agency's State-of-the-Science Evaluation of Nonmonotonic Dose–Response Relationships as They Apply to Endocrine Disruptors. Washington, DC: National Academies Press; 2014.
19. 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. Washington, DC: 2017 2017. Report No.
20. 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.
21. 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.
22. 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.
23. 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.

24. 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.
25. 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
26. 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.
27. 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.
28. 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.
29. 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.
30. 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.
31. Higgins JPT, Altman DG, Sterne JAC. Chapter 8: Assessing risk of bias in included studies. In: Higgins J, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions Version 510 [Updated March 2011]: The Cochrane Collaboration. Available from http://www.cochrane-handbook.org.; 2011.
32. 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.
33. 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.
34. 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.
35. 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.
36. 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.

37. 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.
38. Herbstman JB, Sjödin A, Kurzon M, Lederman SA, Jones RS, Rauh V, Needham LL, Tang D, Niedzwiecki M, Wang RY, Perera F. Prenatal exposure to PBDEs and neurodevelopment. Environ Health Perspect. 2010.
39. 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.
40. 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.
41. White J, Bero LA. Corporate manipulation of research: Strategies are similar across five industries. . Stanford Law & Policy Review. 2010;21((1)):105-34. .
42. 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.
43. 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.
44. 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.
45. 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.
46. 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.
47. 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.

## Appendix 1: 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 | Apparently applied to 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.**<br>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.**<br>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.<br>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** |

| | | | | | | |
|---|---|---|---|---|---|---|
| **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.** age 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** |
| **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<br>**Also choose applicable health effect tags in next section "Methylene Chloride (DCM) Health Effect Tags"<br>EXCLUDE: Occupational studies that do not specify specific solvent exposure<br>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<br>controlled study) and case studies (e.g.individual case report on accidental exposure to chemical) Acute, subchronic, and chronic exposures<br>**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.<br><br>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)."<br><br>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)."<br><br>(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.**<br><br>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** |
| **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.**<br><br>**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).<br><br>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”. | 1 | | | | |
| | | 1 | 0 | 0 | 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.**<br><br>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<br>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.**<br><br>Table 3-1 states that EPA will " Extract data/information using pre-defined templates."<br><br>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:<br>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.**<br><br>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** |

| | | | | | |
|---|---|---|---|---|---|
| | 2 | 0 | 0 | 2 | 0 |
| | Consistent with IOM | Inconsistent with IOM | Not Mentioned | Unclear | Not mentioned/unclear but apparently applied to First 10 TSCA chemicals |
| **TOTALS** | 12 | 2 | 6 | 7 | 5 |

**Appendix B:** 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 C:** Analysis of EPA Ratings Changes

| 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 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Test substance identity | 2* | 2* | 1 | 2* | 2* | 2* | 2* | 1 | 2* | 2* | 2* | 1 | 2* | 1 | 1 | |
| Test substance source | 3 | 3* | 2* | 3 | 3* | 3 | 3* | 1 | 3 | 3* | 3 | 2 | 3* | 2 | 1 | |
| Test substance purity | 3 | 3* | 2* | 3 | 3* | 3 | 3* | 1 | 3 | 3* | 3 | 2 | 3* | 2 | 1 | |
| Negative and vehicle controls | 3* | 3* | NR | 2* | 4* | 3* | 3* | 1 | 2 | 2 | 1 | NR | 1 | 1 | 1 | |
| Postive controls | NR* | NR | NR | NR | NR | NR | NR | NR | NR | NR | NR | NR | NR | NR | 1 | |
| Randomized allocation | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 2 | NR* | NR* | NR | NR | NR | NR | 3 | |
| Preparation and storage of test substance | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 1 | 3 | 3 | 3* | 3* | 3* | 3* | 2 | |
| Consistency of exposure administration | 3 | 3 | 2 | 4* | 4* | 3 | 3 | 1 | 3 | 3 | 1 | 1 | 1 | 1 | 1 | |
| Reporting of doses / concentrations | 1 | 1 | 1 | 4* | 4* | 1 | 1 | 1 | 3 | 3 | 1 | 1 | 1 | 1 | 1 | |
| Exposure frequency and duration | 1 | 1 | 1 | 3* | 3* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | |
| Number of exposure groups and dose spacing | 1 | 1 | 1 | 3* | 3 | 1 | 1 | 1 | 1 | 1 | 1* | 1 | 1 | 1 | 1 | |
| Exposure route and method | 1 | 1 | 1 | 4* | 4* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | |
| Test animal characteristics | 2* | 2* | 2 | 3 | 3 | 3* | 3* | 1 | 2 | 1 | 3 | 2 | 3 | 2 | 2 | |
| Adequacy and consistency of animal husbandry conditions | 3* | 3* | 1 | 3* | 3* | 3* | 3* | 1 | 3* | 2 | 3* | 1 | 3* | 1 | 1 | |
| Number per group | 1 | 1 | 1 | 2* | 3* | 1 | 1 | 1 | 3 | 3 | 2* | 1 | 1 | 1 | 1 | |
| Outcome assessment methodology | 2* | 2* | 1 | 3* | 3* | 2* | 2* | 1 | 3* | 3* | 2* | 1 | 2* | 1 | 1 | |
| Consistency of outcome assessment | 2 | 2 | 1 | 3* | 3* | 3* | 3* | 1 | 1 | 1 | 2* | 1 | 2* | 1 | 1 | |
| Sampling adequacy | 1 | 1 | 1 | 2* | 2* | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | |
| Blinding of assessors | NR* | NR* | NR* | NR* | NR* | NR* | NR* | NR | NR* | NR* | NR | NR | NR | NR | NR | |
| Negative control response | NR | NR | NR | 3* | 4* | NR | NR | 1 | NR* | NR* | 1 | NR | 1 | 1 | 1 | |
| Confounding variables in test setup and procedures | 2 | 2 | 2 | 3* | 3* | 3* | 3* | 1 | 2 | 2 | 3* | 2* | 3* | 1 | 1 | |
| Health outcomes unrelated to exposure | 3* | 3* | 1 | 3* | 3* | 3* | 3* | 1 | 3* | 3* | 3* | 1 | 3* | 1 | 1 | |
| Statistical methods | NR* | NR* | 1 | NR* | NR* | NR* | NR* | 1 | NR* | NR* | NR* | 1 | NR* | 1 | 1 | |
| Reporting of data | 2 | 2 | 1 | 3* | 4* | 3* | 2 | 1 | 1* | 1* | 1 | 1 | 1 | 1 | 1 | |
| Overall Quality Level, 2018[2] | High | High | High | Medium | Medium | High | High | High | Medium | Medium | High | High | High | High | High | |
| Overall Quality Level, 2019[3] | Medium | Medium | High | Unacceptable | Unacceptable | Low | Low | High | Medium | Medium | Medium | High | Medium | High | High | |
| | | | | | | | | | | | | | | | | **Total Changes** |
| **Changes (Reported)** | 8 | 7 | 2 | 18 | 18 | 10 | 9 | 0 | 8 | 8 | 7 | 2 | 8 | 1 | 0 | 106 |
| **Changes (Not reported)** | 1 | 3 | 1 | 0 | 1 | 1 | 3 | 0 | 1 | 2 | 3 | 0 | 2 | 0 | 0 | 18 |
| **Total** | 9 | 10 | 3 | 18 | 19 | 11 | 12 | 0 | 9 | 10 | 10 | 2 | 10 | 1 | 0 | 124 |

**Legend**

| | |
|---|---|
| 1 | High |
| 2 | Medium |
| 3 | Low |
| 4 | Unacceptable |
| NR | Not Rated |
| * | Changes were made between previous and current version |
| | **not reported** |

**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 2019