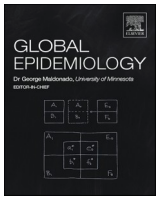
Global Epidemiology 4 (2022) 100093

Global Epidemiology

journal homepage: www.sciencedirect.com/journal/global-epidemiology



Commentary

Lost in the woods: Finding our way back to the scientific method in systematic review

Heather N. Lynch [a,*], Kenneth A. Mundt [a], Dirk Pallapies [b], Paolo F. Ricci [c]

[a] Stantec ChemRisk, Boston, MA, United States of America
[b] The Institute for Prevention and Occupational Medicine of the German Social Accident Insurance, Bochum, Germany
[c] University of Bologna, Bologna, Italy

ARTICLE INFO

Keywords
Systematic review
Regulatory risk assessment
Evidence integration

ABSTRACT

Systematic review has become the preferred approach to addressing causality and informing regulatory and other decision-making processes, including chemical risk assessments. While advocates of systematic reviews acknowledge that they hold great potential for increasing objectivity and transparency in assessments of chemicals and human health risks, standardizing and harmonizing systematic review methods have been challenging. This review provides a brief summary of the development of systematic review methods and some of the frameworks currently in use in the US and Europe. We also provide an in-depth evaluation and comparison of two "competing" US EPA systematic review frameworks, informed by the constructively critical recommendations from the US National Academies of Science, Engineering and Medicine. We conclude with suggestions for moving forward to harmonize systematic review methods, as we believe that further criticism of individual available frameworks likely will be unproductive. Specifically, we issue a call to action for an international collaboration to work toward a blueprint that embraces the most scientifically critical elements common to most systematic review frameworks, while necessarily accommodating adaptations for specific purposes. Despite the array of available systematic review methods, it is clear that there is a shared goal and desire to promote objective assessment and synthesis of scientific evidence informing globally important issues regarding disease causality and human health risk evaluation.

1. Introduction

Scientific research generates evidence, and scientific evidence accumulates. How can health scientists and public health decision-makers make any sense of the piles of published data, studies, and interpretations?

Health scientists – including toxicologists, epidemiologists, exposure experts (including industrial hygienists) and risk assessors – constantly search, screen, review and interpret published scientific studies to understand what causes human diseases such as cancers. We scientists, as well as colleagues in regulatory, policy, and decision-making roles, generally agree that the most valid interpretations and conclusions on which important decisions should be based will derive from carefully critically reviewing, synthesizing and integrating the body of relevant scientific evidence. However, while agreeing in principle, the health sciences community has not arrived at any harmonized way of performing such reviews and often does not address the central activity of evidence integration.

The lack of consistent definitions or a harmonized approach in chemical hazard assessment methods has given rise to differing results and interpretations across reviews of the same substances – sometimes largely based on the same primary studies [29]. One modern classic example may be the evaluation of glyphosate, which the International Agency for Research on Cancer (IARC) classified as "probably carcinogenic to humans" [16] whereas EPA's evaluation indicated that "glyphosate is not likely to be carcinogenic to humans" [10]. Similarly, the European Food Safety Authority (EFSA) concluded that glyphosate is "unlikely" to cause cancers in humans [5].

The glyphosate hazard classifications coincided with an increasing focus among health agencies on the *methods* being used to evaluate the potential human health effects of chemicals, as well as the communication about these methods. Stakeholders recognized that the systematic review (SR) process as applied for more than a decade in health care intervention research (e.g., the Cochrane Collaboration, founded in

* Corresponding author at: Stantec ChemRisk, 607 Boylston St, Ste 301, Boston, MA 02116, United States.
E-mail address: heather.lynch@stantec.com (H.N. Lynch).

AB36-Comm-147-4

https://doi.org/10.1016/j.gloepi.2022.100093
Received 16 June 2022; Received in revised form 20 October 2022; Accepted 1 November 2022
Available online 9 November 2022

1992) also could be adapted – at least in principle – to the assessment and synthesis of results of observational studies of occupational and environmental exposures and risk of diseases including cancers. Thus, numerous agencies in the United States and globally have begun embracing SR objectives and methods.

While there is no consensus regarding the definition of SR or SR methods – even within its origins in health care – systematic review is defined by Cochrane as an "explicit, systematic" method intended to minimize bias through "a clearly stated set of objectives with predefined eligibility criteria for the studies; an explicit, reproducible methodology; a systematic search that attempts to identify all the studies that would meet the eligibility criteria; an assessment of the validity of the findings of the included studies, for example through the assessment of the risk of bias; and a systematic presentation, and synthesis, of the characteristics and findings of the included studies" [12,30]. As noted in Krnic Martinic et al. [33], commonly used definitions are helpful, but terms such as "explicit" and "systematic" remain vague in a practical sense and subject to individual interpretation.

Currently, groups including public health researchers, regulators, and policy- and other decision-makers continue to use different approaches and methods for performing SRs, some more formally than others. Some organizations including government agencies have generated guidance for identifying, screening, selecting, interpreting, and critically assessing research studies. Additionally, a few have attempted to frame methods for integrating or synthesizing scientific evidence in support of causal conclusions. While evidence integration is considered by some to be separate and outside of the framework of SR, arguably, and as implemented by some EPA offices, evidence integration is the phase of review in which clearly articulated and consistent approaches are most needed [8,9,28]. Stakeholders will have higher confidence in a SR in support of a causal conclusion if standard evaluation methods are applied not only to individual studies in each line of evidence, but also to the way in which these lines of evidence are integrated – i.e., "the implementation of a prespecified and structured approach to reach a decision about a potential hazard, exposure or risk associated with chemicals" [28]. More controversial discussions on human health hazards arise when lines of evidence contradict another, e.g., mechanistic studies indicate that a carcinogenic MOA can be substantiated, there is strong animal evidence in one strain of rats but not mice or hamsters, and the epidemiology demonstrates no clear association. Thus, methods for integrating complex and conflicting evidence are crucial to valid and reproducible assessments.

Some of the organizations leading the challenge to develop or standardize SR methods include the following:

- International Agency for Research on Cancer (IARC) Monographs (a World Health Organization [WHO] organization) [13]
- WHO Chemical Risk Assessment Network [28]
- US EPA Office of Research and Development (ORD) Integrated Risk Information System (IRIS) Toxicological Reviews (US government) [9]
- US EPA Office of Pollution Prevention and Toxics (OPPT) under the Toxic Substances Control Act (TSCA) (US Government) [8,11]
- US National Institute for Environmental Health Sciences (NIEHS) Report on Carcinogens (RoC) (US Government; [21])
- National Toxicology Program (NTP) Office of Health Assessment and Translation (OHAT) [20,21]
- US Agency for Toxic Substances and Disease Registry (ATSDR) Toxicological Profiles (US Government) [32]
- US National Academies of Science, Engineering and Medicine (NASEM) (US private, nonprofit institutions that serve as Advisors to the US Congress and the nation)
  - o National Academy of Medicine (formerly the Institute of Medicine, or IOM), one of three academies of NASEM
- Cochrane Collaboration (UK charitable organization) [12]
- Ottawa Hospital/University of Oxford/Monash University's Preferred Reporting Items for Systematic Reviews (PRISMA) [23]

Additionally, several organizations have incorporated "weight of evidence" and other SR concepts into their chemical assessment frameworks and guidance including the European Food Safety Authority [4,6], the European Chemicals Agency (ECHA) [2], and the Organization for Economic Co-Operation and Development (OECD) [22]. While the application of "weight of evidence" methods varies across frameworks, it generally is interpreted as the process of first synthesizing within different lines of evidence (e.g., animal, human, and mechanistic studies) and then integrating across these lines of evidence to determine the answer to a research question [4]. Weight of evidence is employed within SRs, but when used as a standalone method, it does not necessarily include a similarly structured system for planning the review and collecting information. Suter et al. [27], for example, contrast "weight of evidence" with SR methods and propose a model for integrating them.

Some of these agencies explicitly set out general SR guidance for evaluating clinical and non-clinical literature (e.g., IOM, Cochrane, PRISMA [reporting guidelines only]), whereas others incorporated key elements of SR into their existing scientific review processes (e.g., ATSDR, EPA IRIS, EPA OPPT, and IARC). On the other hand, some of the European frameworks (e.g., ECHA, European Medicines Agency [EMA], OECD) promote SR, but do not define specific methods or consistently incorporate SR approaches across their programs [1]. While considerable methodological improvements and standardization have been achieved, less attention has been paid to the important differences between clinical experimental research and purely observational epidemiological studies. As a result, some well-intended efforts to review and summarize published evidence (especially those developed in the context of experimental science, including meta-analyses) are often mechanical and somewhat detached from concerns about evidence quality (especially confounding) and even the basic principles rooted in the scientific method.

The array of SR guidance promulgated by various organization reflects important basic SR themes and elements such as evaluating research and evidence on the basis of "quality" – defined and implemented at least slightly differently in each [17] – as well as making SR more objective, transparent and therefore reproducible. Collectively, these efforts have awakened a new sensibility regarding the importance and utility of valid SR, and increased urgency in advancing but also harmonizing SR methods. We now need to assure that SR is motivated by articulated and biologically sound hypotheses and follows basic research principles inherent to the scientific method.

2. European and international systematic review frameworks

Several European and international agencies have incorporated SR principles into their chemical hazard assessments; however, these generally are not statutorily required, and SR methods often are applied inconsistently. In 2012, ECHA released its *Guidance on Information Requirements and Chemical Safety Assessment*. Part B (Hazard Assessment) includes guidance for identifying and evaluating data for relevance, reliability, and adequacy – primarily intended for chemical companies and consortia to use when preparing and submitting data and dossiers under the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH) regulation. Whether REACH registrants actually and consistently follow such guidance, however, remains unclear.

In 2010, EFSA released guidance entitled, *Application of Systematic Review Methodology to Food and Feed Safety Assessments to Support Decision Making*, which serves as a framework for applying SR in evaluating the safety of food products and animal feed [6]. In 2017, EFSA published a "weight of evidence" guidance document that details important SR steps, including evidence integration, but forgoes the more formal

protocol steps of standardized individual study quality evaluation generally required for SR [4]. In particular, the EFSA guidance provides strategies for integrating evidence within and across lines of scientific inquiry (Section 4.4), including a conceptual model to visually demonstrate how evidence of differing weight is combined in the hazard assessment. It also recommends that the EFSA working group members "take account of any dependencies between different pieces and/or lines of evidence," as they "impact how evidence should be integrated" ([4], p 21).

However, EFSA may not yet consistently apply this guidance. In the EFSA [3] scientific assessment of per- and polyfluoroalkyl substances (PFAS) in food, for example, the "Literature search and appraisal of studies" section is under one page long, and largely focused on the search terms used to identify relevant studies. With regard to evaluating evidence, the document simply states:

> *"The information retrieved was subsequently reviewed by the CONTAM working group (WG) on PFASs in food, and has been used for the present assessment based on expert judgment. Selection of the scientific papers for inclusion or exclusion was based on consideration of the extent to which the study was relevant to the assessment and general study quality considerations" ([3], p 23–24).*

The specific quality assessment methods are not described, nor is the EFSA [4] guidance cited. Nonetheless, the assessment appears to be comprehensive, and it is hard to imagine that the review methods were not driven by basic SR principles. The updated [31] 2020 EFSA review of PFAS also neither referred to nor cited the EFSA 2010 or 2017 guidance.

IARC, an intergovernmental agency under the WHO, also incorporated SR principles into the last two iterations of the Preamble to the IARC Monographs [13,15]. The impetus for revising the Preamble was based on an IARC Advisory group recommendation "at a time when significant shifts are occurring in the scientific evidence that contributes to the understanding of carcinogenicity, as well as in approaches to information gathering and evidence assessment and integration" ([14], p.1). The Preamble in part functions as guidance for preparing IARC Monographs. While many of the key SR principles are apparent in the 2019 update, discussion of the study evaluation methods is quite brief, and technically the approach falls short of standard SR approaches as there is no protocol for systematically evaluating study quality. Furthermore, each of the three lines of scientific inquiry (i.e., epidemiological, animal and mechanistic studies) is reviewed and synthesized by different subgroups of scientists who may be assessing the evidence in different (and possibly subjective) ways.

3. Systematic review at US EPA

The US EPA clearly recognized the need to develop and incorporate SR processes into its human health hazard evaluations of chemicals. Most notably, both EPA's IRIS and TSCA programs embraced the challenge, but generated and subsequently used separate SR approaches.

Nevertheless, the road toward implementing SR has been neither straight nor smooth for either IRIS or TSCA. Not the least of the critics have been separate expert National Academies of Science, Engineering, and Medicine (NASEM, previously NAS (NASEM, previously known as NAS and its operating arm the National Research Council [NRC])) committees that independently reviewed at various points over the last two decades both the IRIS and TSCA draft guidance for performing SRs, and some criticisms have been remarkably direct. Among the most direct was the 2011 NRC review of the EPA IRIS Toxicological Review of Formaldehyde (Draft):

> *"Overall, the committee noted some recurring methodologic problems in the draft IRIS assessment of formaldehyde. Many of the problems are similar to those which have been reported over the last decade by other NRC committees tasked with reviewing EPA's IRIS assessments for other chemicals. Problems with clarity and transparency of the methods appear to be a repeating theme over the years, even though the documents appear to have grown considerably in length. In the roughly 1,000-page draft reviewed by the present committee, little beyond a brief introductory chapter could be found on the methods for conducting the assessment. Numerous EPA guidelines are cited, but their role in the preparation of the assessment is not clear. In general, the committee found that the draft was not prepared in a consistent fashion; it lacks clear links to an underlying conceptual framework; and it does not contain sufficient documentation on methods and criteria for identifying evidence from epidemiologic and experimental studies, for critically evaluating individual studies, for assessing the weight of evidence, and for selecting studies for derivation of the RfCs and unit risk estimates" ([19], p. 4).*

Another NRC committee recommendation was for EPA to scope reviews around biological plausibility. NRC specifically recommended that EPA "select outcomes [for assessment] on the basis of available evidence and understanding of mode of action" and to consider whether the scientific evidence indicated that a hypothesized exposure-disease association was biologically plausible ([19], p. 164). This specification of potential causal models on which SR may be based generally has been lacking in most SR guidance.

In the years following these NRC critiques, EPA has updated the IRIS evaluation process to incorporate more key SR elements. Until recently, these changes appeared within specific IRIS assessments. However, in November 2020, EPA's Office of Research and Development (ORD) released a draft *ORD Staff Handbook for Developing Integrated Risk Information System (IRIS) Assessments* [9]. The Handbook makes great strides toward addressing many of the NRC and NASEM committees' and others' criticisms. However, the pendulum might have swung too far in the other direction with the level of detail and prescriptiveness. For example, Section 4, which describes processes for literature search and screening, is nearly 40 pages and details numerous tools and programs available to identify relevant literature. Similarly, Section 8 of the Handbook *Extraction and Display of Study Results of Health Effects and Toxicities from Epidemiology and Toxicology Studies* spans 20 pages and includes recommended templates for data extraction. Although these reflect areas of substantial investment by EPA, of greater concern is that the "clear links to an underlying conceptual framework" – including potential causal models and testable hypotheses – remain underdeveloped.

On nearly the tenth anniversary of the release of the NRC committee review of the IRIS Toxicological Review of Formaldehyde (Draft), another NASEM committee released its review of the 2018 TSCA SR guidance. This committee, too, was critical of the approach [18]:

- The TSCA systematic review approach does not meet the criteria of "comprehensive, workable, objective and transparent."
- The phases of scoping, problem formulation, and protocol development are merged and unclear.
- There is a lack of clear, refined research questions or a documented protocol for each risk evaluation.
- The numerical scoring used in the data quality evaluations is not justified.
- A documented approach to evidence synthesis and integration is lacking.

Interestingly, many of the NASEM committee criticisms of the TSCA systematic review reiterated those raised a decade earlier pertaining to the IRIS framework, and some still have not been addressed by either group. The 2021 Draft TSCA SR protocol [8] addressed many of the NASEM concerns and adopted substantive portions of the IRIS Handbook. However, differences – and gaps – remain in both guidance documents. Indeed, in the April 2022 Science Advisory Committee on Chemicals (SACC) public meeting regarding peer review of the TSCA SR protocol, the Committee noted that while the TSCA SR Protocol has

improved since the 2018 guidance, there are remaining areas requiring clarity, and, in some cases, additional improvements [7]. This raises key questions: Are in fact the two EPA approaches to SR all that different? Are these two EPA guidance documents superior to other competing SR frameworks? How critical is the "systematic" in SR?

4. Review of the literature: crosswalk of EPA systematic review frameworks and NASEM Committee comments

We reviewed the available US frameworks for SR in chemical risk assessment, including the EPA *Draft Staff Handbook for Conducting IRIS Assessments* [9] as well as the EPA OPPT's *Application of Systematic Review in TSCA Risk Evaluations* (2018) and the *Draft Systematic Review Protocol Supporting TSCA Risk Evaluations* that supersedes it [8], comparing the key steps in each framework against the best practices recommendations outlined in the NRC and NASEM committees' reviews of IRIS [19] and TSCA [18], respectively. We also reviewed the NTP OHAT *Handbook for Conducting Systematic Reviews for Health Effects Evaluations* (also adopted by the Agency for Toxic Substances and Disease Registry [ATSDR]), which was first developed in 2015 and was used to guide the development of both EPA frameworks. A high-level crosswalk of the frameworks is provided in Table 1 and a detailed description of each framework and their coverage of key elements of systematic review in Supplemental Table 1. These documents demonstrate substantial progress institutionalizing and improving SR within multiple EPA programs, but also highlight the opportunity to harmonize these frameworks.

Many of the 2011 NRC committee recommendations for the IRIS SR process appeared in the 2018 revised EPA TSCA framework and were implemented in the 2021 draft Handbook. The notable differences between the IRIS and TSCA frameworks largely reflect the discrete purpose and pace of reviews that each program requires. More specifically, the 2018 draft TSCA guidance and the 2021 TSCA SR protocol appear to have been tailored to meet the statutory mandates of using "best available science" and "weight of the scientific evidence", but also to be workable under relatively tight statutory deadlines (i.e., evaluations must be completed within 3–3.5 years of designating a chemical as a high-priority substance). EPA thoroughly developed much of the early SR process phases for its risk evaluations, focusing on scoping and problem formulation efforts, literature identification, and a detailed data quality evaluation system to assess human health, exposure and environmental data. In the 2018 guidance, later phases of the TSCA SR process, notably evidence synthesis and evidence integration for hazard classification, lacked explicit guidance. EPA expected these methods would be developed over the course of the first ten risk evaluations, and the 2018 framework updated accordingly. However, as noted by the NASEM committee (2021), there remains no clear framework for these critical steps, and little consistency in the associated narratives in the risk evaluations for the first ten chemical evaluations completed under the amended TSCA. The 2021 TSCA SR protocol, which supersedes the 2018 guidance, addresses much of the NASEM criticism regarding a clear framework for evidence synthesis and cross-discipline evidence integration, which the TSCA SR framework adopted almost word-for-word from the IRIS Handbook. However, the TSCA SR Protocol remains silent on how it plans to integrate the findings of its SR into later risk evaluation steps – i.e., selecting/deriving toxicity benchmarks and determining whether a chemical may cause unreasonable risk to human health or the environment.

The draft IRIS Handbook essentially functions as the Agency's response to the NRC committees' (and others') critiques of the IRIS

Table 1
Comparison of NRC/NASEM Committee Review Recommendations and Areas Addressed in TSCA and IRIS Systematic Review Frameworks.

Principles of Systematic Review	NRC Committee 2011 Review of IRIS	Draft IRIS Handbook (2020)	NTP OHAT [20]/ ATSDR (2014)	TSCA 2018 Process	NASEM Committee Review of TSCA	2021 TSCA SR Protocol
Early assessment planning and problem formulation	No recommendations	Addressed	Addressed	Addressed	Recommended	Addressed
Standardization of review and evaluation approaches across reviewers	Recommended	Partially addressed	Addressed	Partially addressed	Recommended	Partially addressed
Transparent and well documented methods	Recommended	Addressed	Addressed	Addressed	Recommended OHAT, IRIS, and Navigation Guide methods	Addressed
Comprehensive and well documented literature review and article selection process	Recommended	Addressed	Addressed	Addressed	Recommended	Addressed
Incorporation of Mode of Action (MOA) into hazard assessment approach (e.g., selection of outcomes of concern)	Recommended	Partially addressed	Partially addressed	Partially addressed	Recommended organization around PECO	Partially addressed
Standardized approach to data abstraction/tabular summary	Recommended	Addressed	Addressed	Addressed	No recommendations	Addressed
Standardized study quality evaluation approach	Recommended	Addressed, qualitative	Addressed	Addressed, quantitative	Recommended, qualitative	Partially addressed, qualitatively
Clear and consistent evidence synthesis method and narrative	Recommended	Addressed	Partially addressed	Partially addressed	Recommended	Partially addressed
Clear evidence integration using Bradford Hill or similar causal approach	Recommended	Partially addressed	Addressed	Partially addressed	Recommended NAAQS approach	Partially addressed
If conducting quantitative dose-response analysis, consider quality and limitations of studies	Recommended	Addressed	Not applcable	Addressed	Recommended	Addressed
Clear guidelines for selection of studies and derivation of toxicity values/ data pooling approaches	Recommended	Partially addressed	Partially addressed (meta-analysis)	Not addressed	Recommended	Not addressed
Well described and justified use of dose-response models	Recommended	Not addressed	Not applicable	Not addressed	Recommended	Not addressed
Thorough discussion of uncertainty and variability in all steps of review	Recommended	Partially addressed	Partially addressed	Partially addressed	Recommended w/ quantitative characterization	Partially addressed

program. The Handbook is intended to clarify and improve IRIS "operating procedures" in accordance with, and without altering any existing EPA guidance [9]. EPA noted that the overall IRIS assessment development process has not changed, but has been "supplemented" by improved SR approaches. The draft Handbook is nearly 300 pages, but much of the detail centers around somewhat mechanical tasks of literature searching, screening, and inventory phases of the analysis. In some parts, the document is so highly prescriptive in its treatment of specific databases, keywords, and machine learning tools that it appears to have omitted vital sections on how to evaluate and interpret mixed bodies of literature in the evidence synthesis and evidence integration phases. Less clear are improvements regarding the objectives of the reviews, and the specific hypotheses that motivate them and in principle help determine an appropriate review methodology. In other words, it is not clear that the SR process necessarily must adhere to the principles of the scientific method (as with primary research studies) including formulating and specifying testable hypotheses.

EPA has outlined a general approach to weighing evidence and assigning strength to individual streams of evidence in deriving causal determinations; however, at several points scientific (and perhaps subjective) judgment may be required, and therefore it remains to be seen whether a truly "objective" framework can be applied consistently across IRIS assessments. On page 11–3 of the IRIS Handbook, EPA describes the process for overall strength of evidence determinations, stating:

> *"To add transparency and improve clarity in the systematic process, a standardized set of terms is used to describe the strength of the human and animal evidence for each assessed health effect. The terms associated with the different strength of evidence judgments are robust, moderate, slight, and indeterminate, which are differentiated by the quantity and quality of information available to rule out alternative explanations. Additionally, a judgment of compelling evidence of no effect may be used in rare instances". (p. 11–5).*

EPA provides some guidance around considerations within each of these strength of evidence judgments, i.e., they provide a framework for documenting decision-making. Nonetheless, given that the guidance is short and somewhat vague, their uniform application across assessments has been challenging.

In fact, subjective determinations appear to be necessary throughout the SR process, and therefore transparency might be the higher goal. Further, little discussion addresses the later steps including evidence integration and the application of the SR results to inform toxicity value derivation and quantitative risk assessment. Despite the 2011 NRC committee's recommendation that EPA use clear guidelines for study selection for toxicity value derivation, EPA has maintained that because dose-response is outside of the traditional systematic review paradigm, it is outside of the scope of providing a dose-response assessment framework within the SR protocol.

It also is intriguing that in the review of the 2018 TSCA framework the NASEM committee recommended that EPA adopt the IRIS framework (and develop a TSCA Handbook), in addition to considering the NTP OHAT approach to which the TSCA, IRIS, the Navigation Guide and other recent SR guidance documents all are highly referential. Indeed, the draft TSCA Systematic Review Protocol, while still retaining some of its unique methods (i.e., the qualitative portion of its quality evaluation system) has adopted much of the approach described in the IRIS Handbook. Nevertheless, while the specific NASEM committee criticisms differ (largely as a function of the stage of development of each EPA group's SR methods), their overarching recommendations regarding SR best practices remained the same. This outcome points to a promising pathway out of the forest, but will require cooperation rather than competition, compromise over assertion, inquisitiveness over the need-to-prove, and open-mindedness in formulating a harmonized SR methodology that aligns with basic scientific method.

5. Revisiting the core principles of the scientific method in systematic review

The US EPA's pursuit of SR within IRIS and TSCA possibly represents the largest effort expended globally toward developing standardized SR methods. Ironically, the parallel SR development by two separate EPA programs has both helped reinforce some key concepts, but also has contributed to SR method heterogeneity. Comparing the IRIS and TSCA SR frameworks and their respective NASEM committee reviews and the NTP OHAT framework on which they are based illuminates three consistent findings:

1) There is general agreement on the critical high-level steps of SR;

2) Where procedural and methodological differences exist, they often reflect differing stakeholder needs and Agency/program priorities; and.

3) Neither guidance currently could serve, on its own, as the "end all, be all" guidance document for SR within EPA or more broadly.

Specific agencies and organizations around the world may require a high degree of specificity and depth in their SR methods, especially in the interest of consistency and for meeting the specific needs of stakeholders. In contrast, some groups, such as US EPA's OPPT, are held to strict, relatively short regulatory deadlines and thus must use more fit-for-purpose methods. However, regardless of the level of documentation and the need for specific databases, software, and tools, these frameworks are united by the same underlying SR principles. As demonstrated in Table 1, neither the IRIS, nor OHAT, nor TSCA guidance fully describes all of the major steps in SR in a sufficiently simple, clear and actionable manner.

We therefore raise a call to action in which an international, multidisciplinary and inter-agency/entity collaborative team would revisit the foundational SR conceptual framework and crystallize a set of basic definitions and elements that are (or should be) common to SRs for use in chemical risk assessment. While seemingly a lofty goal, the objective would be to create as pointed and concise a document as possible, allowing for individual agencies to build upon it to meet regulatory and programmatic needs. Systematic review can and should be conducted in the same rigorous manner as a scientific study, and the editorial boards of top medical and health journals embrace related standards for such research. Doing so is paramount to, and probably no different from, any other scientific exercise, in that the guideline should conform to basic principles of the scientific method as conceived in the 17th century and universally adopted to this day. In brief, the basic steps of the scientific method are: 1) formulating key questions; 2) formulating the hypothesis (es); 3) testing the hypothesis with transparent and reproducible methods; 4) analyzing the data; 5) drawing conclusions; and 6) communicating results. In the SR context, the "testing" methods are aimed at describing and evaluating primary studies in a way that avoids or reduces bias, to the extent possible.

In a nutshell, SR should begin with formulating specific research questions and hypotheses guided by specified causal models reflecting current understanding of biological plausibility, a transparent review method, and a clear rationale for interpreting results. An adaptable, standard framework (or "blueprint") would not compromise the principles of systematic review, but would require that the reviewer define the purpose and rationale for deviating (if needed and justified) from the framework. While scientific judgment – and therefore some element of subjectivity – is unavoidable in any SR, with baseline requirements for full documentation of methods and scientific judgments made throughout the review process, stakeholders can pinpoint specific areas of disagreement about which further scientific discussion is warranted.

Our increasingly complicated regulatory and decision-making environment would benefit immediately from a shared, streamlined core SR framework based on a solid conceptual framework inspired by the foundations of the scientific method. As needed, an agency or program could supplement the core framework – preferably transparently – with additional guidance in a fit-for-purpose manner. Based on the NASEM

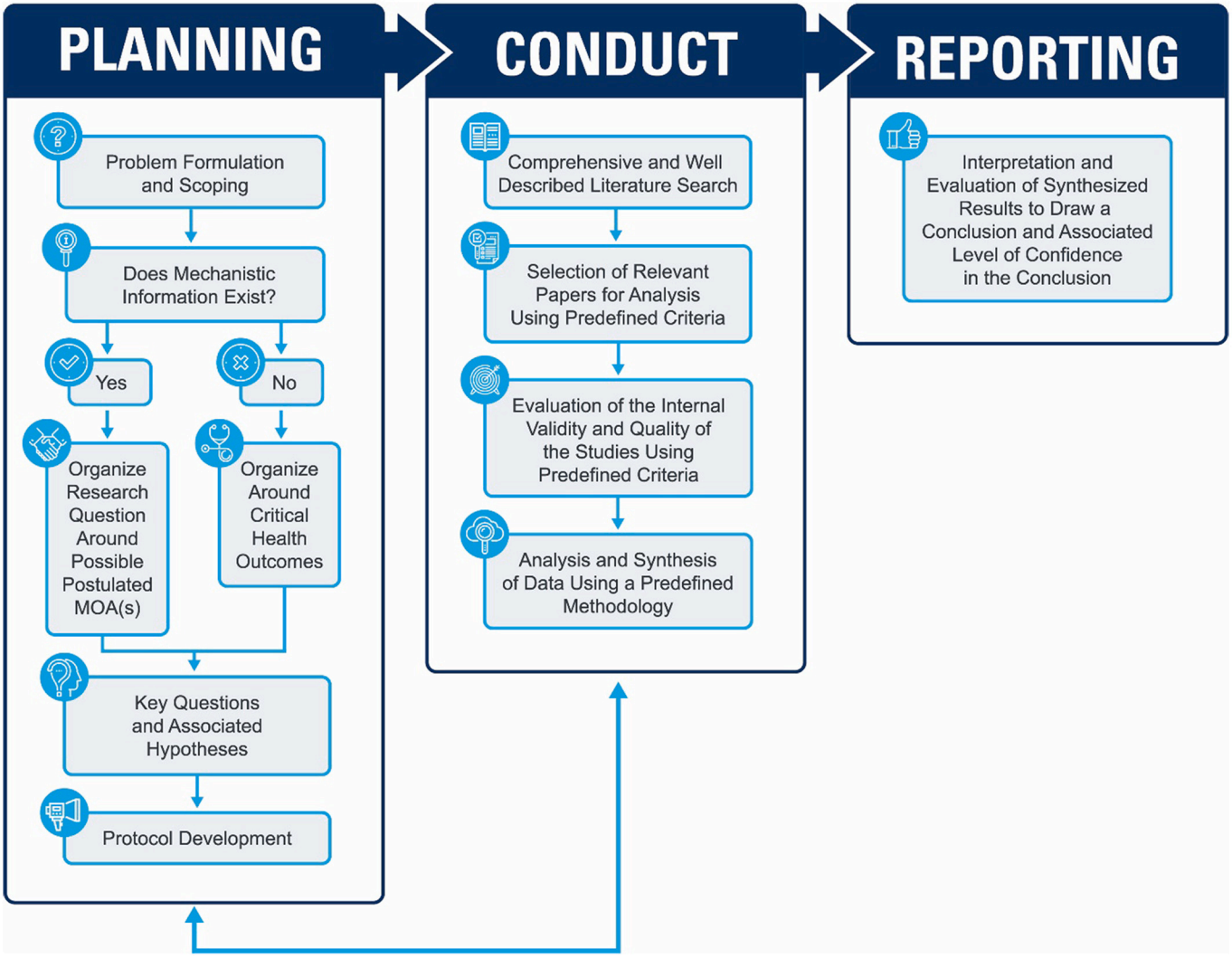


Fig. 1. Core principles and processes of systematic review.

committee critiques and numerous peer-reviewed publications regarding best practices for weight of evidence and SR in chemical risk assessment ([24–26]; [29]; [4]), the core SR framework elements are as outlined in Fig. 1 below. The early problem formulation and scoping exercises are critical for establishing a foundation that advances science in ways relevant to stakeholders. Scientific hypothesis then can be framed around what may be known or unknown about likely modes of action underlying the potential effects of interest.

6. Future directions

Developing sound SR core methods has risen to an appropriate level of importance and urgency. It is encouraging that many organizations and agencies around the globe require at least some elements of SR in chemical risk evaluations and some offer approaches and methodological options; however, it also presents us with the dizzying task of wading through various frameworks in hopes of finding the "best" method. The best methods likely are already in hand – spread across multiple frameworks – but need to be identified, selected and refined based on a new blueprint that starts with a testable hypothesis or research question informed by a biological understanding of potential exposure-disease associations. This blueprint should include the basic structural elements of a research study but remain flexible enough to be responsive to the specific research questions at hand as well as to the unique needs of stakeholders. In other words, a "new" method is not needed; rather, a more concise and broadly applicable framework for SR in chemical risk evaluation could be developed by drawing from the most consistently emphasized principles and strongest guidance in the existing frameworks discussed in Table 1.

A multi-disciplinary collaborative team of key professionals, including scientists, advisory and regulatory agency representatives and professional society members in the US and Europe would be well positioned to set forth, harmonize and reach agreement on the fundamental principles upon which to base valid SRs. We believe that further criticism of existing specific frameworks likely will be unproductive. On the other hand, working toward more concise and clearly articulated SR guidance would enhance consistency across public health agencies globally and lead to broader confidence in the chemical risk evaluations generated. Leadership and cooperation will be essential to success.

Funding

HNL and KAM are employees of Stantec ChemRisk, which provide services to clients such as governments, corporations, law firms and various scientific/professional organizations. The content and the conclusions of the manuscript are exclusively those of the authors. This manuscript was supported by funding from the Formaldehyde Panel of the American Chemistry Council, an industry trade association.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.gloepi.2022.100093.

References

[1] Ågerstrand M, Beronius A. Weight of evidence evaluation and systematic review in EU chemical risk assessment: foundation is laid but guidance is needed. Environ Int 2016;92-93:590–6. https://doi.org/10.1016/j.envint.2015.10.008. Epub 2015 Dec 9. PMID: 26682868.

[2] ECHA (European Chemicals Agency). Guidance on information requirements and chemical safety assessment. Part B: Hazard assessment. Helsinki, Finland, ECHA-11-G-16-EN. Accessed online at, https://echa.europa.eu/documents/10162/17235/information_requirements_part_b_en.pdf/7e6bf845-e1a3-4518-8705-c64b17cecae8; 2011.

[3] EFSA (European Food Safety Authority) CONTAM Panel. Scientific opinion on the risk to human health related to the presence of perfluorooctane sulfonic acid and perfluorooctanoic acid in food. EFSA J 2018;16(12):284. https://doi.org/10.2903/j.efsa.2018.5194. 5194.

[4] EFSA (European Food Safety Authority), Hardy A, Benford D, Halldorsson T, Jeger MJ, Knutsen HK, et al. Scientific opinion on the guidance on the use of the weight of evidence approach in scientific assessments. EFSA J 2017;15(8):69. https://doi.org/10.2903/j.efsa.2017.4971. 4971.

[5] EFSA (European Food Safety Authority). Conclusion on the peer review of the pesticide risk assessment of the active substance glyphosate. EFSA J 2015;13(11):107. https://doi.org/10.2903/j.efsa.2015.4302. 4302.

[6] EFSA (European Food Safety Authority). Application of systematic review methodology to food and feed safety assessments to support decision making. EFSA J 2010;8(6):1637 [90 pp.], https://doi.org/10.2903/j.efsa.2010.1637.

[7] EPA (United States Environmental Protection Agency). Peer Review of the Draft TSCA Systematic Review Protocol. Accessed online at: https://www.epa.gov/tsca-peer-review/peer-review-draft-tsca-systematic-review-protocol; 2022.

[8] EPA (United States Environmental Protection Agency). Draft Protocol for Systematic Review in TSCA Risk Evaluations. Washington, DC: Office of Chemical Safety and Pollution Prevention, U.S. EPA; 2021. EPA Document# EPA-D-20-031. Accessed online at, https://www.epa.gov/assessing-and-managing-chemicals-under-tsca/draft-protocol-systematic-review-tsca-risk-evaluations.

[9] EPA (United States Environmental Protection Agency). ORD Staff Handbook for Developing IRIS Assessments (Public Comment Draft, Nov 2020). Washington, DC: U.S. EPA Office of Research and Development; 2020. EPA/600/R-20/137, 2020. Accessed online at, https://cfpub.epa.gov/ncea/iris_drafts/recordisplay.cfm?deid=350086.

[10] EPA (United States Environmental Protection Agency). Revised Glyphosate Issue Paper: Evaluation of Carcinogenic Potential. Office of Chemical Safety and Pollution Prevention, Office of Pesticide Programs, U.S. EPA, Environmental Protection; 2017. DP Barcode: D444689 TXR#: 0057688.

[11] EPA (United States Environmental Protection Agency). Application of Systematic Review in TSCA Risk Evaluations. EPA Document # 740-P1–8001. Washington, DC: Office of Chemical Safety and Pollution Prevention, U.S. EPA, Environmental Protection; 2018. Accessed online at, https://www.epa.gov/sites/production/files/2018-06/documents/final_application_of_sr_in_tsca_05-31-18.pdf.

[12] Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al. Cochrane handbook for systematic reviews of interventions version 6.2 (updated February 2021). Cochrane 2021:2021. Available from, www.training.cochrane.org/handbook.

[13] IARC (International Agency for Research on Cancer). IARC Monographs on the Identification of Carcinogenic Hazards to Humans: Preamble. Lyons, France: IARC; 2019. Accessed online at, https://monographs.iarc.who.int/wp-content/uploads/2019/09/Preamble-AG-Report_final-web.pdf.

[14] IARC (International Agency for Research on Cancer). Report of the Advisory Group to Recommend an Update to the Preamble to the IARC Monographs. Lyons, France: IARC; 2019. Accessed online at.

[15] IARC (International Agency for Research on Cancer). IARC monographs on the identification of carcinogenic hazards to humans: preamble. Lyons, France: IARC; 2006. Accessed online at, https://monographs.iarc.who.int/wp-content/uploads/2018/06/CurrentPreamble.pdf.

[16] IARC (International Agency for Research on Cancer). Some organophosphate insecticides and herbicides IARC monographs on the evaluation of carcinogenic risks to humans. In: IARC monographs on the evaluation of the carcinogenic risk of Chemicals to man. Volume 112. Lyon: International Agency for Research on Cancer; 2016. Accessed online at, https://publications.iarc.fr/549.

[17] Lynch HN, Goodman JE, Tabony JA, Rhomberg LR. Systematic comparison of study quality criteria. Regul Toxicol Pharmacol 2016;76:187–98. https://doi.org/10.1016/j.yrtph.2015.12.017.

[18] NASEM (National Academies of Sciences, Engineering, and Medicine). The use of systematic review in EPA's Toxic Substances Control Act risk evaluations. Washington, DC: The National Academies Press; 2021.

[19] NRC (National Research Council). Review of the Environmental Protection Agency's draft IRIS assessment of formaldehyde. Washington, DC: The National Academies Press; 2011. https://doi.org/10.17226/13142.

[20] NTP (National Toxicology Program). Handbook for Conducting a Literature-Based Health Assessment Using OHAT Approach for Systematic Review and Evidence Integration. March 4, 2019. Research Triangle Park, NC Office of Health Assessment and Translation (OHAT), Division of the National Toxicology Program, National Institute of Environmental Health Sciences; 2019.

[21] NTP (National Toxicology Program). Handbook for preparing report on carcinogens monographs. Office of Report on Carcinogens, NTP, NIEHS; 2015. Accessed online at, https://ntp.niehs.nih.gov/ntp/roc/handbook/roc_handbook_508.pdf.

[22] OECD. Guiding principles and key elements for establishing a weight of evidence for chemical assessment. Series Test Assess 2019;No. 311 [Environment, Health and Safety Division, Environment Directive].

[23] Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71. https://doi.org/10.1136/bmj.n71.

[24] Rhomberg LR, Bailey LA, Goodman JE, Hamade AK, Mayfield D. Is exposure to formaldehyde in air causally associated with leukemia?–A hypothesis-based weight-of-evidence analysis. Crit Rev Toxicol 2011;41(7):555–621. https://doi.org/10.3109/10408444.2011.560140 [Epub 2011 Jun 2].

[25] Rhomberg LR, Goodman JE, Bailey LA, Prueitt RL, Beck NB, Bevan C, et al. A survey of frameworks for best practices in weight-of-evidence analyses. Crit Rev Toxicol 2013;43(9):753–84.

[26] Stephens ML, Betts K, Beck NB, Cogliano V, et al. The emergence of systematic review in toxicology. Toxicol Sci 2016;152(1):10–6.

[27] Suter G, Nichols J, Lavoie E, Cormier S. Systematic review and weight of evidence are integral to ecological and human health assessments: they need an integrated framework. Integr Environ Assess Manag 2020;16(5):718–28. https://doi.org/10.1002/ieam.4271.

[28] WHO (World Health Organization). Framework for the use of systematic review in chemical risk assessment. World Health Organization; 2021. https://apps.who.int/iris/handle/10665/347876 [License: CC BY-NC-SA 3.0].

[29] Whaley P, Halsall C, Ågerstrand M, Aiassa E, Benford D, Bilotta G, et al. Implementing systematic review techniques in chemical risk assessment: challenges, opportunities and recommendations. Environ. Int. 2016;92-93:556–64. https://doi.org/10.1016/j.envint.2015.11.002.

[30] Clarke, M., & Chalmers, I. Discussion sections in reports of controlled trials published in general medical journals: islands in search of continents?. *JAMA*, 1998,*280*(3), 280–282. 10.1001/jama.280.3.280.

[31] EFSA (European Food Safety Authority) CONTAM Panel. Scientific opinion on the risk to human health related to the presence of perfluorooctane sulfonic acid and perfluorooctanoic acid in food. EFSA J 2020;18(9):6223. https://doi.org/10.2903/j.efsa.2020.6223.

[32] Murray HE, Thayer KA. Implementing systematic review in toxicological profiles: ATSDR and NIEHS/NTP collaboration. J Environ Health 2014;76(8):34–5.

[33] Krnic Martinic M, Pieper D, Glatt A, Puljak L. Definition of a systematic review used in overviews of systematic reviews, meta-epidemiological studies and textbooks. BMC Med Res Methodol. 2019;19(1):203. https://doi.org/10.1186/s12874-019-0855-0. Published 2019 Nov 4.