ENVIRONMENTAL PROTECTION NETWORK COMMENTS
"Application of Systematic Review in TSCA Risk Evaluation"
August 16, 2018

Introduction

The Environmental Protection Network (EPN), a volunteer organization of EPA alumni and others who work to preserve the nation's bipartisan progress toward clean air, water, land and climate protections, believes that the process followed to develop the TSCA systematic review process is seriously flawed. The guidance should not be applied to the risk evaluation of chemicals under TSCA or any other environmental statute until it has been properly evaluated and deemed to be at least as good as the Integrated Risk Information System (IRIS) systematic review process. Our concern stems not only from procedural irregularities, but specifics of the guidance that we believe would result eliminate important evidence of public health impacts from consideration, or give these impacts only limited weight. Its use could also result in accepted scientific findings about chemical risks and regulatory controls being reversed, and the weakening of public health and environmental protections.

TSCA requires that EPA make decisions about chemical risks based on the "best available science" and the "weight of the scientific evidence." EPA's risk evaluation rule (40 CFR Section 8702.33) 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." EPN is greatly concerned that EPA has released for public comment a new systematic review process for TSCA that does not build on the four years of progress in developing the IRIS systematic review process that has been endorsed by the National Academy of Sciences.[1] The new process for the TSCA program described in the guidance document is incomplete, has not been developed in a transparent manner with the scientific community, and departs significantly from accepted scientific principles for systematic review supported by the Institute of Medicine and adopted by the National Toxicology Program.[2,3]

[1] 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.
[2] Institute of Medicine. Finding What Works in Health Care. Standards for Systematic Review. Washington, D.C.: The National Academies Press; 2011.
[3] 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.

Several critical steps are missing from the process to adopt the "TSCA systematic review" approach. We provide the Benchmark Dose (BMD) methodology as an example of how the review process should be undertaken. In the case of BMD, EPA conducted research, held workshops, published scientific papers, sought public comment, created public domain software for practitioners to use, and wrote guidance documents – all under the auspices of the appropriate external scientific peer review process. The EPA BMD methodology is now recognized internationally because of the thorough vetting of the approach in the scientific and regulatory community. In contrast, this draft TSCA guidance has not been the subject of workshops, scientific papers, or external scientific peer review.

EPN provides specific comments in three sections below: 1) on EPA's failure to follow the proper procedures in developing this guidance, 2) on general flaws associated with the entire process as described, and 3) on critical flaws identified in assessing individual studies, using epidemiology studies as examples. (Appendix H of EPA's guidance).

1. Procedural Failures

This TSCA guidance qualifies as a "Highly Influential Scientific Assessment" as defined in the EPA Peer Review Handbook, and as such should have been subject to a comprehensive external peer review with public participation.[4] The fact that it departs substantially from current recommendations on systematic review principles indicates that the TSCA guidance is a novel approach requiring an expert panel to evaluate its scientific validity. In addition, a cross-program EPA review should have taken place under the agency's Action Development Process so that the TSCA process could have been compared to and evaluated with accepted scientific principles of systematic review. Following that rigorous internal EPA review, a federal interagency review should have been conducted under Executive Order 13422 to allow the National Toxicology Program's systematic review experts, among others, to critique the draft TSCA approach. Since none of these reviews were conducted on this draft guidance, it was inappropriate to use this guidance to evaluate the 10 chemicals currently undergoing TSCA review, as well as chemical reviews conducted under other environmental statutes.

The risk evaluation rule requires that a systematic review for these purposes "use[s] a pre-established protocol to comprehensively, objectively, transparently, and consistently identify and evaluate each stream of evidence." This draft guidance does not meet this criterion; therefore its use in evaluating the 10 TSCA chemicals is in clear violation. It also raises the question of why the existing IRIS systematic review process was not used.

[4] U.S. Environmental Protection Science and Technology Council. Agency Peer Review Handbook 4th Edition; October 2015.

2. Guidance Flaws

EPN describes three critical flaws in the draft TSCA guidance: 1) failure to include protocols or guidance to synthesize evidence within each of the seven evidence domains, and to combine the evidence from all domains into a coherent summary, 2) use of an arbitrary quantitative scoring system for assessing individual studies, with no validation, and 3) failure to adopt adequate implementation procedures for conducting the systematic review. EPN also describes how one of the agency's systematic review processes (used for IRIS) has none of these critical flaws.

a. The TSCA guidance fails to include a protocol for synthesizing the body of evidence selected for inclusion in the systematic review

The Institute of Medicine identified five steps in conducting systematic reviews: 1) formulating the topic, 2) developing the systematic review protocol, 3) finding and assessing individual studies, 4) synthesizing the body of evidence, and 5) providing a detailed comprehensive final report.[5] The TSCA draft guidance document acknowledges all five steps but provides details only for steps one through three, focusing most heavily on assessing individual data sources and studies for inclusion in a systematic review. The TSCA guidance on "Data Integration and Summary of Findings" (p. 26) states that this critically important step will be done but provides no information on how it will be done. The TSCA guidance lacks any protocol for determining the strengths and relevance of the selected individual studies, grouping them into streams of evidence from each of the seven domains, and integrating the findings from all domains into a coherent summary with a set of judgments about the weight of the evidence as a whole. This omission of critical steps in systematic review disqualifies the guidance from use because it does not meet the TSCA risk evaluation rule requirement; the systematic review must use 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."

While the IRIS Handbook is in the process of being updated to describe in detail its systematic review process, EPA presented the key elements to the National Academy of Sciences at a workshop on February 1-2, 2018. Unlike the TSCA draft guidance, the IRIS systematic review process covers all five steps identified by the Institute of Medicine. The IRIS approach applies the principles of systematic review to identify pertinent studies of animal and human health effects, to evaluate the strengths of study methods and quality, to synthesize the body of evidence, to integrate evidence for each health outcome, and to select studies for

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

derivation of toxicity values. The IRIS systematic review process for TSCA chemical risk evaluations would provide a more comprehensive approach than use of the incomplete draft TSCA guidance.

b. The guidance uses an arbitrary quantitative scoring system for assessing individual studies

The second critical flaw in the draft TSCA guidance is the use of an arbitrary and untested numerical scoring system which assigns, based on the professional judgment of one or two reviewers, numerical values for quality domains and then sums up those values to decide whether a study is high, medium, low, or unacceptable quality. None of the widely accepted systematic review methodologies in use today employ numerical scoring systems, and both the Cochrane Collaboration and National Academy of Sciences (NAS) recommend strongly against such scoring systems because they are arbitrary and not science-based.[6,7] The Cochrane Collaboration, founded in 1993, is an international non-profit, independent organization which includes the world's most authoritative expertise on systematic review methods. The Cochrane Collaboration warns that calculating a score involves choosing appropriate weights for each subcomponent of a study, and such scaling is nearly impossible to justify. The NAS explains that in order to assign a scientifically justified measure the reviewer would have to know how much each risk of bias domain contributes to study quality, and the domains would have to be independent of each other. The Cochrane Collaboration further explains that scoring systems inappropriately mix criteria that assess risk of bias with criteria that reflect the quality of reporting. That is a concern with this TSCA guidance, which lacks any commitment to request additional information from the authors of relevant studies, only mentioning that such requests might be made after the initial screen of the literature. Risk of bias reflects study-design characteristics that can introduce a systematic error that might affect the magnitude and even the direction of the apparent effect. Potential biases must be assessed to determine how confidently conclusions can be drawn from a study. A critical flaw of the draft TSCA guidance is its focus on reporting limitations that do not negate a study's value in demonstrating health risks. A study might be well designed to eliminate bias, which would make it valuable for consideration; however, because the study failed to report details in the publication under review the TSCA guidance would assign it a low score or deem it unacceptable. Reporting requirements are known to vary among technical journals which have different allowances for

[6] 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; 2011. https://us.cochrane.org

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

details based on the expected audience and space limitations. The TSCA scoring system for study quality and the formula for calculating a composite score lack empirical support, nor have they been evaluated or "ground truthed," as is the common practice in developing scoring approaches.

EPN notes that the IRIS systematic review process followed the recommendations of the National Academy of Sciences and does not include a numerical scoring system.[8] Instead, the IRIS approach provides detailed criteria for assessing the quality of data sources and studies, which are appropriately focused on identifying the risk of biases rather than reporting limitations. For example, IRIS evaluation of epidemiology studies is based on the Cochrane risk of bias approach, modified for environmental and occupational exposures.[9] While the IRIS systematic review process identified similar domains for epidemiology studies as the draft TSCA guidance, the IRIS approach deems a study unacceptable only when there is a bias that would produce a substantive change in the estimated effect estimate.

c. <u>The guidance fails to adopt adequate implementation procedures for systematic reviews.</u>

The third critical flaw in the TSCA draft guidance is the failure to adopt adequate implementation procedures for the systematic review. The Cochrane Collaboration requires that at least two reviewers with appropriate expertise assess each study to minimize bias, and recommends that a conflict resolution process include an additional reviewer to resolve differences in ratings between the reviewers. The draft TSCA guidance does not identify the expertise needed to review studies in any of the seven topics for which it provides a numerical scoring system: physical-chemical properties; environmental fate; occupational exposure and release; exposures to the general population, consumers and the environment; ecological hazard studies; animal toxicity and *in vitro* toxicity; and epidemiology studies. Further, the guidance states that only one or at most two reviewers will be employed at any phase of the review, and it is vague about conflict resolution among reviewers, indicating only that the reviewers will seek consensus. A further concern about implementation procedures is the lack of emphasis on the need to query authors for additional information if necessary data are not reported in the publication under review. It should be standard practice that EPA give authors of relevant studies an opportunity to provide additional information beyond that provided in a publication.

EPN notes once again that the IRIS systematic review process does not suffer from any of these implementation failings. It is clear in that process that a minimum of two reviewers will

[8] NRC 2014.

[9] Sterne, Hernan, et al. ROBINS-I: a tool for assessing risk of bias in non-randomized studies of interventions. BMJ 2016; 355: i4919.

be used with appropriate expertise, and it is standard practice to ask authors of relevant studies to provide additional information if needed to evaluate the study quality and risk of bias.

3. Flaws in TSCA guidance that could eliminate reliable and relevant data from inclusion in systematic review

EPN believes that the application of this draft TSCA guidance will result in the exclusion of quality research in all seven of the topic areas covered. We provide detailed comments below on the evaluation of epidemiologic studies, as we believe this area may be the most affected.

The draft guidance provided for assessing epidemiologic studies is intended to cover the following study designs: controlled exposure, cohort, case-control, cross-sectional, and case crossover. Studies are to be evaluated in six data quality domains: study participation, exposure characterization, outcome assessment, potential confounding/variability control, analysis, and other/consideration for biomarker selection and measurement. Each of the six domains is evaluated using two to seven metrics for a total of 19 metrics. In addition, differential weights are assigned to each metric. According to the guidance, studies with even one metric scored as unacceptable will be excluded from use in a chemical's risk evaluation.

a. General comment on scoring

The assignment of equal weight to each of the "evaluation domains" is arbitrary and not based on evidence. Within each category, the assignment of "metric weighting factor" is also arbitrary, and each metric is limited to two values (X or 2X) (a similar scheme is used for the animal and in vitro studies), with the values dependent on the number of metrics in the category. The validity of this approach is untested and, given the arbitrary input values, may or may not be an accurate reflection of study quality. Also, the metrics mix study quality and reporting quality, as noted earlier, is discouraged by other systematic review expert advice.

b. General comment on information missing from published reports

The reasons for "unacceptable" ratings for nearly all items include information "not reported." While the possibility of contacting authors to obtain additional information is mentioned in the body of the report, there is no acknowledgment in the tables of such filling in of information. There are many reasons for information not to appear in a published report but to be nonetheless available. If the aim is to base decisions on the totality of the reliable evidence, considerable effort should be placed on filling in gaps where possible. (See earlier discussion of this point.)

c. Comment on using STROBE criteria for reporting

Many of the criteria for epidemiologic studies cite the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement. STROBE provides widely respected guidance on the reporting of the types of observational studies that could be included in TSCA reviews. The STROBE developers state:

> We emphasize that the STROBE Statement was not developed as a tool for assessing the quality of published observational research. Such instruments have been developed by other groups and were the subject of a recent systematic review [28]. In the Explanation and Elaboration paper, we used several examples of good reporting from studies whose results were not confirmed in further research--the important feature was the good reporting, not whether the research was of good quality.[10]

This clarifies the distinction that the STROBE criteria relate to the quality of study *reporting*, but not necessarily the quality of the *research*. Appreciation of this distinction is lacking in guidance document. EPN is not opposed to considering the quality of reporting, but we do not believe that a missing data item should form the basis for excluding studies. Also, using these criteria are likely to handicap older studies that precede the 2007 publication of the STROBE criteria.

d. Comments on Study Participation

The evaluation domains and metrics listed are generally appropriate, but are not well differentiated or explained. These examples illustrate problems in the "study participation" evaluation domain.

i. Three metrics are listed under "study participation": participant selection, attrition, and comparison group. However, the comparison group are also participants and should be subsumed under participant selection, leaving just two categories. This would affect the arbitrary scoring in this category. If the authors intended to separate cases and controls, or exposed and unexposed into two metrics, they should state this clearly. This does not appear to be the case. This error may suggest a lack of understanding of the variety of epidemiologic study designs.

ii. "Participant selection" is chosen as one of the critical metrics, with this given as the rationale:

> The participants selected for the study must be representative of the target population. Differences between participants and nonparticipants determines the amount of bias present, and differences should be well-described. (Galea and Tracy 2007)

[10] Elm, E. Von, Altman, D. G., Egger, M., Pocock, S. J., Gøtzsche, P. C., & Vandenbroucke, J. P. (2008). The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement : guidelines for reporting observational studies, *61*, 344–349. https://doi.org/10.1016/j.jclinepi.2007.11.008

This is presented as a critical metric for "participant selection." We agree that participants (cases/controls, exposed/unexposed, exposed vs. unexposed time periods for case-crossover studies) should be carefully selected for all study types. Bias, however, as suggested in the second sentence, is a result of many factors, not just "nonparticipants." The paper referenced as support for this metric is largely about participants and nonparticipants in surveys and prospective studies, which make up only a portion of the study types (e.g., most studies relying on retrospective records would not have "nonparticipants," but still be subject to bias, which should be assessed). This also suggests a lack of appreciation for the differences among epidemiologic study designs.

e. Comments on Potential Confounding/Variability Control

i. "Variability control" is not a standard epidemiologic term, suggesting a possible lack of familiarity with epidemiologic terminology. It should be defined, deleted, or changed to a meaningful term.

ii. Two of the three metrics in this evaluation domain are the same or similar: Co-exposure Confounding/Moderation/Mediation and Covariate Adjustment. The point of covariate adjustment is to reduce or eliminate bias or confounding from any source. A covariate may be a personal characteristic, an exposure, or some other feature. Without further explanation, it would be difficult to apply these metrics independently.

Conclusion

EPN recommends that EPA continue to develop and evaluate this draft guidance and commit to submitting it to a thorough scientific and interagency review before applying it in regulatory reviews. EPA should use the IRIS systematic review process for evaluation of chemical risks under TSCA, including for the 10 chemicals currently under consideration. The IRIS protocol can be applied immediately because it has already been peer reviewed and endorsed by the National Academy of Sciences.

Respectfully submitted,

Michelle Roos
Executive Director, Environmental Protection Network
michelle.roos@environmentalprotectionnetwork.org

These comments were prepared by Betsy Southerland, Hellen Gelband, Robert Kavlock, Trish Komen, and Rita Schoney, on behalf of the Environmental Protection Network. Questions should be addressed to Betsy Southerland, easydee420@gmail.com.

For further information:

Visit our website at www.environmentalprotectionnetwork.org
Email EPN at info@environmentalprotectionnetwork.org or call 202-656-6229