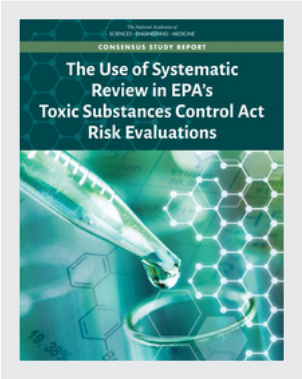

The Use of Systematic Review in EPA's Toxic Substances Control Act Risk Evaluations (2021)

DETAILS

82 pages | 8.5 x 11 | PAPERBACK

ISBN 978-0-309-68386-9 | DOI 10.17226/25952

CONTRIBUTORS

Committee to Review EPA's TSCA Systematic Review Guidance Document; Board on Environmental Studies and Toxicology; Division on Earth and Life Studies; National Academies of Sciences, Engineering, and Medicine

SUGGESTED CITATION

National Academies of Sciences, Engineering, and Medicine. 2021. *The Use of Systematic Review in EPA's Toxic Substances Control Act Risk Evaluations*. Washington, DC: The National Academies Press. https://doi.org/10.17226/25952.

The Use of Systematic Review in EPA's Toxic Substances Control Act Risk Evaluations

Committee to Review EPA's TSCA Systematic Review Guidance Document

Board on Environmental Studies and Toxicology

Division on Earth and Life Studies

A Consensus Study Report of

The National Academies of

SCIENCES • ENGINEERING • MEDICINE

THE NATIONAL ACADEMIES PRESS

Washington, DC

www.nap.edu

THE NATIONAL ACADEMIES PRESS 500 Fifth Street, NW Washington, DC 20001

This activity was supported by a contract between the National Academy of Sciences and the U.S. Environmental Protection Agency. Any opinions, findings, conclusions, or recommendations expressed in this publication do not necessarily reflect the views of any organization or agency that provided support for the project.

International Standard Book Number-13: 978-0-309-68386-9
International Standard Book Number-10: 0-309-68386-6
Digital Object Identifier: https://doi.org/10.17226/25952

Additional copies of this publication are available from the National Academies Press, 500 Fifth Street, NW, Keck 360, Washington, DC 20001; (800) 624-6242 or (202) 334-3313; http://www.nap.edu.

Printed in the United States of America

Suggested citation: National Academies of Sciences, Engineering, and Medicine. 2021. *The Use of Systematic Review in EPA's Toxic Substances Control Act Risk Evaluations*. Washington, DC: The National Academies Press. https://doi.org/10.17226/25952.

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

The **National Academy of Sciences** was established in 1863 by an Act of Congress, signed by President Lincoln, as a private, nongovernmental institution to advise the nation on issues related to science and technology. Members are elected by their peers for outstanding contributions to research. Dr. Marcia McNutt is president.

The **National Academy of Engineering** was established in 1964 under the charter of the National Academy of Sciences to bring the practices of engineering to advising the nation. Members are elected by their peers for extraordinary contributions to engineering. Dr. John L. Anderson is president.

The National Academy of Medicine (formerly the Institute of Medicine) was established in 1970 under the charter of the National Academy of Sciences to advise the nation on medical and health issues. Members are elected by their peers for distinguished contributions to medicine and health. Dr. Victor J. Dzau is president.

The three Academies work together as the **National Academies of Sciences, Engineering, and Medicine** to provide independent, objective analysis and advice to the nation and conduct other activities to solve complex problems and inform public policy decisions. The National Academies also encourage education and research, recognize outstanding contributions to knowledge, and increase public understanding in matters of science, engineering, and medicine.

Learn more about the National Academies of Sciences, Engineering, and Medicine at **www.nationalacademies.org**.

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

Consensus Study Reports published by the National Academies of Sciences, Engineering, and Medicine document the evidence-based consensus on the study's statement of task by an authoring committee of experts. Reports typically include findings, conclusions, and recommendations based on information gathered by the committee and the committee's deliberations. Each report has been subjected to a rigorous and independent peer-review process and it represents the position of the National Academies on the statement of task.

Proceedings published by the National Academies of Sciences, Engineering, and Medicine chronicle the presentations and discussions at a workshop, symposium, or other event convened by the National Academies. The statements and opinions contained in proceedings are those of the participants and are not endorsed by other participants, the planning committee, or the National Academies.

For information about other products and activities of the National Academies, please visit www.nationalacademies.org/about/whatwedo.

COMMITTEE TO REVIEW EPA'S TSCA SYSTEMATIC REVIEW GUIDANCE DOCUMENT

Members

Jonathan M. Samet (NAM) (*Chair*), Colorado School of Public Health, Aurora
Deborah H. Bennett, University of California, Davis
Bryan W. Brooks, Baylor University, Waco, TX
Jessica L. Myers, Texas Commission on Environmental Quality, Austin
Kristi Pullen Fedinick, Natural Resources Defense Council, Washington, DC
Karen A. Robinson, Johns Hopkins University School of Medicine, Baltimore, MD
Joseph V. Rodricks, Ramboll, Arlington, VA
Katya Tsaioun, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD
Yiliang Zhu, University of New Mexico, Albuquerque

Staff

Elizabeth Barksdale Boyle, Project Director
Clifford Duke, Director, Board on Environmental Studies and Toxicology
Andrea Hodgson, Senior Program Officer
Tamara Dawson, Program Coordinator

Sponsor

U.S. Environmental Protection Agency

BOARD ON ENVIRONMENTAL STUDIES AND TOXICOLOGY

Members

William H. Farland (*Chair*), Colorado State University, Fort Collins
Dana Boyd Barr, Emory University, Atlanta, GA
Ann M. Bartuska, U.S. Department of Agriculture (retired), Washington, DC
E. William Colglazier, American Association for the Advancement of Science, Washington, DC
Francesca Dominici, Harvard University, Boston, MA
George Gray, The George Washington University, Washington, DC
R. Jeffrey Lewis, ExxonMobil Biomedical Sciences, Inc., Annandale, NJ
Germaine M. Buck Louis, George Mason University, Fairfax, VA
Linsey C. Marr, Virginia Polytechnic Institute and State University, Blacksburg
R. Craig Postlewaite, U.S. Department of Defense, Burke, VA
Reza J. Rasoulpour, Corteva Agriscience, Indianapolis, IN
Ivan Rusyn, Texas A&M University, College Station
Deborah L. Swackhamer, University of Minnesota, St. Paul
Joshua Tewksbury, Future Earth, Boulder, CO
Sacoby M. Wilson, University of Maryland, College Park

Staff

Clifford Duke, Director
Raymond A. Wassel, Scholar and Director of Environmental Studies
Kaley Beins, Associate Program Officer
Laura Llanos, Finance Business Partner
Tamara Dawson, Program Associate
Cesar Segovia, Program Assistant

Reviewers

This Consensus Study Report was reviewed in draft form by individuals chosen for their diverse perspectives and technical expertise. The purpose of this independent review is to provide candid and critical comments that will assist the National Academies of Sciences, Engineering, and Medicine in making each published report as sound as possible and to ensure that it meets the institutional standards for quality, objectivity, evidence, and responsiveness to the study charge. The review comments and draft manuscript remain confidential to protect the integrity of the deliberative process.

We thank the following individuals for their review of this report:

Lisa Bero, University of Colorado
Anna Beronius, Karolinska Institutet
Carol J. Burns, Burns Epidemiology Consulting
Anne Fairbrother, Exponent
Elaine Faustman, University of Washington
R. Jeffrey Lewis, ExxonMobil Biomedical Sciences
David M. Michaels, The George Washington University
Ivan Rusyn, Texas A&M University
Paul Whaley, Lancaster University

Although the reviewers listed above provided many constructive comments and suggestions, they were not asked to endorse the conclusions or recommendations of this report nor did they see the final draft before its release. The review of this report was overseen by **David Eaton (NAM)**, University of Washington, and **Gary Ginsberg**, New York State Department of Health. They were responsible for making certain that an independent examination of this report was carried out in accordance with the standards of the National Academies and that all review comments were carefully considered. Responsibility for the final content rests entirely with the authoring committee and the National Academies.

Preface

In 2016, the Frank R. Lautenberg Chemical Safety for the 21st Century Act ("Lautenberg Act") (Pub L No. 114-182) was signed to overhaul the Toxic Substances Control Act (TSCA), 40 years after the original act was passed. The Lautenberg Act requires the U.S. Environmental Protection Agency's (EPA's) Office of Pollution Prevention and Toxics (OPPT) to evaluate chemicals existing before the original 1976 TSCA was amended. Given that several committees of the National Academies have recommended that EPA use systematic review to improve transparency and objectivity of risk-based decisions, it is commendable that OPPT has begun to apply systematic review methods in the risk evaluations produced under TSCA. However, OPPT is under unique challenges in embarking on the use of systematic review in TSCA risk evaluations, due to the ambitious statutory deadlines, the diverse evidence streams considered, and the need to consider many different uses of the chemicals that undergo the evaluations.

In this report, the Committee to Review EPA's TSCA Systematic Review Guidance Document offers practical recommendations that EPA's OPPT could use to improve use of systematic review and more generally evidence-based practices within the risk evaluations.

The committee gratefully acknowledges the following for their presentations: Yousuf Ahmad, Stanley Barone, Amy Benson, Susanna Blair, Francesca Branch, Iris Camacho, Marcy Card, Kellie Fay, Tala Henry, Ariel Hou, Kara Koehrn, Yadi Lopez, Amelia Nguyen, Chantel Nicolas, Nerija Orentas, Katherine Philips, Tameka Taylor, Amina Wilkins, and Eva Wong from EPA OPPT, who described and answered questions on the processes used in TSCA risk evaluations. Others who provided presentations and public testimony include Julie Goodman, Gradient; Suzanne Hartigan and Steve Risotto, American Chemistry Council; Patricia Koman, University of Michigan School of Public Health; Jennifer McPartland, Environmental Defense Fund; Robert Sussman, Sussman & Associates; Anthony Tweedale, R.I.S.K. Consultancy; Daniele Wikoff, ToxStrategies; and Tracey Woodruff, University of California, San Francisco.

The committee is also grateful for the assistance of the National Academies' staff in preparing this report. Staff members who contributed to this effort are Elizabeth Boyle, project director; Clifford Duke, director of the Board on Environmental Studies and Toxicology; Andrea Hodgson, senior program officer; and Tamara Dawson, program coordinator.

I would especially like to thank the committee members for their efforts throughout the development of this report.

Jonathan M. Samet, *Chair*
Committee to Review EPA's TSCA Systematic
Review Guidance Document

Contents

FIGURES

TABLES

Summary[1]

Exposures to industrial chemicals in food, water, air, and consumer products can cause harm to human health and the environment. Risk assessment is a key public policy tool to inform decision making to protect public health and ecological receptors[2] from unsafe environmental exposures to chemicals. In recent years, there has been a trend to apply systematic review for gathering evidence within the risk assessment process to increase transparency, objectivity, and reproducibility. Consequently, the U.S. Environmental Protection Agency (EPA) has been adopting systematic review within its risk assessment processes since the 2011 National Research Council review of the Integrated Risk Information System (IRIS) Program's formaldehyde assessment. In 2016, when the Frank R. Lautenberg Chemical Safety for the 21st Century Act ("Lautenberg Act") (Pub L No. 114-182) was signed to overhaul the Toxic Substances Control Act (TSCA) after 40 years, many stakeholders called for the adoption of systematic review within these important risk assessments.

The Lautenberg Act provides EPA's Office of Pollution Prevention and Toxics (OPPT) with increased authority to regulate chemicals existing before the original 1976 TSCA was amended. To exert this new authority, EPA needed to promulgate rules to implement the new requirements and responsibilities under the law. The conduct of risk evaluations is determined by the Procedures for Chemical Risk Evaluation Under the Amended Toxic Substances Control Act, often referred to as the "Risk Evaluation Rule" (40 CFR Part 702, 82 FR 33726). The statute requires that EPA must establish by rule a process for risk evaluation and that the risk evaluation will contain a scope or problem formulation, a hazard assessment, an exposure assessment, a risk characterization, and a determination of unreasonable risk.[3] The new authority afforded to EPA came with a timetable that imposed tight deadlines on OPPT as it assembled teams, promulgated rules, and drafted the guidance documents and operating procedures that prescribe how OPPT exerts its new authority. The Lautenberg Act also required strict statutory deadlines of 9 to 12 months for chemical prioritization, the process to determine which chemicals should undergo risk evaluations. The first 10 high-priority chemicals then underwent risk evaluations, which were to be completed within 3 years of initiation. An additional 20 were to follow.

Within the Risk Evaluation Rule, EPA chose only to define terms used within the statute, including "best available science" and "weight of the scientific evidence." The term "systematic review" appears within the definition of weight of the scientific evidence; EPA chose to leave a reference to systematic review in the preamble of the Risk Evaluation Rule but not to codify a definition. Furthermore, EPA states that it will use a systematic review approach not only for the hazard assessment but also throughout the risk evaluation process.

As defined by the 2011 Institute of Medicine report *Finding What Works in Health Care: Standards for Systematic Reviews*, systematic review is "a scientific investigation that focuses on a specific question and uses explicit, prespecified scientific methods to identify, select, assess, and summarize the findings of similar but separate studies." Systematic review has become the foundation for assessing evidence to be used for decision making in a variety of health contexts, including health care and public health.

[1]This Summary does not include references. Citations for the findings presented in the Summary appear in the subsequent chapters.

[2]Ecological receptors include any living organisms other than humans, the habitat that supports such organisms, or natural resources that could be adversely affected by environmental contaminations resulting from a release at or migration from a site.

[3]In TSCA, the risk assessments are termed risk evaluations because they contain the risk determination, an element that is traditionally outside the risk assessment process.

Well-conducted systematic reviews methodically identify, select, assess, and synthesize the relevant body of research, and clarify what is known and not known about the potential benefits and harms of the exposure being researched.

In 2018, after beginning the first 10 chemical risk evaluations under the Lautenberg Act, OPPT released the document *Application of Systematic Review in TSCA Risk Evaluations* to guide the agency's selection and review of studies. OPPT did not directly draw on existing methods, such as those being used and developed by the European Food Safety Authority, the Office of Health Assessment and Translation (OHAT) of the National Toxicology Program, the Navigation Guide, the Texas Commission on Environmental Quality, the World Health Organization, and the International Labour Organization, for occupational exposures. These methods, however, apply systematic review only to the hazard assessment portion of a risk assessment. Instead, OPPT developed a new approach that applies systematic review to the hazard assessment, the exposure assessment, data on physical and chemical properties, and other components for which systematic review is not generally applied. The approach taken by OPPT is presumably based on the definition of weight of evidence (WOE) in the Risk Evaluation Rule and the decision by OPPT to apply a systematic review method for evaluation of the evidence streams outside those included in the hazard assessment.

THE COMMITTEE'S APPROACH

EPA requested that the National Academies of Sciences, Engineering, and Medicine convene a committee to review EPA's 2018 guidance document on *Application of Systematic Review in TSCA [Toxic Substances Control Act] Risk Evaluations* and associated materials (see Box S-1; the full Statement of Task is included in Chapter 1). The committee considered public comments on the document, EPA's responses to public comments, and enhancements to the systematic review process reflected in documentation of the first 10 chemical risk evaluations. The committee was also asked to make recommendations for enhancements to EPA's 2018 guidance document.

BOX S-1 The Committee's Approach to the Evaluation

- The committee read and critiqued *Application of Systematic Review in TSCA Risk Evaluations*.
- To consider the enhancements made to the systematic review approach in the first 10 risk evaluations, the committee reviewed the Draft Risk Evaluation for Trichloroethylene (TCE) and the Risk Evaluation for 1-Bromopropane (*n*-Propyl Bromide) (1-BP). These assessments were chosen because at the first meeting, OPPT suggested that the 1-BP risk evaluation was a "typical risk evaluation," and later OPPT suggested that the TCE risk evaluation was the "best example of integration." The committee also critiqued the systematic review of the toxicology and epidemiology studies within the TCE risk evaluation using the "assessment of multiple systematic reviews" (AMSTAR-2) measurement tool.
- To consider enhancements to the TSCA systematic review process beyond the first 10 risk evaluations, the committee considered oral and poster presentations provided by OPPT at the committee's public virtual meetings.
- The committee posed questions to OPPT following the first committee meeting to clarify aspects of the use of systematic review in the risk evaluations.
- The committee then considered all of this information to determine whether the TSCA systematic review process is comprehensive, workable, objective, and transparent—the core evaluation measures of the Statement of Task. The committee also provides its recommendations with regard to steps that could be taken to improve the evaluation process as described in the 2018 guidance document and further elaborated in the evaluations considered by the committee.

SYSTEMATIC REVIEW

Over the past decade, several approaches to applying systematic review for assessing evidence on risks of environmental agents have been elaborated, drawing on methods developed in other fields. These approaches to systematic review for chemical risk assessments have been typically applied to research questions about hazards to humans or ecosystems.

Figure S-1 provides a schema for how systematic review can be conducted to inform hazard assessment and make risk determinations. Figure S-2 illustrates the OPPT approach to systematic review within TSCA risk evaluations, which differs to an extent from the generic approach in Figure S-1 in that OPPT applies systematic review to all elements of the risk evaluation. Prior to the conduct of a systematic review, planning and problem formulation should take place. The planning and problem formulation step should include stakeholder engagement, broad literature searching to map the evidence on the topic, and identification of the most important questions and the best approach for answering such questions. The research questions and the approach should inform the first step of the systematic review, the development of the protocol. Each research question entails, in a sense, a separate systematic review; thus, the protocol should contain a Population (including animal or plant species), Exposure, Comparator, and Outcome (PECO) statement, the inclusion and exclusion criteria, and the plans for the synthesis of the data for each research question. Evidence identification, which includes searching and screening the literature and finding the reports as prescribed in the protocol, is the next step. Evidence evaluation follows. This step includes evaluation of the internal validity (i.e., Is the study biased?) of the individual studies, usually by using an accepted and evaluated risk-of-bias assessment tool appropriate to the design of the studies considered. Evidence synthesis should consist of a qualitative evaluation of the evidence, which can be complemented by a quantitative pooling of the data or a meta-analysis. The synthesis is completed by evaluation of the confidence in the overall body of evidence for a given data stream (e.g., human, animal or ecological receptors, or mechanistic) and its endpoints. This synthesis of the various, specific streams of evidence is followed by hazard assessment with integration of the multiple evidence streams of human, animal or ecological receptors, and mechanistic. Questions about human and ecological exposures could also be evaluated with systematic review, but systematic review tools for gathering and evaluating exposure data are not well developed. Exposure and hazard data are integrated to characterize risk (see Figure S-1).

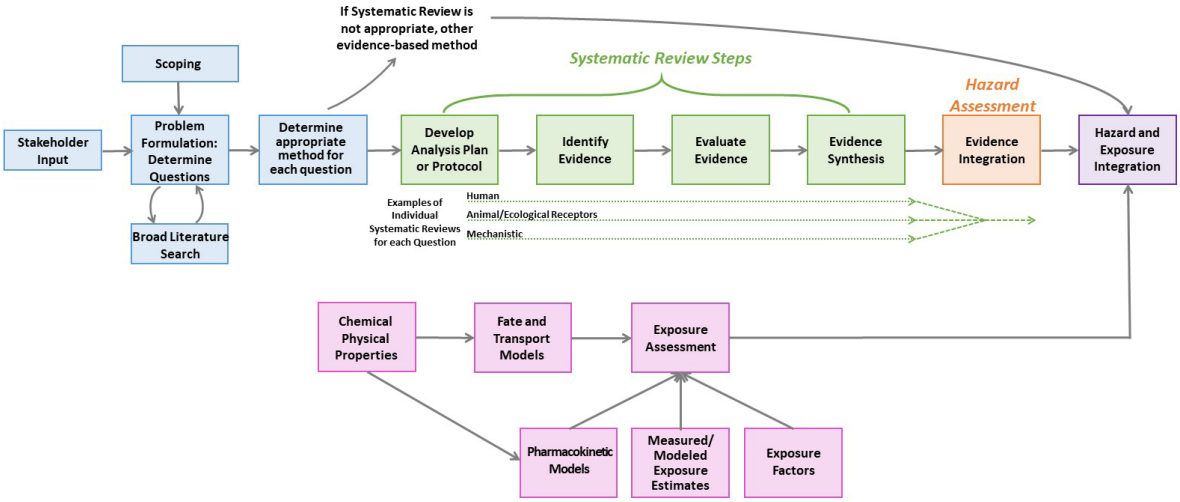

FIGURE S-1 Example approach of systematic review in the context of risk assessment. The blue boxes refer to steps that are conducted prior to the systematic review, green denotes the systematic review process, orange denotes the hazard assessment, and purple is the integration of hazard and exposure. The pink boxes refer to the exposure assessment, which is conducted outside of the systematic review but is used to make the final risk characterization.

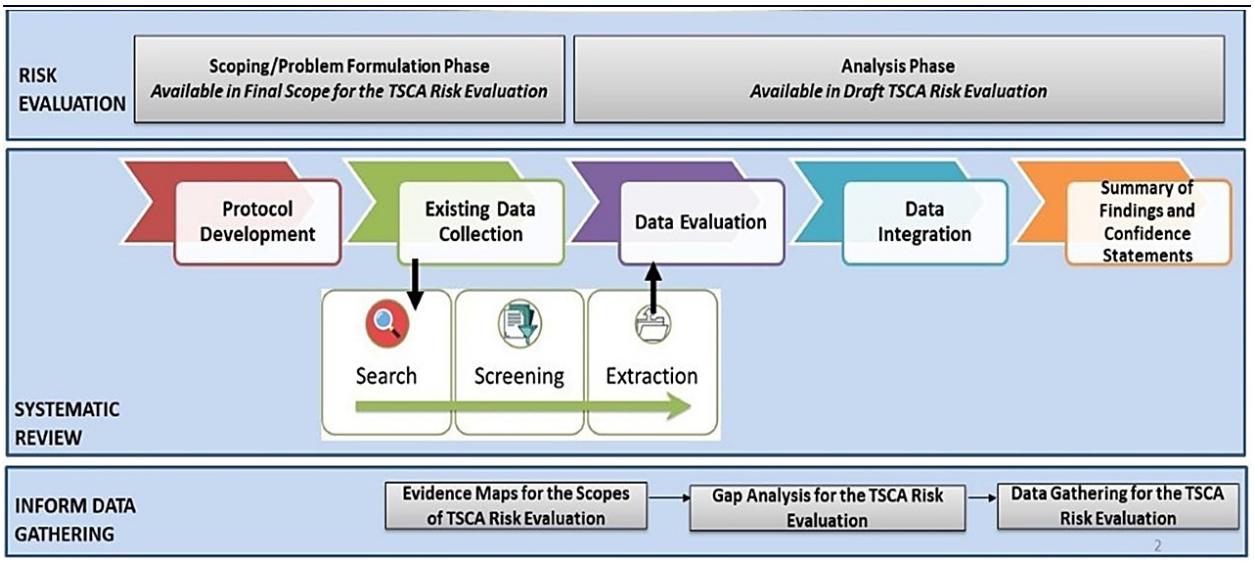

FIGURE S-2 The systematic review process for TSCA risk evaluations. SOURCE: U.S. Environmental Protection Agency, presentation to the committee, June 19, 2020.

Systematic review approaches have been widely used to assemble the evidence needed to assess human health and ecological receptors. Yet, the use of systematic review to collect, evaluate, and synthesize evidence streams that contribute to the exposure assessment of human and ecological receptors is not established and there is very little precedent for applying systematic review to these streams of evidence. Within the agency, the guidance that dictates how exposure, fate and transport, and physical chemical property data should be assembled for decisions about risks to human health and ecological receptors is contained in the Guidelines for Human Exposure Assessment, the Guidelines for Ecological Risk Assessment, and the operating procedures for the use of the ECOTOXicology (ECOTOX) knowledgebase.

Figure S-2 illustrates OPPT's approach to systematic review, which differs to an extent from the above description and includes the systematic review as part of the broader process of risk evaluation. The OPPT process has problem formulation and scoping occurring somewhat in parallel with the protocol development and data collection process for both the hazard assessment and the exposure assessment. Within this step, OPPT uses a variety of software tools and approaches to conduct broad searching and to map the available evidence. OPPT uses exhaustive search strategies that include major scientific databases, backward searching for studies in previous chemical risk assessments, additional gray literature sources, studies submitted under TSCA, and studies identified in peer review. OPPT then screens the titles and abstracts against its list of needs for the evaluation (although it was unclear if the PECO statements are always used or if a list of data needs is used, as for the TCE evaluation). Next, it conducts a full-text screening of the papers to determine relevance prior to extraction for evaluation.

To evaluate the evidence, OPPT has developed an extensive de novo critical appraisal tool, termed TSCA's "fit-for-purpose evaluation framework," which is applied to human, animal or ecological receptors, mechanistic, exposure, fate, and chemical–physical property studies. OPPT has stated that the evaluation strategies were developed after review of various qualitative and quantitative scoring systems. The critical appraisals for different types of studies use different domains, and within each domain there are several metrics or questions.

After evaluation, OPPT merges the steps of evidence synthesis within a stream, and the step of evidence integration across streams makes it difficult to determine the general approach followed at this point. In the 2018 guidance document, OPPT notes that the evidence integration step has three phases.

Planning involves developing a strategy for analyzing and summarizing data across studies within each evidence stream and a strategy for weighing and integrating evidence across those streams. The execution phase involves the implementation of the strategies developed in the planning phase and the development of WOE conclusions. The third phase involves a check on the quality of the data used. Ultimately, this leads to a summary of findings and confidence statements.

Due to the decision to apply systematic review beyond the evidence streams of hazard assessment, OPPT developed approaches to apply systematic review to data types such as exposure, fate and transport, and chemical and physical properties.

CRITIQUING THE OPPT APPROACH

Looking at the core review elements of the Statement of Task, which address whether the TSCA approach to systematic review is "comprehensive, workable, objective, and transparent," the committee finds that the approach presented by OPPT could be broadly improved to better meet these characteristics for the major review steps. The committee notes that its review was complicated by the challenge posed by inadequate documentation, itself an indication of failing at being comprehensive, workable, objective, and transparent. This discussion provides an overview of the critique of OPPT's approach within Chapter 2, which also provides recommendations that would improve these aspects of OPPT's approach.

Comprehensive

The committee found that the OPPT approach was not comprehensive at each step. The approach to problem formulation and protocol development did not result in refined research questions or a documented protocol for how the review should be conducted. This failing had implications for the number of studies identified and evaluated and resulted in challenges to integration across evidence streams. While the OPPT approach for identifying the evidence is comprehensive in regard to searching for literature in many databases, it is less clear how comprehensive the searches are for data that support models for ecological assessment and human health exposure assessment. In the TCE evaluation, for example, the hydrology data and product use information were both decades old.

The OPPT approach also does not give guidance on how physiologically based pharmacokinetic models will be evaluated, and, while it provides how the exposure models will be scored, little seems to be done to actually evaluate the models. With regard to synthesis, the approach does not contain elements important to addressing the research question. Finally, the WOE determination or evidence integration process was only detailed for one specific example, but it is uncertain whether this process represents a limited use for a specific endpoint for TCE or represents a method that will be used (in its current form) for future risk assessments. The committee finds that OPPT has merged aspects of important elements of evidence-based methods: evaluating individual studies, evaluating a body of evidence (i.e., strength or certainty of evidence for a conclusion), and evaluating level of confidence in a recommendation or determination of causation. In addition, the synthesis of the evidence within a specific data stream is not differentiated from integration of the evidence across the data streams. This condensation of steps makes it difficult to assess the extent to which the procedures used are comprehensive.

Workable

Considering whether the OPPT approach is workable, the report notes several concerns at each step. The current approach taken to problem formulation and protocol development is adding to a laborious process for searching, screening, and evaluating the literature. Completing a scoping review prior to the development of the PECO statements could narrow the search to appropriate studies and help

in selecting the appropriate evidence-based methodology for the review. The broad PECO statements led to inclusion and exclusion criteria that allowed inclusion of studies that may not be relevant. The broad searches for exposure data may have provided some efficiency, as it would be difficult to conduct the searches for chemicals that have many exposure scenarios relating to the conditions of use. Although OPPT is using a number of validated artificial intelligence–based tools to help make the process of screening hundreds of references more efficient, their use requires that precise and explicit inclusion and exclusion criteria are used consistently by all reviewers.

The evidence evaluation step includes items that do not assess risk of bias, most notably relevance. Relevance should be handled prior to study evaluation. Later, relevance can also be addressed in the evaluation of the body of evidence. The use of numerical scoring in critical appraisal does not follow standards for the conduct of systematic reviews and no justification is provided for the weighting of the specific metrics within the domains to create the overall quality score, making it hard to determine if the weights are appropriate.

Lastly, without a clear, documented approach to evidence synthesis and to integration, the risk evaluation process becomes unworkable because staff have to decide on approaches for these critical steps for each new evaluation rather than relying on a protocol or guidance.

Objective

The committee found the OPPT approach to be lacking objectivity at each step, from not using a defined approach to documenting how the problem formulation and protocol are developed. Further examples include inclusion and exclusion criteria that are too broad to identify the evidence, inherent subjectivity within the metrics that make up the evaluation score for study quality (without providing evidence that the metrics had been validated or tested for reliability as well as allowing a single reviewer to override them), and the lack of a consistent approach for documenting the objectives or methods for synthesis and evidence integration. The committee found that many of these concerns were related to the absence of a protocol a priori or the combination of the traditionally discrete and distinct steps of a systematic review. For example, OPPT does not clearly distinguish among scoping, problem formulation, and protocol development and merges the steps of evidence synthesis and integration.

Another problematic element of the TSCA evaluation framework is that the studies that are scored unacceptable are excluded from further analyses. Any fatal flaws in the methodology or conduct that preclude including a study should be used as eligibility criteria during the screening process. Once a study is determined to be eligible, the study should be included in the synthesis and the risk-of-bias assessment, with its limitations accounted for in any qualitative or quantitative synthesis. Given the large number of metrics scored for these data types, the possibility that a single unsatisfactory rating could completely nullify the use of a particular study from synthesis is problematic as it may lead to a biased review. Statistical power and statistical significance are not markers of risk of bias or quality. Statistical significance is not a measure of association or strength of association and should not be used to evaluate studies. In fact, combining multiple small, low-powered but similar studies in a synthesis is one of the benefits of systematic review.

Transparent

The committee found that transparency of the entire risk evaluation process is compromised across all of its elements. Neither clear questions nor protocols have been developed for the systematic reviews. Consequently, the review process is not documented from its start, and clarity is lacking when the review is finished and published. Overall, the committee found that the lack of information and details about the specific processes used for the identification of evidence reduced confidence in the findings. The OPPT processes and practices are not consistent with the standard of practice for systematic review.

Information about the search process was scattered across multiple documents within the docket for TCE and 1-BP, making the identification of details laborious and time consuming. The use of numeric scores for the evaluation of the studies obscures key differences between studies and prevents users of the reviews from making their own determinations about important strengths and limitations in study methods. OPPT reported to the committee that it generally follows standard practice by using two independent reviewers, but it was unclear how discrepancies between the two reviewers were handled. Furthermore, the committee noted that there were changes in study scores from one document to another. Evaluation of evidence synthesis was also complicated by the merger of evidence synthesis (within an evidence stream) and evidence integration (across multiple evidence streams) into a concurrent process, further reducing transparency and consistency. Confusing terminology used in the various documents (and sometimes even the variations in use within one document), the lack of information presented to describe the process, and the lack of documentation to explain deviations from the processes that were documented all contributed to the lack of transparency.

GENERAL FINDINGS

The committee finds that the process outlined in the 2018 guidance document, and as elaborated and applied in the example evaluations, does not meet the criteria of "comprehensive, workable, objective, and transparent." The committee's evaluation was made difficult by the incomplete and hard-to-follow documentation of many details of the process—adequacy of documentation is requisite for achieving transparency, objectivity, and replicability. In the committee's judgment, the specific and general problems in TSCA risk evaluations are partially due to the decision to develop a largely de novo approach, rather than starting with the foundation offered by approaches that were extant in 2016. OPPT was further challenged by the statutory schedule for completing assessments.

As a general finding, the committee judged that the systematic reviews within the draft risk evaluations considered did not meet the standards of systematic review methodology. Given that systematic review is well established for application to the hazard stream of evidence, the committee reviewed the hazard component of the Draft Risk Evaluation for Trichloroethylene, applying a tool used to assess bias in systematic reviews (AMSTAR-2). The appraisal process was unnecessarily complicated due to insufficient and unclear documentation. Despite this barrier to applying the AMSTAR-2 instrument, the committee found that the TCE hazard assessment did not perform positively on the vast majority of AMSTAR-2 questions.

Another crosscutting finding relates to terminology and specifically to the interchangeable use of the terms "weight of evidence" and "systematic review." The Risk Evaluation Rule (40 CFR Part 702) specifies that weight of the scientific evidence "means a systematic review method." However, the language may not in and of itself require a systematic review. The Draft Risk Evaluation for Trichloroethylene refers to a "weight of the evidence analysis" for an individual outcome involving successive determinations: first, considering individual lines of evidence; and second, integrating these disparate lines of evidence. Throughout the report, the committee comments on the consequences of the definition of WOE in the Risk Evaluation Rule and the conflation of "weight of the scientific evidence" with "a systematic review method" as necessitated by the Risk Evaluation Rule. The committee understands that the definition of WOE within the Risk Evaluation Rule is difficult to change but suggests that OPPT adopt a different specific term to be used during the evidence integration step, such as "strength of evidence" or "certainty of evidence" as utilized in the Grading of Recommendations Assessment, Development and Evaluation process.

Regardless of terminology, a narrative description should be provided that describes the basis for the determination of the strength of evidence during the evidence integration step for all applicable data streams. We urge the use of standard descriptors for the strength of evidence as with the Integrated Science Assessments for the National Ambient Air Quality Standards.

OVERALL RECOMMENDATIONS

The committee was in strong consensus that the processes used by OPPT do not meet the evaluation criteria specified in the Statement of Task (i.e., comprehensive, workable, objective, and transparent). The committee recognizes that OPPT faced substantial challenges in implementing review methods on the schedule required by the Lautenberg Act. Those challenges have not yet been successfully met. The committee makes a number of specific recommendations in Chapter 2 as to how to improve the methods for assessments, both in general and with reference to particular elements of the evaluation process.

The general recommendations are summarized as follows:

- The OPPT approach to systematic review does not adequately meet the state of the practice. The committee suggests that OPPT comprehensively reevaluate its approach to systematic review methods, addressing the comments and recommendations provided in Chapter 2.
- With regard to hazard assessment for human and ecological receptors, the committee comments that OPPT should step back from the approach that it has taken and consider components of the OHAT, IRIS, and Navigation Guide methods that could be incorporated directly and specifically into hazard assessment.
- The committee finds that OPPT's use of systematic review for the evidence streams, for which systematic review has not been previously adapted, to be particularly unsuccessful. Given these novel applications of systematic review, the committee suggests that OPPT should elaborate plans for continuing the refinement of methods, ideally in collaboration with internal and external stakeholders. The committee also suggests that OPPT evaluate how the existing OHAT, IRIS, and Navigation Guide methods could be modified for the other evidence streams. In addition, OPPT should use existing guidance within the agency such as the Guidelines for Human Exposure Assessment, the Guidelines for Ecological Risk Assessment, and the operating procedures for the use of the ECOTOXicology knowledgebase. Following these existing guidelines would improve transparency of the assessments.
- The committee recommends that a handbook for TSCA review and evidence integration methodology be put together that details the steps in the process. Throughout this report, the committee points to problems of documentation. The committee believes that the effort of developing and publicly vetting a handbook will pay off in the long run by making the process more straightforward, transparent, and easier to follow.

There is an ongoing cross-sector effort to develop and validate new tools and approaches for exposure, environmental health, and other new areas of application of systematic review. The committee strongly recommends that OPPT staff engage in these efforts. The approaches used for TSCA evaluation would benefit from the substantial external expertise available as well as additional transparency and acceptance by the different stakeholders and society in general as these tools are developed. The refinements recommended by this committee would boost the ability of actions taken under the Lautenberg Act to advance the mission of EPA: "to protect human health and the environment."

1

Introduction

Exposures to industrial chemicals in food, water, air, and consumer products can cause harm to human health and the environment. The 1976 Toxic Substances Control Act (TSCA; Pub L No. 94-469) provided the U.S. Environmental Protection Agency (EPA) Office of Pollution Prevention and Toxics (OPPT) with the authority to regulate chemicals in commerce. Subsequently, some public health and environmental groups concluded that TSCA was not sufficiently effective, particularly with regard to the regulation of existing chemicals (Bergeson 2016). In 2016 with bipartisan support, President Barack Obama signed the Frank R. Lautenberg Chemical Safety for the 21st Century Act ("Lautenberg Act") (Pub L No. 114-182), overhauling TSCA after 40 years. The Lautenberg Act provides the agency with increased authority to regulate chemicals existing before the 1976 TSCA was amended. For example, the Lautenberg Act requires evaluation of existing chemicals. Additionally, the agency must adhere to clear and enforceable deadlines. Chemicals are assessed against a risk-based safety standard. The risk evaluations consider both human health and ecological risks and must consider risks to susceptible and highly exposed populations; unreasonable risks identified in the risk evaluation must be eliminated; and the agency is provided expanded authority to more quickly require development of chemical information when needed.

The new authority afforded to the agency also put tremendous time pressure on OPPT to assemble teams, promulgate rules, and draft the guidance documents and related operating procedures that prescribe how OPPT responds to its mandate. The Lautenberg Act also required strict statutory deadlines of 9 to 12 months for chemical prioritization, the process to determine which chemicals should undergo risk evaluations. The first 10 high-priority chemicals then underwent risk evaluations, which were to be completed within 3 years of initiation (Susanna Blair, presentation to the committee, February 28, 2020).

The Lautenberg Act required the agency to promulgate several rules to implement the new requirements and responsibilities under the law. These rules provide the agency a framework to carry out its authority under the Lautenberg Act, including the Procedures for Chemical Risk Evaluation Under the Amended Toxic Substances Control Act, referred to as the "Risk Evaluation Rule" (40 CFR Part 702, 82 FR 33726). The statute requires that the agency must establish by rule a process for risk evaluation and that the risk evaluation will contain a scope or problem formulation, a hazard assessment, exposure assessment, a risk characterization, and the determination of unreasonable risk. The draft risk evaluations also must use the best available science (see Table 1-1); integrate and assess reasonably available information

TABLE 1-1 Terms as Defined in Procedures for Chemical Risk Evaluation Under the Amended Toxic Substances Control Act

Term	Definition in Rule
Best available science	Science that is reliable and unbiased.
Weight of the scientific evidence	Means a systematic review method, applied in a manner suited to the nature of the evidence or decision, that uses a pre-established protocol to comprehensively, objectively, transparently, and consistently identify and evaluate each stream of evidence, including strengths, limitations, and relevance of each study and to integrate evidence as necessary and appropriate based on strengths, limitations, and relevance.
Systematic review	No codified definition within the rule.

on hazards and exposures for the conditions of use, including information on specific risks of injury to health or the environment and information on potentially exposed or susceptible subpopulations; describe the ecological receptors that EPA plans to evaluate, whether aggregate or sentinel exposures were considered, and the basis; account for the likely duration, intensity, frequency, and number of exposures under the conditions of use; describe the weight of the scientific evidence for the identified hazard and exposure; be developed without consideration of cost or other non-risk factors; be published in the *Federal Register*; and have at least a 30-day public comment period.

Within the Risk Evaluation Rule, the agency chose only to define terms used within the statute, including "best available science" and "weight of the scientific evidence" (see Table 1-1). The term "systematic review" appears within the definition of weight of the scientific evidence, but the term is not codified within the Risk Evaluation Rule. Page 33734 of the preamble to the Risk Evaluation Rule states that the agency asked commenters about the application of systematic review to the hazard identification.

Commenters both supported and opposed the use of systematic review; thus, the agency chose to leave a reference to systematic review in the preamble of the Lautenberg Act but not codify a definition. Furthermore, the preamble states that it will not limit the use of a systematic review approach solely to the hazard assessment but will use it throughout the risk evaluation process.

As defined by the 2011 Institute of Medicine report *Finding What Works in Health Care: Standards for Systematic Reviews*, systematic review is "a scientific investigation that focuses on a specific question and uses explicit, prespecified scientific methods to identify, select, assess, and summarize the findings of similar but separate studies. The goal of systematic review methods is to ensure that the review is complete, unbiased, reproducible, and transparent" (IOM 2011, p. 1). Systematic review has become the foundation for assessing evidence to be used for decision making in a variety of health contexts, including medical care and public health. Well-conducted systematic reviews methodically identify, select, assess, and synthesize the relevant body of research, and clarify what is known and not known about the potential benefits and harms of the exposure being researched (Higgins et al. 2019). Key elements of systematic review include the following:

- Clearly stating objectives (defining the question);
- Developing a protocol, which a priori describes the specific criteria and approaches that will be used throughout the review process;
- Applying the search strategy to identify relevant evidence;
- Selecting the relevant studies (papers) using predefined criteria;
- Evaluating the internal validity (i.e., Are the study results at risk of bias?) and the quality of the studies using predefined criteria;
- Analyzing and synthesizing the data using the predefined methodology; and
- Interpreting and evaluating the synthesized results to draw a conclusion and specify level of confidence in that conclusion (Stephens et al. 2016).

In recent years, systematic review has been increasingly applied for gathering evidence within the risk assessment process to increase transparency, objectivity, and reproducibility. EPA has been adopting systematic review within the Office of Research and Development's Integrated Risk Information System (IRIS) Program since the 2011 National Research Council review of the draft IRIS formaldehyde assessment (NRC 2011). That report suggested that systematic review would remedy problems identified in the processes used to develop the draft assessment. Systematic review approaches are now being used and developed by the European Food Safety Authority, the Office of Health Assessment and Translation of the National Toxicology Program, the Texas Commission on Environmental Quality, and the Navigation Guide (Morgan et al. 2018; OHAT 2019; Schaefer and Meyers 2017; Woodruff and Sutton 2014). Additionally, the World Health Organization and the International Labour Organization have collabo-

rated to develop a risk-of-bias tool for assessing data on prevalence of exposure. The newly developed methods are based on review of existing methods, influenced by consultation with experts, tested for validity and reliability, and published in the peer-reviewed literature (Pega et al. 2020). Methods have also been proposed for applying systematic review methods to risk evaluations for ecological receptors in a framework integrated with human health risk evaluations (Suter et al. 2020).

Many stakeholders called for the adoption of systematic review within TSCA risk evaluations, which is why EPA asked for public comment on systematic review when developing the Risk Evaluation Rule. In 2018, OPPT developed the document *Application of Systematic Review in TSCA Risk Evaluations* to guide the agency's selection and review of studies (EPA 2018a). OPPT did not draw directly from the methods being developed in the IRIS Program in this document, or from other methods in development, and instead put together a new approach for evaluating evidence.

In these other existing frameworks, systematic review is typically applied to reviewing research for hazard identification (Whaley et al. 2016). OPPT has used systematic review to evaluate all lines of research that are integral to chemical risk assessment, such as exposure assessment and chemical and physical properties of the agent of concern (see Figure 1-1). The need to apply systematic review to types of studies to which it has not previously been applied has been offered by OPPT as the reason for developing its own approach (Francesca Branch, presentation to the committee, July 23, 2020).

THE COMMITTEE, ITS TASK, AND ITS APPROACH

EPA requested that the National Academies of Sciences, Engineering, and Medicine convene a committee to review the use of systematic review in TSCA risk evaluations. The committee included expertise in toxicology, epidemiology, risk assessment, exposure science, statistics, modeling, evidence integration, and systematic review; see Appendix A for biographical information on the committee. The verbatim statement of the committee's task is provided in Box 1-1. To address its task, the committee

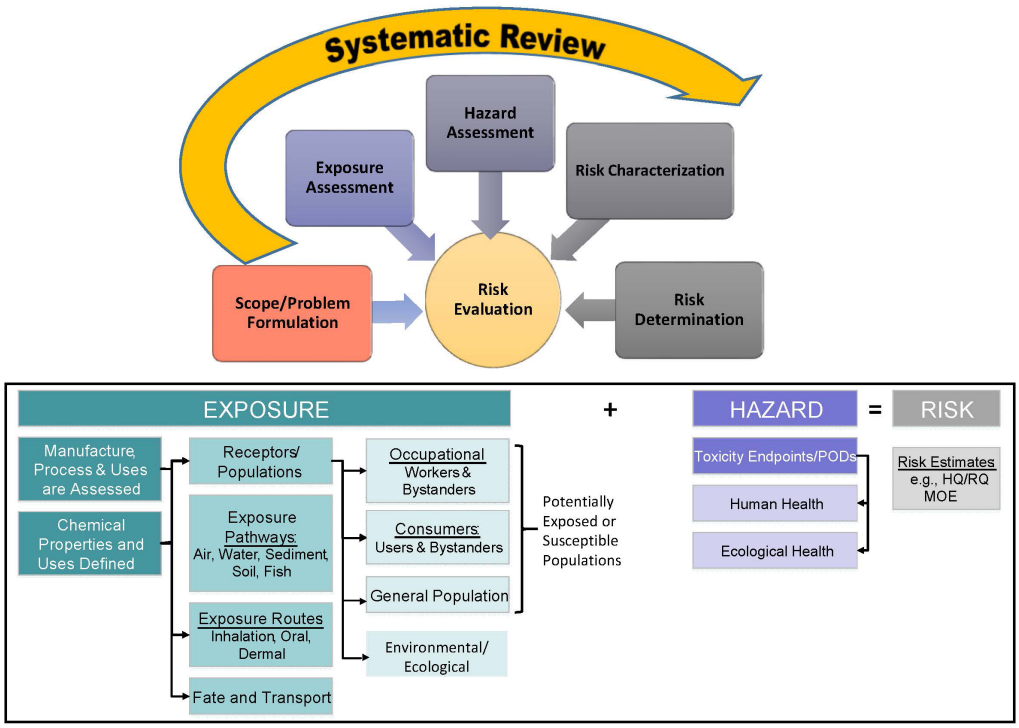

FIGURE 1-1 The components that feed into TSCA risk evaluations. NOTE: HQ, hazard quotients; POD, point of departure; RQ, risk quotient. SOURCE: U.S. Environmental Protection Agency, presentation to the committee, August 24, 2020.

BOX 1-1 Statement of Task

An ad hoc committee of the National Academies of Sciences, Engineering, and Medicine will prepare a review strategy for evaluating EPA's guidance document on Application of Systematic Review in TSCA [Toxic Substances Control Act] Risk Evaluations (EPA 2018a) and associated materials. The committee will consider public comments on the document, EPA's responses to public comments, and enhancements to the systematic review process reflected in documentation of the first 10 chemical risk evaluations. The committee will use the strategy to make a determination about whether EPA's process is comprehensive, workable, objective, and transparent. Recommendations for enhancements to EPA's 2018 guidance document will be made.

held a half-day public meeting during which OPPT presented an overview of systematic review under TSCA. OPPT also participated in three virtual public committee meetings, during which OPPT staff described tools used in searching and screening the literature, updates to their data evaluation and scoring procedures, and approaches to evidence integration supporting the exposure and hazard assessments. The presentations and posters described advancements that the agency has made since the publication of Application of Systematic Review in TSCA Risk Evaluations in 2018 (EPA 2018a). Stakeholders were allowed to provide written input to the committee throughout the review. In addition, opportunities for oral testimony were provided during two meetings.

During the first public meeting, OPPT clarified to the committee that it wanted the committee not only to consider the approach described in Application of Systematic Review in TSCA Risk Evaluations (herein the 2018 guidance document) and enhancements to the systematic review process reflected in documentation of the first 10 chemical risk evaluations but also how OPPT is planning to improve the use of systematic review in TSCA risk evaluations. These were broadly described at the committee's three virtual meetings, which occurred in June, July, and August 2020. The committee took the following approach to the Statement of Task given the agency's needs:

- The committee read and critiqued the 2018 guidance document.
- To consider the enhancements made to the systematic review approach in the first 10 risk evaluations, the committee reviewed the Draft Risk Evaluation for Trichloroethylene (TCE) and the Risk Evaluation for 1-Bromopropane (n-Propyl Bromide) (1-BP) (EPA 2020a,c). These evaluations were chosen because at the first meeting, EPA suggested that the 1-BP risk evaluation was a "typical risk evaluation," and later OPPT suggested that the TCE risk evaluation was the "best example of integration." The committee also critiqued the systematic review of the toxicology and epidemiology studies within the TCE risk evaluation using the "assessment of multiple systematic reviews" (AMSTAR-2) measurement tool (Shea et al. 2017). During its review of the TCE and 1-BP evaluations (including public comments, reviews from EPA's Science Advisory Committee on Chemicals, and EPA's response to the comments or reviews), the committee found significant overlap between the procedural steps taken in the two systematic review processes. Given that the TCE risk evaluation occurred later in time than the 1-BP assessment, and because it included a more detailed outline of the integration process, the committee chose to focus more heavily on the TCE risk evaluation in this report. Specific examples for 1-BP are used if there was a significant deviation from the process used in TCE.
- To consider enhancements to the TSCA systematic review process beyond the first 10 risk evaluations, the committee considered oral and poster presentations provided by OPPT at the committee's public virtual meetings.

- The committee posed questions to OPPT following the first committee meeting to clarify aspects of the use of systematic review in the risk evaluations.

The committee then considered all of this information to determine whether the TSCA systematic review process is comprehensive, workable, objective, and transparent—the core evaluation measures of the Statement of Task. The committee also provides its recommendations with regard to steps that could be taken to improve the evaluation process as described in the 2018 guidance document and further elaborated in the evaluations.

ORGANIZATION OF THE REPORT

This report is organized into three chapters and three appendixes. Chapter 2 presents a critique of the TSCA approach to systematic review. It begins with a short overview of how systematic review has been adopted within chemical risk assessments and provides some general observations concerning the OPPT approach. For each systematic review step, the committee provides a brief overview of the state of the practice, describes how the step is applied in TSCA risk evaluations, critiques the approach used in TSCA risk evaluations, and then provides recommendations for improvement for each step. Chapter 3 addresses crosscutting and more general issues related to the use of systematic review in TSCA risk evaluations. Appendix A provides biographical information on the committee. Appendix B provides the meeting agendas and Appendix C provides a list of the documents reviewed by the committee.

2

Evaluation of the TSCA Systematic Review Approach

INTRODUCTION

As scientific evidence on the risks of chemicals has grown, both in scope and in the contributing disciplines, the need for evidence-based approaches to addressing risks to human and ecosystem health has been increasingly recognized. Methods first developed and used in clinical medicine and other areas for assembling and evaluating bodies of evidence have now been extended to assessing risks to human health and the environment. Evidence-based methods—with their transparency, objectivity, comprehensiveness, and reproducibility—serve as a foundation of modern clinical practice. Evidence-based methods include protocols and comprehensive documentation of assumptions and decisions in compiling evidence, which allows for tracing of every step of the evaluation. This transparency is one primary advantage of using these practices. The principal tool for evaluating the evidence base on a topic is systematic review. Consequently, over the past decade, several approaches to applying this methodology to assessing evidence on risks of environmental agents have been elaborated, such as by the European Food Safety Authority (EFSA), the Office of Health Assessment and Translation (OHAT) of the National Toxicology Program (NTP), the Texas Commission on Environmental Quality, and the Navigation Guide (Morgan et al. 2018; OHAT 2019; Schaefer and Meyers 2017; Woodruff and Sutton 2014). Additionally, the World Health Organization and the International Labour Organization have collaborated to develop a risk-of-bias tool for assessing data on prevalence of exposure. The newly developed methods are based on review of existing methods, influenced by consultation with experts, tested for validity and reliability, and published in the peer-reviewed literature (Pega et al. 2020). Methods have also been proposed for applying systematic review methods to risk evaluations for ecological receptors in a framework integrated with human health risk evaluations (Suter et al. 2020).

A chemical risk assessment includes an initial problem formulation as well as hazard assessment and dose-response assessment, exposure assessment, and risk characterization, as appropriate (NRC 2009). Figure 2-1 provides a schema for how systematic review is applied within chemical risk assessments. Prior to the conduct of a systematic review, planning and problem formulation should take place (see blue boxes in Figure 2-1). The planning and problem formulation should include stakeholder engagement and broad literature searching to find the evidence on the topic, and end with the identification of the most important questions and the best approach for answering such questions. The research questions and the approach should inform the first step of the systematic review—the development of the protocol. The planning and problem formulation should also determine if another evidence-based method could be applied to answer the research question required for the risk evaluation.

If a systematic review is the appropriate approach, development of the protocol is the first step of the review (see green boxes in Figure 2-1). Evidence identification, which includes searching and screening the literature and finding the reports as prescribed in the protocol, is the next step. Evidence evaluation follows. This step includes evaluation of the internal validity of the individual studies (i.e., Are the study results at risk of bias?), using the appropriate risk-of-bias tool for the type of study being reviewed. Synthesis consists of a qualitative evaluation of the evidence and can be complemented by a quantitative pooling—a meta-analysis.

This synthesis of the various specific streams of evidence is followed by hazard assessment (see orange box in Figure 2-1) with integration of the multiple evidence streams of human, animal and other

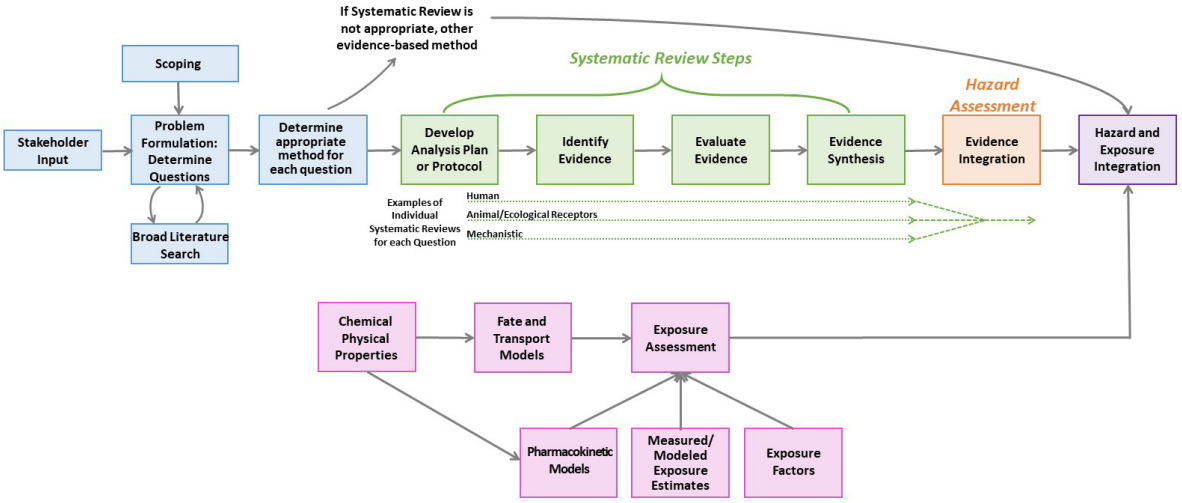

FIGURE 2-1 Example approach of systematic review in the context of risk assessment. The blue boxes refer to steps that are conducted prior to the systematic review, green denotes the systematic review process, orange denotes the hazard assessment, and purple is the integration of hazard and exposure. The pink boxes refer to the exposure assessment, which is conducted outside of the systematic review but is used to make the final risk characterization.

ecological receptors, and mechanistic findings. Gathering information on human and ecological exposures could also be approached with systematic review, but systematic review tools for those questions are not yet well developed (see pink boxes in Figure 2-1). For this reason, these steps are not shown as including the systematic review steps of protocol development, evidence identification, evaluation, and synthesis. Exposure and hazard data are integrated to characterize risk (see purple box in Figure 2-1).

Figure 2-2 illustrates the Office of Pollution Prevention and Toxics' (OPPT's) approach to systematic review, which differs to an extent from the description in Figure 2-1 and includes systematic review as part of the broader process of risk evaluation. The committee has organized this chapter by the steps that are integral to systematic review as used generally (i.e., matching the current standards outlined by the Institute of Medicine [IOM] report *Finding What Works in Health Care: Standards for Systematic Reviews* [Bero et al. 2018; IOM 2011]) rather than following the specific steps outlined by OPPT. These steps, per the IOM report, are protocol development, evidence identification, evidence evaluation, and evidence synthesis. The systematic review protocol is informed by a problem formulation, which guides the development of a protocol. The evidence identified and evaluated for the systematic review is synthesized and then the evidence from the various streams is brought together for the data integration step to determine if a hazard exists. If the review conducted by OPPT is indeed a systematic review, then the committee expects to find each of the steps of a systematic review (even if not specifically called out as such). Additionally, the committee would expect that all systematic reviews should meet the definition and principles from the IOM (2011) report—explicit prespecified methods to identify, select, and synthesize the evidence from studies. When reviewing the methods and assessments used by OPPT, the committee compared all streams of evidence that OPPT specified as applying to the systematic review process to this definition. To determine the approach to systematic review being used within the Toxic Substances Control Act (TSCA) risk evaluations, the committee reviewed the Draft Risk Evaluation for Trichloroethylene (TCE) and the Risk Evaluation for 1-Bromopropane (n-Propyl Bromide) (1-BP) (EPA 2020a,c) and considered comments on the evaluations from the public and from peer-review evaluations (e.g., the Science Advisory Committee on Chemicals), as well as OPPT's responses to the comments.

Systematic review approaches have been used widely and increasingly to assemble the evidence needed to assess human health and ecological receptors. The number of such systematic reviews focusing on environmental risks has doubled from 2016 to 2019 (Whaley et al. 2020). Yet, the use of systematic review to collect, evaluate, and synthesize the non-hazard evidence streams required for TSCA risk

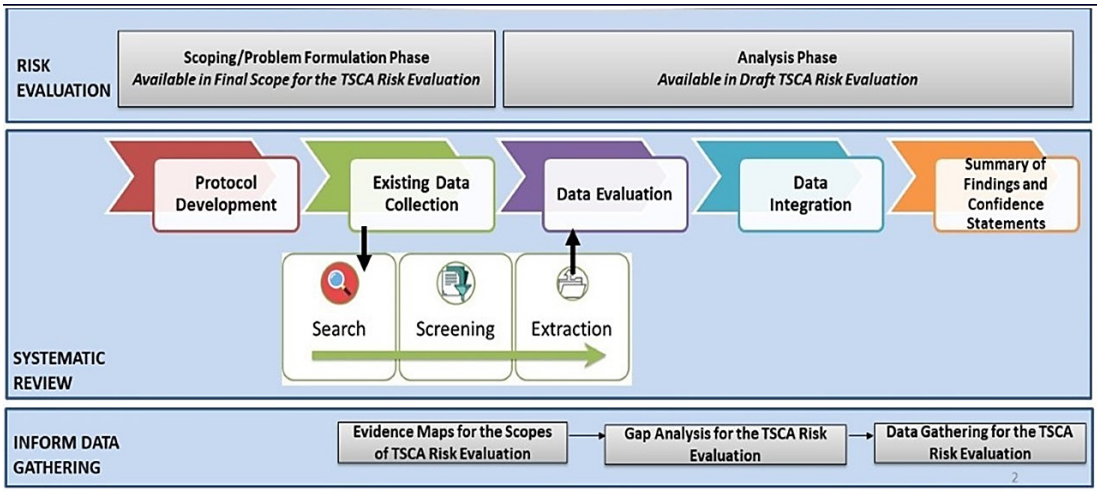

FIGURE 2-2 The systematic review process for TSCA risk evaluations. SOURCE: U.S. Environmental Protection Agency, presentation to the committee, June 19, 2020.

evaluations, such as data on exposure, fate and transport, and chemical and physical properties, is not established and very little precedent exists for applying systematic review to these streams of evidence. Within the U.S. Environmental Protection Agency (EPA), the *Guidelines for Human Exposure Assessment, the Guidelines for Ecological Risk Assessment*, and the operating procedures for the use of the ECOTOXicology (ECOTOX) knowledgebase (EPA 1998, 2019b, 2020b) dictate how exposure, fate and transport, and physical chemical property data should be assembled for decisions about risks to human health and ecological receptors.

The agency is beginning to consider how to advance systematic review of these other streams of data. A recent paper from EPA's Office of Research and Development (ORD) describes considerations required to advance systematic review for exposure science (Cohen Hubal et al. 2020). The authors describe how the tools for searching and organizing the literature can be applied to evaluating exposures. The article discusses Population (including animal or plant species), Exposure, Comparator, and Outcome (PECO) statements but stops short of addressing how PECO could be used to assemble exposure data. A checklist is provided for evaluating exposures, but, as written, it is proposed for evaluating exposures within the context of environmental epidemiology. The paper argues that in scientific literature, exposure information should be presented in a way that facilitates the determination of applicability to a PECO statement. Another paper authored by scientists in ORD proposes that systematic review and weight of evidence (WOE) are integral to ecological and human health assessments and provides an integrated framework, but the paper does not detail a process for how systematic review should be applied to all streams of evidence that are integrated to make a final risk determination (Suter et al. 2020).

The committee evaluated systematic review of the exposure and data on exposure, fate and transport, and chemical and physical properties similarly to how it evaluated other streams of evidence, such as whether there were explicit prespecified methods to identify, select, and synthesize the evidence from studies. The committee also considered whether OPPT followed the appropriate agency guidelines for these evidence streams (i.e., *Guidelines for Human Exposure Assessment*, the *Guidelines for Ecological Risk Assessment*, and the operating procedures for the use of the ECOTOX knowledgebase [EPA 1998, 2019b, 2020b]). The committee chose to use these agency sources as a reference because the Guidelines for Human Exposure Assessment are cited in the approach described for estimating consumer exposures in TSCA evaluations on the EPA website,[1] and the *Guidelines for Ecological Risk Assessment* state that they are intended to be agency-wide.

[1]See https://www.epa.gov/tsca-screening-tools/approaches-estimate-consumer-exposure-under-tsca, accessed November 13, 2020.

GENERAL FINDINGS

In planning its approach, the committee intended to review all systematic reviews (e.g., those focused on hazards and exposures) conducted within the Draft Risk Evaluation for Trichloroethylene and the Risk Evaluation for 1-Bromopropane (n-Propyl Bromide) with a tool used to assess bias in systematic reviews (AMSTAR-2) (Shea et al. 2017). Given that systematic review has been applied within the human health hazard assessment, the committee piloted this approach by reviewing the hazard evaluation of the TCE evaluation. The committee's strategy was to test whether the use of the AMSTAR-2 instrument would be helpful and, if so, then the instrument would also be applied to assess the OPPT approach for the other streams of evidence in the TCE and the 1-BP evaluations. In this pilot, the committee found that the TCE evaluation was lacking adequate descriptions of many elements of a systematic review—the details about review methods that are needed to apply the AMSTAR-2 tool were either absent or provided in a disparate and varied set of documents, not all of which were referred to in the risk evaluation.

In this evaluation with the AMSTAR-2, four committee members reviewed the hazard stream in the TCE risk evaluation. In that pilot attempt only the element "Does this review contain the elements of a patient-intervention-comparison-outcome (PICO) statement?" received a positive evaluation by all reviewers. All reviewers gave a partial yes to the elements "Did the review authors use a comprehensive literature search strategy?" and "Did the review authors describe the included studies in adequate detail?" For all other elements of the AMSTAR-2 tool, reviewers gave a "no" response. The overall confidence in the results of the TCE hazard review would be considered "Critically Low," indicating that the review "should not be relied on to provide an accurate and comprehensive summary of the available studies." Given that the TCE risk evaluation was released in February 2020, about 1.5 years after the first risk evaluation under TSCA, the committee judged this document to be a reasonable exemplar of the program's processes for the first 10 risk evaluations. Additionally, the committee also concluded that the quality of the systematic review was unlikely to be better for the exposure, chemical properties, and fate and transport streams of evidence, as systematic review is not as well developed in those areas. Consequently, the approaches to these evidence streams were given a general review and the AMSTAR-2 tool was not used.

Another crosscutting finding, discussed in detail in the evidence integration section of this chapter, relates to terminology and specifically to the interchangeable use of the terms "weight of evidence" and "systematic review." The Procedures for Chemical Risk Evaluation under the Amended Toxic Substances Control Act, referred to as the "Risk Evaluation Rule" (40 CFR Part 702), specify that weight of the scientific evidence "means a systematic review method." However, this definition may not be intended to mean a systematic review as defined by the IOM. Furthermore, the Draft Risk Evaluation for Trichloroethylene refers to a "weight of the evidence analysis" for an individual outcome involving successive determinations: first, considering individual lines of evidence; and second, integrating these disparate lines of evidence. The committee considers these two steps as data synthesis and data integration, respectively, and notes that the integration step is outside of systematic review. It is worth noting that a 2014 National Research Council (NRC) report found that the terms "systematic review" and "weight of evidence analysis" have been used interchangeably, leading to confusion. That report distinguished systematic review as including "protocol development, evidence identification, evidence evaluation, and an analytic summary of the evidence" while WOE analysis is a judgment-based process to infer causation that follows the systematic review (NRC 2014, p. 4). That report found "evidence integration to be more useful and more descriptive of the process that occurs after completion of systematic reviews" (NRC 2014, p. 4).

The remainder of this chapter is organized by systematic review step (problem formulation and protocol development, evidence identification, evidence evaluation, evidence synthesis, and evidence integration). For each step, the committee describes "the state of the practice," or how the step is generally conducted; describes how the committee thinks the step is being conducted within TSCA risk evaluations; offers a critique of the approach being used in TSCA; and makes recommendations to improve

the approach to the step. The committee notes that describing what OPPT does in order to critique the process was challenged by the lack of documentation of the systematic review approaches used by OPPT.

PROBLEM FORMULATION AND PROTOCOL DEVELOPMENT

Planning the review and carrying out problem formulation generally precede the systematic review process, typically comprising a scoping review and engagement of stakeholders. Many groups consider planning the review and subsequent problem formulation as the most important steps in a review because a high-level view is taken at this stage that sets the approach and protocol for the review.

The product of problem formulation is a well-defined question, appropriately guiding the selection of the methodology during the planning stage (e.g., systematic scoping review, systematic map, or systematic review). If a systematic review, systematic map, or systematic scoping review is determined to be the appropriate method to address the research question, a protocol is then developed and registered. Systematic review protocols need to report the following: (1) the research question, (2) the sources that will be searched with a reproducible search strategy, (3) the explicit inclusion and exclusion criteria for study selection described in an unambiguous and replicable way, (4) the methods used to select primary studies, (5) tools for critical appraisals of the risk of bias or quality in the included primary studies, and (6) information about the approaches for evidence synthesis with sufficient clarity to support replication of these steps (Krnic Martinic et al. 2019).

A protocol makes the methods and the process of the review transparent, provides the opportunity for peer review of the methods, and stands as a record of the review process. Having a protocol minimizes the potential for bias in many steps throughout the systematic review, such as in evidence identification, by ensuring that inclusion of studies in the review does not depend on the findings of the studies (NRC 2014). The risk evaluation process is typically much broader than that of systematic review alone, and generally not every stream of evidence is evaluated using systematic review, while other evidenced-based approaches may be used. This section discusses the state of the practice for development of systematic review research questions and protocols and describes OPPT's approach.

STATE OF THE PRACTICE

Planning the Review

Because minimizing bias is a guiding principle of systematic reviews, even the initial planning should be conducted as rigorously, objectively, and transparently as possible. This step may involve iterative consideration of sponsor and stakeholder needs, scoping of the topic—including considerations of feasibility—and input and participation from a multidisciplinary team sharing a variety of roles.

Once the plan to conduct a review assumes shape, a decision is made as to which type of review to perform. In some cases, a narrative approach may be chosen for any of a variety of reasons, including limited access to data or to express an expert opinion. However, if the goal is to provide an objective and comprehensive summary of how the evidence on a certain topic answers a specific research question, a systematic review should be conducted. The committee would like to emphasize that only a review conducted following all the steps in the framework is a systematic review.

Various motivations exist to conduct a systematic review in toxicology. In the frameworks that have been created by agencies, including NTP, EPA, and EFSA, the motivation for doing so is driven by the respective public health mandates and needs of the agency conducting the review (Morgan et al. 2018). Whether conducted by an agency or not, a systematic review may seek to clarify the human health or ecological effects of an evidence-rich chemical. In other cases, a systematic review may be undertaken

when evidence is scarce, to identify data gaps, or to assess the accuracy of a toxicological test method (NASEM 2017b).

While not necessarily required, scoping the literature on the topic is typically done to assess the need for a systematic review. This approach is particularly useful in fields in which little is known regarding the current state of the literature and when systematic reviews have not been performed previously. Scoping may range from a simple non-systematic search in one or two databases to a more formalized, resource-intensive scoping review (described in Levac et al. [2010] and Peters et al. [2015]). The findings of a scoping exercise may reveal that the question has already been adequately addressed or may confirm that better understanding of the evidence could provide clarity. A scoping search can inform the planning process by revealing important details such as the expertise required, the stakeholders interested in the topic, and the resources needed. Scoping may be conducted before or after a review team is formed, but the approach used should be transparent and objective. By the end of the planning stage, the decision as to whether or not to conduct the review will have been confirmed. The resources and the timeframe will have been established, and the review team and advisory group will be in place.

One major challenge in the planning phase of systematic reviews in toxicology, environmental health, and ecological toxicology and exposure is to compose a skilled review team, as experience among these disciplines in systematic review is still being built. Until sufficient systematic review capacity is built, clinical or preclinical systematic review experts may need to be engaged. Similarly, ecological risk assessment is still evolving. Consequently, expertise from those familiar with applying systematic review to human health should be brought in to develop systematic review protocols for data streams supporting ecological risk assessment. Following the planning stage and once the type of review needed is established, the stakeholder group begins the work on problem formulations.

Developing and Refining a Research Question

The ultimate goal of a systematic review is to address a specific question. The question should be developed a priori as it will shape many of the steps in the remainder of the review, giving form to the literature search and screening strategies.

In many contexts, narrowing the research question to the health or ecological outcomes that are of potential concern may be challenging, as there may be an array of outcomes, extending from markers of injury to specific diseases. In addition, traditional narrative literature reviews supporting risk assessments might include many different outcomes to determine the most sensitive endpoint on which to base the risk characterization (NRC 2009, 2014). Reviews that encompass large topics may benefit from an initially broad PECO statement to identify the body of literature and then multiple narrow and well-defined PECO statements to tease out the individual questions within the review. Scoping reviews can be extremely useful in identifying the reach and breadth of a systematic review prior to developing a research question. Assessing the breadth of the body of literature before defining the research question can allow for selection of specific endpoints, evidence streams, routes of exposure, or developmental stages. The problem formulation exercise, including scoping review results and input from stakeholders, can also be used to prioritize outcomes, study designs, and populations of interest. Moreover, scoping reviews can significantly improve the quality and usefulness of the inclusion and exclusion criteria for the literature search and selection. For example, which type of mechanistic evidence is relevant to the research question being addressed by a systematic review should be determined in advance based on an understanding of the pathogenesis for the outcome of concern. This advance determination will allow the creation of inclusion and exclusion criteria that eliminate mechanistic evidence not relevant to humans. If the reviewers have a baseline knowledge of the available literature database, it is easier to define and refine the criteria needed to exclude studies outside the interest of the review. NRC's *Review of EPA's Integrated Risk Information System (IRIS) Process* provides an approach for how to narrow a research question using a scoping review (NRC 2014).

For a health hazard assessment, the research question is typically focused on a PECO statement. This varies slightly from the health-based reviews which use a PICO statement, which were primarily focused on interventions (I) rather than exposures (E) (Rooney et al. 2014). The PECO statement is used to define the objective of the review and lay out the framework for the research question to be answered. Groups should enlist experts in various fields relevant to the review (i.e., epidemiologists, information specialists, and data analysts) and/or stakeholders with an interest in the outcome of the review to provide input in the PECO statement (OHAT 2019). Similarly, expertise within fields of importance to ecological risk assessment (e.g., ecotoxicology, ecology, environmental chemistry, and environmental engineering) should be engaged. For example, assessment endpoints, which clearly link to and support ecological protection goals, and associated measures of effect must be identified during problem formulation. Thus, an ecologically focused research question and PECO statement need to encompass these assessment endpoints. Once finalized, the PECO statement becomes the primary source of information used during the literature search and for the inclusion and exclusion criteria during literature screening.

While secondary PECO review question(s) may be necessary for complex risk-based assessments and listed as sub-questions, a clear single primary review question should drive the formulation of the review. Because this question will be the systematic review's guiding element and principal goal, defining it precisely and appropriately is of crucial importance. A properly framed review question will facilitate all the review's subsequent steps, including the definition of the eligibility criteria and the literature search, how the evidence will be collected, and how the results will be presented and synthesized. In particular, the question should help define the criteria for the inclusion and exclusion of research studies in a way that ensures that all relevant evidence is included to answer a particular question. For example, the review question could focus on a specific study type, such as chronic toxicity studies in animals, and would thus exclude any other study type, such as acute or sub-acute toxicity studies.

Determining the Appropriate Approach

The approach to answer the questions of interest is determined based on the results of the planning, question refinement, and scoping review processes. This decision should be based on the most appropriate methods to address the question as well as issues of transparency and efficiency.

Different approaches can be used for conducting the hazard assessment, including a systematic review, a narrative review with a protocol, an assessment by an authoritative body, or an update to a high-quality narrative or systematic review (NASEM 2019). Use of an existing systematic review may be an efficient and transparent way to address a question, particularly when the scoping review identified potentially relevant reviews. In the approach of using an existing systematic review, there is a search to identify current relevant reviews (i.e., those that match the PECO elements of the question to be addressed), the relevant review(s) are evaluated to identify a trustworthy review (i.e., using ROBIS or AMSTAR-2), and a bridge search (and, as needed, screening, risk-of-bias assessment, and synthesis) is conducted to update the search results. This process was recommended and demonstrated in a prior National Academies report, *Using 21st Century Science to Improve Risk-Related Evaluations* (NASEM 2017a).

Developing a Systematic Review Protocol

After the approach to answering the question is determined, the next step is to develop the protocol. The protocol is a detailed plan or set of steps that should describe the methods that will be used to conduct all the steps of the systematic review from evidence identification through evidence synthesis. Protocols are critical in de novo systematic reviews and can also be used in other evidenced-based literature searching methods, such as scoping reviews and narrative reviews. The PECO elements of the refined questions determine the search strategy and the eligibility criteria. These elements should be as

specific as possible, including listing specific outcomes of interest. Furthering transparency, protocols should be publicly registered, such as on PROSPERO or Open Science Framework (Booth et al. 2012; Foster and Deardorff 2017), or posted on a public website where they can be reviewed. Protocols should be posted for review prior to the start of the systematic review for public comment and peer review. Any revisions to the protocol should be as an amendment to the protocol. All versions of the protocol will remain available upon request, although the evaluation will usually proceed according to the most updated version of the protocol.

Standards of practice for applying evidenced-based methods for developing a research question, planning an approach, and developing a protocol are not well established for questions about exposures and ecological risks. The *Guidelines for Human Exposure Assessment* suggest that the problem formulation for human exposure assessment should include the identification of the individual, life stage(s), and group(s) or population(s) of concern; a conceptual model presenting the anticipated pathway(s) of the agent from the source(s) to receptor(s) of concern; and an analysis plan that charts the approach for conducting the assessment. Additionally, the *Guidelines for Human Exposure Assessment* suggest that aggregate exposures resulting from all potential uses of the compound should be calculated unless not needed per the specific research question (EPA 2019b).

The *Guidelines for Ecological Risk Assessment* state that the problem formulation for ecological risk assessments should depend on high-quality assessment endpoints and include conceptual models and an analysis plan. Assessment endpoints, and specific measures of effect, should be linked directly to the ecological protection or management goals, as framed within the conceptual models, along with exposure characteristics, ecosystems, and species of particular concern and at potential risk (EPA 1998).

Committee Description of the Approach in TSCA Risk Evaluations

OPPT is using a variety of software tools and approaches to conduct broad searching and to map the available evidence. The process has problem formulation and scoping occurring somewhat in parallel with the protocol development and data collection process (see Figure 2-2). Systematic review approaches use problem formulation to determine the protocol (see Step 1 in Figure 2-1). In advance of the risk evaluation, OPPT publishes scope documents that describe what it expects to consider in its risk evaluation pursuant to TSCA section 6(b). The TSCA website states the following:[2]

> The scope of a risk evaluation will include the hazards, exposures, conditions of use, and the potentially exposed or susceptible subpopulations the Administrator expects to consider. The scope will also include:
>
> - A Conceptual Model, which will describe the relationships between the chemical, under the conditions of use, and humans and the environment.
> - An Analysis Plan, which will identify the approaches and methods EPA intends to use to assess exposures and hazards.
> - "Conditions of use" under TSCA means "the circumstances, as determined by the Administrator, under which a chemical substance is intended, known, or reasonably foreseen to be manufactured, processed, distributed in commerce, used or disposed of." For purposes of prioritization, the Administrator may determine that certain uses fall outside the definition of "conditions of use." During the risk evaluation scoping process, EPA may decide to narrow the scope of the risk evaluation further, potentially excluding conditions of use that present low risk.

[2]See https://www.epa.gov/assessing-and-managing-chemicals-under-tsca/risk-evaluations-existing-chemicals-under-tsca#determination, accessed November 13, 2020.

The committee reviewed the scope documents that accompanied the TCE and 1-BP risk evaluations and found that the scoping and problem formulation documents merged the steps of problem formulation, protocol development, and the conduct of systematic review (see EPA 2017a,b). For these two assessments, the scope documents did not include PECO statements, but the problem formulation documents that were released after the evaluation was started did contain PECO statements. The committee notes that some of the more recently released scope documents contain more detailed information, such as the evidence identification methods and PECO statements[3] (see Box 2-1). Additionally, the scope documents did not contain protocols as typically defined in systematic review, as the scope documents did not prespecify the approaches for conducting each step of the systematic review process.

Although the 1-BP evaluation states that a "preliminary review of the health effects literature" was conducted, it is difficult to determine how that review affected the PECO statement, other than in the outcomes considered. For the human health hazard assessment, health outcomes were limited to several specific organ systems, although there is a note stating that other outcomes may be assessed as necessary (see Table 2-1). For the population element, specific species could have been limited to those that were either relevant to the outcome or that were available in the literature. For the exposure element in human epidemiology studies, a very broad search was conducted including not only the chemical of interest but also any metabolites or mixtures and studies with only qualitative estimates of exposure, which would have limited use in a dose-response assessment.

In the exposure assessment, the problem formulation includes occupational and non-occupational scenarios, as well as the environmental releases. As with the problem formulations for hazard assessment, the problem formulation, scoping, and data collection processes are also merged for these streams. The determination of what exposures are relevant to the risk evaluation is dictated by the "conditions of use." "For purposes of prioritization, the Administrator may determine that certain activities fall outside the definition of 'conditions of use.' During the risk evaluation scoping process, EPA may decide to narrow the scope of the risk evaluation further, potentially excluding conditions of use that present low risk."[4] For example in the TCE scope documents, all indoor studies of exposure are included and evaluated, but in the final risk evaluation, measurements of indoor home exposures were determined not to be related to a specific condition of use and thus were not included in determining exposures. In the search for data, sources that provide measured concentrations, as well as sources related to models, and all potential model inputs, are included. Additionally, in both the 1-BP and TCE evaluations if the manufacturer provided safety data sheets stipulating that workers will be provided personal protective equipment (PPE), EPA assumes that workers will be given proper PPE by their employer, that they will be trained to use it correctly, and that they will have no medical conditions precluding that use.

BOX 2-1 OPPT Uses Different Terms to Define PECO Statements for Different Evidence Streams

PECO: Population, Exposure, Comparator, and Outcome, used for environmental and human health hazards and consumer, environmental, and general population exposures (PECO health and PECO exposure)

PESO: Pathways and Processes, Exposure, Setting or Scenario, and Outcomes used for the fate and transport evidence

RESO: Receptor, Exposure, Scenario/setting, and Outcome used for the occupational exposure evidence and environmental releases discipline

[3]See, for example, https://www.epa.gov/sites/production/files/2020-09/documents/casrn_117-81-7_di-ethylhexyl_phthalate_final_scope.pdf, accessed November 13, 2020.

[4]See https://www.epa.gov/assessing-and-managing-chemicals-under-tsca/risk-evaluations-existing-chemicals-under-tsca#determination, accessed November 13, 2020.

For the ecological risks, OPPT describes the environmental fate and transport, releases to and occurrence in the environment, and probable environmental hazards (ecotoxicology) that will be considered in the scope documents. In the analysis plan of the scope documents, OPPT identifies its objectives, the conditions of use, and data types associated with physical-chemical properties, environmental fate parameters, and ecotoxicology information. Conceptual models are used to account for exposure pathways.

TABLE 2-1 Health Hazard Assessment PECO Statement from the 1-BP Risk Evaluation

PECO Element	Evidence Stream	Features Included
Population	Human	Any population All life stages Study designs: – Controlled exposure, cohort, case-control, cross-sectional, case-crossover – Case studies and case series that are related to deaths from acute exposure
	Animal	All non-human, whole-organism mammalian species All life stages
Exposure	Human	Exposure based on administered dose or concentration of 1-BP, biomonitoring data (e.g., urine, blood or other specimens), environmental or occupational-setting monitoring data (e.g., air, water levels), job title, or residence Primary metabolites of interest as identified in biomonitoring studies Exposure identified as or presumed to be from oral, dermal, or inhalation routes Any number of exposure groups Quantitative, semi-quantitative or qualitative estimates of exposure Exposures to multiple chemicals/mixtures only if 1-BP or related metabolites were independently measured and analyzed
	Animal	A minimum of two quantitative dose or concentration levels of 1-BP plus a negative control group Acute, subchronic, or chronic exposure from oral, dermal, or inhalation routes Exposure to 1-BP only (no chemical mixtures) Quantitative and/or qualitative relative/rank-order estimates of exposure
Comparator	Human	A comparison population (not exposed, exposed to lower levels, or exposed below detection) for endpoints other than death from acute exposure
	Animal	Negative controls that are vehicle-only treatment and/or no treatment
Outcome	Human Animal	Endpoints described in the 1-BP scope document: – Kidney toxicity – Liver toxicity – Neurotoxicity – Reproductive toxicity – Developmental toxicity – Cancer Other endpoints

SOURCE: EPA 2020c.

OPPT created PECO statements that guided three large literature searches to gather exposure data (under a PECO exposure statement); occupational exposure and information on industrial uses (under a Receptor, Exposure, Scenario/setting, and Outcome [RESO] statement); and information on chemical properties and factors related to environmental fate, transport, and ecological exposures and hazards (under a Pathways and Processes, Exposure, Setting or Scenario, and Outcome [PESO] statement) (see Box 2-1).

The very broad PECO-exposure (see Box 2-2) and RESO statements (see Box 2-3), one for consumer exposures and one for occupational exposures, have the goal of assembling all the potentially relevant literature to identify any measured levels of exposure in consumer or occupational settings. Additionally, the PECO statements include any models and model inputs necessary for modeling estimates of exposures. In addition to the actual PECO-exposure or RESO statements, OPPT also creates a list of all specific uses and scenarios for which there is a need to gather data, based on the knowledge obtained about the use of the chemical in prior steps.

Critique of the TSCA Approach

Looking at the core review elements of the Statement of Task, which is to address whether the TSCA approach to systematic review is "comprehensive, workable, objective, and transparent," the committee finds that the approach to problem formulation and protocol development could be improved broadly to better meet these characteristics.

BOX 2-2 PECO Statement for General Exposures in the TCE Risk Evaluation

Population
- Human: Consumers (i.e., receptors who use a product directly) and bystanders (i.e., receptors who are non-product users that are incidentally exposed to the product or article), such as infants, children, pregnant women, lactating women, women of childbearing age, and high-end consumers
- Ecological: Aquatic exposure species

Exposure Expected: Primary Exposure Sources, Pathways, Routes
- Sources: Consumer uses in the home producing releases of TCE to air and dermal contact; industrial and commercial activities producing releases to surface water
- Pathways: Indoor air and dermal contact with TCE in consumer products; surface water
- Routes of Exposure: Inhalation via indoor air (consumer and bystander populations) and dermal exposure via direct contact with consumer products containing TCE; surface water and sediments, soils, and groundwater

Comparator (Scenario)
- Human: Consumer and bystander exposure via use of TCE-containing consumer products in the home
- Ecological: Aquatic species and plants exposed via releases to or presence in surface water

Outcomes for Exposure Concentration or Dose
- Acute, subchronic, and/or chronic external dose estimates (mg/kg/day); acute, subchronic, and/or chronic air concentration estimates (µg/m3, mg/m3). Both external potential dose and internal dose based on biomonitoring and reverse dosimetry mg/kg/day will be considered
- Ecological: A wide range of ecological receptors will be considered (range depending on available ecotoxicity data, including mechanistic ecotoxicology information) using surface water concentration(s) (µg/L, mg/L)

SOURCE: EPA 2017c.

BOX 2-3 RESO Statement Example from the TCE Risk Evaluation

Receptors—Humans, workers, including occupational non-users.

Exposure—Worker exposure to and relevant occupational releases of the chemical substance of interest.
- Dermal and inhalation exposure routes.
- Any relevant media/pathway as indicated in the conceptual model.

Setting or Scenario—Any occupational setting or scenario resulting in worker exposure and relevant environmental releases including a predetermined list of processes.

Outcomes—Quantitative estimates of worker exposures and of relevant environmental releases from occupational settings.

General information and data related and relevant to the occupational estimates.

SOURCE: EPA 2017c.

Comprehensive

The approach to problem formulation and protocol development is not comprehensive as it did not result in refined research questions or a documented approach to how the reviews required to support the risk evaluations should be conducted. The ill-defined questions within TSCA risk evaluations hinder the necessary prespecification of systematic review methods, notably the eligibility criteria for studies. Failing to adequately refine the focus of a systematic review leads to overly broad questions, in turn leading to the identification of heterogeneous studies, to more complicated analysis, and to challenges in integrating across evidence streams to draw conclusions.

With respect to the ecological assessments, the conceptual models are not consistently accounting for all exposure pathways. For example, within the environmental conceptual model in the TCE evaluation, land application of wastewater effluent is not considered, yet this practice from centralized and decentralized (e.g., advanced aerobic systems) wastewater treatment plants introduces chemical contaminants to soils. Similarly, though the range of instream dilution considerations for point source discharges is important for predicting exposure scenarios, the TCE evaluation uses 10- to 15-year-old Exposure and Fate Assessment Screening Tool (E-FAST) models in the TCE assessment. Consequently, there may be underestimation of surface water exposure levels in regions experiencing decreased flows due to climate change (e.g., prolonged droughts) and increased water extraction (see EPA 2020a, p. 98). Additionally, the documents do not prespecify all cut-off values for environmental fate parameters, such as those used for bioconcentration factors and bioaccumulation. As with the human health hazard assessments, the TCE evaluation does not include sufficient protocols that prespecify how the systematic review for ecological outcomes will be conducted.

For the more typical exposure factors, TSCA assessments do rely on the *Exposure Factors Handbook*, which is a well-documented and regularly updated source of information. However, there are some activities not covered in the handbook, and searches for those data streams should be included in the data-needs list in the problem formulation. Examples noted relevant to TCE consumer use would be activity pattern information on the frequency with which the average gun owner cleans his or her gun, other product-specific activities not included in the handbook, or air exchange rates associated with horse stables (EPA 2011). Additionally, in the TCE and the 1-BP evaluation, searches are needed on the actual use, type, and effectiveness of PPE for the different occupational uses of the products. The as-

sumption that PPE would be used consistently and by all workers is overly optimistic, a criticism that the committee noted in the public comments on the TCE and 1-BP risk evaluations. Additionally, breathing rates during occupational activities where the products of interest are used to estimate chemical intake, should also be included in the data needs list. Inclusion of these additional search terms in the product formulation would result in an exposure assessment that is more consistent with the *Guidelines for Human Exposure Assessment*.

Workable

The current approach taken to problem formulation and protocol development is adding to a laborious process for searching, screening, and evaluating the literature. Completing a scoping review prior to the development of the PECO statements could narrow the search to appropriate studies.

Objective

OPPT is using a variety of software tools and approaches to conduct broad searching and to map the available evidence; however, it is not using those approaches in an objective way to determine the research question. This may be because the TSCA process has problem formulation and scoping occurring somewhat in parallel with the protocol development and data collection process (see Figure 2-2).

Transparent

The process for problem formulation is not transparent. It is not well documented in any of the risk evaluations or related scope documents reviewed for this report, and procedures for problem formulation are not included in *Application of Systematic Review in TSCA Risk Evaluations* (herein 2018 guidance document). Moreover, the transparency of the entire risk evaluation is compromised because in addition to not developing clear questions for the systematic reviews, there are no protocols for the reviews or to guide the synthesis step. Consequently, the review process is not documented or prespecified from its start, and clarity is lacking when the review is finished and published.

Specifically, it is unclear how OPPT is determining the list of data needs that will inform the human exposure assessment. The TSCA webpage states that OPPT is utilizing EPA's *Guidelines for Human Exposure Assessment* (EPA 2019b) for exposure assessments. If this is the case, the committee notes several inconsistencies with the guidelines. First, the guidelines specify that exposure calculations should be aggregate exposures, resulting from all potential uses of the compound (EPA 2019b). For example, a typical consumer's exposure includes both day-to-day exposures occurring indoors as well as increased exposure resulting from product use, which may occur on a semi-regular basis. TSCA assessments do include indoor concentrations that result from aggregate exposure in the problem formulation statement for consumer exposure. However, later steps determine that the indoor exposures could not be linked to any individual consumer product, and those exposures are omitted from the final exposure assessment.

The data needs list does include many of the parameters needed to run the existing EPA models but does seem to exclude some necessary model inputs that should be included in the data search based on the guidelines. OPPT could improve these assessments and make them better align with agency guidelines if clear questions on frequency of use for consumer products that may contain the chemical of interest were identified in the problem formulation.

Recommendations

In order to improve these issues with TSCA's approach to problem formulation and protocol development, the committee recommends the following:

- Scoping and mapping exercises and stakeholder engagement should be used to conduct a full problem formulation prior to the conduct of the systematic review.
- The results from problem formulation should include refined questions and an approach for each research question. A systematic review may not be required for every stream of evidence that is part of a risk evaluation. The full problem formulation and understanding of the literature base for an evaluation should allow OPPT to determine which research questions may be evaluated with a systematic review and which questions should be evaluated with a different evidenced-based approach. Ecological research questions should be linked to assessment endpoints, and ecological receptors of concern should be identified (e.g., algae, aquatic macrophytes, invertebrates, fish, amphibians, and threatened and endangered species). When chemicals are identified as bioaccumulative, other receptors (e.g., birds and mammals) should also be included. When there is no adequate literature base to answer questions, OPPT should be transparent about its alternative approaches. Regardless of the approach taken, OPPT should ensure that the reviews are comprehensive, objective, transparent, and consistent. OPPT should also highlight and explain areas in which deviations from a systematic review process occur.
- Evidence streams should be clearly defined to facilitate the determination of which evidence-based methods should be followed for each stream. Such a definition is especially critical for the exposure and non-hazardous streams.
- Potential redundancies should be reduced by explicitly considering appropriate methods to address questions that may include updating an existing and adequate systematic review rather than conducting a de novo review.
- A systematic review protocol that details the prespecified methods, including eligibility and critical appraisal criteria, and that is peer-reviewed and publicly posted before the review commences should be prepared. Ideally, this would be one document or, if multiple documents are needed, there should be clear crosswalks between documents.
- The problem formulation for the exposure assessment should more closely follow the *Guidelines for Human Exposure Assessment* and include inputs on frequency of product use, exposure factors related to specific uses, breathing rates, and use of PPE.

EVIDENCE IDENTIFICATION

Evidence identification is the next step in the systematic review process (see Figure 2-1). This step includes searching for the evidence related to the particular question and strategies for selecting both the evidence to be considered and the evidence to be excluded from consideration. As noted in *Finding What Works in Health Care: Standards for Systematic Reviews* (IOM 2011), this step presents the first opportunity for bias to enter the review. Without a comprehensive search to identify evidence informing the PECO statement, the resulting systematic review "will reflect and possibly exacerbate existing distortions in the biomedical literature" (IOM 2011, p. 81). Therefore, it is critical to have pilot search strategies and to have quality assurance (QA) and quality control (QC) measures during the evidence identification step of the systematic review process (IOM 2011).

STATE OF THE PRACTICE

Searching for the Evidence

Searching for the evidence starts with the design of a search plan that is aligned with the PECO question(s). This plan needs to be sensitive enough that it does not inadvertently exclude evidence relevant to the review question, without returning an unmanageably large amount of irrelevant information. The search plan should be specified in protocol, include databases to be searched and search

strategies (for at least one database), discuss gray literature that will be searched, and may also include other methods of search such as snowball searching, scanning references of included studies, and using existing systematic reviews. The date of the searches that are planned for updating should also be documented. A comprehensive search strategy should be guided by the PECO question in the selection of search terms, be in line with the pertinent inclusion criteria (publication date or language(s) to be considered), strike a balance between sensitivity (the ability to identify relevant evidence) and specificity (the ability to exclude irrelevant information), and be appropriately documented in the protocol and made publicly available.

Selecting the Evidence

Literature searches can yield thousands of records. In order to prevent subjectivity and reduce bias in the evidence selection, systematic reviews prespecify eligibility criteria based on the PECO question in the study protocol. The protocol should also specify how the quality of the selection is controlled, usually by independent duplicate review (i.e., requiring that two screeners, or one screener and an artificial intelligence [AI] tool, independently carry out the selection, with a procedure to resolve disagreements). In addition, it should provide instruction on how the selection is documented to allow its replication by others.

The selection is usually carried out in two stages: (1) title and abstract and (2) full text. At the first stage, all identified records are screened on the basis of the title and abstract in order to exclude the records that are obviously beyond the scope of the review. Studies rejected at this stage of the process will either be completely off-topic or fail to meet one or more eligibility criteria.

The second stage of the selection is a full-text review, during which reasons for the exclusion of each study need to be documented. Detailed documentation of the decision(s) made in the selection process is essential for the transparency of the review. Screeners' assessments should be captured, as well as the solutions in case of disagreements. All full texts retrieved should be kept in a database for the systematic review.

A widely accepted tool for summarizing the selection process is the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses[5]) flow diagram presented in Moher et al. (2009). Koustas et al. (2014) provide a practical example of its use for a toxicological systematic review.

In order to handle the vast amount of identified records and appropriately document the selection process, various software applications are now available that allow reviewers to implement an efficient and transparent record management process. Van der Mierden et al. (2019) provided an overview of informatics solutions to support various processes of a systematic review, and the systematic review toolbox contains more than 100 software tools for a broad range of systematic review tasks.[6]

The committee notes that there are no standards of practice to search for the evidence in the streams that are typically included in the exposure assessment (see Figure 2-1). The *Guidelines for Human Exposure Assessment and the Guidelines for Ecological Risk Assessment* do provide guidelines as to what should be included in an exposure assessment (EPA 1998, 2019b), and the approach described for searching for evidence for other streams is similar to the hazard assessment.

COMMITTEE DESCRIPTION OF THE APPROACH IN TSCA RISK EVALUATIONS

Searching for the Evidence

OPPT uses exhaustive search strategies that include major scientific databases, backward searching for studies in previous chemical risk assessments, additional gray literature sources, studies submit-

[5]See www.prisma-statement.org, accessed November 13, 2020.

[6]See http://systematicreviewtools.com/advancedsearch.php, accessed November 13, 2020.

ted under TSCA, and studies identified in peer review (see Figure 2-3). The terms used and databases searched were found to be exhaustive (see Table 2-2).

In the TCE evaluation, the searches conducted are wide-ranging as a result of the broadly framed questions. For example, to support the review of the occupational data, OPPT uses a grouped data acquisition strategy. Rather than conducting a separate search for each exposure to be determined, exposures are grouped into three data acquisition strategies. Specifically, one search looks for information that can be used to assess consumer exposures; a second assesses environmental release data for occupational exposures; and a third gathers physical chemical property data that can be used to run the models to calculate exposure. The search for the needed chemical properties was included in the ecological search, but data were used for all types of models. This approach improves the efficiency of the data search but makes following and evaluating the process difficult. However, it is evident that the process includes areas such as monitoring data, models, and completed exposure assessments. Addi-

TABLE 2-2 Search Strategies and Terms in TSCA Risk Evaluations

	Physical/Chemical Properties	Conditions of Use	Fate, Engineering, Exposure, Human Health	Environmental Health
Search terms	CAS RN Chemical name Chemical structure	CAS RN Chemical name Synonyms Trade names Common misspellings	CAS RN Chemical name Synonyms *AND* Use terms *OR* Exposure, Engineering and Fate terms *OR* Health Effect Terms	CAS RN Chemical name Synonyms as identified by STN International and Pesticide Action Network
Sources	STN REAXYS ChemSpider	Information reported to EPA Trade publications Open literature reports Citations in other assessments Safety data sheets EPA's Chemical and Product Categories NIH's household product database Company websites	Existing assessments Peer-reviewed sources (Pub Med, Web of Science, Toxline) Gray literature • Google -first 100 sites • Web scraping (e.g., ATSDR and NIOSH documents • databases (ChemView, NHANES)	ECOTOX Knowledgebase approach: • Science Direct • Agricola • Toxline • Scifinder • Proquest

NOTE: ATSDR, Agency for Toxic Substances and Disease Registry; CAS RN, CAS Registry Number; ECOTOX, ECOTOXicology; EPA, U.S. Environmental Protection Agency; NHANES, National Health and Nutrition Examination Survey; NIH, National Institutes of Health; NIOSH, National Institute for Occupational Safety and Health.
SOURCE: U.S. Environmental Protection Agency, presentation to the committee, June 19, 2020.

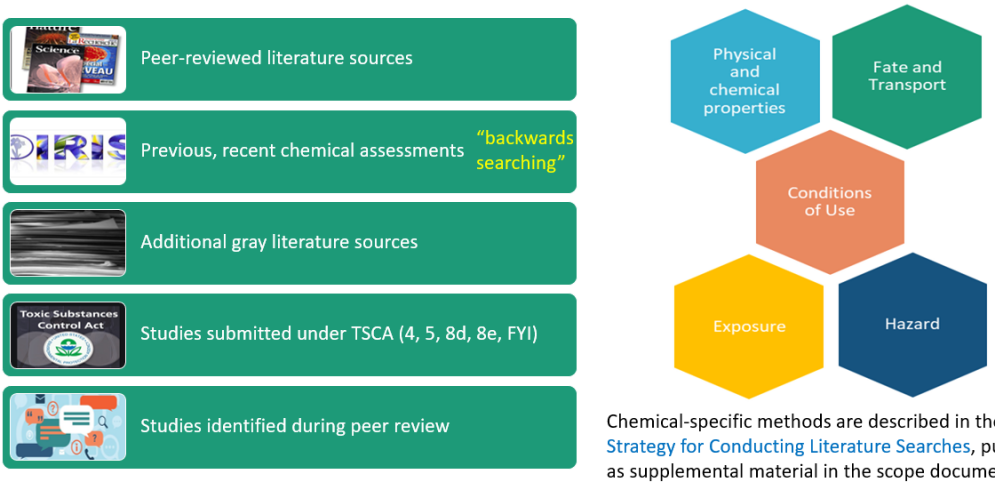

FIGURE 2-3 Literature searching process for TSCA risk evaluations. SOURCE: U.S. Environmental Protection Agency, presentation to the committee, June 19, 2020.

tionally, the OPPT process is relatively thorough in using a wide range of search terms and considering both the published literature and the gray literature. Gray literature searching included the first 100 sites on Google, web scraping (e.g., Agency for Toxic Disease Registry and National Institute for Occupational Safety and Health documents), and databases (e.g., ChemView). The process allows for the inclusion of other data that are submitted during peer review. It was noted in the TCE evaluation that additional reports were suggested in the public comments; those reports were screened and in some cases included. These additional materials may be government reports or other gray literature, which can be difficult to search for in a comprehensive fashion.

OPPT collects information on physical-chemical properties and environmental fate parameters. These routine physical-chemical properties (e.g., water solubility, vapor pressure, log Kow, and Henry's Law Constant) provide indications of environmental compartments (e.g., surface water, groundwater, sediment, and air) where exposure may occur and are required to support environmental modeling efforts. OPPT conducts literature searches to populate a database for further review. Similarly, OPPT also reviews the literature and the U.S. Geological Survey, EPA, and the U.S. Department of Agriculture databases for ambient surface water exposure data from the United States and other countries. In addition to identifying empirical datasets, OPPT uses the Estimation Programs Interface (EPI) Suite modeling for physical-chemical property and environmental fate information. Such practices are not surprising. As there are approximately 350,000 chemicals and chemical mixtures registered for commercial use around the world (Wang et al. 2020), empirical data on physical-chemical properties are not consistently available, and environmental fate parameters are relatively limited. Environmental fate modeling is thus necessary, though it remains challenging to cover the range of environmental exposure scenarios and compartments with the existing tools. More recent information is available through the National Hydrography Dataset. Its use would be advantageous to improve dilution expectations and thus aquatic exposure predictions, particularly since stream flow datasets within E-FAST 2014 are reported to be 15 to 30 years old (Card et al. 2017).

OPPT also relies on the Ecological Structure Activity Relationships (ECOSAR) program to estimate ecotoxicity data when empirical data are limited. It is not clear whether or how ECOSAR is consistently being used to predict acute toxicity information for fish, aquatic and terrestrial invertebrates, and algae during each risk evaluation. Similarly, it is not clear whether physical chemical property information is being evaluated a priori to ensure it is captured within ECOSAR applicability domains. When empirical ecotoxicology data are lacking, another ORD tool, the Web-based Interspecies Correlation Estimation model, is available to support cross-species predictions. In addition, EPA ORD has advanced adverse outcome pathway (AOP) conceptual models to support mechanistic ecotoxicology data integration within risk evaluations. It has developed Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS), another innovative tool that presents bioinformatic opportunities to advance toxicity extrapolation efforts across species. However, it does not appear that these models have been identified during systematic reviews.

Many databases support the human exposure assessments, but the process for searching for these data is unclear. For example, in the TCE risk evaluation, OPPT relied on a consumer product use database that is more than 30 years old and may not reflect current usage patterns. However, a process could not be identified for obtaining more relevant and recent data. Information is not readily available on what chemicals are in particular products, as databases with such information are lacking both in quantity and quality.

It is also unclear whether the specific search statements are intended to identify factors that may be important for the exposure calculations for the conditions of use. For example, one pathway considered for TCE was related to hoof polish for horses. It was assumed that the duration of use and the mass of polish used was the same as for shoe polish. It was assumed that the barn where the product was being applied was the same size as a garage but that the air exchange rate was higher, with reference to a sin-

gle report from Pennsylvania State University supporting that value.[7] It is unclear if a systematic search was conducted to obtain data for this parameter or if the search was limited to finding a source of data.

Selecting the Evidence

To select evidence, OPPT then screens the titles and abstracts against a list of needs for the evaluation. However, the committee was provided conflicting information on how this step was conducted. The 2018 guidance document states that OPPT uses PECO at title and abstract screening; the Strategy for Conducting Literature Searches for TCE: Supplemental Document to the TSCA Scope Document states that OPPT uses a list of data needs. Next, a full-text screening of the papers is conducted. As noted in the problem formulation section, OPPT uses PECO or PECO-like statements to compare articles to determine eligibility. For occupational data, there is a RESO statement to gather information on potential occupational exposures, and for the TCE and 1-BP evaluations such statements are used to inform the full-text screening. This process, as carried out for the evaluations of TCE and 1-BP, is illustrated in Figure 2-4. The broad PECO/PECO-like statements lead to unclear and shifting eligibility criteria and to an unclear or questionable selection process (i.e., changes in the process are allowed and reasons are not specified). In more recent evaluations, OPPT is using a number of AI-based tools to help with the large number of references (Kellie Fay, poster presentation to committee, June 19, 2020). These tools aim to make the screening of articles more efficient by automatically prioritizing articles by using user feedback to push the most relevant articles to the top of the list (Howard et al. 2016).

The literature flow diagram for the TCE human health hazard assessment shows that more than 6,000 studies were identified from the initial search using key words, 5,954 studies went through title and abstract screening, and 180 studies were evaluated during the full-text screen (including 95 key studies from previous assessment activities) (see Figure 2-5). Of the 180 studies identified for human health hazard assessment data evaluation, 119 studies were selected for full-text evaluation for animal and mechanistic data (see Risk Evaluation for TCE Systematic Review Supplemental file: Data Quality Evaluation of Human Health Hazard Studies—Animal and Mechanistic Data). No information was supplied in the risk evaluation documents identifying how these studies were selected. While it appears that the studies were selected from the general human health hazard assessment pool, the lack of details on how the agency arrived at the particular subset of animal and mechanistic studies makes it impossible to determine the process by which the studies were identified. Additionally, no information is supplied related to the excluded references (e.g., which studies were excluded and why).

[7]See https://extension.psu.edu/horse-stable-ventilation, accessed November 13, 2020.

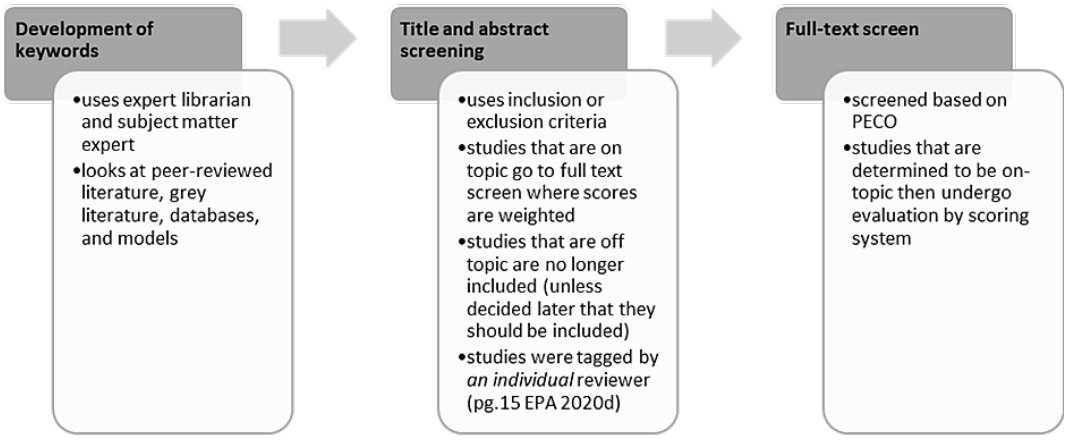

FIGURE 2-4 Committee's interpretation of the OPPT approach to identifying and selecting evidence in the TCE and 1-BP risk evaluations.

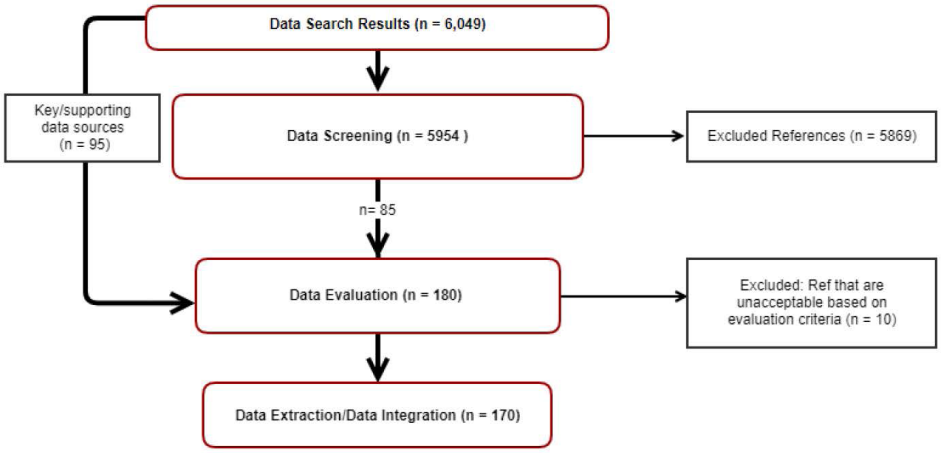

FIGURE 2-5 Literature flow diagram for human health hazard from TCE risk evaluation. SOURCE: EPA 2020a, p. 66.

Using the inclusion/exclusion criteria based on the PECO statement for TCE (see Table 2-3), the populations identified in the PECO statement include any population and all life stages, and although fetal cardiotoxicity was an important outcome for the TCE evaluation, no justification was given for limiting outcomes of animal and in vitro studies to developmental toxicity.

Critique of the TSCA Approach

Looking at the core review elements of the Statement of Task, which address whether the TSCA approach to systematic review is "comprehensive, workable, objective, and transparent," the committee finds that the TSCA approach to searching for evidence identification could be improved on all of these characteristics.

Comprehensive

TSCA risk evaluations include searches for evidence in most major scientific databases, backward searching for studies in previous chemical risk assessments, additional gray literature sources, studies submitted under TSCA, and studies identified in peer review (see Figure 2-3). The terms used and databases searched were found to be exhaustive. The ecological assessment and the human health exposure assessment rely often on databases that support the models for those evaluations. The committee found that the process for searching for these types of data may not be comprehensive. For example, the hydrology data used in the TCE evaluation were not the most recent, and the product use information was more than 30 years old.

Workable

The TSCA approach could be more efficient, as the broad PECO statements led to inclusion and exclusion criteria that allowed inclusion of studies that may not be relevant. OPPT is using a number of AI-based tools to help make the process of screening hundreds of references more efficient. Many of these tools have been validated on dozens of systematic reviews, and the committee is supportive of their use (Gartlehner et al. 2019; Howard et al. 2020; Van der Mierden et al. 2019). However, these tools will only work well when precise and explicit inclusion and exclusion criteria are used consistently by all screeners.

TABLE 2-3 Inclusion and Exclusion Criteria for the TCE Risk Evaluation

PECO Element	Evidence Stream	Papers/Features Included	Papers/Features Excluded
Population	*Animal*	• All non-human, whole-organism mammalian species • All lifestages	• Non-mammalian species
	Mechanistic/Alternative Methods	• Human or animal cells (including nonmammalian model systems), tissues, or biochemical reactions (e.g., ligand binding assays) with in vitro exposure regimens; bioinformatics pathways of disease analysis; or high throughput screening data	
Exposure	*Animal*	• A minimum of 2 quantitative dose or concentration levels of TCE plus a negative control group[a] • Acute, subchronic, chronic exposure from oral, dermal, inhalation routes • Exposure to TCE only (no chemical mixtures) • Quantitative and/or qualitative relative/ rankorder estimates of exposure	• Only 1 quantitative dose or concentration level in addition to the control[a] • Route of exposure *not* by inhalation, oral or dermal type (e.g., intraperitoneal, injection) • No duration of exposure stated • Exposure to TCE in a chemical mixture
	Mechanistic/Alternative Methods	• A minimum of 2 quantitative concentrations of TCE plus a negative control group[a] • Exposure to TCE only (no chemical mixtures)	• Only 1 quantitative dose or concentration level in addition to the control[a] • Exposure to TCE in a chemical mixture
Comparator	*Animal*	• Negative controls that are vehicle-only • treatment and/or no treatment	• Negative controls *other than* vehicle-only treatment or no treatment
	Mechanistic/Alternative Methods	• Negative controls that are vehicle-only • treatment and/or no treatment	• Negative controls *other than* vehicle-only treatment or no treatment
Outcome	*Animal*	• Endpoints described in the methylene chloride scope document:[b] o Acute toxicity o Liver toxicity o Kidney toxicity o Reproductive/developmental Toxicity o Neurotoxicity o Immunotoxicity o Sensitization o Cancer • Other endpoints[c]	
	Mechanistic/Alternative Methods	• All data that may inform mechanisms of • developmental toxicity	• Data that inform mechanisms of toxicity for endpoints *other than* developmental toxicity
General Considerations		**Papers/Features Included**	**Papers/Features Excluded**
		• Written in English[d] • Reports primary data or meta-analysis a • Full-text available • Reports both TCE exposure *and* a health outcome or mechanism of action	• Not written in English[d] • Reports secondary data (e.g., review papers)[a] • No full-text available (e.g., only a study description/abstract, out-of-print text) • Reports a TCE-related exposure *or* a health outcome/mechanism of action, but not both (e.g. incidence, prevalence report)

[a]Some of the studies that are excluded based on the PECO statement may be considered later during the systematic review process. For TCE, EPA will evaluate studies related to susceptibility and may evaluate, toxicokinetics and physiologically based pharmacokinetic models after other data (e.g., human and animal data identifying adverse health outcomes) are reviewed. EPA may also review other data as needed (e.g., animal studies using one concentration, review papers).

[b]EPA will review key and supporting studies in the IRIS assessment that were considered in the dose-response assessment for non-cancer and cancer endpoints as well as studies published after the IRIS assessment.

[c]EPA may screen for hazards other than those listed in the scope document if they were identified in the updated literature search that accompanied the scope document.

[d]EPA may translate studies as needed.

SOURCE: EPA 2018b.

Objective

The process for searching and selecting the evidence lacked objectivity, because the inclusion and exclusion criteria were broad and thus less objective. A benefit of systematic review is that clear, predefined inclusion or exclusion criteria increase objectivity of the process for selecting the evidence.

Transparent

Overall, the committee found that the lack of information about the specific processes used for the identification of evidence reduced confidence in the findings and were inconsistent with systematic review practices. Information about the search process was scattered across multiple documents within the docket for TCE, making the identification of details laborious and time consuming. The committee recommends organizing the information in one main document with clear references to supporting documents.

In the TSCA evaluation process, eligibility criteria are not predefined in the protocols and shift during the systematic review process. Outcomes specified are frequently too broad for true systematic review and would have been focused in scoping exercises. These shifts in inclusion and exclusion criteria are particularly problematic when used in building machine learning models as the shifting exclusion criteria may mislead and confuse the algorithm, resulting in exclusion of relevant studies. The committee also noted that the outcomes to be assessed are not specifically outlined in the protocol. If not the case, the systematic review and its conclusions are at risk of bias from incomplete reporting.

Recommendations

In order to improve these issues with OPPT's approach to evidence identification, the committee recommends the following:

- Registering the protocol for each risk evaluation is important: That protocol should include an explicit search strategy, and search strategies for each database should be consistently listed in the appendix to the risk evaluation.
- OPPT could improve the evidence identification process by requesting information from manufacturers, such as ingredients for products, and from organizations that have provided data previously during the peer-review stage. Such requests made earlier in the process could lead to more complete data gathering. TSCA provides OPPT with authority to collect information on chemical manufacturing, processing, and use, which could be used to collect information in advance of the risk evaluation (TSCA section 8(a) Reporting Requirements, 15 U.S.C. 2607).
- Machine learning and AI-based tools should be used for searching and screening, especially if the tools are validated by the developer and users or there are publications available that document this validation.
- Eligibility criteria need to be based on PECO statements that are formulated in a standard way and need to be predefined in the protocol. The eligibility of outcomes needs to be carefully considered a priori to prevent a systematic exclusion of outcomes that could bias the results, such as excluding studies that have findings counter to those anticipated for the included outcomes.
- Documentation of all studies identified in searches should be more clear with the provision of a list of included studies, detailed evidence tables of included studies, and documentation of excluded studies with reasons for exclusion.
- OPPT should specify the methods by which the screening will be conducted. Examples include the number of reviewers (e.g., two screeners or one screener and an AI tool) and how disagreements are handled.

EVALUATION OF THE EVIDENCE

Following evidence identification, the next step in the systematic review process is evaluation of the evidence (see Figure 2-1). In a systematic review, the individual studies are critically appraised using predefined criteria and then the body of evidence (i.e., all of the included studies for a particular question and outcome) is synthesized (qualitatively and/or quantitatively) and evaluated to draw a conclusion and specify a level of confidence in that conclusion. The systematic review should assess the strengths and limitations of the evidence so that decision makers and stakeholders can judge whether the data and results of the included studies are valid (IOM 2011).

State of the Practice

Individual human, animal and other ecological receptors, and mechanistic studies are assessed for internal validity (commonly referred to as "risk of bias" in systematic review) by considering aspects relevant to the type of study (OHAT 2019). Bias is a systematic error that leads to study results that differ from the truth. Bias can lead to an observed effect when in truth there is not one, or to no observed effect when there is a true effect. Risk of bias is the appropriate term, as a study may be unbiased despite a methodological flaw. The risk-of-bias assessment differs from an assessment of study quality, which is the appraisal of included studies to evaluate the extent to which study authors conducted their research to the highest possible standards (Higgins et al. 2011). Some tools assess risk of bias and study quality separately because the risk of bias addresses how valid the individual studies are; a study can be of high quality and still have a high risk of bias. Many markers of a high-quality study (e.g., whether a study's investigator has performed a sample size calculation and whether the study is reported adequately or has received appropriate ethical approvals) are unlikely to have any direct implication for the potential for a study to be affected by bias.

There are many tools for assessing risk of bias, such as those used by the Navigation Guide, OHAT, and the IRIS Program, and there is no consensus on the best tool for risk-of-bias analysis. However, there are best practices. For example, tools are preferred that rely on the evaluation of individual domains rather than the creation of overall quality scores (Eick et al. 2020). Such tools provide a structured framework within which to make qualitative decisions on the overall quality of studies and to identify potential sources of bias. Overall quality scores may not adequately distinguish between studies with high and low risk of bias in meta-analyses (Herbison et al. 2006). Importantly, there is also a lack of empirical evidence on the use of quality scores (Jüni et al. 1999).

While there is inevitably variation in the internal validity and risk of bias across individual studies, it is standard practice to include all studies, even the studies with a high risk of bias into the evidence synthesis. The most appropriate method to exclude studies from evidence synthesis is based on predefined exclusion criteria that should preclude an irrelevant study from being evaluated.

Although there is not a specific standard of practice for evaluating exposure data, the agency *Guidelines for Human Exposure Assessment* discuss the importance of critically reviewing data for use in an exposure assessment. To address the quality of analytical methods, the *Guidelines for Human Exposure Assessment* suggest a series of questions that should be asked when reviewing data for use in an exposure assessment: Has an authoritative body adopted these (and other considerations about whether the exposure data are useful for the research question being addressed in the exposure assessment)? Were the study objectives and designs suitable for the purpose of the exposure assessment? When evaluating the study data, consideration needs to be given to potential bias in the exposure data, which may be selective for high or low exposures; for example, some occupational monitoring data focus on the most highly exposed workers (EPA 2019b). For data on human exposures that are generated by mathematical models, the *Guidelines for Human Exposure Assessment* also discuss methods for model evaluation to test whether the analytical results from the model are of sufficient quality to serve as a basis for decisions.

To complete a model evaluation, the model operation and results are verified both qualitatively and quantitatively through calibration, or the process of adjusting selected model parameters within an expected range until the differences between model predictions and field observations meet selected criteria. Then, important sources of uncertainty, including measurement error, statistical sampling error, non-representativeness of data, and structural uncertainties in scenarios and formulations of models, are checked. Sensitivity analysis may also be conducted to determine the extent to which estimates are dependent on variability and uncertainty in the parameters within the model. The guidelines for model evaluation apply to several different types of models included in exposure assessment, such as physiologically based pharmacokinetic (PBPK) modeling and fate and transport models (EPA 2019b).

As yet, there is not a complementary tool matching the NTP's OHAT Risk-of-Bias tool for application to ecotoxicology studies. One increasingly used method for assessing the quality of ecotoxicity studies is the Criteria for Reporting and Evaluating Ecotoxicity Data (CRED) (Moermond et al. 2016). CRED, which was built from Klimisch et al. (1997), presents a comprehensive and state-of-the-practice approach for evaluation of ecotoxicological information. CRED includes four reliability categories: reliable without restrictions, reliable with restrictions, not reliable, and not assignable. These are used for 20 reliability criteria falling into the categories of test set-up, test compound, test organisms, exposure conditions, and statistical design and biological response. CRED was developed, in part, to ensure that high-quality ecotoxicology information, including mechanistic evidence, is not excluded a priori from regulatory assessment processes simply because a study was not performed according to a standardized protocol using a standardized model species. The reliability of CRED as an assessment approach was determined from a ring trial that compared it to the Klimisch method (Kase et al. 2016). Results showed that the CRED evaluation method was a more detailed and transparent evaluation of reliability and relevance than the Klimisch method. Ring test participants perceived it to be less dependent on expert judgment, more accurate and consistent, and practical regarding the use of criteria and time needed for performing an evaluation.

Committee Description of the Approach in TSCA Risk Evaluations

OPPT has developed an extensive de novo critical appraisal tool, termed TSCA's "fit-for-purpose evaluation framework," which is applied to human, animal, ecological receptors, mechanistic, exposure, fate, and physical chemical property studies. OPPT has stated that the evaluation strategies were developed after review of various qualitative and quantitative scoring systems. OPPT considered items such as NTP's OHAT Risk-of-Bias tool, CRED, and EPA ORD's draft IRIS handbook. These tools were not adopted because they do not encompass the entirety of TSCA's scope and specifically do not include either exposure assessment or fate and transport assessment (Francesca Branch, presentation to the committee, July 23, 2020).

The critical appraisals for different types of studies use different domains (see Table 2-4), but in general the method of study evaluation is the same. The data quality domains are based on a variety of sources, but they were not directly adopted from any one source for any domain. For example, the domains for ecological hazards are based on CRED and criteria from the ECOTOX knowledgebase. Within each domain there are several metrics or questions. For the epidemiologic studies there are 7 domains, and within each domain there are between 2 and 7 metrics for a total of 22 metrics. For each metric, options include "high" (a quantitative score of 1), "medium" (score of 2), "low" (score of 3), and "unacceptable" (score of 4). The metrics generally include items that assess elements of study quality, risk of bias, reporting quality, and relevance. For example, statistical power is included within the test organism data quality domain used to assess the ecological hazard and toxicity studies and the study participation quality domain of the epidemiologic studies. These metrics are weighted with another numeric value (1 or 2), and then a multi-metric numeric value is identified, based on the sum of the weighted metric scores divided by the total number of metric weighting factors. This final study quality score is then cat-

TABLE 2-4 TSCA Data Quality Domains by Data Stream

Data Stream	Data Sources or Types of Studies	Data Quality Domain
Physical chemical property data	Not listed	Representativeness, appropriateness, evaluation/review, reliability/unbiased (method objectivity), reliability/analytic method
Fate data	Experimental data, field studies, modeling data, monitoring data	Test substance, test design, test conditions, test organisms, outcome assessment, confounding/variable control, data presentation and analysis
Occupational exposure and environmental release data	Monitoring data, environmental release data, published models for exposures or releases, completed exposure or risk assessments, reports for data or information other than exposure or release data	Reliability, representativeness, accessibility/clarity, variability and uncertainty
Consumer, general population, and environmental exposure data	Monitoring data, modeling data, survey-based data, epidemiological data, experimental data, completed exposure assessments and risk characterizations, database sources not unique to a chemical	Reliability, representativeness, accessibility/clarity, variability and uncertainty
Ecological hazard	Acute and chronic toxicity to aquatic invertebrates and fish (e.g., freshwater, saltwater, and sediment-based exposures); toxicity to algae, Cyanobacteria, and other microorganisms; toxicity to terrestrial invertebrates; acute oral toxicity to birds; toxicity to reproduction of birds; toxicity to terrestrial plants; toxicity to mammalian wildlife	Test substance, test design, exposure characterization, test organism, outcome assessment, confounding/variable control, and data presentation and analysis
Animal and in vitro toxicity studies	Animal: oral, dermal, and inhalation routes; lethality, irritation, sensitization, reproduction, fertility, developmental, neurotoxicity, carcinogenicity, systemic toxicity, metabolism, pharmacokinetics, absorption, immunotoxicity, genotoxicity, mutagenicity, endocrine disruption In vitro: irritation, corrosion, sensitization, genotoxicity, dermal absorption, phototoxicity, ligand binding, steroidogenesis, developmental, organ toxicity, mechanisms, high throughput, immunotoxicity	Test substance, test design, exposure characterization, test organism/test model, outcome assessment, confounding/variable control, and data presentation and analysis
Epidemiological studies	Controlled exposure, cohort, case-control, cross-sectional, case-crossover	Study participation, exposure characterization, outcome assessment, potential confounding/variability control, analysis

SOURCE: Derived from EPA 2018a.

egorized as high (≥1 to <1.7), medium (≥1.7 to <2.3), low (≥2.3 to ≤3), or unacceptable (4). It is worth noting that if any of the metrics are scored as "unacceptable" then the final study quality score is also deemed unacceptable. Strengths and limitations are considered when assigning a quality rating for each relevant metric. With proper justification, a reviewer may adjust the overall quality rating to capture professional judgment not originally included in the metric criteria (Francesca Branch, poster presentation to the committee, August 24, 2020). Unacceptable studies are then excluded from further analysis. The details of TSCA's "fit-for-purpose evaluation framework" are included as appendices to the 2018 guidance document. Generally, the approach is tiered with the following steps:

- A check for relevance of data;
- A quality evaluation that considers all reasonably available data deemed potentially relevant to the risk evaluation;
- Reporting quality and risk-of-bias elements integrated in the review process;
- Elimination of unacceptable studies from further consideration;
- Weighting of criteria for key elements, in some cases; and
- The planning, execution, and QA/QC assessment activities supporting the data evaluation.

Critique of the TSCA Approach

Looking at the core review elements of the Statement of Task, which address whether the OPPT approach to systematic review is "comprehensive, workable, objective, and transparent," the committee finds the TSCA "fit-for-purpose evaluation framework" could improve on all these elements.

Comprehensive

TSCA's "fit-for-purpose evaluation framework" is comprehensive in that it contains many domains and each domain has many metrics. Yet, the committee found that although the 2018 guidance document discusses the use of PBPK models in risk assessments, and that OPPT will use evaluation strategies for animal and in vitro toxicity data to assess the quality of the data supporting the model, the document does not give guidance as to how these models will be evaluated. The committee could not find evidence of this practice being followed in the draft TCE risk evaluation document and supplemental materials.

Similarly, exposure models are scored only based on reliability, representativeness, accessibility or clarity, and variability or uncertainty, but little seems to be done to actually evaluate the model. For example, a model could score "high" in the reliability category simply for being published in a peer-reviewed journal, when in fact the reliability of the model may never have been evaluated.

Workable

The committee notes that TSCA's "fit-for-purpose evaluation framework" may not produce the desired results. It includes items that do not assess risk of bias, such as relevance and incomplete reporting. In systematic review, study relevance should be addressed with predefined eligibility criteria, which are used during screening to select relevant studies and exclude irrelevant studies from the evaluation step. For example, if the systematic review is focused on a certain life stage, then studies not including that life stage would be excluded in ether title-abstract or full-text screening. Incomplete reporting can be a challenge in evaluating a study, but it is not a marker of the validity of the study findings.

The reliance on numeric quality scores is problematic because scores do not distinguish between high- and low-quality studies, and the relationship between quality scores and an association or effect is inconsistent and unpredictable (Greenland and O'Rouke 2001; Herbison et al. 2006; Jüni et al. 1999).

More generally, the use of numerical scoring in critical appraisal does not follow standards for the conduct of systematic reviews. Additionally, there was no justification provided for the weighting of specific metrics within the domains to create the overall quality score, making it difficult to determine if the weights are appropriate. The committee notes that many public comments also discussed these problems with using numeric scores to evaluate studies.

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

Objective

All evaluations of studies have an inherent subjectivity, which is not overcome with numeric scoring. No data were shared with the committee showing that the TSCA evaluation framework and numeric scoring schema had been validated or tested for reliability. Allowing a reviewer to override the score after it has been applied is another threat to objectivity.

Another problematic element of TSCA's "fit-for-purpose evaluation framework" is that the unacceptable studies are excluded from further analyses. Any fatal flaws in the methodology or conduct should be included in the exclusion criteria applied during the screening process. Once a study is determined to be eligible, the study could be included in the synthesis and the risk-of-bias assessment and its limitations accounted for in any qualitative or quantitative synthesis. Given the large number of metrics scored for these data types, the possibility that a single unsatisfactory rating could nullify the use of a particular study from synthesis is problematic, as it may lead to a biased review. In the synthesis step, low-quality studies may be excluded as a sensitivity analysis, but it is inappropriate to leave them out of synthesis completely.

Statistical power and statistical significance are not markers of risk of bias or quality. Statistical significance is not a measure of association or strength of association and should not be used to evaluate studies. In fact, combining multiple small, low-powered but similar studies in a synthesis is one of the potential benefits of systematic review. The committee notes that this critique was also shared in many public comments reviewed and was published in a commentary in the *American Journal of Public Health* (Singla 2019).

Transparent

The committee found that several facets of TSCA's "fit-for-purpose evaluation framework" lacked transparency. The use of numeric scores prevents users of the reviews from making their own determinations about important strengths and limitations in study methods based on results presented in the review. The process of critical appraisal of the studies is not documented. OPPT reported to the committee that it generally follows standard practice by using two independent reviewers, but it was unclear how discrepancies between the two reviewers are handled.

The committee also found that the documentation of the study scores was hard to follow. In the TCE evaluation, the environmental releases and occupational exposure study scores are tracked in one document and the data extraction is tracked in another. There were studies listed as having been extracted that were not included in the tables of extracted studies. One study was excluded because it only had occupational exposure data, but there was no indication as to whether it was then evaluated as a source of occupational data. There were studies for which the score changed from the scoring document to the extraction document without documentation as to why.

Recommendations

In order to improve these issues with TSCA's "fit-for-purpose evaluation framework," the committee recommends the following:

- Do not use numeric scores to evaluate studies; replace them with domain-based scoring as is done in the tools used in the Navigation Guide and OHAT.
- Use established tools for assessing risk of bias and study quality such as those developed for use by OHAT or the Navigation Guide, or, at a minimum, remove inappropriate appraisal criteria from the current tools.
- Do not exclude studies based on risk of bias, study quality, or reporting quality.
- The CRED approach is robust and should continue to be employed during TSCA risk evaluations for ecotoxicology information. Clarify this interface with the ECOTOX knowledgebase. A risk-of-bias instrument should be developed for ecotoxicology studies, building from the OHAT Risk-of-Bias approach.
- Develop a method for clearly tracking how articles are handled through the steps of extraction and evidence evaluation, so those reading the risk evaluation can follow how articles are handled in each step.
- Improve documentation of how the evaluation is applied, including at what points in the process quality determinations may be changed and on what basis.

EVIDENCE SYNTHESIS

Evidence synthesis represents the next step in a systematic review. It involves a qualitative analysis as well as a quantitative analysis, when feasible and appropriate, of the results of the individual studies within a stream of evidence (e.g., fetal cardiac effects of TCE in animal models).

State of the Practice

According to the *Cochrane Handbook for Systematic Reviews of Interventions*, evidence synthesis is a process of bringing together data from a set of included studies with the aim of drawing conclusions about a body of evidence (Higgins et al. 2019). The process consists of summarizing study characteristics, quality, and effects, and combining results and exploring differences among the studies (e.g., variability of findings and uncertainties), using qualitative and/or quantitative methods. The choice of methods for data synthesis depends on data completeness and the hypothesis to be addressed. Whereas statistical methods, such as meta-analysis, are often preferred for combining results quantitatively, other methods, including graphic displays, may be used when a statistical method is not feasible or appropriate due to incomplete or incompatible data extracted from different studies. To ensure transparency and consistency of a systematic review, approaches and methods for evidence synthesis should be prespecified in the protocol. Following evidence synthesis, evidence integration across multiple evidence streams—an essential step in the TSCA risk evaluation process—is done but is not a part of the traditional systematic review process.

There is much research and a growing consensus on how certainty in a body of evidence should be determined in the field of toxicology. For systematic reviews for environmental health assessments, Rooney et al. (2014) proposed a system for rating certainty in the overall body of evidence across outcomes based on GRADE (Grading of Recommendations Assessment, Development and Evaluation) (Balshem et al. 2011; Guyatt et al. 2011). The GRADE environmental health working group is adapting the original GRADE approach to environmental health. For each outcome, the body of evidence

is assessed to determine a final certainty in evidence by rating high certainty, moderate certainty, low certainty, or very low certainty. Conventionally, bodies of evidence for outcomes informed by randomized controlled trials (RCTs) start at high initial certainty and non-randomized studies start at low initial certainty to account for the lack of a prognostic balance (present in well-done RCTs). This initial level of certainty may be decreased or increased if certain attributes are present in the studies. The criteria that can lower the certainty are as follows: (1) limitations of detailed design and execution (overall risk of bias), (2) inconsistency (or heterogeneity), (3) indirectness to PECO and applicability to main PECO, (4) imprecision (number of events and confidence intervals), and (5) publication bias. It is necessary to be transparent about considerations that informed this judgment, and the ranking is normally done by two reviewers with a conflict resolution process (defined in the protocol) in place. This certainty rating is considered helpful as it increases the transparency of the final conclusions.

The approach developed by Morgan et al. (2016) and the approach used by the Navigation Guide (Woodruff and Sutton 2014) are similar to the original GRADE approach, but they use slightly different criteria for setting initial levels of certainty and for upgrading and downgrading. For example, in the Navigation Guide observational studies start at a moderate rather than a low level of confidence, and the NTP uses consistency across species as an extra upgrading criterion. These examples illustrate that there is not yet consensus in the toxicology field regarding evidence synthesis, but there is some convergence on common baseline methods, such as the GRADE approach, to bring consistency. Moreover, the committee would like to emphasize that all of the above mentioned methods have undergone pilot testing, stakeholder vetting, and peer review and have been made public.

Within the agency, the *Guidelines for Human Exposure Assessment*, the *Guidelines for Ecological Risk Assessment*, and the operating procedures for the use of the ECOTOX knowledgebase dictate how exposure, fate and transport, and physical chemical property data should be synthesized for decisions about risks to human health and ecological receptors. The *Guidelines for Human Exposure Assessment* indicate that aggregate exposures integrated across sources and routes of exposure should be estimated and that probabilistic exposure models are preferred to those that are based on a single data point, such as measure of central tendency. Sensitivity analyses are also suggested to account for the most important sources of uncertainty in the model (EPA 1998, 2019b, 2020b). The standard of practice for species sensitivity distributions (SSDs) in risk evaluations is the SSD Toolbox.[8]

Committee Description of the Approach in TSCA Risk Evaluations

Section 3.4 of the 2018 guidance document does not separate evidence synthesis from evidence integration and instead combines the two into a single step of data integration (EPA 2018a, p. 26):

> Data integration is the stage where the analysis, synthesis and integration of data/information takes place by considering quality, consistency, relevancy, coherence and biological plausibility. It is in this stage where the weight of the scientific evidence approach is applied to evaluate and synthetize multiple evidence streams in order to support the chemical risk evaluation.

A TSCA risk evaluation involves multiple research questions, which are combined into three core elements—hazard assessment, exposure assessment, and dose-response assessment—which are then integrated for risk characterizations. To determine whether and how a systematic review approach can be followed for evidence synthesis, the specific stream of evidence that needs to be synthesized must first be identified. As noted throughout this chapter, the goals and objectives of the systematic reviews conducted within TSCA risk evaluations are unclear, resulting in lack of clarity on what should be synthesized.

[8]See https://www.epa.gov/chemical-research/species-sensitivity-distribution-ssd-toolbox, accessed November 13, 2020.

For risk evaluation of TCE, OPPT developed a problem formulation and scoping document (EPA 2018b), which indicates four categories of evidence: physical and chemical properties, conditions of use, exposures, and hazards. Each of these categories may consist of multiple evidence streams. For example, exposures may be divided into environmental release, fate and transport, environmental exposure, and human exposure (occupational versus consumers). Evidences in environmental release and fate and transport are upstream events to environmental and human exposure. Hazards can be divided into human health hazards and environmental hazards. OPPT developed a PESO statement for environmental release and fate, a RESO statement for engineering and occupational exposure, a PECO statement for consumers and ecological receptors, and a PECO statement for human hazards.

The PECO statement for human hazards indicates three parallel streams of evidence: human, animal, and mechanistic/alternative methods. Taking into consideration species, route of exposure, type of health outcome (e.g., cancer versus non-cancer), and organ or body system, there could be a large number of research questions that could be evaluated with a systematic review and ultimately synthesized. The committee recognizes that in the context of TSCA risk evaluation, defining a sub-stream of evidence is not always trivial and may not be unique but is a prerequisite. It finds that, in the absence of an explicit definition of an evidence stream, it is difficult to assess whether OPPT should follow a systematic review approach for evidence synthesis and, if so, how evidence synthesis should be conducted. Making that judgment is further complicated by OPPT's merger of evidence synthesis within a stream and the step of evidence integration across streams. The following are a few examples from the committee's review of the draft TCE evaluation.

Environmental release and aquatic exposure are two components that are important for evaluating questions about ecological risks. OPPT assessed aquatic exposure based on monitoring data, but where monitoring data were unavailable, it used model-predicted surface water concentrations with environmental release data as input to the E-FAST Version 2014 model, but it did not provide a plan for assessing sensitivity and validity of such a model-based synthesis. The draft TCE evaluation stated that OPPT followed a systematic review process for the monitoring data stream, although the problem formulation document failed to indicate so. Evidence synthesis for both environmental release and monitoring data could be made more quantitative for each scenario because the data appear adequate for analysis of range and distribution.

For the human exposure stream, OPPT calculated the central tendency and a high-end estimate of exposure for each occupational or consumer exposure scenario, using both model-based estimates and, when available, measured data. However, the committee could find no information presented in the draft TCE evaluation on how to combine data from multiple datasets. When evaluating how this defined process is applied in the case of concentrations in dry cleaners, for example, the committee found that documentation was not transparent. Concentrations are not presented for the individual studies selected, and there is no mention as to how they were combined.

In synthesizing the environmental hazard stream, OPPT adopted statistical methods to derive an SSD for aquatic species (Etterson 2020). SSDs, which are routinely used by the agency to derive water quality criteria for protection of aquatic life and to assess ecological risks of pesticides (Posthuma et al. 2001), are more advantageous than deterministic approaches, including hazard quotients, when data are adequate. For example, SSDs fit common toxicity values (e.g., no observable effect concentration, EC50) from all included aquatic studies along a probabilistic distribution, from which the 5-percentile or HC05 can be selected to quantify hazard on the basis of the entire stream of evidence. The SSDs are innovative because the distribution, hence HC05, effectively accounts for variability (e.g., species sensitivity at different trophic position) and uncertainty (e.g., between-experiment variation). NRC made a recommendation in relation to the 2010 IRIS assessment of TCE to encourage the agency to fit a probabilistic distribution for any comparable values of the point of departure (POD), such as benchmark dose level (NRC 2010). SSDs are common probabilistic hazard assessment tools utilized elsewhere by EPA (e.g., Office of Water and Office of Pesticide Programs) that when coupled with probabilistic environmental

exposure distributions (EEDs) of measured or predicted chemical concentrations, provide robust inputs for performing probabilistic ecological risk assessments. Thus, these EEDs and SSDs represent useful implementation opportunities for meta-analysis within the context of risk assessment.

OPPT followed a systematic review approach within evidence streams of human, animal and other ecological receptors, and mechanistic models, respectively, as indicated in the problem formulation document (EPA 2018b). It is less clear whether OPPT considered all non-cancer endpoints as a single research question or developed more refined questions for each organ or body system. Evidence synthesis for non-cancer endpoints was qualitative, in which the key study or studies were chosen to advance to later steps. This approach can be improved by using more quantitative methods such as fitting a probabilistic distribution for associated PODs, as illustrated and recommended by NRC (2010). The use of SSDs for environmental hazards also illustrates such approaches. For cancer endpoints, the TCE risk evaluation included a meta-analysis to synthesize evidence, combining three cancers—non-Hodgkin lymphoma, kidney cancer, and liver cancer—in human studies. Although OPPT followed a systematic review approach for evidence from mechanistic models, the number of mechanistic studies was limited and the evidence synthesis was narrative, concluding that a genotoxic mode of action is highly plausible for kidney cancer.

The committee notes that OPPT's risk evaluation has consistently discussed sources of uncertainty, occasionally described the range of an uncertainty factor, but rarely quantified the impact of uncertainties on downstream events along the risk assessment process. For example, distributional properties such as percentiles can be derived for aquatic concentrations of TCE within an occupational or consumer exposure scenario or across scenarios. The application of SSDs for environmental hazard represents a good example of using probabilistic distributions to quantify uncertainty and variability, but they are dependent on sufficient data availability. Probabilistic distribution–based approaches for acute and chronic datasets should be utilized when feasible and appropriate. The committee also notes that in the 2018 guidance document (EPA 2018a), OPPT does not provide guidance to identify specific research questions for which data should be synthesized.

OPPT conducted a dose-response assessment for human cancer and non-cancer endpoints in its evaluations of 1-BP and TCE. For non-cancer endpoints, OPPT included the studies appropriate for the dose-response assessment and organized them by exposure route, chronic or acute exposure, and target organ. OPPT then estimated a POD using either benchmark dose software (BMDS), no observed adverse effect level, or lowest observed adverse effect level for every included study after fitting every dose-response model implemented in the BMDS (EPA 2012a), and converted the PODs to human equivalent concentrations (HECs) applying relevant uncertainty factors. Finally, OPPT synthesized these studies by selecting the most sensitive POD and associated HEC both within and across the organ or exposure groups. The committee also observed that in the evaluation of 1,4-Dioxane (EPA 2019a), OPPT chose the POD associated with the best statistically fit dose-response model (EDF 2020).

OPPT conducted a similar dose-response assessment for cancer endpoints. However, the dose-response assessment of cancer endpoints was inconsistent across different evaluations. In the evaluation of TCE, for example, OPPT used three models for each cancer endpoint: a multistage model; an average (frequentist) model over multistage, log-probit, and Weibull models; and a Bayes' average model, included in the BMDS. OPPT then chose the most sensitive POD and associated risk estimate. In the evaluation of 1,4-Dioxane, however, OPPT used a different approach to synthesize across multiple cancers by looking at cancer risk at any sites in addition to cancer at an individual site. This approach was implemented in the statistical model MS-Combo in BMDS, which assumes that cancer risk at one site is independent of another. This composite cancer risk approach has the statistical property of being more sensitive than estimating the risk of any individual cancer. As a result, the risk estimate from the MS-Combo becomes the default outcome if the objective is to determine the most conservative risk estimate. There is a lack of consistency in current TSCA practice with respect to cancer dose-response modeling of 1,4-Dioxane and TCE. The model-average or MS-Combo approach is not yet state of the practice in risk

assessment. The systematic review protocol in these evaluations did not include choice and justification of adopting these approaches.

Risk assessment in general, and evidence synthesis and integration in particular, can could benefit from emerging methodologies such as key characteristic approaches andt the the framework of Adverse Outcome Pathways (AOP). The key characteristics approaches are designed to facilitate the organization and utilization of mechanistic information in hazard identification (Arzuga et al. 2019; La Merrill et al. 2020; Smith et al. 2016). An AOP pieces together key chemical exposure induced biological events, from molecular and cellular level activities to tissues and organ level effects and to population health along plausible pathways to inform relevance and causality of adverse outcomes (Villeneuve et al. 2014a,b). As reviewed in the 2017 National Academies report *Using 21st Century Science to Improve Risk-Related Evaluations* (NASEM 2017), the AOP framework is still evolving and has not become the state-of-practice for risk-based evaluation. The committee encourages OPPT to track these and other evolving methodologies.

Critique of the TSCA Approach

OPPT failed to clearly identify evidence streams, especially sub-streams, for which it determined to follow a systematic review approach for evaluation. Looking at the core review elements of the Statement of Task, which address whether the TSCA approach to systematic review is "comprehensive, workable, objective, and transparent," the committee finds that the TSCA approach to evidence synthesis could be improved with regard to each of these four desired characteristics.

Comprehensive

The OPPT approach for evidence synthesis is not comprehensive, as it does not contain elements that are important to addressing the research question. OPPT failed in the TCE draft evaluation to include all eligible datasets, after a critical appraisal, to describe the variability and uncertainty when data were adequate and relevant. OPPT could have addressed the level of confidence in the body of evidence more comprehensively and more quantitatively if taking a more probabilistic and quantitative approach. The lack of definition of evidence streams that require a data synthesis also makes a judgment of comprehensiveness difficult.

Workable

Without a clear definition of evidence streams or documented approaches to evidence synthesis within each evidence stream, it is difficult to assess and reproduce the results of evidence synthesis. Concurrent implementation of evidence synthesis and evidence integration further makes evidence synthesis approaches of TSCA risk evaluations less workable.

Objective

The committee did not find a consistent pattern in documenting the objectives or methods of choice for synthesis in the draft risk evaluation of TCE or the support documents. Characterizing variability and uncertainty within the body of evidence is an essential objective of data synthesis. In situations in which data appeared adequate to support a more quantitative discussion of uncertainty and variability analysis, OPPT offered only a qualitative discussion of the sources of uncertainty. Examples include surface water concentration estimates and human hazards of non-cancer endpoints. When adequate and appropriate, a quantitative (e.g., a probabilistic distribution such as SSD) approach for evidence synthesis, especially in view of variability and uncertainty, is preferred. This approach also applies to the syn-

thesis of PODs after dose-response assessment using the benchmark dose approach in contrast to the current approach that chooses either the most sensitive risk value, or the best fitted statistical model, or average. This preference holds whether or not OPPT determines to adopt a systematic review approach for an evidence stream.

Transparent

The absence of a well-documented protocol reduces transparency and consistency of evidence synthesis. Lack of documentation and justification for the use of average modeling or composite risk modeling for cancer dose-response assessment in the evaluation of 1,4-Dioxane and TCE raises concern about consistency in TSCA risk evaluations. While the committee recognizes the value of exploring new methodologies in this case, careful examination and full documentation of the underlying assumptions and requirements of these methods are necessary to ensure transparency and consistency. The MS-Combo, for example, assumes that cancers at different sites are independent, an assumption that needs justification (EPA 2012b). Violation of this assumption likely leads to a downward bias in risk estimation. For the model-averaging approach, including every model offered in the BMDS as OPPT did in the 1-BP evaluation for the Bayes average cancer model requires strong justification. OPPT did not document inclusion criteria or provide justification regarding selection of the underlying model components. Evaluation of evidence synthesis was also complicated and made less transparent and consistent by the fact that OPPT carried out evidence synthesis (within an evidence stream) and evidence integration (across multiple evidence streams) concurrently. Public comments concurred with the committee's observation of a lack of transparency, as this step was not noted within the TSCA risk evaluations or the 2018 guidance document.

Recommendations

OPPT's adaptation of systematic review to support its chemical risk evaluation creates both challenges and flexibilities. In order to improve these issues with TSCA's process for evidence synthesis, the committee recommends the following:

- Document the synthesis methods and data requirements in the study protocol for each evidence stream. This should be done for every evidence stream whether or not a systematic review approach will be taken.
- Separate evidence synthesis from evidence integration. Evidence synthesis deals with more homogeneous data within a single stream, and evidence integration deals with more heterogeneous data from multiple streams.
- Develop guidance for evidence synthesis, considering objectives, data requirements, strengths, and limitations. Seek stakeholders' input, incorporate feedback, document changes, and amend the 2018 guidance document. Refer to this amended guidance document in the protocol(s).
- Strive for more quantitative analysis of uncertainty and variability by using a probabilistic distribution approach (e.g., SSDs) or at a minimum describing distributional properties as recommended by NRC (2010).
- Strengthen analysis of uncertainty and variability by focusing on the propagation and aggregation of uncertainty and variability from upstream to downstream of the risk assessment process. For example, it is important to understand how uncertainty and variability in environmental release of TCE affect the uncertainty and variability of model-predicted concentrations in surface water and how estimated concentrations affect environmental or human hazards.
- Develop a clear approach for extracting data from occupational exposure studies and clearly document values extracted from each study. Then, clearly document how values from various

studies are synthesized into a single estimate of the central tendency and high-end exposure estimates.

- Dose-response assessment is amenable to systematic reviews of hazard assessment. The process of dose-response assessment should be planned, justified, and prespecified in the systematic review protocol, including eligibility criteria for study and endpoint selection (e.g., number of dose levels, group size), model selection (e.g., biological plausibility, statistical appropriateness), and methods for synthesis (e.g., probabilistic distribution of PODs and the most conservative risk estimate). Upon completion of the dose-response assessment and associated evidence synthesis, results should be reported in accordance to the criteria specified in the protocol.

EVIDENCE INTEGRATION

State of the Practice

Evidence integration is typically considered outside of the systematic review process itself (see Figure 2-1) and, in the context of risk evaluations, only employed when an evaluation reviews different evidence streams that have to be reconciled and, as the name suggests, integrated. The outcome of the integration step is an overall conclusion that is based on the holistic consideration of the various evidence streams. Risk evaluations that only contain data from a single stream of evidence (e.g., a systematic review on hypospadias and dibutyl phthalate in rats) would not have other evidence streams to integrate and could therefore be finalized after the evidence synthesis or move straight into hazard and exposure integration (see Figure 2-1). For evaluations with multiple streams of evidence (i.e., human and animal data), the highest level of evidence conclusions for a specific health outcome are integrated to inform a hazard assessment (OHAT 2019). This applies whether or not the data support the health effect conclusion or provide no evidence of effect. A qualitative descriptor may be applied to each stream of evidence based on a specific health effect or group of effects (e.g., whether there is sufficient evidence or suggestive evidence of an effect of anti-androgenic toxicity due to dibutyl phthalate exposure in rats), and then an overall conclusion is drawn from consideration of the streams together (i.e., the integration). As demonstrated in Figure 2-1, there is a process for integrating across evidence streams to reach the exposure assessment (see pink boxes) and a process for the hazard assessment; these two parts are then also integrated to inform the risk characterization.

Expert judgment is typically used to assign descriptor categories that describe a final conclusion, although there are various guidelines for bringing evidence together from the different streams (e.g., GRADE, adapted for animal studies based on the historical Bradford-Hill considerations (Hill 1965; Hooijmans et al. 2018; OHAT 2019; Woodruff and Sutton 2014). Although these examples are more in line for hazard identification, similar principles might also apply when integrating other types of evidence, such as exposure data (NRC 2014). In the end, the evidence integration step is generally tasked with answering the question of causation (NRC 2014). This should be done for each health effect that the data support through a narrative, qualitative assessment of all evidence streams (NRC 2014).

Committee Description of the Approach in TSCA Risk Evaluations

According to Table 3-1 in the 2018 guidance document, the evidence integration step has three phases. The first phase (planning) involves the development of a strategy for analyzing and summarizing data across studies within each evidence stream (discussed previously by the committee as synthesis) and a strategy for weighing and integrating evidence across those streams. Phase two (execution) involves the implementation of the strategies developed in the planning phase as well as the development of WOE conclusions. The table notes that documented professional judgments may be invoked in some

of these analyses. The third phase involves a check on the quality of the data used and discussion of uncertainties.

During phase two, OPPT conducts a WOE evaluation, as the term is generally used, because it is here that EPA makes judgment-based decisions to infer causation. The results of the evidence integration are applied to support the risk evaluation that is the basis for decision making. The Risk Evaluation Rule requires TSCA risk decisions to be based on such WOE evaluations (EPA 2018a, pp. 26–27 and Table 3-1). However, the committee notes some confusion around the terminology used to describe the steps, as detailed in Box 2-4.

BOX 2-4 Weight of the Scientific Evidence and Evidence Integration

In discussing OPPT's approach to evidence integration, the committee reviewed the definition of "weight of the scientific evidence" and its application within the evidence integration stage. While the definition of weight of the scientific evidence was introduced in Chapter 1, given that it is a key element of this stage, a deeper discussion is included here. According to the 2018 guidance document (EPA 2018a, pp. 26–27),

> the weight of the scientific evidence is defined as "*a systematic review method, applied in a manner suited to the nature of the evidence or decision, that uses a pre-established protocol to comprehensively, objectively, transparently, and consistently identify and evaluate each stream of evidence, including strengths, limitations, and relevance of each study and to integrate evidence as necessary and appropriate based upon strengths, limitations, and relevance*." 40 C.F.R. 702.33. In other words, WOE will involve assembling the relevant data and evaluating the data for quality and relevance, followed by synthesis and integration of the evidence to support conclusions (EPA, 2016). The significant issues, strengths, and limitations of the data and the uncertainties that require consideration will be presented, and the major points of interpretation will be highlighted. Professional judgment will be used at every step of the process and will be applied transparently, clearly documented, and to the extent possible, follow principles and procedures that are articulated prior to conducting the assessment (EPA, 2016).

This definition of WOE seems to say that the TSCA systematic review is *itself* a WOE evaluation. As such, the agency's legal obligation to conduct a WOE evaluation is fulfilled by the fact that systematic review is the basis for TSCA evaluations. The 2018 guidance document also states that the WOE approach is *applied* at the data integration stage "to evaluate and synthesize multiple evidence streams." Thus, under this description, WOE is *applied at one stage* in the systematic review process, whereas under the interpretation described above, the systematic review process is itself a WOE evaluation. The committee also notes that Table 3-1 of the 2018 guidance document describes WOE as a component of data integration,[a] separate from the data evaluation and the other components of systematic review.

However, not all of the presentations made by OPPT refer to WOE as a "systematic review method," but rather they adopt language closer to more traditional WOE definitions. The presentation of Drs. Barone and Wong, titled "Evidence Integration Supporting Exposure and Hazard Assessments for TSCA Risk Evaluations" (Stanley Barone and Eva Wong, presentation to the committee, July 23, 2020),[b] defines "Integration—weight of the scientific evidence" as "judgments regarding the strength of evidence for exposure or health effect developed by looking across evidence streams." The presentation had no description of how a WOE evaluation is conducted. This is perhaps due to the confusion that the systematic review process is, as a whole, a WOE evaluation.

Some of the lack of clarity in the 2018 guidance document and in the July 2020 presentation was remedied in a follow-up poster presentation (Koehrn et al. presentation to the committee, August 23, 2020) and in a second presentation by Dr. Barone (August 24, 2020). These presentations make it clear that the agency considers that the TSCA systematic review process constitutes the required WOE evaluation, although the WOE is still somewhat confusingly said to be applied at the evidence integration stage. "Strength of evidence" and "confidence level for risk estimation" are defined and are described as "qualitative judgments" for both hazard and exposure.

---

[a] It is worth noting that the 2018 guidance document uses data integration to mean the same as evidence integration. Evidence and data are not necessarily synonymous and should be distinguished.

[b] The committee notes that this presentation refers to evidence integration, rather than data integration.

In the 2018 guidance document, the descriptions of evidence integration lack details and specificity. As noted previously, the committee examined TSCA risk evaluations for 1-BP and TCE to develop opinions on the clarity and appropriateness of EPA's approach, and OPPT provided the committee with several presentations and posters to elaborate on the processes. In those presentations, OPPT presented the evidence integration step as comprised of an evaluation of individual evidence streams by summarizing the strength of the evidence for both hazard and exposure information separately (see Figure 2-6a) and examining the coherence across the bodies of evidence by making inference across evidence streams. These lead to the evidence integration summary (see Figure 2-6b). While a specific process was not described, the presentation did provide some considerations that went into the evaluation and inference-making and pointed to some examples where OPPT had conducted evidence integration for the committee to further consider.

While the WOE/evidence integration details are scant for most areas of the 1-BP evaluation, the TCE evaluation document did provide some methodological details on the WOE evaluation for a single hazard endpoint—congenital heart defects. This evaluation presents three levels of WOE determinations "made in succession, first for subsets of a line of evidence, then for the full lines of evidence, and then for the overall database, each building on the assessments that came before" (EPA 2020a, p. 612). The process detailed in Appendix G-2 of the TCE evaluation is an adapted version of the EPA Risk Assessment Forum 2016 Weight of Evidence in Ecological Assessment document (EPA 2016) that utilizes symbols (ranging from ---, --, -, 0, +, ++, +++ depending on the specific area [i.e., reliability, outcome/strength, and relevance] for the purposes of scoring at each level of the assessment [i.e., individual studies, across studies within one stream/line of evidence, and across lines of evidence]). Using these metrics and steps, the TCE evaluation concluded that the epidemiological studies provided "suggestive evidence" (+). Various animal studies were judged to be "weakly positive" (0/+) and "positive" (for TCE metabolites) (+), and the inhalation studies were judged to be "negative evidence" (-). The mechanistic studies were described as providing "strong and consistent supporting information" (++). When combined to a summary score of (+), these scores are said to represent "positive overall evidence that TCE may produce cardiac defects in humans" (EPA 2020a, p. 621). A mathematical average was used to integrate evidence areas for all studies that appears to be a translation of the symbols into numbers with the examples equating ++ to a value of 2 and 0/- scores to a value of -0.5; however, how this mathematical average is used is unclear.

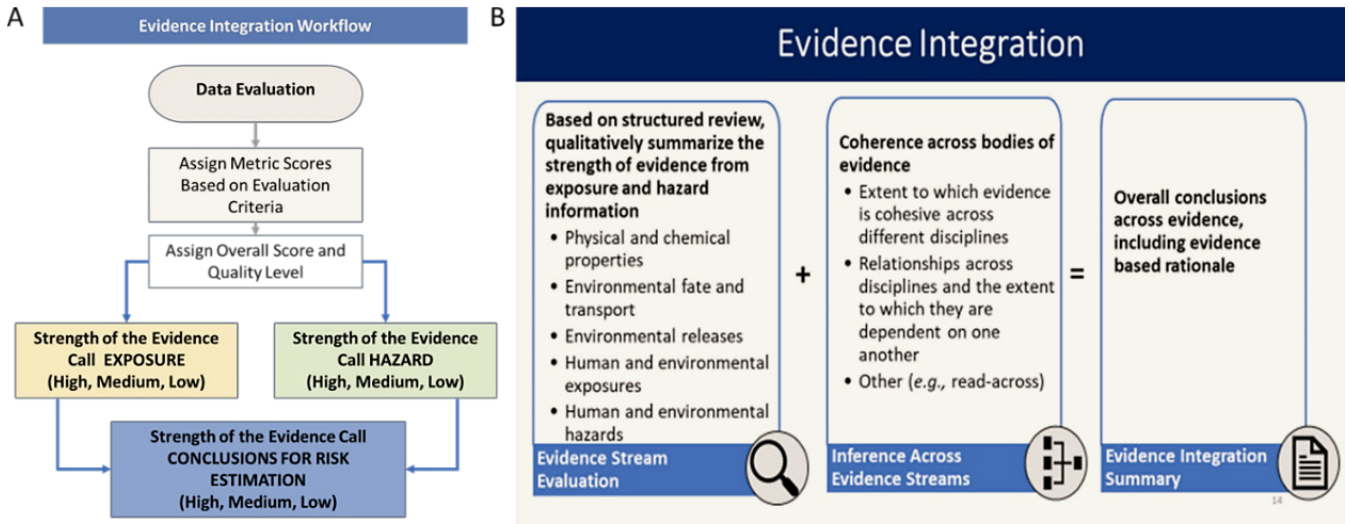

FIGURE 2-6 (A) Evidence integration workflow used in TSCA risk evaluations. SOURCE: Stanley Barone, presentation to the committee, August 24, 2020. (B) Considerations for evidence integration within TSCA risk evaluations. SOURCE: U.S. Environmental Protection Agency, presentation to the committee, July 23, 2020.

With regard to exposures, TSCA risk evaluations include calculations of occupational exposures, exposures from the use of consumer products, ecological exposures, and exposures to the general population as part of the systematic review process using either measured data or through models (with models requiring chemical properties, release data both to the environment and within the context of specific occupational activities or consumer uses, and exposure factors). Calculating these exposures is clearly a required step in making a risk determination yet is not part of a traditional systematic review. One could argue that in a systematic review, the next step is to integrate data from multiple sources and determine the most likely exposure value. These resulting exposures are then combined with the hazard numbers to make a risk determination (see purple box in Figure 2-1). For an exposure assessment, this activity would most logically combine measured distributions to develop an overall distribution of exposure and compare that to estimates from multiple available models, considering sources of uncertainty and variability. However, the two data streams were not integrated in the assessments reviewed by the committee. Additionally, there was no integration of occupational exposure through inhalation and dermal exposure. All exposures were separately evaluated in the risk assessment.

Critique of the TSCA Approach

The committee has not found the 2018 TSCA systematic review document, the several presentations made to the committee by OPPT, or the 1-BP and TCE evaluations sufficiently detailed to provide the information needed to assess the methodology and appropriateness of the framework for evidence integration. Therefore, when considering the core review elements of the Statement of Task, "comprehensive, workable, objective, and transparent," the committee finds that the evidence integration step is lacking in all respects. Where possible, the committee elaborates on processes that specifically impact each of the core review elements.

Comprehensive

Given that the WOE determination or evidence evaluation for heart defects in the TCE evaluation was the only detailed example provided, it is uncertain as to whether this process represents a limited use for a specific endpoint for TCE or represents a method that will be used (in its current form) for future risk assessments. The committee finds that OPPT has conflated aspects of three important systematic review elements: evaluating individual studies, a body of evidence (i.e., strength or certainty of evidence for a conclusion), and a level of confidence in a recommendation or determination of causation. Furthermore, the condensation of synthesis, evidence integration, and integration of hazard and exposure into one step makes it difficult to assess the extent to which the procedures used are comprehensive.

Workable

Without examples of how well the WOE process described in the TCE evaluation could be applied for other endpoints or evidence streams it is difficult to judge if it is sufficiently robust to be applied broadly and provide the desired outcome.

Objective

The lack of a true protocol, determined at the start of the evaluation, or a documented standard for evidence integration calls into question the objectivity of the process. There was also confusion relat-

ed to the apparent repetition of data quality evaluations at this stage of the process when evaluation of individual study quality and relevance should have been dealt with during the data evaluation stage. It is unclear if this sequencing impacts the objectivity of the process.

Transparent

The committee has the most concern with regard to the transparency of the evidence integration steps, much of which stems from confusion related to the terminology used in the various documents (and sometimes even the variations in use within one document), the lack of information presented to describe the process, and the lack of documentation to explain deviations from the little process that was documented. This confusion was also manifest within the public comments to TSCA assessments as the approach to integration of the evidence was not included in a protocol and also was not described in the 2018 guidance document.

With regard to the terminology, terms such as "strength of evidence," "sufficient and insufficient evidence," "confidence" in the evidence, and a number of other descriptors such as "summary scores" and "positive overall evidence" serve only to create confusion because some of these terms have not been well defined or are used differently throughout different assessment documents, within the presentations, or within the 2018 guidance document.

When comparing the work done in the 1-BP and the TCE evaluations, very little consistency can be observed between the evidence integration steps other than the lack of transparency as to what methods were used and how decisions were made. Very little information is provided about the process used to integrate the available data, and although qualitative statements are provided, the data do not appear to actually be integrated. The term "data integration" is mentioned in several of the other evidence streams as well, including exposure and environmental fate, but it is not clear as to how the data were selected or integrated into the assessment. Significant improvement in transparency and consistency is needed to fully understand EPA's process of evidence integration within this assessment.

The lack of documented process and lack of documentation to explain deviations greatly impacts the transparency of the evidence integration. The committee notes here a number of examples in which a lack of explanation was problematic:

- The evaluation of heart defects in the TCE evaluation followed a new, "fit-for-purpose" modification of the EPA Risk Assessment Forum 2016 Weight of Evidence in Ecological Assessment document (EPA 2016). It is unclear from the documentation provided in the risk assessment docket, and from the presentations by the agency, why this deviation existed and why it was needed.
- The use of two different methods within the evidence integration step (i.e., true mathematical average versus semi-qualitative grouping) is troubling, particularly due to the lack of rationale provided for this divergence. This runs completely counter to the guidance provided by the risk assessment forum in the Weight of Evidence in Ecological Assessment document which states,

  > Symbols are preferable to numerical scores because their use implies that they cannot be numerically combined. Two strongly supporting laboratory tests (++ and ++) are not equal to four somewhat supporting field tests (+, +, +, +). For a test result, a – score for study design and a + score for replication of the test do not sum to 0, because they are not commensurable. They simply signify different results for the different qualitative properties. Adding numerical scores generates a number with no units that signifies no quantity in particular. (EPA 2016, p. 29)

- In reviewing the supplemental data table for congenital heart defects in the TCE evaluation, the method of data integration across subsets did not seem consistent with the true average method discussed in the risk assessment document. There were some instances (e.g., toxicological [in vivo animal]—inhalation and toxicological [in vivo animal]—other) in which the overall subset value was greater than the sum of the contributing factors. While there is no issue with upgrading overall quality, the reasons for these deviations from the stated protocol were unclear.

Recommendations

In order to improve these issues with TSCA's process for evidence integration, the committee recommends the following:

- Although the integration of different evidence streams is not part of the systematic review process, it is desirable that this process is conducted in a structured, transparent, and prespecified way.
- Develop guidance for evidence integration of multiple streams. Currently, the guidance within the agency for integrating evidence from different streams is lacking and needs to be developed. The guidance should include frameworks for how to best consider different lines of evidence in many types of information that informs the evaluation, such as the PBPK models that integrate different streams of hazard evidence (i.e., human, animal, and mechanistic) and exposure models that integrate across sources and routes of exposure. The guidance should also include processes for documenting decisions and engaging stakeholders.
- As has been noted in the committee's reviews of the completed TCE and 1-BP risk evaluations, there is significant lack of clarity in the language used to describe the integration process, for both hazard and exposure, and the lack of clarity is not limited to the evidence descriptors. It is difficult to understand why and how certain steps were taken, particularly regarding the inclusions and exclusions of certain data.
- Integration methods should be documented a priori and applied consistently for each assessment unless a strong argument for a deviation exists, and then a detailed justification should be included at each step.

3

Crosscutting Issues with the TSCA Approach

INTRODUCTION

Chapter 2 examined the systematic review approach taken by the U.S. Environmental Protection Agency (EPA) in its Toxic Substances Control Act (TSCA) risk evaluations, based on *Application of Systematic Review in TSCA Risk Evaluations* (herein the 2018 guidance document), and further elaborations of its processes as documented in the specific evaluations examined by the committee—the Draft Risk Evaluation for Trichloroethylene (TCE) and the Final Risk Evaluation for 1-Bromopropane (n-Propyl Bromide) (EPA 2020a,c). This chapter builds on Chapter 2, addressing crosscutting and more general issues related to the use of systematic review in TSCA risk evaluations. The committee first offers overall findings and then turns to specific topics.

Additionally, in framing the committee's overall evaluation in this chapter, the report notes that EPA and specifically its Office of Pollution Prevention and Toxics (OPPT) have faced the herculean challenges of developing their risk assessment processes while meeting the schedule for evaluations specified in the 2016 amended TSCA. As noted in Chapter 1, the agency also needs to operate with the definition of weight of evidence (WOE) provided in the Procedures for Chemical Risk Evaluation Under the Amended Toxic Substances Control Act, referred to as the "Risk Evaluation Rule" (40 CFR Part 702, 82 FR 33726). The definition in the Risk Evaluation Rule deviates from the more conventional meaning of the term WOE and application of approaches for WOE. Suggestions are made with regard to better aligning terminology in TSCA assessments with usual practice.

OVERALL FINDINGS

Restating the Statement of Task, this committee was asked to "review EPA's guidance document on *Application of Systematic Review in TSCA [Toxic Substances Control Act] Risk Evaluations* and associated materials to determine whether the process is comprehensive, workable, objective, and transparent." In summary, the committee finds that the process outlined in the 2018 guidance document, and as elaborated and applied in the example evaluations, does not meet the criteria of "comprehensive, workable, objective, and transparent." The committee's evaluation was made difficult by the incomplete and hard-to-follow documentation of many details of the process—adequacy of documentation is requisite for achieving transparency, objectivity, and replicability.

The committee found that the systematic reviews within the draft risk evaluations considered did not meet the standards of systematic review methodology. The committee applied the critical appraisal tool for systematic review "assessment of multiple systematic reviews" (AMSTAR-2) to the hazard assessment in the draft TCE risk evaluation and found the appraisal process to be unnecessarily complicated due to insufficient and unclear documentation. Despite this barrier to applying the AMSTAR-2 instrument, the committee found that the TCE hazard assessment did not perform positively on the vast majority of AMSTAR-2 questions. Hence, the committee concluded that the hazard assessment within the TSCA TCE risk evaluation was of critically low quality, meaning that the review had "more than one critical flaw and should not be relied on to provide an accurate and comprehensive summary of the available studies" (Shea et al. 2017, p. 6). Consequently, the committee suggests that the OPPT team comprehensively reevaluate its approach so as to achieve the state of the practice for systematic review.

In the committee's judgment, the specific and general problems in TSCA risk evaluations are partially due to the decision to develop a largely de novo approach, rather than starting with the foundation offered by approaches that were extant in 2016. OPPT was challenged by the statutory schedule for completing assessments. Nonetheless, looking forward, the committee strongly recommends that OPPT reconsider its overall strategy. Further guidance for moving forward is offered below.

THE USE OF SYSTEMATIC REVIEW FOR THE RISK EVALUATIONS

From the point of view of the broad range of stakeholders involved with TSCA risk evaluations, the evaluations need to be developed with methods that are rigorous, reproducible, valid, and transparent. Systematic review is a method that meets these requirements, if carried out correctly. As discussed in Chapter 1, the statute itself, however, did not specifically require systematic review but discussed that systematic review would be used.

Systematic review methods are already established for the evidence streams (e.g., human, animal or ecological receptors, or mechanistic) contributing to hazard assessment for human health and ecological receptors. However, comparable systematic review approaches were not available for the other evidence streams included in TSCA assessments (see Figure 1-1). As a result, OPPT staff embarked on extending methods of systematic review to chemical properties, fate and transport, and exposures of the population generally and in occupational settings—components that are broadly related to the human and ecological exposure assessments. Consequently, the resulting processes for these components of the evaluations are still evolving according to OPPT and, up to now, have required substantial effort for their development and implementation on the part of the OPPT TSCA team. The approaches for these evidence streams are not yet fixed and lack rigorous evaluation by OPPT staff or external committees. This committee understands the enormity of the task of carrying out TSCA evaluations but notes that such innovations, in order to be broadly accepted, need rigorous evaluation and testing by multiple stakeholders.

Under the amended TSCA, EPA is required to evaluate the "weight of the scientific evidence" with the definition beginning with "means a systematic review method" (40 CFR Part 702, 82 FR 33726). The committee notes that "systematic review method" is left undefined. Without a clear definition of "a systematic review method," the committee inferred from the presentations by OPPT staff and the 2018 guidance document that the TSCA program interpreted the rule to mean "systematic review" as conventionally defined. With this interpretation of the rule in mind and considering the full set of requirements for what constitutes systematic review, the committee has found that OPPT has not been performing "systematic review" for the various evidence streams in a way that meets current standards.

The committee finds that any evidence-based methods applied to each evidence stream should meet the required characteristics of the process for "weight of the scientific evidence" because per the amended TSCA this process "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." In the committee's reading, the requirement is for a protocol for addressing each stream of evidence. Systematic review, as defined by this committee and others, is not specifically described.

While the state of the practice for gathering and reviewing evidence is not as thoroughly specified for the evidence streams other than those contributing to hazard assessment, the committee did not find that approaches were being used that are consistent with best practices for exposure assessment and for fate and transport. Experience to date shows that developing de novo a systematic review approach that can be applied across all evidence streams has proved challenging, and perhaps impracticable, and strains resources. Review strategies are needed that meet the criteria of being "comprehensive, workable, objective, and transparent." While each study in all evidence streams may need to be identified,

examined, and synthesized for the elements of hazard assessment, such review methodology does not necessarily directly extend to running a fate and transport model, for example.

Additionally, under some circumstances there may be reasonable alternatives to carrying out a de novo systematic review; for example, the relevant literature may be non-existent or too limited in scope or there may be a recent systematic review that meets quality standards. In some cases, it may be possible to use an alternative approach to systematic review as long as it meets the transparency, consistency, reproducibility, and comprehensiveness requirements of evidence-based methodologies. When utilizing an alternative evidence-based methodology in lieu of systematic review, however, the rationale for the deviation from the systematic review should be explicitly stated in the risk evaluation scoping document.

DEVELOPMENT OF NEW METHODS OR UTILIZATION OF EXISTING METHODS

Beginning with its initial assessments, OPPT developed review approaches that were particular to its mandate under TSCA. As OPPT began to develop its approach following the 2016 amended TSCA, methods were extant for carrying out comprehensive, transparent assessments of the hazards posed by environmental agents (e.g., those elaborated by the Office of Health Assessment and Translation [OHAT], EPA's Integrated Risk Information System [IRIS] Program, and the Navigation Guide). Additionally, instruments were available for assessing risk of bias and certainty of evidence, and versions thereof were being tailored for environmental agents. Nonetheless, OPPT did not adopt these general methodologies nor use available tools that are directly applicable to the TSCA charge and instead embarked on the arduous task of developing new methods and instruments.

The committee suggests that OPPT step back from the approach that it has taken and consider what components of the OHAT, IRIS, or Navigation Guide methods could be incorporated directly and specifically into hazard assessment (NASEM 2018; OHAT 2019; Woodruff and Sutton 2014). These methods have a trajectory of use and community acceptance and reflect the state of the practice. The committee also suggests that OPPT evaluate the ways that these existing methods could be modified for other evidence streams. In addition, OPPT should use existing guidance within the agency such as the *Guidelines for Human Exposure Assessment* and the *Guidelines for Ecological Risk Assessment*.

Similarly, as described in Chapter 2, there are already established tools for assessing risk of bias and certainty that have been used extensively and that could be used in TSCA risk evaluations. The committee recommends that OPPT give full consideration to existing approaches related to all evidence streams before continuing on the track of developing new instruments that may not be needed.

WEIGHT OF EVIDENCE

Throughout this report, there are comments on the consequences of the definition of WOE in the Risk Evaluation Rule and the conflation of "weight of the scientific evidence" with "a systematic review method" as necessitated by the Risk Evaluation Rule. This deviation from convention comes with consequences for describing the application of WOE principles in evidence integration.

OPPT is cognizant of the discrepancy between the definition in the Risk Evaluation Rule and usual practice; for example, during a presentation on February 28, 2020, OPPT staff described one step with regard to risk evaluation/risk characterization: "Describe the weight of the scientific evidence for the identified hazard and exposure" (Susanna Blair, presentation to the committee, February 28, 2020).

The committee understands that the definition of WOE within the Risk Evaluation Rule is not easily changed but suggests that OPPT adopt a specific term to describe the WOE evaluation during the evidence integration step, other than the term "weight of evidence," to avoid the semantic clash between the definition in the Risk Evaluation Rule and its application during the evaluation process. Alternatives might include "strength of evidence" or "certainty of evidence" as utilized in the Grading of Recommendations Assessment, Development and Evaluation process (Whaley et al. 2020).

Regardless of terminology, a narrative description should be provided that describes the basis for the determination of the strength of evidence during the evidence integration step for all applicable data streams. The committee proposes the use of standard descriptors for the strength of evidence as with the Integrated Science Assessments for the National Ambient Air Quality Standards (EPA 2019c).

ENHANCING CLARITY OF DOCUMENTATION OF THE ASSESSMENT METHODS

The committee carefully examined the 2018 guidance document, considered several TSCA evaluations, and was briefed by OPPT staff on multiple occasions, including with presentations and poster sessions. Nonetheless, the committee's work to complete the Statement of Task was limited by a lack of complete, coherent, and readily accessible material on how the assessments were conducted, particularly as the approach evolved across the assessments from the first 10 to the 20 now in progress. The committee anticipates that the many stakeholders who use these documents are similarly struggling. In fact, with this lack of clarity, the transparency requirement of the WOE definition in the rule is not adequately met, complicating a determination as to whether the requirements for the WOE assessment (i.e., a process that is comprehensive, objective, transparent, and consistent) are achieved. Moreover, one of the advantages of systematic review is that it is reproducible and updateable. The committee finds that TSCA evaluations would be extremely difficult if not impossible to reproduce or update with new information. Standardization of documentation of every step in the process is a critical step toward this goal.

Consequently, the committee suggests the development of enhanced documentation that would facilitate both development and use of TSCA evaluations. Systematic review methodology itself provides structure and recommendations on planning, conducting, and reporting (Moher et al. 2009).

The committee recommends that a handbook of TSCA systematic review and evidence integration methodology be put together that details the steps in the process. Throughout this report, the committee points to problems of documentation. For example, there needs to be a record of decisions made for each publication identified during the systematic review or other evidence-gathering processes. Such a handbook would likely need internal and external review and require substantial time for its development. However, in the committee's view and drawing on its experience in writing this report, the 2018 guidance document is not adequate as a stand-alone document to describe how systematic reviews are carried out in TSCA risk evaluations. The committee believes that the effort of developing and publicly vetting a handbook will pay off in the long run by making the process more straightforward and transparent as well as easier to follow.

Recognizing that preparation of a comprehensive new document describing procedures will need substantial time for planning and development, OPPT staff should assemble an "evergreen" compilation of how reviews are carried out, ideally documenting evolution of practice from the 2018 document to the present. The inclusion of a comprehensive glossary of key terms, including all the terms used throughout the review process, will be important. Such a document should capture changes made to the review process subsequent to the 2018 guidance document and describe what changes have been made, the rationale for the changes, and the risk assessments to which they apply.

SUMMARY

The committee was in strong consensus that the processes used by OPPT do not meet the evaluation criteria specified in the Statement of Task (i.e., comprehensive, workable, objective, and transparent). OPPT faced substantial challenges in integrating review methods on the schedule required by the Lautenberg Act. Those challenges have not yet been successfully met. Chapter 2 includes a number of specific recommendations as to how to improve the methods for assessments, both in general and with reference to particular elements of the evaluation process.

Chapter 3 covers broad issues in relation to the Statement of Task. The general recommendations from this chapter are summarized as follows:

- The OPPT approach to systematic review does not adequately meet the state of the practice. The committee suggests that OPPT comprehensively reevaluate its approach to systematic review methods, addressing the comments and recommendations of Chapter 2.
- With regard to hazard assessment for human and ecological receptors, the committee comments that OPPT should step back from the approach that it has taken and consider components of the OHAT, IRIS, and Navigation Guide methods that could be incorporated directly and specifically into hazard assessment.
- The committee finds that OPPT's use of systematic review for the evidence streams for which it has not been previously adapted to be particularly unsuccessful. Given these novel applications of systematic review, the committee suggests that OPPT elaborate plans for continuing the refinement of methods, ideally, in collaboration with internal and external stakeholders. The committee also suggests that OPPT evaluate the ways that existing OHAT, IRIS, and Navigation Guide methods could be modified for the other evidence streams. In addition, OPPT should use existing guidance within the agency such as the Guidelines for Human Exposure Assessment, the Guidelines for Ecological Risk Assessment, and the operating procedures for the use of the ECOTOXicology knowledgebase, as following existing guidelines would improve transparency of the assessments.
- The committee recommends that a handbook for TSCA review and evidence integration methodology be put together that details the steps in the process. Throughout this report, the committee points to problems of documentation. The committee believes that the effort of developing and publicly vetting a handbook will pay off in the long run by making the process more straightforward, transparent, and easier to follow.

There is an ongoing cross-sector effort on developing and validating new tools and approaches for exposure, ecotoxicology, environmental health, and other new areas of application of systematic review. The committee strongly recommends that OPPT staff engage in these efforts. The approaches used for TSCA evaluation would benefit from the substantial external expertise available as well as additional transparency and acceptance by the different stakeholders and society in general as these tools are developed. The refinements recommended by this committee would help boost the ability of actions taken under the Frank R. Lautenberg Chemical Safety for the 21st Century Act to advance the mission of EPA: "to protect human health and the environment."

References

Balshem, H., M. Helfand, H. J. Schünemann, A. D. Oxman, R. Kunz, J. Brozek, G. E. Vist, Y. Falck-Ytter, J. Meerpohl, and S. Norris. 2011. GRADE guidelines: 3. Rating the quality of evidence. *Journal of Clinical Epidemiology* 64(4):401–406.

Bergeson, L. L. 2016. TSCA reform: Is it still in our future? *Industrial Biotechnology* 12(2):75–77.

Bero, L., N. Chartres, J. Diong, A. Fabbri, D. Ghersi, J. Lam, A. Lau, S. McDonald, B. Mintzes, P. Sutton, J. L. Turton, and T. J. Woodruff. 2018. The risk of bias in observational studies of exposures (ROBINS-E) tool: Concerns arising from application to observational studies of exposures. *Systematic Reviews* 7(1):242.

Booth, A., M. Clarke, G. Dooley, D. Ghersi, D. Moher, M. Petticrew, and L. Stewart. 2012. The nuts and bolts of PROSPERO: An international prospective register of systematic reviews. *Systematic Reviews* 1(1):1–9.

Card, M. L., V. Gomez-Alvarez, W. H. Lee, D. G. Lynch, N. S. Orentas, M. T. Lee, E. M. Wong, and R. S. Boethling. 2017. History of EPI Suite™ and future perspectives on chemical property estimation in US Toxic Substances Control Act new chemical risk assessments. *Environmental Science: Processes & Impacts* 19(3):203–212.

Cohen Hubal, E. A., J. J. Frank, R. Nachman, M. Angrish, N. C. Deziel, M. Fry, R. Tornero-Velez, A. Kraft, and E. Lavoie. 2020. Advancing systematic-review methodology in exposure science for environmental health decision making. *Journal of Exposure Science and Environmental Epidemiology* 30:906–916.

EDF (Environmental Defense Fund). 2020. *Comments on 1,4-Dioxane; Supplemental Analysis to the Draft Toxic Substances Control Act (TSCA) Risk Evaluation*. http://blogs.edf.org/health/files/2020/12/EDF-Comments-on-14-dioxane-DRE-Supplement-FINAL-12-10-20.pdf, accessed November 13, 2020.

Eick, S. M., D. E. Goin, N. Chartres, J. Lam, and T. J. Woodruff. 2020. Assessing risk of bias in human environmental epidemiology studies using three tools: Different conclusions from different tools. *Systematic Reviews* 9(1):249.

EPA (U.S. Environmental Protection Agency). 1998. Guidelines for Ecological Risk Assessment. EPA/630/R-95/002F. *Federal Register* 63(93):26846–26924.

EPA. 2011. *Exposure Factors Handbook 2011 Edition (Final Report)*. EPA/600/R-09/052F. U.S. Environmental Protection Agency, Washington, DC.

EPA. 2012a. *Benchmark Dose Technical Guidance*. U.S. Environmental Protection Agency, Washington, DC.

EPA. 2012b. *Benchmark Dose Software (BMDS) Development and Maintenance Multitumor (MS Combo) Model*. U.S. Environmental Protection Agency, Washington, DC.

EPA. 2016. *Weight of Evidence in Ecological Assessment*. U.S. Environmental Protection Agency, Washington, DC.

EPA. 2017a. *Scope of the Risk Evaluation for Trichloroethylene*. Office of Chemical Safety and Pollution Prevention, Washington, DC.

EPA. 2017b. *Scope of the Risk Evaluation for 1-Bromopropane*. Office of Chemical Safety and Pollution Prevention, Washington, DC.

EPA. 2017c. *Strategy for Conducting Literature Searches for Trichloroethylene (TCE): Supplemental Document to the TSCA Scope Document CASRN: 79-01-6*. Office of Chemical Safety and Pollution Prevention, Washington, DC.

EPA. 2018a. *Application of Systematic Review in TSCA Risk Evaluations*. Office of Chemical Safety and Pollution Prevention, Washington, DC.

EPA. 2018b. *Problem Formulation of the Risk Evaluation for Trichloroethyene*. Office of Chemical Safety and Pollution Prevention and Office of Pollution Prevention and Toxics, Washington, DC.

EPA. 2019a. *Final Risk Evaluation for 1,4-Dioxane*. U.S. Environmental Protection Agency, Washington, DC.

EPA. 2019b. *Guidelines for Human Exposure Assessment*. Risk Assessment Forum, U.S. Environmental Protection Agency, Washington, DC.

EPA. 2019c. *Integrated Science Assessment (ISA) for Particulate Matter*. U.S. Environmental Protection Agency, Washington, DC.

EPA. 2020a. *Draft Risk Evaluation for Trichloroethylene*. Office of Chemical Safety and Pollution Prevention, Washington, DC.

EPA. 2020b. *ECOTOX User Guide: ECOTOXicology Database System*, Version 5.3. U.S. Environmental Protection Agency, Washington, DC.

EPA. 2020c. *Risk Evaluation for 1-Bromopropane (n-Propyl Bromide)*. Office of Chemical Safety and Pollution Prevention, Washington, DC.

Etterson, M. 2020. *Technical Manual: SSD Toolbox Version 1.0.* EPA/600/R-19/104. Great Lakes Toxicology and Ecology Division, Center for Computational Toxicology and Exposure, Office of Research and Development, U.S. Environmental Protection Agency, Duluth, MN.

Foster, E. D., and A. Deardorff. 2017. Open science framework (OSF). *Journal of the Medical Library Association* 105(2):203–206.

Gartlehner, G., G. Wagner, L. Lux, L. Affengruber, A. Dobrescu, A. Kaminski-Hartenthaler, and M. Viswanathan. 2019. Assessing the accuracy of machine-assisted abstract screening with DistillerAI: A user study. *Systematic Reviews* 8(1):277.

Greenland, S., and K. O'Rouke. 2001. On the bias produced by quality scores in meta-analysis, and a hierarchical view of proposed solutions. *Biostatistics* 2(4):463–471.

Guyatt, G. H., A. D. Oxman, R. Kunz, J. Woodcock, J. Brozek, M. Helfand, P. Alonso-Coello, P. Glasziou, R. Jaeschke, and E. A. Akl. 2011. GRADE guidelines: 7. Rating the quality of evidence—inconsistency. *Journal of Clinical Epidemiology* 64(12):1294–1302.

Herbison, P., J. Hay-Smith, and W. J. Gillespie. 2006. Adjustment of meta-analyses on the basis of quality scores should be abandoned. *Journal of Clinical Epidemiology* 59(12):1249–1256.

Higgins, J. P., D. G. Altman, P. C. Gøtzsche, P. Jüni, D. Moher, A. D. Oxman, J. Savovic, K. F. Schulz, L. Weeks, and J. A. Sterne. 2011. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ* 343:d5928.

Higgins, J. P., J. Thomas, J. Chandler, M. Cumpston, T. Li, M. J. Page, and V. A. Welch. 2019. *Cochrane Handbook for Systematic Reviews of Interventions*. Chichester, UK: John Wiley & Sons.

Hill, A. 1965. The environment and disease: Association or causation? *Proceedings of the Royal Society of Medicine* 58(5):295–300.

Hooijmans, C. R., R. B. M. de Vries, M. Ritskes-Hoitinga, M. M. Rovers, M. M. Leeflang, J. IntHout, K. E. Wever, L. Hooft, H. de Beer, T. Kuijpers, M. R. Macleod, E. S. Sena, G. Ter Riet, R. L. Morgan, K. A. Thayer, A. A. Rooney, G. H. Guyatt, H. J. Schünemann, and M. W. Langendam. 2018. Facilitating healthcare decisions by assessing the certainty in the evidence from preclinical animal studies. *PLoS One* 13(1):e0187271.

Howard, B. E., J. Phillips, K. Miller, A. Tandon, D. Mav, M. R. Shah, S. Holmgren, K. E. Pelch, V. Walker, A. A. Rooney, M. Macleod, R. R. Shah, and K. Thayer. 2016. SWIFT-Review: A text-mining workbench for systematic review. *Systematic Reviews* 5:87.

Howard, B. E., J. Phillips, A. Tandon, A. Maharana, R. Elmore, D. Mav, A. Sedykh, K. Thayer, B. A. Merrick, and V. Walker. 2020. SWIFT-Active Screener: Accelerated document screening through active learning and integrated recall estimation. *Environment International* 138:105623.

IOM (Institute of Medicine). 2011. *Finding What Works in Health Care: Standards for Systematic Reviews*. Washington, DC: The National Academies Press.

Jüni, P., A. Witschi, R. Bloch, and M. Egger. 1999. The hazards of scoring the quality of clinical trials for meta-analysis. *Journal of the American Medical Association* 282(11):1054–1060.

Kase, R., M. Korkaric, I. Werner, and M. Ågerstrand. 2016. Criteria for Reporting and Evaluating ecotoxicity Data (CRED): Comparison and perception of the Klimisch and CRED methods for evaluating reliability

and relevance of ecotoxicity studies. *Environmental Sciences Europe* 28(1):1–14.

Klimisch, H.-J., M. Andreae, and U. Tillmann. 1997. A systematic approach for evaluating the quality of experimental toxicological and ecotoxicological data. *Regulatory Toxicology and Pharmacology* 25(1):1–5.

Koustas, E., J. Lam, P. Sutton, P. I. Johnson, D. S. Atchley, S. Sen, K. A. Robinson, D. A. Axelrad, and T. J. Woodruff. 2014. The Navigation Guide—evidence-based medicine meets environmental health: Systematic review of nonhuman evidence for PFOA effects on fetal growth. *Environmental Health Perspectives* 122(10):1015–1027.

Krnic Martinic, M., D. Pieper, A. Glatt, and L. Puljak. 2019. Definition of a systematic review used in overviews of systematic reviews, meta-epidemiological studies and textbooks. *BMC Medical Research Methodology* 19(1):203.

Levac, D., H. Colquhoun, and K. K. O'Brien. 2010. Scoping studies: Advancing the methodology. I*mplementation Science* 5(1):69.

Mayo-Bean, K., K. Moran-Bruce, J. V. Nabholz, W. M. Meylan, P. H. Howard, and L. Cassidy. 2017. *Operation Manual for the Ecological Structure-Activity Relationship Model (ECOSAR) Class Program*. Risk Assessment Division, U.S. Environmental Protection Agency, Washington, DC.

Moermond, C. T., R. Kase, M. Korkaric, and M. Ågerstrand. 2016. CRED: Criteria for Reporting and Evaluating ecotoxicity Data. *Environmental Toxicology and Chemistry* 35(5):1297–1309.

Moher, D., A. Liberati, J. Tetzlaff, and D. G. Altman. 2009. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine* 6(7):e1000097.

Morgan, R. L., K. A. Thayer, L. Bero, N. Bruce, Y. Falck-Ytter, D. Ghersi, G. Guyatt, C. Hooijmans, M. Langendam, and D. Mandrioli. 2016. GRADE: Assessing the quality of evidence in environmental and occupational health. *Environment International* 92:611–616.

Morgan, R. L., P. Whaley, K. A. Thayer, and H. J. Schünemann. 2018. Identifying the PECO: A framework for formulating good questions to explore the association of environmental and other exposures with health outcomes. *Environment International* 121(Pt 1):1027–1031.

NASEM (National Academies of Sciences, Engineering, and Medicine). 2017a. *Using 21st Century Science to Improve Risk-Related Evaluations*. Washington, DC: The National Academies Press.

NASEM. 2017b. *Application of Systematic Review Methods in an Overall Strategy for Evaluating Low-Dose Toxicity from Endocrine Active Chemicals*. Washington, DC: The National Academies Press.

NASEM. 2018. *Progress Toward Transforming the Integrated Risk Information System (IRIS) Program: A 2018 Evaluation*. Washington, DC: The National Academies Press.

NASEM. 2019. *Review of DOD's Approach to Deriving an Occupational Exposure Level for Trichloroethylene*. Washington, DC: The National Academies Press.

NRC (National Research Council). 2009. *Science and Decisions: Advancing Risk Assessment*. Washington, DC: The National Academies Press.

NRC. 2010. *Review of the Environmental Protection Agency's Draft IRIS Assessment of Tetrachloroethylene*. Washington, DC: The National Academies Press.

NRC. 2011. *Review of the Environmental Protection Agency's Draft IRIS Assessment of Formaldehyde*. Washington, DC: The National Academies Press.

NRC. 2014. *Review of EPA's Integrated Risk Information System (IRIS) Process*. Washington, DC: The National Academies Press.

OHAT (Office of Health Assessment and Translation). 2019. *Handbook for Conducting a Literature-Based Health Assessment Using OHAT Approach for Systematic Review and Evidence Integration*. https://ntp.niehs.nih.gov/ntp/ohat/pubs/handbookmarch2019_508.pdf, accessed November 13, 2020.

Pega, F., S. L. Norris, C. Backes, L. A. Bero, A. Descatha, D. Gagliardi, L. Godderis, T. Loney, A. Modenese, R. L. Morgan, D. Pachito, M. B. S. Paulo, P. T. J. Scheepers, V. Schlünssen, D. Sgargi, E. K. Silbergeld, K. Sørensen, P. Sutton, T. Tenkate, D. Torreão Corrêa da Silva, Y. Ujita, E. van Deventer, T. J. Woodruff, and D. Mandrioli. 2020. RoB-SPEO: A tool for assessing risk of bias in studies estimating the prevalence of

exposure to occupational risk factors from the WHO/ILO Joint Estimates of the Work-related Burden of Disease and Injury. *Environment International* 135:105039. doi: 10.1016/j.envint.2019.105039.

Peters, M. D., C. M. Godfrey, H. Khalil, P. McInerney, D. Parker, and C. B. Soares. 2015. Guidance for conducting systematic scoping reviews. *International Journal of Evidence-based Healthcare* 13(3):141–146.

Posthuma, L., G. W. Suter II, and T. P. Traas. 2001. *Species Sensitivity Distributions in Ecotoxicology*. Boca Raton, FL. CRC Press. https://doi.org/10.1201/9781420032314.

Rooney, A. A., A. L. Boyles, M. S. Wolfe, J. R. Bucher, and K. A. Thayer. 2014. Systematic review and evidence integration for literature-based environmental health science assessments. *Environmental Health Perspectives* 122(7):711–718.

Schaefer, H. R., and J. L. Myers. 2017. Guidelines for performing systematic reviews in the development of toxicity factors. *Regulatory Toxicology and Pharmacology* 91:124–141.

Shea, B. J., B. C. Reeves, G. Wells, M. Thuku, C. Hamel, J. Moran, D. Moher, P. Tugwell, V. Welch, E. Kristjansson, and D. A. Henry. 2017. AMSTAR 2: A critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ* 358:j4008.

Singla, V. I., P. M. Sutton, and T. J. Woodruff. 2019. The Environmental Protection Agency Toxic Substances Control Act systematic review method may curtail science used to inform policies, with profound implications for public health. *American Journal of Public Health* 109(7):982–984.

Stephens, M. L., K. Betts, N. B. Beck, V. Cogliano, K. Dickersin, S. Fitzpatrick, J. Freeman, G. Gray, T. Hartung, J. McPartland, A. A. Rooney, R. W. Scherer, D. Verloo, and S. Hoffmann. 2016. The emergence of systematic review in toxicology. *Toxicological Sciences* 152(1):10–16.

Suter, G., J. Nichols, E. Lavoie, and S. Cormier. 2020. Systematic review and weight of evidence are integral to ecological and human health assessments: They need an integrated framework. *Integrated Environmental Assessment and Management* 16(5):718–728.

Van der Mierden, S., K. Tsaioun, A. Bleich, and C. H. Leenaars. 2019. Software tools for literature screening in systematic reviews in biomedical research. *Altex* 36(3):508–517.

Villeneuve, D. L., D. Crump, N. Garcia-Reyero, M. Hecker, T. H, Hutchinson, C. A. LaLone, B. Landesmann, T. Lettieri, S. Munn, M. Nepelska, and M. A. Ottinger. 2014a. Adverse outcome pathway (AOP) development I: Strategies and principles. *Toxicological Sciences* 142(2):312–320.

Villeneuve, D. L., D. Crump, N. Garcia-Reyero, M. Hecker, T. H. Hutchinson, C. A. LaLone, B. Landesmann, T. Lettieri, S. Munn, M. Nepelska, and M. A. Ottinger. 2014b. Adverse outcome pathway development II: Best practices. *Toxicological Sciences* 142(2):321–330.

Wang, Z., G. W. Walker, D. C. G. Muir, and K. Nagatani-Yoshida. 2020. Toward a global understanding of chemical pollution: A first comprehensive analysis of national and regional chemical inventories. *Environmental Science & Technology* 54(5):2575–2584.

Whaley, P., C. Halsall, M. Ågerstrand, E. Aiassa, D. Benford, G. Bilotta, D. Coggon, C. Collins, C. Dempsey, R. Duarte-Davidson, R. FitzGerald, M. Galay-Burgos, D. Gee, S. Hoffmann, J. Lam, T. Lasserson, L. Levy, S. Lipworth, S. M. Ross, O. Martin, C. Meads, M. Meyer-Baron, J. Miller, C. Pease, A. Rooney, A. Sapiets, G. Stewart, and D. Taylor. 2016. Implementing systematic review techniques in chemical risk assessment: Challenges, opportunities and recommendations. *Environment International* 92–93:556–564. doi: 10.1016/j.envint.2015.11.002.

Whaley, P., E. Aiassa, C. Beausoleil, A. Beronius, G. Bilotta, A. Boobis, R. de Vries, A. Hanberg, S. Hoffmann, N. Hunt, C. F. Kwiatkowski, J. Lam, S. Lipworth, O. Martin, N. Randall, L. Rhomberg, A. A. Rooney, H. J. Schunemann, D. Wikoff, T. Wolffe, and C. Halsall. 2020. Recommendations for the conduct of systematic reviews in toxicology and environmental health research (COSTER). *Environment International* 143:105926.

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

Appendix A

Biographical Information on the Members of the Committee to Review EPA's TSCA Systematic Review Guidance Document

Jonathan M. Samet (NAM) (*Chair*) is a pulmonary physician and epidemiologist. He is the Dean of the Colorado School of Public Health. Dr. Samet's research has focused on the health risks posed by inhaled pollutants. He has served on numerous committees concerned with public health: the U.S. Environmental Protection Agency's Clean Air Scientific Advisory Committee; committees of the National Academies, including chairing the Biological Effects of Ionizing Radiation (BEIR) VI Committee, the Committee on Research Priorities for Airborne Particulate Matter, the Committee to Review EPA's Draft IRIS Assessment of Formaldehyde, the Committee to Review the IRIS Process, and the Board on Environmental Studies and Toxicology, among others; and the National Cancer Advisory Board. He is a member of the National Academy of Medicine. Dr. Samet received his MD from the University of Rochester School of Medicine and Dentistry and a master's degree in epidemiology from the Harvard T.H. Chan School of Public Health.

Deborah H. Bennett is a professor in the Division of Environmental and Occupational Health at the University of California, Davis, School of Medicine. Her research focuses on the measurement and modeling of organic compounds in the indoor environment. She has served on various U.S. Environmental Protection Agency Science Advisory boards, panels, and advisory committees related to the Exposure Factors Handbook and Exposure Metrics for the National Children's Study. She has served as the estimation associate editor for the *Journal of Exposure Science and Environmental Epidemiology*. She has served as an elected councilor, treasurer, and chair of the Awards Committee for the International Society of Exposure Assessment. She has an MS and a PhD from the University of California, Berkeley.

Bryan W. Brooks is a distinguished professor of environmental science and biomedical studies at Baylor University. His scholarship incorporates laboratory and field studies in environmental toxicology and chemistry, environmental health, hazard and risk assessment, and water resources. He leads harmful algal blooms research for the Center for the Assessment and Prediction of the Interactions of Climate Change on Oceans and Human Health, a National Institute of Environmental Health Sciences Center based at the University of South Carolina. Dr. Brooks serves as the editor-in-chief of *Environmental Science and Technology Letters*. Dr. Brooks has an MS from the University of Mississippi and a PhD from the University of North Texas.

Jessica L. Myers is a toxicologist and risk assessor. She is currently working at the Texas Commission on Environmental Quality where she has drafted guidance on the development of systematic reviews for toxicity factors. She has a bachelor's degree and a PhD in cell and molecular biology from The University of Texas at Austin.

Kristi Pullen Fedinick is a senior scientist and the director of science and data in the Healthy People & Thriving Communities Program at the Natural Resources Defense Council (NRDC). She also serves as part-time faculty in the Department of Environmental and Occupational Health of the Milken Institute

School of Public Health at The George Washington University. Dr. Pullen Fedinick's research career includes experience in environmental health and policy; molecular, structural, and computational biology; biochemistry; and population health. Prior to joining NRDC, she worked as a scientist for a Chicago-based environmental nonprofit, where she focused on air and drinking water quality, science communications, and environmental justice. Her current work focuses on the use of high-throughput technologies, predictive toxicology, and computational approaches to chemical risk assessments. Additional work includes the geospatial and statistical analysis of chemicals in the environment, with a particular emphasis on drinking water and on the disproportionate impact of chemical exposures in vulnerable populations. She holds a bachelor's degree in biochemistry and molecular biology from the University of Maryland, Baltimore County, and a PhD in molecular and cell biology with a focus on structural biology and biochemistry from the University of California, Berkeley. She was a Robert Wood Johnson Foundation Health and Society Scholar at the Harvard T.H. Chan School of Public Health.

Karen A. Robinson is a professor of medicine at the Johns Hopkins University School of Medicine. She is also the director of the Johns Hopkins University Evidence-based Practice Center and is a member of the core faculty in the Center for Clinical Trials and Evidence Synthesis at the university's Bloomberg School of Public Health. Dr. Robinson's research focuses on evidence-based health care and evidence-based research. She conducts systematic reviews that are used to develop clinical practice guidelines and to inform other health decisions. She served on the National Academies' Committee on Endocrine-Related Low-Dose Toxicity, the Committee to Review Advances Made to the IRIS Process, the Committee to Review DOD's Approach to Deriving an Occupational Exposure Level for Trichloroethylene, and the Committee to Review EPA's IRIS Assessment Plan for Inorganic Arsenic. Dr. Robinson received an MSc in health sciences from the University of Waterloo, Ontario, and a PhD in epidemiology from the Johns Hopkins Bloomberg School of Public Health.

Joseph V. Rodricks is a founding principal of ENVIRON (now Ramboll) and an internationally recognized expert in toxicology and risk analysis. He has consulted for hundreds of manufacturers, new product developers, and government agencies in the evaluation of health risks associated with human exposure to chemical substances of all types. Dr. Rodricks came to consulting after a 15-year career as a scientist at the U.S. Food and Drug Administration (FDA). In his last 4 years at FDA, he served as the associate commissioner for health affairs. His experience extends from pharmaceuticals, medical devices, consumer products and foods, to occupational chemicals and environmental contaminants. He has served on the National Academies' Board on Environmental Studies and Toxicology and on more than 40 boards and committees of the National Academies, including the committees that produced the seminal works *Risk Assessment in the Federal Government: Managing the Process* (1983) and *Science and Decisions: Advancing Risk Assessment* (2009). Most recently he served on the National Academies' committee that issued *Guiding Principles for Developing Dietary Reference Intakes Based on Chronic Disease*. He has more than 150 scientific publications and has received 11 honorary awards from professional societies and other academic and non-academic institutions. He is author of the widely used text *Calculated Risks*, now in its second edition, published by Cambridge University Press, and has presented more than 300 lectures in countries around the world. Dr. Rodricks earned his PhD in biochemistry from the University of Maryland, College Park.

Katya Tsaioun is the director of the Evidence-based Toxicology Collaboration at the Johns Hopkins Bloomberg School of Public Health. The collaboration's mission is to bring together the international toxicology community to facilitate the use of the evidence-based toxicology to inform regulatory, environmental, and public health decisions. She received her PhD in human nutrition science from the Tufts University Friedman School of Nutrition Science and Policy.

Yiliang Zhu is a professor in the Division of Epidemiology, Biostatistics, and Preventive Medicine in the School of Medicine at the University of New Mexico (UNM). He directs the biostatistics, epidemiology, and research design cores for the Clinical and Translational Research Center of UNM and for the Mountain West Clinical and Translational Research Infrastructure Network, a consortium of 13 universities in 7 states. His research focuses on quantitative methods in health risk assessment, including integrative modeling of biological systems, dose-response modeling, benchmark-dose methods, and uncertainty quantification. He also conducts research in biostatistics methods, clinical- and health-outcome evaluation, and impact assessment of health care systems and policies in northwestern rural China. Before joining UNM, Dr. Zhu was a professor at the University of South Florida College of Public Health, where he directed the Biostatistics PhD program and the Center for Collaborative Research. Dr. Zhu has served on several National Academies' committees, including the Committee on EPA's Exposure and Human Health Assessment of Dioxin and Related Compounds, the Committee on Tetrachloroethylene, the Committee to Review EPA's Draft IRIS Assessment of Formaldehyde, and the Committee to Review the IRIS Process. He received a PhD in statistics from the University of Toronto.

Appendix B

Public Meeting Agendas

First Committee Meeting for the Committee to Review EPA's TSCA Systematic Review Guidance Document

National Academy of Sciences
2101 Constitution Avenue, NW
Lecture Room
Washington, DC 20418

PUBLIC AGENDA

FRIDAY, FEBRUARY 28, 2020 – 9:15 am–12:00 pm (Eastern)

MEETING OBJECTIVES

- Hear presentations from the U.S. Environmental Protection Agency (EPA) on systematic review under TSCA
- Learn about systematic review principles

FRIDAY, FEBRUARY 28, 2020

OPEN SESSION (NAS, Lecture Room)

9:15 **Purpose of Open Session and Introduction of Committee Members**
Jonathan Samet
Chair, Committee to Review EPA's TSCA Systematic Review Guidance Document
Dean, Colorado School of Public Health

9:30 **Systematic Review Under TSCA**
Susanna Blair and Iris Camacho, EPA Office of Pollution Prevention and Toxics

11:30 **Opportunity for Public Comments to National Academies Committee**
Each speaker has a maximum time limit of 5 minutes. Accompanying written materials are encouraged.

12:00 **Adjourn**

Public Agenda for Virtual Meeting 2.1
for the Committee to Review EPA's
TSCA Systematic Review Guidance Document

FRIDAY, JUNE 19, 2020 – 10:00 am–1:00 pm (Eastern)

MEETING OBJECTIVE

- Learn EPA's innovations in literature searching and screening to identify relevant scientific studies

FRIDAY, JUNE 19, 2020

10:00 **Purpose of Open Session and Introduction of Committee Members**
Jonathan Samet
Chair, Committee to Review EPA's TSCA Systematic Review Guidance Document
Dean, Colorado School of Public Health

10:05 **Overview of TSCA Risk Evaluation Process**
Stan Barone, Deputy Director, Risk Assessment Division, EPA Office of Pollution Prevention and Toxics, Office of Chemical Safety and Pollution Prevention

10:20 **Innovations in Searching and Screening Literature**
Kellie Fay, Risk Assessment Division, EPA Office of Pollution Prevention and Toxics, Office of Chemical Safety and Pollution Prevention

10:50 **Committee Discussion**

11:20 **Break**

11:25 **Automated Literature Prioritization Methods in SWIFT Review**
Chantel Nicolas, EPA Office of Chemical Safety and Pollution Prevention, Office of Pollution Prevention and Toxics

Electronic Screening in DistillerSR and SWIFT ActiveScreener
Francesca Branch, EPA Office of Chemical Safety and Pollution Prevention, Office of Pollution Prevention and Toxics

11:50 **Committee Discussion**

12:00 **Break**

12:05 **Interactive Breakout Sessions**

- These breakout sessions will discuss videos that are prerecorded and posted on the TSCA Systematic Review study website. Viewing of these videos prior to participating in the session is necessary to fully participate. Discussion moderators will pose questions from the committee and the public during the breakout session to the poster presenters.

Breakout Session 1: The Role of PECO Statements, Search Criteria, and Templates in Searching and Screening
Moderated by Karen Robinson

Videos discussed:

- Evidence Mapping of Gray Literature Under TSCA: A Gray Literature Decision Tree Framework
  Yousuf Ahmad, Risk Assessment Division, EPA Chemical Safety and Pollution Prevention, Office of Pollution Prevention and Toxics
- Strategy for Developing Literature Search Strings to Identify Publications Containing Environmental Fate and Transport and Physical-Chemical Property Data
  Amina Wilkins, EPA Office of Research and Development

Breakout Session 2: Evidence Mapping Through Prioritizing, Pre-Screening, and Beyond

Moderated by Katya Tsaioun

Videos discussed:

- Evidence Mapping for Engineering and Exposure: Literature Search, Prioritization, and Pre-Screening Strategy
  Katherine Phillips, EPA Office of Research and Development
- Evidence Mapping for Engineering and Exposure: Part B—Title/Abstract and Full-Text Screening
  Yadi Lopez, Risk Assessment Division, EPA Office of Pollution Prevention and Toxics, Office of Chemical Safety and Pollution Prevention
- Evidence Mapping of Environmental and Human Health Hazard Evidence: Phthalic Anhydride Example
  Kellie Fay, Risk Assessment Division, EPA Office of Pollution Prevention and Toxics, Office of Chemical Safety and Pollution Prevention

12:55 **Report Back from Breakouts**

1:00 **Adjourn Public Session**

Public Agenda for Virtual Meeting 2.2
for the Committee to Review EPA's
TSCA Systematic Review Guidance Document

THURSDAY, JULY 23, 2020 – 2:30 pm–5:20 pm (Eastern)

MEETING OBJECTIVE

- Provide a summary of EPA's innovations in the TSCA data evaluation and evidence integration process

THURSDAY, JULY 23, 2020

2:30 **Purpose of Open Session and Introduction of Committee Members**
Jonathan Samet
Chair, Committee to Review EPA's TSCA Systematic Review Guidance Document
Dean, Colorado School of Public Health

2:35 **Introductions from EPA**

2:40 **Data Evaluation Process and How Scoring Is Applied**
Francesca Branch, EPA Office of Chemical Safety and Pollution Prevention, Office of Pollution Prevention and Toxics

3:10 **Comparison of Different Data Evaluation Procedures**
Tracey Woodruff, Program on Reproductive Health and the Environment, University of California, San Francisco

3:20 **Committee Discussion**

3:50 **Evidence Integration Supporting Hazard and Exposure Assessments**
Stan Barone, Deputy Director, Risk Assessment Division, EPA Office of Pollution Prevention and Toxics, Office of Chemical Safety and Pollution Prevention

Eva Wong, Risk Assessment Division, EPA Office of Pollution Prevention and Toxics, Office of Chemical Safety and Pollution Prevention

4:10 **Committee Discussion**

4:40 **Public Comments (each presenter has a 3-minute limit)**
Suzanne Fluharty, Yurok Tribe Environmental Program
Julie Goodman, Gradient
Ross Henderson, CAREM LLC
Patricia Koman, self
Stefanus Muryanto, UNTAG University in Semarang, Indonesia
Steve Risotto, American Chemistry Council
Robert Sussman, Sussman & Associates
Anthony Tweedale, R.I.S.K. Consultancy
Daniele Wikoff, ToxStrategies

5:20 **Adjourn Open Session**

Public Agenda for Virtual Meeting 2.3
for the Committee to Review EPA's
TSCA Systematic Review Guidance Document

MONDAY, AUGUST 24, 2020 – 2:30 pm–4:55 pm (Eastern)

MEETING OBJECTIVE

- To hear details about EPA's innovations in the TSCA data evaluation and evidence integration process

MONDAY, AUGUST 24, 2020

2:30 **Purpose of Open Session and Introduction of Committee Members**
Jonathan Samet
Chair, Committee to Review EPA's TSCA Systematic Review Guidance Document
Dean, Colorado School of Public Health

2:35 **Discussion of Questions About the TSCA Process**
Stan Barone, Deputy Director, Risk Assessment Division, EPA Office of Pollution Prevention and Toxics, Office of Chemical Safety and Pollution Prevention

2:50 **Discussion and Clarifications on Approach from EPA (Discussion time 35 minutes)**

3:25 **Break** (10 minutes)

3:35 **Interactive Breakout Sessions**

- These breakout sessions will discuss videos which are prerecorded and posted on the TSCA Systematic Review study website. Viewing of these videos prior to participating in the session is necessary to fully participate. Discussion moderators will pose questions from the committee and the public during the breakout session to the poster presenters.

Breakout Session 1: Evaluation and Scoring
Moderated by Bryan Brooks

Videos Discussed:

- Data Evaluation for Physical-Chemical and Fate Properties Under TSCA
  Tameka Taylor, Risk Assessment Division, EPA Office of Pollution Prevention and Toxics, Office of Chemical Safety and Pollution Prevention
- Data Evaluation for Exposure and Engineering Studies Under TSCA
  Nerija Orentas, Risk Assessment Division, EPA Office of Pollution Prevention and Toxics, Office of Chemical Safety and Pollution Prevention
- Data Evaluation for Environmental Hazard Studies Under TSCA
  Amelia Nguyen, Risk Assessment Division, EPA Office of Pollution Prevention and Toxics, Office of Chemical Safety and Pollution Prevention
- Data Evaluation for Animal Toxicity and In Vitro Studies to Support Human Health Hazard Under TSCA
  Amy Benson, Risk Assessment Division, EPA Office of Pollution Prevention and Toxics, Office of Chemical Safety and Pollution Prevention
- Data Evaluation for Epidemiological Studies Under TSCA
  Francesca Branch, EPA Office of Chemical Safety and Pollution Prevention, Office of Pollution Prevention and Toxics

Breakout Session 2: Evidence Integration
Moderated by Jessica Myers

Videos discussed:

- Evidence Integration of Physical-Chemical and Fate Property Data Under TSCA
  Marcy Card, Risk Assessment Division, EPA Office of Pollution Prevention and Toxics, Office of Chemical Safety and Pollution Prevention
- Evidence Integration of Exposure Data Under TSCA
  Eva Wong and Ariel Hou, Risk Assessment Division, EPA Office of Pollution Prevention and Toxics, Office of Chemical Safety and Pollution Prevention
- Evidence Integration of Environmental and Human Health Hazard Data Under TSCA
  Kara Koehrn, Risk Assessment Division, EPA Office of Pollution Prevention and Toxics, Office of Chemical Safety and Pollution Prevention

4:20 **Discussion Following Breakouts Groups, Recap of Breakouts**

4:55 **Adjourn Public Session**

Appendix C

Documents Reviewed by the Committee

In order to complete its review, the committee reviewed several documents, which are listed below. Many are available through the provided links and others are available through the National Academies' Public Access File (as are all presentations made to the committee; see the agendas for the public meetings in Appendix B). In order to review these documents, please email paro@nas.edu for more information.

1. EPA (U.S. Environmental Protection Agency). 2018. Application of Systematic Review in TSCA Risk Evaluations. Office of Chemical Safety and Pollution Prevention (OCSPP), Washington, DC. https://www.epa.gov/sites/product Office of Chemical Safety and Pollution Prevention (OCSPP) ion/files/2018-06/documents/final_application_of_sr_in_tsca_05-31-18.pdf, accessed November 5, 2020.

2. EPA. 2020. Draft Risk Evaluation for Trichloroethylene. Office of Chemical Safety and Pollution Prevention (OCSPP), Washington, DC. https://www.epa.gov/sites/production/files/2020-02/documents/1_draft_risk_evaluation_for_trichloroethylene_tce_public.pdf, accessed November 5, 2020.
   a. Draft evaluation main page: https://www.epa.gov/assessing-and-managing-chemicals-under-tsca/draft-risk-evaluation-trichloroethylene.
   b. Docket entry: https://beta.regulations.gov/document/EPA-HQ-OPPT-2019-0500-0001.
   c. Scope, Literature Search Strategy, and Bibliography Documents: https://www.epa.gov/assessing-and-managing-chemicals-under-tsca/trichloroethylene-tce-scope-document-and-supplemental.
   d. Problem Formulation Document: https://www.epa.gov/assessing-and-managing-chemicals-under-tsca/trichloroethylene-tce-problem-formulation.

3. EPA. 2020. Risk Evaluation for 1-Bromopropane (n-Propyl Bromide). Office of Chemical Safety and Pollution Prevention (OCSPP), Washington, DC. https://www.epa.gov/sites/production/files/2020-08/documents/risk_evaluation_for_1-bromopropane_n-propyl_bromide.pdf, accessed November 5, 2020.
   a. Evaluation main page: https://www.epa.gov/assessing-and-managing-chemicals-under-tsca/final-risk-evaluation-1-bromopropane.
   b. Docket entry: https://beta.regulations.gov/document/EPA-HQ-OPPT-2019-0235-0001.
   c. Scope, Literature Search Strategy, and Bibliography Documents: https://www.epa.gov/assessing-and-managing-chemicals-under-tsca/1-bromopropane-1-bp-scope-document-and-supplemental.
   d. Problem Formulation Document: https://www.epa.gov/assessing-and-managing-chemicals-under-tsca/1-bromopropane-1-bp-problem-formulation.

4. Seema Schappelle, PhD, Office of Chemical Safety and Pollution Prevention, U.S. EPA. Text document with answer to committee questions sent on June 10, 2020. Available at paro@nas.edu.