Environmental Defense Fund Comments on
Application of Systematic Review in TSCA Risk Evaluations
EPA-HQ-OPPT-2018-0210
83 Federal Register 26998 - 27000 (Monday, June 11, 2018)
Submitted Thursday August 16, 2018

Environmental Defense Fund (EDF) appreciates the opportunity to provide comments to the Environmental Protection Agency (EPA) regarding its *Application of Systematic Review in TSCA Risk Evaluations.*[1]

TABLE OF CONTENTS

[1] U.S. EPA, *Application of Systematic Review in TSCA Risk Evaluations* (May 2018), https://www.epa.gov/sites/production/files/2018-06/documents/final_application_of_sr_in_tsca_05-31-18.pdf.

1. Introduction

In May 2018, EPA's Office of Chemical Safety and Pollution Prevention (OCSPP) released its *Application of Systematic Review in TSCA Risk Evaluations* (hereafter "TSCA systematic review document"). This document provides details regarding the Office of Pollution Prevention and Toxics's (OPPT) development of a proposed "systematic review" approach, and the application of this approach to chemical risk evaluations under the Toxic Substances Control Act (TSCA). EPA states that it will apply this approach to the first ten chemicals undergoing risk evaluation under TSCA.

OPPT indicates that it has developed a systematic review approach in order to meet the TSCA requirement that "EPA use data and/or information (hereinafter referred to as data/information) in a manner consistent with the best available science and that EPA base decisions on the weight of the scientific evidence." (p. 14) In the final rule *Procedures for Chemical Risk Evaluation Under the Amended Toxic Substances Control Act*, the agency defines weight of the scientific evidence as "a systematic review method, applied in a manner suited to the nature of the evidence or decision, that uses a pre-established protocol to comprehensively, objectively, transparently, and consistently identify and

evaluate each stream of evidence, including strengths, limitations, and relevance of each study and to integrate evidence as necessary and appropriate based upon strengths, limitations, and relevance."[2]

In addition to being required by the agency's risk evaluation rule, applying a systematic review framework to chemical risk evaluation is consistent with the recommendations of the National Academy of Sciences (NAS)[3] and leading chemical assessment initiatives across government[4,5] and academia.[6]

However, the process that OPPT has outlined in this document omits key aspects of what is entailed in a systematic review – even by the agency's own definition.[7] Among other aspects of systematic review that are missing, the TSCA systematic review document does not describe a general approach to protocol development or data integration.[8] To be consistent with the systematic review, EPA should have developed a protocol for each chemical undergoing risk evaluation. EPA has not developed protocols for any of the first 10 chemicals undergoing risk evaluation.

---

[2] Procedures for Chemical Risk Evaluation Under the Amended Toxic Substances Control Act, 82 Fed. Reg. 33726, 33733, 33748 (Jul. 20, 2017) (codified at 40 C.F.R. § 702.33), https://www.regulations.gov/document?D=EPA-HQ-OPPT-2016-0654-0108 (hereinafter "Risk Evaluation Rule").

[3] Nat'l Research Council, *Review of EPA's Integrated Risk Information System (IRIS) Process* (2014), https://www.ncbi.nlm.nih.gov/books/NBK230072/.

[4] INTEGRATED RISK INFORMATION SYSTEM, https://www.epa.gov/iris (last visited Aug. 15, 2018).

[5] U.S. Dep't of Health & Human Servs., Nat'l Toxicology Program, *Handbook for Conducting a Literature-Based Health Assessment Using OHAT Approach for Systematic Review and Evidence Integration* (Jan. 2015), https://ntp.niehs.nih.gov/ntp/ohat/pubs/handbookjan2015_508.pdf.

[6] NAVIGATION GUIDE, https://prhe.ucsf.edu/navigation-guide (last visited Aug. 15, 2018); Tracey J. Woodruff, et al., *An Evidence-Based Medicine Methodology to Bridge the Gap Between Clinical and Environmental Health Sciences, 30:5 HEALTH AFFAIRS 931* (May 2011), https://www.healthaffairs.org/doi/pdf/10.1377/hlthaff.2010.1219?siteid=healthaff&keytype=ref&ijkey=z58MCEPW2X49.&.

[7] The Risk Evaluation Rule uses the Institute of Medicine's definition of systematic review: "systematic review 'is a scientific investigation that focuses on a specific question and uses explicit, pre-specified scientific methods to identify, select, assess, and summarize the findings of similar but separate studies. The goal of systematic review methods is to ensure that the review is complete, unbiased, reproducible, and transparent.'" 82 Fed. Reg. at 33734. The rule further states that "Key elements of systematic review include:

> —A clearly stated set of objectives (defining the question);
> —Developing a protocol which describes the specific criteria and approaches that will be used throughout the process;
> —Applying the search strategy criteria in a literature search;
> —Selecting the relevant papers using predefined criteria;
> —Assessing the quality of the studies using predefined criteria;
> —Analyzing and synthesizing the data using the predefined methodology;
> —Interpreting the results and presenting a summary of findings." *Id.*

[8] For the purpose of these comments, EDF uses "data integration" and "evidence integration" synonymously.

Additionally, the one aspect of systematic review OPPT *has* addressed – evaluation of individual study quality – deviates in several significant ways from established best practices in systematic review.[9,10] EPA has not provided any empirical evidence or other justification for why these deviations are reasonable, necessary, or scientifically sound. Indeed, EPA has provided no indication that it has even attempted to test its approach on a robust set of actual studies to determine what effect its approach to individual study evaluation will have on study inclusion, evidence integration, and the risk evaluation process more generally.

In sum, the TSCA systematic review document deviates significantly from best practices in systematic review—practices that are empirically based and have been scientifically reviewed, vetted, and instituted by other agencies and authoritative scientific bodies. EPA should substantially revise its TSCA systematic review document and subject it to peer review by qualified external experts in the field.

2. The TSCA Systematic Review Document is not consistent with TSCA § 26

EPA's proposed approach will lead to violations of EPA's science obligations under TSCA § 26(h), (i), and (k). These directives require that EPA must consider all reasonably available information, and that EPA then must make decisions reflecting the "best available science" and "weight of the scientific evidence" based on the body of evidence *as a whole*.[11] EPA's proposed approach erroneously tries to apply these directives at the level of individual studies, and the result is that EPA may exclude reasonably available information on the grounds that an individual piece of evidence is somehow imperfect, even when it contributes to the "best available science" or adds to the "weight of the scientific evidence" when available information is considered as a whole.

These statutory commands in TSCA repeatedly emphasize that EPA must make decisions based on the information that is "available," and the courts have recognized that such a duty requires action on the basis of available information even if that information is imperfect.[12] EPA cannot craft its systematic review process to incrementally exclude available information study-by-study, with the possibility of prohibiting use of the best available science simply because one or more of the underlying studies is imperfect in some manner. While certain systematic review approaches in exceptional cases may exclude from further consideration some studies because they entail a substantial risk of bias or have

[9] The Cochrane Collaboration, *Cochrane Handbook for Systematic Reviews of Interventions* (Mar. 2011), http://handbook-5-1.cochrane.org/; U.S. Dep't of Health & Human Servs., Nat'l Toxicology Program, *supra* note 5; Univ. of Cal. San Francisco, Program on Reprod. Health & the Env't, *supra* note 6; Inst. of Med. of the Nat'l Acads., *Finding What Works in Health Care: Standards for Systematic Reviews* (Mar. 2011), https://www.ncbi.nlm.nih.gov/pubmedhealth/PMH0079447/.

[10] *See* 82 Fed. Reg. at 33734.

[11] 15 U.S.C. § 2625(k), (h), (*i*).

[12] *See, e.g.*, *Southwest Ctr. for Biological Diversity v. Babbitt*, 215 F.3d 58, 60 (D.C. Cir. 2000) ("Even if the available scientific and commercial data were quite inconclusive, [the agency] may—indeed must—still rely on it at that stage.") (quoting *City of Las Vegas v. Lujan*, 891 F.2d 927, 933 (D.C. Cir. 1989)).

severe methodological shortcomings,[13] EPA's proposed scoring approach appears to allow or require EPA to frequently exclude studies based solely on reporting flaws or other flaws that do not rise to the level of these exceptions.

As described more below, EPA's approach will also exclude certain reasonably available information on the basis that it does not meet EPA's preset expectations. For example, for monitoring data, environmental release data, completed exposure or risk assessments, and reports containing other exposure or release data, EPA plans to rate as "unacceptable" any data derived from occupational or non-occupational scenarios that do not precisely correspond to an occupational scenario EPA has identified within the scope of a given risk evaluation. Pp.75-76, 79-80, 86-87. The far more appropriate response to discovering reasonably available information revealing scenarios outside the scope of the risk evaluation would be for EPA to consider whether it needs to expand the scope of the risk evaluation and potentially the protocol (where any such changes would be clearly documented); nothing in TSCA authorizes or requires EPA to simply ignore that reasonably available information on the basis that it does not meet EPA's preset expectations.

3. The TSCA Systematic Review Document is not representative of systematic review

EPA's TSCA systematic review document is not representative of a true systematic review method as required by EPA's own risk evaluation rule, which requires inclusion of a "pre-established protocol" that addresses, among other things, how EPA will "integrate evidence".[14]

Born out of the clinical sciences, systematic review employs structured approaches to evidence identification, evaluation, and synthesis in a manner that promotes scientific rigor, consistency, transparency, objectivity, and reduction of bias. Indeed, systematic review transformed the field of medicine—serving today as the method for evaluating the effectiveness of interventions and diagnostic tools.[15]

Prominent systematic review methods and tools in medicine, particularly Cochrane[16] and GRADE,[17] have been shaped and refined over several decades based on empirical evidence and experience in application. Appropriately, leading systematic review approaches that have emerged in environmental

[13] Nat'l Research Council, *Review of EPA's Integrated Risk Information System (IRIS) Process* at chp. 5, p. 72 (2014), https://www.ncbi.nlm.nih.gov/books/NBK230063/.
[14] 40 C.F.R. § 702.33 (definition of "weight of scientific evidence").
[15] Tracey J. Woodruff & Patrice Sutton, *The Navigation Guide Systematic Review Methodology: A Rigorous and Transparent Method for Translating Environmental Health Science into Better Health Outcomes*, 122:10 ENVTL. HEALTH PERSPECTIVES 1007 (Oct. 2014), https://ehp.niehs.nih.gov/1307175/.
[16] *See* ABOUT COCHRANE REVIEWS, https://www.cochranelibrary.com/about/about-cochrane-reviews (last visited Aug. 15, 2018).
[17] *See* GRADE APPROACH, https://training.cochrane.org/grade-approach (last visited Aug. 15, 2018).

health, including the UCSF Navigation Guide[18] and the National Toxicology Program's literature-based reviews,[19] have modeled themselves from these methods.

Bizarrely, EPA correctly cites authoritative sources on systematic review and at points describes processes that generally align with best practices, but then deviates substantially from those established best practices in detailing its specific plans for systematic review. Further, EPA provides no explanation or justification for its deviations.

A. OPPT's approach to systematic review lacks a generally linear progression, inconsistent with the conduct of true systematic review

In section three, *Integration of Systematic Review Principles Into TSCA Risk Evaluation*, EPA includes key excerpts from the preamble to the final rule *Procedures for Chemical Risk Evaluation Under the Amended Toxic Substances Control Act*:

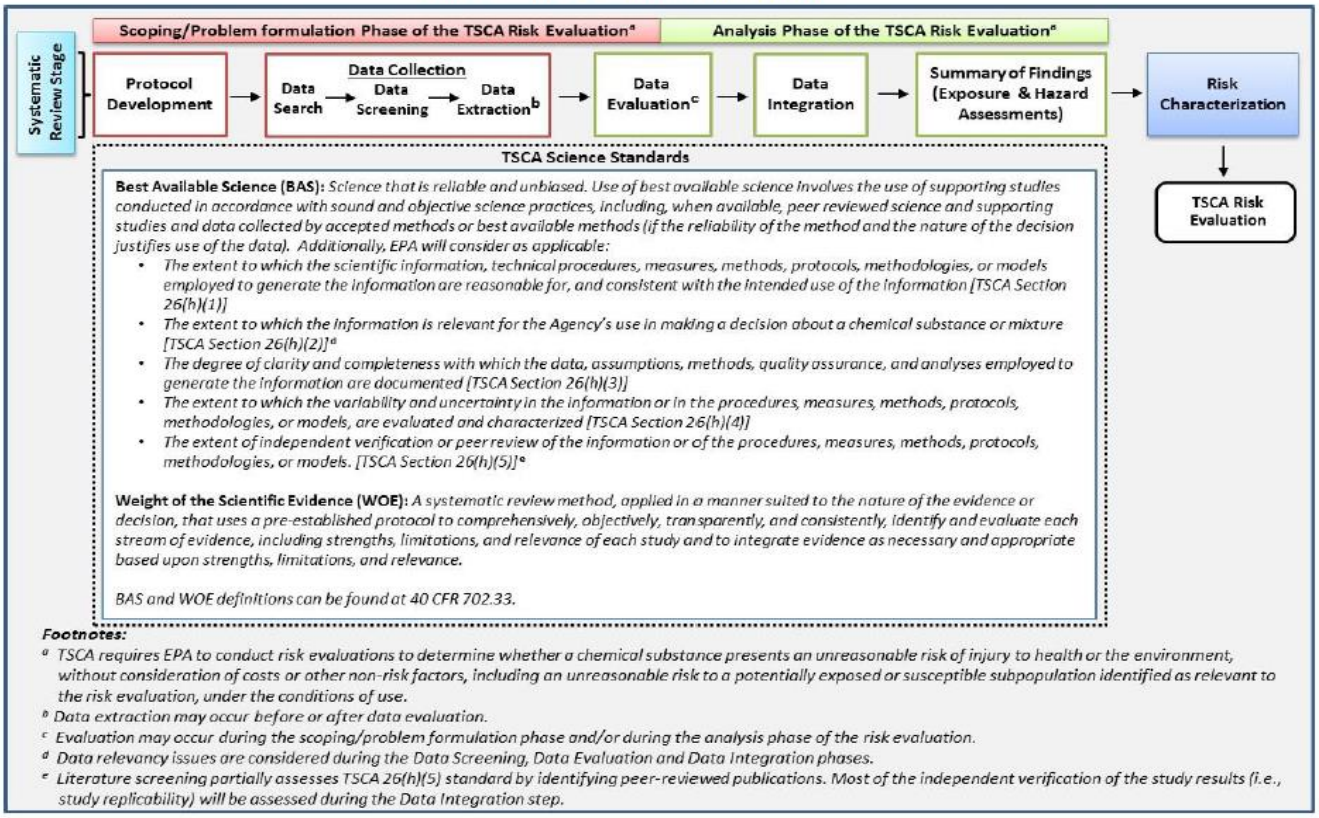
> As defined by the Institute of Medicine, systematic review "is a scientific investigation that focuses on a specific question and uses explicit, pre-specified scientific methods to identify, select, assess, and summarize the findings of similar but separate studies" (National Academy of Sciences, 2017). The goal of systematic review methods is to ensure that the review is complete, unbiased, reproducible, and transparent (Bilotta et al., 2014).****Key elements of systematic review include: a clearly stated set of objectives (defining the question); developing a protocol that describes the specific criteria and approaches that will be used throughout the process; applying the search strategy in a literature search; selecting the relevant papers using predefined criteria; assessing the quality of the studies using predefined criteria; analyzing and synthesizing the data using the predefined methodology; [and] interpreting the results and presenting a summary of findings. (p. 13-14).

These excerpts by and large reflect core tenets and approaches of a systematic review framework. However, the TSCA systematic review document makes evident that EPA has no interest in authentically applying systematic review. Indeed, it would be wrong to call what EPA has developed a systematic review framework, method, or tool. This becomes very evident in Figure 3-1 of the document, TSCA Systematic Review Process, copied below.

[18] Tracey J. Woodruff, et al., *An Evidence-Based Medicine Methodology to Bridge the Gap Between Clinical and Environmental Health Sciences*, 30:5 HEALTH AFFAIRS 931 (May 2011), https://www.healthaffairs.org/doi/pdf/10.1377/hlthaff.2010.1219?siteid=healthaff&keytype=ref&ijkey=z58MCEPW2X49.&.

[19] *See* OHAT SYSTEMATIC REVIEW, https://ntp.niehs.nih.gov/pubhealth/hat/review/index-2.html (last visited Aug. 15, 2018) (describing the National Toxicology Program's work on systematic review).

Figure 3-1. TSCA Systematic Review Process[4]



[4] 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).

The graphic portion of the figure illustrates a generally linear process in alignment with a true systematic review framework. However, an examination of the footnotes makes evident that the figure is a mirage:

- Footnote b: Data extraction may occur before or after data evaluation.
- Footnote c: Evaluation may occur during the scoping/problem formulation phase and/or during the analysis phase of the risk evaluation.
- Footnote d: Data relevancy issues are considered during the Data Screening, Data Evaluation and Data Integration phases.
- Footnote e: ***Most of the independent verification of the study results (i.e., study replicability) will be assessed during the Data Integration Step.

The effect of the footnotes is to undermine the basic premise and purpose of systematic review—to provide consistency, objectivity, transparency, and reduction of bias in the identification, evaluation, and integration of evidence, as foundationally supported by the development of a pre-defined protocol that articulates how these elements are to work. While it is difficult to parse out the specific meaning of

EPA's footnotes, it is evident that the agency intends to jumble the process to such an extent that it is no longer a systematic review.

Also deeply concerning is EPA's use of "replicability" as a standard for independent verification. This is wholly inappropriate as it suggests that a study must be repeated in order to be considered valid or of high quality. A study's validity or quality is not dependent on whether the study and its findings have been repeated as discussed extensively in EDF's comments on EPA's proposal, *Strengthening Transparency in Regulatory Science*.[20] EPA must strike this language in Figure 3-1 and anywhere else it may appear.

B. OPPT has failed to develop individual protocols for the first 10 chemicals undergoing risk evaluation.

In the TSCA systematic review document, EPA states:

> Protocol development is intended to *pre-specify* the criteria, approaches and/or methods for data collection, data evaluation and data integration. It is important to plan the systematic review approaches and methods *in advance* to reduce the risk of introducing bias into the risk evaluation process (p. 19, emphases added)

EPA has appropriately emphasized the importance of protocol development in systematic review – including its development at the outset. Authoritative sources on systematic review including Cochrane, National Academy of Sciences, the National Toxicology Program's Office of Health Assessment and Translation (OHAT), and the Navigation Guide all stress the import of upfront protocol development:

> *Cochrane Handbook*: Since Cochrane reviews are by their nature retrospective ***, it is important that the methods to be used should be established and documented in advance. Publication of a protocol for a review prior to knowledge of the available studies reduces the impact of review authors' biases, promotes transparency of methods and processes, reduces the potential for duplication, and allows peer review of the planned methods (Light 1984).[21]
>
> *2014 NAS Review*: Critical elements of conducting a systematic review include formulating the specific question that will be addressed (problem formulation) and developing the protocol that specifies the methods that will be used to address the question (protocol development).[22]

[20] EDF Comments on Strengthening Transparency in Regulatory Science, https://www.regulations.gov/docket?D=EPA-HQ-OA-2018-0259.
[21] The Cochrane Collaboration, *Cochrane Handbook for Systematic Reviews of Interventions* at pt. 1, chp. 2.1 (Mar. 2011), http://handbook-5-1.cochrane.org/.
[22] Nat'l Research Council, *Review of EPA's Integrated Risk Information System (IRIS) Process* at 5 (2014), https://www.ncbi.nlm.nih.gov/books/NBK230060/.

> *2014 NAS Review*: When the systematic review questions have been specified, a protocol for each review should be developed. A protocol makes the methods and the process of the review transparent, can provide the opportunity for peer review of the methods, and stands as a record of the review. The protocol also minimizes bias in evidence identification by ensuring that inclusion of studies in the review does not depend on the findings of the studies.[23]
>
> *Navigation Guide*: The application of the Navigation Guide is guided by a detailed protocol developed prior to undertaking the review (Figure 1). In contrast, expert-based narrative review methods do not provide a document that predefines a specific question to be answered and sets up the "rules" of the evaluation. A predefined protocol is a staple of systematic reviews in the clinical sciences because it reduces the impact of review authors' biases, provides for transparency of methods and processes, reduces the potential for duplication, and allows for peer review of the planned methods (Higgins and Green 2011).[24]

Despite EPA's acknowledgement of the importance of upfront protocol development, EPA has failed to develop such protocols for the first 10 chemicals and it is not evident whether EPA plans to do so for future chemical risk evaluations. EPA states:

> The first ten chemical substances were not subject to prioritization, the process through which EPA expects to collect and screen much of the relevant information about chemical substances that will be subject to the risk evaluation process. EPA has 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. (p. 19)

EPA must develop upfront protocols for each chemical undergoing risk evaluation. The National Academy of Sciences in its recent review of the EPA Integrated Risk Information System (IRIS) program, *Progress Toward Transforming the Integrated Risk Information System (IRIS) Program: A 2018 Evaluation*, explained, "***[the] IRIS program is encouraged to complete the public-comment process and finalize the protocol before initiating the systematic review. Doing so will improve transparency in the IRIS process."[25]

---

[23] Nat'l Research Council, *Review of EPA's Integrated Risk Information System (IRIS) Process* at chp. 3, p. 36 (2014), https://www.ncbi.nlm.nih.gov/books/NBK230069/.

[24] Tracey J. Woodruff, et al., *An Evidence-Based Medicine Methodology to Bridge the Gap Between Clinical and Environmental Health Sciences*, 30:5 HEALTH AFFAIRS 931 (May 2011), https://www.healthaffairs.org/doi/pdf/10.1377/hlthaff.2010.1219?siteid=healthaff&keytype=ref&ijkey=z58MCEPW2X49.&.

[25] News Release, The Nat'l Acads. of Scis., Eng'g, Med., EPA's IRIS Program Has Made Substantial Progress, Says New Report (Apr. 11, 2018), http://www8.nationalacademies.org/onpinews/newsitem.aspx?RecordID=25086.

Insufficient time is not an acceptable justification for EPA's failure to develop protocols for the first chemicals undergoing risk evaluation. Upfront protocol development is a fundamental feature of systematic review, which EPA by regulation has explicitly included in its definition of the weight of the scientific evidence. Further, the challenges posed by the time constraints were magnified by EPA's illogical decision not to adopt state-of-the-art approaches to systematic review for chemical assessment that have been peer-reviewed, including by the National Academies, and applied and published (i.e., Navigation Guide, OHAT, and IRIS frameworks). Instead, OPPT has inexplicably chosen to develop *de novo* its own approach to systematic review, the result of which far from resembles a legitimate systematic review.

EPA must develop comprehensive protocols, make them publicly available, and subject them to public comment – prior to initiating subsequent steps of the risk evaluation process. For efficiency, we recommend that EPA simultaneously publish the protocols and chemical scoping documents. This would not be unlike the approach currently taken by the EPA IRIS program, which publishes its assessment plans (scoping and problem formulation) and protocols for public comment in advance of conducting toxicological reviews.[26]

C. OPPT has failed to describe its approach to evidence integration for the first 10 chemicals undergoing risk evaluation.

EPA includes an evidence integration element in its systematic review approach (see Figure 3-1), but has failed to provide any substantive details on how it will execute this phase of the review, leaving a significant aspect of the risk evaluation processes a total black box.

In the problem formulations for the first ten chemicals, EPA refers to the TSCA systematic review document for more details on how data integration will occur. But OPPT indicates in the TSCA systematic review document:

> Data integration activities for the first ten TSCA risk evaluation [*sic*] are anticipated to occur *after* the TSCA Problem Formulation documents are released (Figure 1-1). EPA/OPPT will provide further details about the data integration strategy along with the publication of the draft TSCA risk evaluations. (p. 27, emphasis added)

Beyond the fact that the public review process for the problem formulations did not have the benefit of knowing how EPA would conduct data integration, EPA's plan to describe and implement its approach to evidence integration simultaneously with the publication of the draft risk evaluations is problematic. Specifically, there is a high risk that EPA will inconsistently implement evidence integration across the first 10 chemicals undergoing risk evaluation as different groups of EPA staff concurrently conduct such

[26] U.S. EPA, Office of Research & Dev., National Academy of Science Committee to Review Advances Made to the IRIS Program at slides 22, 24 (Feb. 2018), http://nas-sites.org/dels/files/2018/01/AdIRIS-15.pdf.

evaluations absent a general reference methodology; as well as, significant risk for bias to be introduced in the implementation of evidence integration.

It is antithetical to systematic review to concurrently develop and execute an entire step of the review process. More broadly, the absence of any description of how evidence integration will occur reflects EPA's general failure to develop, publish, and seek comment on upfront protocols for the chemicals undergoing risk evaluation.

At the very least, EPA should immediately describe its general approach to evidence integration, referring to established systematic review approaches, including the OHAT, Navigation Guide, and IRIS methods. EPA should include this general approach in a revised TSCA systematic review document; and going forward, EPA should detail its specific approach to evidence integration in protocols developed for each chemical undergoing risk evaluation

D. OPPT's approach to, and implementation of, systematic review should not provide for excessive iteration.

In the OPPT systematic review document, EPA states:

> Although not shown in Figure 3-1, iteration is a natural component of systematic review and risk evaluation processes. There could be different reasons triggering iteration such as the failure of retrieving relevant data and information after the initial search and screening activities, which would require repeating the data collection stage of the systematic review process, or refinements to the initial search, screening and extraction strategies. (p. 14)

While adjustments during the conduct of a systematic review are acceptable, these adjustments should not be a frequent occurrence. The intent of systematic review is to create a structured, transparent, objective, and consistent approach to identifying, evaluating, and integrating evidence in a manner that reduces bias. Excessive iteration undermines this core purpose and provides a pathway for bias. Indeed Cochrane notes,

> While the intention should be that a review will adhere to the published protocol, changes in a review protocol are sometimes necessary. *** While every effort should be made to adhere to a predetermined protocol, this is not always possible or appropriate. It is important, however, that changes in the protocol should not be made on the basis of how they affect the outcome of the research study. *Post hoc* decisions made when the impact on the results of the research is known, such as excluding selected studies from a systematic review, are highly susceptible to bias and should be avoided.[27]

[27] The Cochrane Collaboration, *Cochrane Handbook for Systematic Reviews of Interventions* at pt. 1, chp. 2.1 (Mar. 2011), http://handbook-5-1.cochrane.org/.

This also exemplifies the problems that arise from EPA's failure to develop upfront protocols. Public comment on the upfront protocols would allow EPA to leverage the larger community in developing a rigorous protocol. A more rigorous protocol upfront would likely reduce the need for iteration. Additionally, in the absence of a protocol, it is impossible for the public to determine when and why EPA has modified its systematic review of a chemical. Documentation of changes to protocols is essential and EPA should provide public access to any changes in the protocol. As Cochrane notes:

> Changes in the protocol should be documented and reported in the 'Differences between protocol and review' section of the completed review, and sensitivity analyses (see Chapter 9, Section 9.7) exploring the impact of deviations from the protocol should be undertaken when possible.[28]

Cochrane systematic reviews are uploaded to PROSPERO, "an international prospective register of systematic reviews in health and social care"[29] that creates a permanent record of protocols and allows changes to be tracked. As of 2013, all Cochrane protocols are automatically registered in PROSPERO. The UCSF Navigation Guide has registered several of its systematic reviews on chemicals in PROSPERO.[30]

4. Use of scoring to evaluate individual study quality is wholly inappropriate and inconsistent with best practices in systematic review

As noted in the systematic review approach document, "EPA/OPPT developed a numerical scoring system to inform the characterization of the data/information sources during the data integration phase" (p. 30). Best practices in systematic review expressly discourage the use of scoring to rate individual studies. The Cochrane handbook for systematic reviews of interventions states:

> The *use of scales for assessing quality or risk of bias is explicitly discouraged* in Cochrane reviews. *While the approach offers appealing simplicity, it is not supported by empirical evidence* (Emerson 1990, Schulz 1995b). Calculating a summary score inevitably involves assigning 'weights' to different items in the scale, and it is difficult to justify the weights

---

[28] *Id.*

[29] PROSPERO, https://community.cochrane.org/editorial-and-publishing-policy-resource/overview-cochrane-library-and-related-content/prospero (last visited Aug. 16, 2018).

[30] *See, e.g.*, Juleen Lam, et al., *Applying the navigation guide systematic review methodology, case study #6: association between formaldehyde exposure and asthma* (2016), https://www.crd.york.ac.uk/PROSPERO/display_record.php?RecordID=38766; Juleen Lam, et al., *Applying the navigation guide systematic review methodology. Case Study #4: association between developmental exposures to ambient air pollution and autism* (2015), https://www.crd.york.ac.uk/PROSPERO/display_record.php?RecordID=17890; Juleen Lam, et al., *Applying the navigation guide systematic review methodology. Case study #5: association between developmental exposures to PBDEs and human neurodevelopment* (2015), https://www.crd.york.ac.uk/PROSPERO/display_record.php?RecordID=19753.

> assigned. Furthermore, *scales have been shown to be unreliable assessments of validity* (Jüni 1999) and they are less likely to be transparent to users of the review. It is preferable to use simple approaches for assessing validity that can be fully reported (i.e. how each trial was rated on each criterion).[31] (emphases added)

Similarly, the National Academies Institute of Medicine (IOM) wrote:

> In recent years, 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 (Jüni, 1999). Quality scoring systems have not been validated. Studies assessed as excellent quality using one scoring method may be subsequently assessed as lower quality using another scoring method (Moher et al., 1996).[32]

In its 2014 review of the IRIS program, the National Academies of Science (NAS) likewise asserted that systematic review methodologies have moved away from calculating quality scores because this type of approach has several fundamental flaws:

> 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).[33]

Despite these warnings and explicit recommendations against applying scores and weights to study evaluation, OPPT has chosen to employ this strategy. Further, EPA has done this without providing any empirical evidence or scientific justification for why such a deviation from best practices in systematic review is reasonable, necessary, and valid. In reality, scientific justification for study scoring in a systematic review framework is scientifically unsound and does not exist.

---

[31] The Cochrane Collaboration, *Cochrane Handbook for Systematic Reviews of Interventions* at pt. 2, chp. 8.3.3 (Mar. 2011), http://handbook-5-1.cochrane.org/.
[32] Inst. of Med. of the Nat'l Acads., *Finding What Works in Health Care: Standards for Systematic Reviews* at chp. 3, p. 132 (Mar. 2011), https://www.ncbi.nlm.nih.gov/pubmedhealth/PMH0079446/.
[33] Nat'l Research Council, *Review of EPA's Integrated Risk Information System (IRIS) Process* at chp. 5, p. 69 (2014), https://www.ncbi.nlm.nih.gov/books/NBK230063/.

The method by which EPA calculates a study's overall quality score highlights the arbitrary nature of the proposed scoring approach. As EPA notes:

> 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)
>
> A study is disqualified from further consideration if the confidence level of one or more metrics is rated as Unacceptable [score of 4]. (p. 33)

The choice of this particular cutoff structure is not science-based. Under this methodology, a study that scores 1.7 is equally weighted relative to a study that scores 2.3, despite the fact that the study with a score of 1.7 was only 0.1 away from being considered a High quality study, whereas the study scoring 2.3 was 0.1 from being considered Low quality. EPA's process amounts to nothing more than an algorithmic exercise lacking any empirical basis.

In addition, collapsing all of a study's individual data quality metrics into a single overall study score presents significant challenges. For example, studies for which many criteria are not applicable can receive higher scores than studies that have more applicable criteria, even if they score the same in overlapping metrics. For instance, EPA gives an example on page 50 of a study within only one domain containing two metrics, "Verification or Plausibility of Results" and "QSAR Models," with weighted metric scores of 2 and 1, respectively, which contribute to an overall study score of 1.5. It is reasonable to assume that another study might also have weighted scores of 2 and 1 for the same two metrics, but in addition might have another separate metric that must be scored. If this additional metric has a weighted score of 2, then this second study will receive a lower score than the first study, despite the fact that they have identical scores on their shared metrics. This means that the presence of a third relevant metric is effectively discounting the scores of the other two metrics, despite the fact that the metrics are not related.

In applying a scoring methodology to study evaluation, EPA is not only deviating from best practices in systematic review, it is deviating from the strategies applied by sources that EPA used to develop this document including IRIS and OHAT. In line with best practices in systematic review, neither of these sources uses a numerical scoring approach to rate study quality. Thus, the very sources that EPA cites as

resources used to develop its study evaluation approach explicitly state that they do not employ a scoring strategy and yet, EPA has chosen to develop a scoring methodology, without explanation or science-based justification.

We strongly urge the agency to do away with a scoring approach to evaluating study quality.

5. OPPT's approach to weighting criteria is inconsistent with best practices in systematic review; lacks empirical evidence and justification; and is entirely arbitrary.

As part of its scoring methodology, OPPT assigns greater weights to metrics that it deems more important than others. EPA refers to these as "critical metrics." However, in its 2014 review of the IRIS program, the NAS wrote that "there is no empirical basis for weighting the different criteria in the scores."[34]

OPPT's metric weights imply that the agency has some scientific basis for the degree to which a given metric criteria affects overall study quality. However, the reality is that there is no evidence to support this approach, while there is empirical evidence suggesting that quality scores and weighting lack validity, can be misleading, and introduce bias.[35]

Disregarding best practices, OPPT provides vague, substantively empty explanations for why it has assigned greater weight to certain metrics. For example, in assigning weights to data quality metrics for occupational exposure and release data, OPPT states that "EPA used expert judgement to determine the importance of a particular metric relative to others," and that "EPA judged applicability and temporal representativeness to be the most important towards overall confidence, and these two metrics were determined to be twice as important as other metrics (weighting factors assigned a value of 2)." EPA's "explanation" amounts to arbitrary, subjective judgment and is particularly dubious because EPA has not interrogated its methodology in practice.

EPA states that "the weighting approach for some of the strategies may need to be adjusted as OPPT tests the evaluation method with different types of studies." This statement highlights the arbitrary

---

[34] Nat'l Research Council, *Review of EPA's Integrated Risk Information System (IRIS) Process* at chp. 5, p. 69 (2014), https://www.ncbi.nlm.nih.gov/books/NBK230063/.

[35] *See, e.g.*, Greenland S. & O'Rourke K., *On the bias produced by quality scores in meta-analysis, and a hierarchical view of proposed solutions*, 2:4 BIOSTATISTICS 463 (Dec. 2001), https://www.ncbi.nlm.nih.gov/pubmed/12933636; Jüni P., et al., *The hazards of scoring the quality of clinical trials for meta-analysis*, 282:11 J. of the Am. Med. Ass'n 1054 (Sept. 1999), https://www.ncbi.nlm.nih.gov/pubmed/10493204; Herbison P., *Adjustment of meta-analyses on the basis of quality scores should be abandoned*, 59:12 J. OF CLINICAL EPIDEMIOLOGY 1249 (Dec. 2006), https://www.ncbi.nlm.nih.gov/pubmed/17098567.

nature of the weighting factors, and more broadly, the outright dismissal of basic tenets of systematic review. In effect, EPA is explicitly allowing a pathway for bias in its study evaluation approach, as the agency will be able to retrospectively favor some study metrics over others and adjust their weights as the results of the study evaluation process unfold—an approach that is antithetical to developing a science-based, systematic review framework.

6. The TSCA systematic review document suggests problematic use of expert judgment

OPPT indicates that expert judgment will be applied throughout its systematic review process:

> Professional judgment will be used at every step of the process and will be applied transparently, clearly documented, and to the extent possible, follow principles and procedures that are articulated prior to conducting the assessment (U.S. EPA, 2016). (p.27)

While expert judgment is certainly part of systematic review, EPA's proposed application of expert judgment raises some concerns.

Most notably, the document states that expert judgment may overrule the overall study score that has been developed through the systematic review process: "After the overall score is applied to determine an overall quality level, professional judgment may be used to adjust the quality level obtained by the weighted score calculation." (p. 34)

OPPT states that "the reviewer must have a compelling reason to invoke the adjustment of the overall score and written justification must be provided," yet few details are given. For example, it is not clear what qualifies as a "compelling reason" to alter the quality score or whose professional judgment can overrule.

While we object to OPPT's use of a scoring methodology to evaluate studies, if there exist legitimate, science-based circumstances that merit changes to a study's "confidence level," they should be factored into the TSCA systematic review document and individual protocols to the extent possible. Further, EPA must, as it has indicated it will do, identify and provide written justification for any adjustment made to overall evaluations of study quality.

7. OPPT erroneously conflates issues of reporting with study quality

OPPT's TSCA Systematic Review document incorrectly and inappropriately conflates study reporting with study quality. In doing, EPA severely jeopardizes use of best available science and weight of the scientific evidence, as the effect of EPA's approach would be to score studies as "low quality" or even exclude studies on the basis of reporting deficiencies rather than actual study quality.

Study reporting pertains to how well study authors describe various aspects of their research, including its design and findings. A well-reported study can be of poor quality and a high-quality study can be insufficiently reported.

In systematic review, individual studies are evaluated primarily for internal validity/risk of bias. The Navigation Guide, OHAT, and IRIS all emphasize internal validity/risk of bias in evaluating individual study quality. The Institute of Medicine describes internal validity as follows:

> An internally valid study is conducted in a manner that minimizes bias so that the results are likely due to a real effect of the intervention being tested. By examining features of each study's design and conduct, systematic reviewers arrive at a judgment about the level of confidence one may place in each study, that is, the extent to which the study results can be believed. Assessing internal validity is concerned primarily (but not exclusively) with an examination of the risk of bias.[36]
>
> The committee chose the term "risk of bias" to describe the focus of the assessment of individual studies and the term "quality" to describe the focus of the assessment of a body of evidence ***.[37]

The OHAT Systematic Review handbook also includes a lengthy discussion of internal validity in describing Step 4 of its systematic review approach, *Assess Internal Validity of Individual Studies*:

> Individual human, animal, and in vitro studies will be assessed for internal validity (commonly referred to as "risk of bias" (RoB) in systematic review) by considering aspects relevant for specific study designs. Assessment of risk of bias is related to but distinguished from the broader concept of assessment of methodological quality (Higgins and Green 2011).
>
> ***Bias*** is a systematic error, or deviation from the truth, in results or inferences. Biases can operate in either direction: different biases can lead to underestimation or overestimation of the true effect. Biases can vary in magnitude: some are small (and trivial compared with the observed effect), and some are substantial (so that an apparent finding may be entirely due to bias). Even a particular source of bias may vary in direction: bias due to a particular design flaw (e.g., lack of allocation concealment) may lead to underestimation of an effect in one study but overestimation in another study. It is usually impossible to know to what extent biases have affected the results of a particular study, although there is good empirical evidence that particular flaws in the design, conduct, and analysis of randomized studies lead to bias. Because the results of a study may in fact be unbiased despite a methodological flaw, it is more appropriate to consider **risk of bias** (Higgins and Green 2011).

---

[36] Inst. of Med. of the Nat'l Acads., *Finding What Works in Health Care: Standards for Systematic Reviews* at chp. 3, p. 125 (Mar. 2011), https://www.ncbi.nlm.nih.gov/pubmedhealth/PMH0079446/.
[37] *Id*. at chp. 3, p. 131.

> ***Quality*** refers to the critical appraisal of included studies to evaluate the extent to which study authors conducted their research to the highest possible standards (Higgins and Green 2011).
>
> Assessment of methodological quality is distinguished from assessment of risk of bias by Cochrane for several reasons, including the following: (1) risk of bias more directly addresses the extent to which results of included studies should be relied on; (2) a study may be performed to the highest possible standards yet still have an important risk of bias (e.g., blinding of subjects or study personnel may not have been conducted or be impossible to achieve); (3) some markers of quality in research, such as obtaining ethical approval, performing a sample-size calculation, and reporting adequately, are unlikely to have direct implications for risk of bias; and (4) an emphasis on risk of bias overcomes ambiguity between the quality of reporting and the quality of the underlying research (Higgins and Green 2011).[38]

Best practices in systematic review strongly advise against conflating issues of reporting and other aspects of study quality when assessing individual studies. While there are some differences across leading systematic review approaches for chemical assessment with how to address reporting issues, its distinction and separation from study quality is clear:

> *OHAT Systematic Review Handbook*: Missing *Information for Risk of Bias Assessment*. OHAT will attempt to contact authors of included studies to obtain missing information considered important for evaluating risk of bias. The product of the evaluation (e.g., monograph, report, or publication) will note that an attempt to contact study authors was unsuccessful if study researchers do not respond to an email or phone request within one month of the attempt to contact. If additional data or information are acquired from study authors, risk of bias judgments will be modified to reflect the updated study information.[39]
>
> *IRIS Draft chloroform assessment protocol*: Authors will be queried to obtain missing critical information, in particular, questions about relationships among variables, missing data, or additional analyses that could address potential limitations. The decision on whether to seek missing information includes consideration of what additional information would be useful, specifically with respect to any information that could result in a re-evaluation of the classification of the domains, and subsequently the overall confidence in the study. Outreach to

[38] U.S. Dep't of Health & Human Servs., Nat'l Toxicology Program, *Handbook for Conducting a Literature-Based Health Assessment Using OHAT Approach for Systematic Review and Evidence Integration* at 33-34 (Jan. 2015), https://ntp.niehs.nih.gov/ntp/ohat/pubs/handbookjan2015_508.pdf (emphases in original).
[39] *Id*. at 41.

study authors will be documented and considered unsuccessful if researchers do not respond within a reasonable amount of time to multiple e-mail or phone requests.[40]

Navigation Guide: Systematic review methodologies distinguish between study-quality criteria that can introduce a systematic error in the magnitude or direction of the result (i.e., risk of bias or "internal validity") from other methodological quality or reporting elements, which are related to important standards by which a study is conducted (e.g., adherence to human subjects and animal welfare requirements) or reported (e.g., complete information provided), but that do not systematically influence study outcomes. A study conducted to the highest methodological standards can still have important risk of bias that will affect the magnitude or direction of a study outcome.[41]

STROBE Statement: 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.[42]

In its TSCA Systematic Review document, OPPT acknowledges the need to delineate between reporting and study quality: "Reporting quality is an important aspect of a study that needs to be considered in the evaluation process. The challenge, in many cases, is to distinguish a deficit in reporting from a problem in the underlying methodological quality of the data/information source. (p. 31)"

However, OPPT then chooses an approach that deviates from this established best practice: "The TSCA evaluation strategies incorporate reporting criteria within the existing domains rather than adding a separate reporting domain as recommended in some evaluation tools/frameworks." (p. 31)

OPPT supports this decision to evaluate these metrics in parallel by stating that the aim of its approach is to "assesses reporting and methodological quality simultaneously with the idea of untangling reporting from study conduct while the reviewer is assessing a particular metric for each domain." Even on its face, this explanation is incoherent: how does assessing the two qualities "simultaneously" lead to a reviewer "untangling" the two? This approach seems likely to achieve precisely the opposite of one of its stated goals.

EPA's decision to conflate reporting issues with study quality and comingle their consideration is significant: It could well lead the agency to not use the best available science and not apply a legitimate

---

[40] U.S. EPA, Office of Research & Dev., *Systematic Review Protocol for the IRIS Chloroform Assessment (Inhalation) [CASRN 67-66-3]* at 18-19 (Jan. 2018), http://ofmpub.epa.gov/eims/eimscomm.getfile?p_download_id=534368.

[41] Tracey J. Woodruff & Patrice Sutton, *The Navigation Guide Systematic Review Methodology: A Rigorous and Transparent Method for Translating Environmental Health Science into Better Health Outcomes*, 122:10 ENVTL. HEALTH PERSPECTIVES 1007, 1009 (Oct. 2014), https://ehp.niehs.nih.gov/1307175/.

[42] STROBE STATEMENT: STRENGTHENING THE REPORTING OF OBSERVATIONAL STUDIES IN EPIDEMIOLOGY, https://www.strobe-statement.org/index.php?id=strobe-aims (last visited Aug. 16, 2018).

weight of the scientific evidence approach. For example, OPPT's scoring methodology contains, for each data quality evaluation domain, a set of "serious flaws" that cause a study to be excluded from further consideration in the review. The methodology includes instances in which reporting issues are considered fatal flaws. One of the fatal flaws for monitoring data from studies on consumer, general population and environmental exposure is that "geographic location is not reported, discussed, or referenced." (p. 99) This is inappropriate as relevant monitoring data may not be associated with a specific geographic location. For example, a consumer market survey that examines product-purchasing behaviors may be useful as proxy for estimating exposure even though it may not include location as a data field or may not publish location information in order to protect respondent privacy. The collected information could very well still be useful in ascertaining chemical exposures. Similarly, a study involving biomonitoring of children at several different childcare facilities would likely not specify the geographic location of the facilities for privacy reasons. Yet again, this information could be incredibly valuable in assessing exposure-response relationships.

Even more egregious is the profusion of reporting quality in metrics used to evaluate epidemiological studies. Insufficiencies in reporting by themselves will frequently result in data quality metric scores of low or even unacceptable. For example, absence of STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) checklist items[43] in epidemiological studies result in a metric score of unacceptable for metrics 2, 3, 4, 6, and 7, and a score of low in metric 15. A score of unacceptable in a single metric across any study quality domain will result in the exclusion of an entire study. This is wholly inconsistent with best practices in systematic review, departs from best available science, and would likely result in EPA not using reasonably available information. It also makes clear that EPA has not meaningfully, if at all, tested its systematic review approach, because if it had it would have found a number of high quality, epidemiological studies would be inappropriately excluded.

OPPT's proposed approach risks excluding scientifically sound studies; again at odds with the use of best available science, weight of the scientific evidence, and EPA's mission to protect human health and the environment.

8. The TSCA systematic review document is fraught with problematic metric criteria beyond conflation of reporting.

OPPT's TSCA systematic review document is fraught with problematic metric criteria that do not support the use of best available science. Limited time to comment has prevented an exhaustive review of all metric criteria, but below we highlight some of the problematic metric criteria identified thus far.

[43] Erik von Elm, et al., *The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies*, 61:4 J. OF CLINICAL EPIDEMIOLOGY 344 (Apr. 2008), https://www.sciencedirect.com/science/article/pii/S0895435607004362?via%3Dihub.

A. Lack of access to underlying study data will downgrade a study's score or eliminate it entirely from consideration.

For some of OPPT's data quality metrics, a study must provide underlying data in order to receive a score of "High" or even be considered. Such a standard mirrors the extensive concerns raised by EPA's *Strengthening Transparency in Regulatory Science* proposed rule,[44] a hugely problematic and widely criticized proposal.[45] As with conflating reporting quality with study quality (see comment section 7), EPA erroneously conflates access to underlying data with study quality—a deeply misguided and misleading treatment of scientific evidence.[46] Listed below are examples of where EPA inappropriately integrates underlying data access to study scoring:

- For studies on consumer, general population, and environmental exposures to receive a score of "High" in Domain 3 (Accessibility/Clarity), Metric 8 (Reporting of Results), it must meet the following standard: "Supplementary or raw data (i.e., individual data points) are reported, allowing summary statistics to be calculated or reproduced." (p. 105) If the supplementary or raw data are not reported, a study's score is automatically downgraded, regardless of its quality.

- For a human epidemiological study to receive a score of "High" in Domain 4 (Potential confounding/variable control), Metric 14 (Reproducibility of analyses), it must meet the following standard: "The description of the analysis is sufficient to understand precisely what has been done and to be reproducible." If an epidemiological study does not meet this standard, EPA will give it a score of "Low."

EPA's invoking of "reproducibility" as a standard to receive a score of "High" in these metrics mirrors similar language in the EPA's censored science proposal, raising serious concerns about the extent to which EPA is effectively requiring that all underlying study data be made publicly available to be meaningfully considered. Also see comments in section 3.A regarding EPA's use of "replicability" as a "verification" standard. EDF incorporates by reference comments submitted by EDF on EPA's proposed rule, *Strengthening Transparency in Regulatory Science*.[47]

B. OPPT will exclude occupational exposure scenarios outside the scope of the risk evaluation.

---

[44] Strengthening Transparency in Regulatory Science, 83 Fed. Reg. 18768 (proposed Apr. 30, 2018), https://www.gpo.gov/fdsys/pkg/FR-2018-04-30/pdf/2018-09078.pdf.

[45] *See, e.g.*, EDF Blogs, EDF Calls on EPA to Withdraw Censored Science Proposal (Jul. 17, 2018), http://blogs.edf.org/health/2018/07/17/edf-calls-on-epa-withdraw-censored-science/; Jennifer Sass, Health Experts Rebut Trump EPA Censoring Science Rule (Jul. 16, 2018), https://www.nrdc.org/experts/jennifer-sass/health-experts-rebut-trump-epa-censoring-science-rule.

[46] *See* EDF Comments on Strengthening Transparency in Regulatory Science, https://www.regulations.gov/docket?D=EPA-HQ-OA-2018-0259.

[47] EDF Comments on Strengthening Transparency in Regulatory Science, https://www.regulations.gov/docket?D=EPA-HQ-OA-2018-0259.

The scoring system also makes clear that OPPT intends to exclude occupational exposure scenarios that are outside the scope of the risk evaluation. For occupational exposure and release data, Domain 2 (Representative), Metric 3 (Applicability) notes that the following will cause a study to be scored "Unacceptable": "The data are from an occupational or non-occupational scenario that does not apply to any occupational scenario within the scope of the risk evaluation." (p. 76)

When EPA discovers studies of occupational or non-occupational scenarios that EPA failed to identify at the scoping stage, EPA must consider whether it needs to revise its approach to the risk evaluation by broadening the scope. TSCA orders EPA to consider "available" and "reasonably available" information in crafting a risk evaluation,[48] and if EPA discovers reasonably available information that reveals the existence of real-world occupational scenarios that EPA failed to identify earlier in the process, TSCA does not authorize EPA to simply ignore that information by labelling the information "unacceptable." Rather, the appropriate resolution is for EPA to consider whether EPA needs to expand the scope to address these real-world exposures. In most circumstances, those circumstances are now "known" to occur and EPA must analyze these known conditions of use.[49]

C. Time of data collection or publication metrics are entirely arbitrary

OPPT's scoring scheme includes data quality metrics that are scored "low" when study data are more than a certain number of years old, but EPA has provided no evidence that older information is *per se* less informative. For example, it appears that EPA intends to give monitoring data studies a low ranking for the temporal representativeness metric if their data are more than 15 or 20 years old. *See* p.77, 103, 110. While EPA provides a cursory explanation that older information is allegedly less representative than more recent information, EPA has not provided any empirical evidence supporting this weighting scheme.

The temporal representativeness metric that is applied to monitoring data from studies of occupational exposure and release highlights the arbitrary nature of OPPT's scoring approach. To receive a "High" confidence level for this metric, the data must have been collected "after the most recent permissible exposure limit (PEL) establishment or update or are generally, no more than 10 years old, whichever is shorter." (p. 77) To receive a "Medium" score, the data must meet the following requirement: "The monitoring data were collected after the most recent PEL establishment or update but are generally more than 10 years old. If no PEL is established, the data are more than 10 years but generally, no more than 20 years old." And finally, the metric is scored "Low" if the data "were collected before the most recent PEL establishment or update or are more than 20 years old if no PEL is established."

There is no empirical basis for favoring data that is fewer than 10 years old more than data that is 20 years old, nor does OPPT even attempt to provide a justification for this distinction. This scoring criteria implies that 9 year-old data is just as valid as 2 year-old data, but is more valid than 11 year-old data.

[48] 15 U.S.C. §§ 2605(b)(4)(F)(i), 2625(k).
[49] *See* 15 U.S.C. § 2602(4).

Furthermore, OPPT provides no clarification for how this metric will be applied. Will studies that are 10 years old at the time of the literature search be included in the systematic review, even if those studies are 11 years old during the data evaluation and data integration phases of the review? For longitudinal studies with multiple years' worth of data, will all of the data – or just the most recent year's data – need to fall within the stated time constraints of a given confidence level? These questions underscore the arbitrariness of the data quality criteria that OPPT's data evaluation strategy employs.

D. EPA inappropriately applies an adverse outcome pathway (AOP) standard to effect biomarkers in epidemiological studies.

For epidemiological studies, Domain 6 (Other (if applicable) Considerations for Biomarker Selection and Measurement), Metric 17 (Effect biomarker (detection/measurement/information biases)) to receive a score of "High" an effect biomarker must be a "[b]ioindicator of a key event in an AOP." (p. 245) To receive a score of "Medium" "[b]iomarkers of effect [must be] shown to have a relationship to health outcomes using well validated methods, but the mechanisms of action is not understood." It is wholly inappropriate to downgrade a study involving biomarkers just because the adverse outcome pathway for an observed effect is unknown.

For many chemicals, the biological processes underlying observed effects are not well understood or may not be understood at all. This is the case even for pharmaceuticals available on the market today. The National Research Council wrote in its 2014 report, *Review of EPA's Integrated Risk Information System (IRIS) Process*, that "if FDA were required to organize drug safety around mechanism, it would be nearly impossible to regulate many important drugs because the mechanism is often not understood, even for drugs that have been studied extensively."[50] Indeed, an earlier 2010 *Nature Medicine* editorial noted:

> It is true that we use many highly prescribed drugs without a clear idea of how they work—which targets they hit, what processes they alter and which of these actions are required for therapeutic efficacy. For instance, lithium, used to treat bipolar disorder, modulates many molecular targets, but which—or how many—of these are required for its beneficial effects is uncertain.[51]

9. The TSCA systematic review document risks discounting non-guideline studies

OPPT claims that its scoring methodology is not meant to systematically favor guideline studies over non-guideline studies:

[50] Nat'l Research Council, *Review of EPA's Integrated Risk Information System (IRIS) Process* at chp. 6, p. 90 (2014), https://www.ncbi.nlm.nih.gov/books/NBK230065/.
[51] Editorial, *Mechanism Matters*, 16:4 Nature Med. 347 (Apr. 2010), https://www.nature.com/articles/nm0410-347.pdf.

> In some cases, reference to study guidelines (in addition to professional judgement) may be helpful in determining the adequacy or appropriateness of certain study designs or analytical methods. This should not be construed to imply that non-guideline studies necessarily have lower confidence than guideline studies. [p. 35]

However, this statement is in itself contradictory. If OPPT is using study guidelines to determine the adequacy and appropriateness of study methods, then guideline studies are likely to receive the highest scores for these data quality metrics because that feature – adherence to a guideline – is used to define the criteria. On the other hand, non-guideline studies, which are more likely to deviate from these standards, will necessarily receive lower scores for these metrics.

Additionally, there are several instances in which the language of the data quality metrics suggests that guideline studies could consistently receive higher scores than non-guideline studies. For example, for experimental data derived from studies on consumer, general population, and environmental exposure (Appendix E), to receive a score of "High" in Domain 1 (Reliability), Metric 1 (Sampling Methodology and Conditions), a study must meet the following standard:

> Samples were collected according to publicly available SOPs, methods, protocols, or test guidelines that are scientifically sound and widely accepted from a source generally known to use sound methods and/or approaches such as EPA, NIST, ASTM, ISO, and ACGIH.
> **OR**
> The sampling protocol used was not a publicly available SOP from a source generally known to use sound methods and/or approaches, but the sampling methodology is clear, appropriate (i.e., scientifically sound), and *similar to widely accepted protocols for the chemical and media of interest*. All pertinent sampling information is provided in the data source or companion source. (p. 131, emphasis added)

Thus, a study must either follow standard protocols or its methods must be similar to standard guidelines for the study to receive the highest score for this metric. This could systematically favor guideline studies over non-guideline studies.

Similarly, the data evaluation criteria for *in vitro* toxicity studies (Appendix G) include language that suggests guideline studies would consistently receive higher scores than non-guideline studies. To receive a score of "High" in Domain 4 (Test Model), Metric 15 (Number per group), a study must satisfy the following requirement: "The number of organisms or tissues per study group and/or number of replicates per study group were reported and were appropriate for the study type and outcome analysis, and consistent with studies of the same or similar type." (p. 215, emphasis added)

Here, "appropriate" directs the reader to current standards and guidelines developed by OECD, EPA, and FDA.[52] On the other hand, a study would receive a score of "Medium" for this metric if it meets the following description:

> The number of organisms or tissues per study group and/or replicates per study group were reported but were lower than the typical number used in studies of the same or similar type (e.g., 3 replicates/strain of bacteria in bacterial reverse mutation assay), but were sufficient for analysis and *unlikely to have a substantial impact on results*. (p. 215, emphasis added)

Here, the basis for scoring a study as "Medium" rather than "High" is that the study did not use a standard methodology. However, to be scored a "Medium," that discrepancy cannot have affected the results significantly. This means that a study that does not use guideline methods is scored lower, despite the fact that the deviation from established methods has not affected the study's results. This would appear to systematically favor guideline studies over non-guideline studies. Similar language is found in Domain 7 (Data Presentation and Analysis), Metric 23 (Data interpretation). (p. 219)

10. OPPT should seek out methodological information that is published in other literature sources

One of the confidence levels that can be given to data quality metrics for any study type is "Not rated/applicable." This category includes instances in which "studies cite a literature source for their test methodology instead of providing detailed descriptions." (p. 33) Reviewers will only look at this cited literature source if the study under consideration "is not [otherwise] classified as 'unacceptable' during the initial review" based on an evaluation of all other data quality metrics.

Given that many of OPPT's data quality metrics focus on reporting quality (which in itself is problematic, as discussed at length in comment section 7), it is reasonable to assume that a study could score "unacceptable" based on reporting issues when, in fact, the information of interest is detailed in another information source referenced by the study authors. Rather than using a "Not rated/applicable" placeholder when a study cites a literature source for its methodology, OPPT should seek out, integrate, and consider all reasonably available information as part of evaluating study quality.

[52] *See, e.g.*, OECD GUIDELINE FOR THE TESTING OF CHEMICALS, SECTION 4, https://www.oecd-ilibrary.org/environment/oecd-guidelines-for-the-testing-of-chemicals-section-4-health-effects_20745788 (last visited Aug. 16, 2018);TEST GUIDELINES FOR PESTICIDES AND TOXIC SUBSTANCES, https://www.epa.gov/test-guidelines-pesticides-and-toxic-substances (last visited Aug. 16, 2018); U.S. Food & Drug Admin., Guidance for Industry and Other Stakeholders Toxicological Principles for the Safety Assessment of Food Ingredients (Jul. 2007), https://www.fda.gov/Food/GuidanceRegulation/GuidanceDocumentsRegulatoryInformation/IngredientsAdditivesGRASPackaging/ucm2006826.htm#TOC.

11. EPA must use at least two screeners for data screening: title/abstract screening, full text screening, and data evaluation.

OPPT notes that "one screener conducted the screening and categorization of titles and abstracts." (p. 24). This is inconsistent with best practices in systematic review, which recommend at least two individuals for all screening steps in order to minimize potential reviewer bias and ensure that all relevant data and studies are captured. As the IOM writes in its standards for systematic review in healthcare, "Without two screeners, SRs may miss relevant data that might affect conclusions about the effectiveness of an intervention. Edwards and colleagues (2002), for example, found that using two reviewers may reduce the likelihood that relevant studies are discarded."[53]

OPPT acknowledges the discrepancy between its approach and best practices in a footnote, stating that a lack of time and resources limited the office to one screener during the title/abstract screening step for the first ten chemicals. However, lack of time and/or resources is not a valid justification for failing to meet systematic review standards that empirically reduce risk of bias. Additionally, OPPT notes that the plan for future reviews is that, "Each article is *generally* screened by two independent reviewers using specialized web-based software." (p. 23, emphasis added) Similarly, for the data evaluation step OPPT states that, "Ideally, each data/information source will be screened by two reviewers, but one reviewer may be used." (p. 26).

The use of two or more independent reviewers for each step of the screening process is not a standard that should be applied generally or only when OPPT can meet ideal targets, it is one that OPPT should adhere to without exception.

12. OPPT must consider financial conflict of interest in the evaluation of both individual studies and the body of evidence.

Absent from the TSCA systematic review document is any consideration of the effect of financial conflict of interest on study results.

Empirical evidence reveals that financial conflict of interest held by study authors or sponsors can influence study results.[54] Leading systematic review organizations recognize and incorporate an

---

[53] Inst. of Med. of the Nat'l Acads., *Finding What Works in Health Care: Standards for Systematic Reviews* at chp. 3, p. 112 (Mar. 2011), https://www.ncbi.nlm.nih.gov/pubmedhealth/PMH0079446/.

[54] *See, e.g.*, Lundh A., et al., *Industry sponsorship and research outcome*, Cochrane Database of Systematic Reviews (Feb. 16, 2017), https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.MR000033.pub3/full; Donna H. Odierna, et al., *The Cycle of Bias in Health Research: A Framework and Toolbox for Critical Appraisal Training*, 20:2 ACCOUNTABILITY IN RESEARCH 127 (2013), https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3726025/; Friedman L. & Friedman M., *Financial Conflicts of Interest and Study Results in Environmental and Occupational Health Research*, 58:3 J. OF OCCUPATIONAL ENVTL. MED. 238 (Mar. 2016), https://www.ncbi.nlm.nih.gov/pubmed/26949873; Paul L. Romain, *Conflicts of interest in research:*

evaluation of financial conflict of interest at some point in the systematic review process. The authors of a 2017 Cochrane systematic review of the impact of industry sponsorship on study outcomes wrote:

> Our analysis suggests that industry sponsored drug and device studies are more often favorable to the sponsor's products than non-industry sponsored drug and device studies due to biases that cannot be explained by standard 'Risk of bias' assessment tools.
>
> ***
>
> Consequently, our data suggest that industry sponsorship should be treated as bias-inducing and industry bias should be treated as a separate domain. There are many subtle mechanisms through which sponsorship may influence outcomes, and an assessment of sponsorship should therefore be used as a proxy for these mechanisms. Interestingly, the AMSTAR tool for methodological quality assessment of systematic reviews includes funding and conflicts of interest as a domain (Shea 2007). Adaptations of Cochrane tools for assessing risk of bias in studies assessing environmental risks have also included funding source and conflicts of interest as a domain (Johnson 2016). Methods for reporting, assessing and handling industry bias and other biases in future systematic reviews must be developed. Specifically, further methodological research should focus on how industry bias is handled in Cochrane reviews.[55]

Indeed, leading scientific journals increasingly require conflict-of-interest disclosures for manuscripts, recognizing the need to have such transparency.[56] These increasingly required publication disclosures facilitate EPA's ability to collect and assess the potential impact conflicts of interest have on study results.

EPA has chosen not to collect such information in its systematic review approach. While EDF opposes conflating reporting issues with study quality, it is worth noting the conspicuous omission from data quality metrics for epidemiological studies of STROBE checklist item #22, "Give the source of funding

---

looking out for number one means keeping the primary interest front and center, 8:2 CURRENT REVIEWS IN MUSCULOSKELETAL MED. 122 (Apr. 2015), https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4596167/; Roy H. Perlis, et al., *Industry Sponsorship and Financial Conflict of Interest in the Reporting of Clinical Trials in Psychiatry*, 162 AM. J. OF PSYCHIATRY 1957 (2005), https://ajp.psychiatryonline.org/doi/pdf/10.1176/appi.ajp.162.10.1957.

[55] Lundh A., et al., *Industry sponsorship and research outcome*, Cochrane Database of Systematic Reviews (Feb. 16, 2017), https://www.cochranelibrary.com/cdsr/doi/10.1002/14651858.MR000033.pub3/full.

[56] COMPETING INTERESTS, https://www.nature.com/authors/policies/competing.html (last visited Aug. 16, 2018); SCIENCE JOURNALS: EDITORIAL POLICIES, http://www.sciencemag.org/authors/science-journals-editorial-policies#conflict-of-interest (last visited Aug. 16, 2018); CONFLICT OF INTEREST, http://www.pnas.org/page/authors/conflict-of-interest (last visited Aug. 16, 2018); JOURNAL POLICIES, https://ehp.niehs.nih.gov/journal-policies/ (last visited Aug. 16, 2018).

and the role of the funders for the present study and, if applicable, for the original study on which the present article is based."[57]

At a minimum, we recommend EPA apply OHAT's approach to considering potential impacts of conflicts of interest on individual studies and the body of evidence:

> Financial conflicts of interest (COI) related to funding source may raise the risk of bias in design, analysis, and reporting (Viswanathan et al. 2012), but there is debate on whether COI should be considered a risk of bias element (Lundh et al. 2012, Viswanathan et al. 2012, Bero 2013, Krauth et al. 2013). Currently, Cochrane recommends collecting and evaluating COI information but it is not considered a specific item in the Cochrane risk of bias tool or ACROBAT-NRSI (Higgins and Green 2011, Sterne et al. 2014) while the Navigation Guide includes COI as a risk of bias element (Johnson et al. 2014b, Koustas et al. 2014). OHAT's practice is not to exclude studies based on funding source and not to consider financial COI as a specific risk of bias element. However, OHAT collects information about funding source during data extraction and considers it at multiple points in the evaluation. 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 (Viswanathan et al. 2012). Funding source should be considered as a potential factor to explain apparent inconsistency within a body of evidence. Also, since many journals now require a COI statement regarding funding, it should be recognized that newer studies may appear to be at greater risk than older studies because of changes in journal reporting standards (Viswanathan et al. 2012).[58]

13. OPPT makes an troubling and potentially inaccurate statement about confidential business information (CBI).

EPA makes a troubling, and potentially inaccurate, assertion about the CBI status of health and safety information in the TSCA systematic review document. In the Data Collection section of the document EPA states:

> EPA/OPPT also plans to search its internal databases for data and information submitted under TSCA (e.g., unpublished industry data). EPA will consider these data in the risk evaluations where relevant and whether or not they are claimed as confidential business information (CBI). If data/information are CBI, EPA/OPPT plans to use it in a manner that protects the confidentiality of the information from public disclosure. (p. 21)

---

[57] Erik von Elm, et al., *The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies*, 61:4 J. OF CLINICAL EPIDEMIOLOGY 344 (Apr. 2008), https://www.sciencedirect.com/science/article/pii/S0895435607004362?via%3Dihub.

[58] U.S. Dep't of Health & Human Servs., Nat'l Toxicology Program, *Handbook for Conducting a Literature-Based Health Assessment Using OHAT Approach for Systematic Review and Evidence Integration* at 40 (Jan. 2015), https://ntp.niehs.nih.gov/ntp/ohat/pubs/handbookjan2015_508.pdf.

Under TSCA section 14(b)(2), health and safety studies and associated information are not eligible for protection from disclosure as CBI (subject to two narrow exceptions). 15 U.S.C. § 2613(b)(2). As with any other health and safety information, such information developed on chemicals to support the development of risk evaluations should be made publicly available.

Health and safety information is not eligible under the law for CBI protection unless it would disclose process or mixture-portionality information. Also, EPA must generally scrutinize CBI claims to ensure that they are valid and substantiated per the requirements set out in TSCA section 14, and make its confidentiality determinations publicly available, *see* 15 U.S.C. § 2625(j)(1). The information referenced in the above quotes from the TSCA systematic review document clearly encompasses "health and safety studies" under TSCA's broad definition of that phrase, TSCA section 14(b)(2), as codified in EPA's regulations at 40 C.F.R. section 716.3. EPA must make this information public. *See, e.g.*, 40 C.F.R. § 720.90(a) ("EPA will deny any claim of confidentiality with respect to information included in a health and safety study" except in limited circumstances).

14. EPA must subject the TSCA Systematic Review document to peer review by established experts in the field.

OPPT must subject the TSCA systematic review document to peer review by established experts in the field given 1) the substantial digression from best practices in systematic review; 2) EPA's decision not to adopt leading systematic review approaches for chemical assessment that have been peer reviewed and developed in consultation with systematic review experts; and 3) the significant uncertainty associated with the outcome of applying its approach, including the implications for risk determination. OPPT must ensure its general approach to protocol development and data integration is included as part of such peer review.