

---

# Framework for the use of systematic review in chemical risk assessment



# Framework for the use of systematic review in chemical risk assessment

Framework for the use of systematic review in chemical risk assessment

ISBN 978-92-4-003448-8 (electronic version)

ISBN 978-92-4-003449-5 (print version)

© World Health Organization 2021

Some rights reserved. This work is available under the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 IGO licence (CC BY-NC-SA 3.0 IGO; <https://creativecommons.org/licenses/by-nc-sa/3.0/igo>).

Under the terms of this licence, you may copy, redistribute and adapt the work for non-commercial purposes, provided the work is appropriately cited, as indicated below. In any use of this work, there should be no suggestion that WHO endorses any specific organization, products or services. The use of the WHO logo is not permitted. If you adapt the work, then you must license your work under the same or equivalent Creative Commons licence. If you create a translation of this work, you should add the following disclaimer along with the suggested citation: "This translation was not created by the World Health Organization (WHO). WHO is not responsible for the content or accuracy of this translation. The original English edition shall be the binding and authentic edition".

Any mediation relating to disputes arising under the licence shall be conducted in accordance with the mediation rules of the World Intellectual Property Organization (<http://www.wipo.int/amc/en/mediation/rules/>).

**Suggested citation.** Framework for the use of systematic review in chemical risk assessment. Geneva: World Health Organization; 2021. Licence: CC BY-NC-SA 3.0 IGO.

**Cataloguing-in-Publication (CIP) data.** CIP data are available at <http://apps.who.int/iris>.

**Sales, rights and licensing.** To purchase WHO publications, see <http://apps.who.int/bookorders>. To submit requests for commercial use and queries on rights and licensing, see <http://www.who.int/about/licensing>.

**Third-party materials.** If you wish to reuse material from this work that is attributed to a third party, such as tables, figures or images, it is your responsibility to determine whether permission is needed for that reuse and to obtain permission from the copyright holder. The risk of claims resulting from infringement of any third-party-owned component in the work rests solely with the user.

**General disclaimers.** The designations employed and the presentation of the material in this publication do not imply the expression of any opinion whatsoever on the part of WHO concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. Dotted and dashed lines on maps represent approximate border lines for which there may not yet be full agreement.

The mention of specific companies or of certain manufacturers' products does not imply that they are endorsed or recommended by WHO in preference to others of a similar nature that are not mentioned. Errors and omissions excepted, the names of proprietary products are distinguished by initial capital letters.

All reasonable precautions have been taken by WHO to verify the information contained in this publication. However, the published material is being distributed without warranty of any kind, either expressed or implied. The responsibility for the interpretation and use of the material lies with the reader. In no event shall WHO be liable for damages arising from its use.

This publication contains the collective views of an international group of experts and does not necessarily represent the decisions or the stated policy of the World Health Organization.

This document was produced with the financial assistance of the European Union. The views expressed herein can in no way be taken to reflect the official opinion of the European Union.

Design and layout by Inis Communication

# Contents

<b>Preface</b>	<b>v</b>
<b>Acknowledgements</b>	<b>vi</b>
<b>Abbreviations</b>	<b>vii</b>
<b>1. Purpose</b>	<b>1</b>
<b>2. Introduction to systematic review and evidence integration</b>	<b>3</b>
2.1 What is systematic review?	3
2.2 Origins of systematic review	4
2.3 Why is systematic review of importance in chemical risk assessment?	7
2.4 Expertise and resources required for conducting a systematic review	9
2.5 Typical steps of a systematic review	10
References	11
<b>3. Problem formulation and protocol development</b>	<b>15</b>
3.1 Introduction	15
3.2 Problem formulation	15
3.3 Protocol development	20
References	22
<b>4. Search, selection and extraction</b>	<b>25</b>
4.1 Identification of records or studies	25
4.2 Screening the literature	30
4.3 Data extraction	31
References	33
<b>5. Appraisal of individual studies</b>	<b>35</b>
5.1 Multiple aspects of critical appraisal	35
5.2 Examples of threats to internal validity and sensitivity	37
5.3 Process for evidence appraisal	38
5.4 Use and reporting of critical appraisal of individual studies in the review	39
References	41
<b>6. Evidence synthesis and evidence integration</b>	<b>45</b>
6.1 Introduction	45
6.2 Evidence synthesis	47
6.3 Evidence integration	51
6.4 Overall uncertainty assessment	52
6.5 Conclusions for hazard identification or risk assessment	53
References	55
<b>7. Expectations for the reporting of systematic reviews</b>	<b>57</b>
Resources to help with reporting systematic reviews	61
<b>8. Future directions</b>	<b>63</b>
8.1 Conducting systematic reviews with limited resources and the automation of systematic review	63
8.2 Evolving applications of systematic review in risk assessment	64
References	64
<b>Annex. WHO Chemical Risk Assessment Network</b>	<b>65</b>

## Figures

Figure 2.1 How systematic review methods integrate with the overall risk assessment process.....	4
Figure 2.2 Stages of chemical risk assessment.....	5
Figure 2.3 Brief timeline of the development of systematic review methods with respect to chemical assessments.....	8
Figure 3.1 Conceptual demonstration of how a general problem formulation conducted for risk assessment may involve multiple questions.....	18
Figure 3.2 Demonstrative example of a PECO-based analytical framework for evaluation of hazard in a risk assessment.....	19
Figure 4.1 Search, selection and data extraction process flow.....	25
Figure 4.2 Example of a literature flow diagram summarizing the results of the screening process.....	31
Figure 5.1 Example of risk of bias heatmap for experimental animal studies in a systematic review .....	40
Figure 6.1 Workflow in evidence synthesis and evidence integration.....	52
Figure 6.2 Schema outlining how systematic review analyses can contribute to overall risk assessment findings.....	54

## Tables

Table 4.1 Examples of bibliographic databases relevant to systematic reviews on chemical risk assessment health topics.....	26
Table 4.2 Example of search string used in PubMed to identify research on PFOA or PFOS and immune effects .....	28
Table 4.3 Example tools for data identification and data extraction.....	29
Table 4.4 Databases and resources, by category.....	32
Table 5.1 Example risk of bias rating and rationale for each rating for DeWitt 2008.....	40
Table 7.1 Systematic review: reporting expectations and explanations.....	58

## Boxes

Box 2.1 Examples of initiatives to utilize systematic review methods in chemical risk assessment processes .....	6
Box 3.1 Questions faced during problem formulation .....	16
Box 3.2 Benefits of making a protocol publicly available.....	21
Box 4.1 Grey literature.....	27
Box 6.1 Definitions of terms for evidence synthesis and evidence integration.....	46

# Preface

The use of systematic review in decision-making for environmental health issues is growing. Systematic review approaches have the potential to improve decision-making in chemical risk assessment, in particular where there is conflicting evidence and where there is significant uncertainty.

This publication uses a high-level overview to provide guidance to chemical risk assessors who are not currently familiar with systematic approaches, without being prescriptive or endorsing any existing published methods. This framework will assist chemical risk assessors to understand assessments conducted by other institutions that have used systematic approaches, and will also assist in understanding the issues, limitations and challenges involved if institutions are considering using systematic review approaches in their own assessments.

# Acknowledgements

This publication was prepared by a drafting group consisting of experts associated with institutions that participate in the World Health Organization (WHO) Chemical Risk Assessment Network. The members of the drafting group were Elisa Aiassa (European Food Safety Authority, Italy); Anna Beronius (Karolinska Institute, Sweden); Brandy Beverly (National Toxicology Program, United States of America); Elaine Faustman (University of Washington, USA); Barbara Glenn (Environmental Protection Agency, USA); Annika Hanberg (Karolinska Institute, Sweden); Andrew Rooney (National Toxicology Program, USA); Nicolas Roth (Swiss Centre for Applied Human Toxicology, Switzerland); Christopher Weis (National Institute of Environmental Health Sciences, USA); Paul Whaley (University of Lancaster, United Kingdom of Great Britain and Northern Ireland); Daniele Wikoff (ToxStrategies, USA, through work undertaken in collaboration with the Office of Health Assessment and Translation of the National Toxicology Program, USA); Martin Wilks (Swiss Centre for Applied Human Toxicology, Switzerland); and Johanna Zilliacus (Karolinska Institute, Sweden). Angelika Tritscher (former WHO staff member) also contributed to this publication. Prior to the initial drafting of the manuscript, a questionnaire was sent to institutions that participate in the WHO Chemical Risk Assessment Network in order to assess the level of knowledge of and interest in systematic approaches within institutions undertaking chemical risk assessments, and the feedback from responses to that questionnaire informed the initial development of this publication (see Annex). Logistical support for developing this publication was provided by MDB Inc., through a contract funded by the PAHO/WHO Collaborating Centre for Environmental Health Sciences at the National Institute of Environmental Health Sciences, National Institutes of Health, USA.

Drafting of the manuscript was assisted by a meeting of drafting group members on 28–30 October 2019, hosted by the PAHO/WHO Collaborating Centre for Environmental Health Sciences at the National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, North Carolina, USA.

The draft publication underwent international peer review from participants in the WHO Chemical Risk Assessment Network and organizations active in the development of systematic approaches. Peer review comments were received from the following: Dinara Kenessary (Kazakhstan Medical University, Kazakhstan); Mary Gulumian (National Institute for Occupational Health, South Africa); Kathryn Guyton and Iciar Indave (International Agency for Research on Cancer, France); Kyriakoula Ziegler-Skylakakis (MAK Commission, Germany); Yadvinder Bhuller, Salma Iqbal, Benny Ling, Sara Mohr, Samira Roufik, Kavita Singh and Alexander Tsertsivadze (Health Canada, Canada); Heather Schaefer (Food and Drug Administration, USA); Homa Kashani and Masud Yunesian (Tehran University of Medical Sciences, Islamic Republic of Iran); Jessica Myers (Texas Commission on Environmental Quality, USA); Biljana Antonijevic (University of Belgrade, Serbia); Bette Meek (University of Ottawa, Canada); Kumiko Ogawa (National Institute of Health Sciences, Japan); Myungsil Hwang (National Institute of Food and Drug Safety Evaluation, Republic of Korea); Ingrid Druwe, Andrew Kraft, Kristina Thayer and George Woodall (Environmental Protection Agency, USA); George Fotakis (European Chemicals Agency, Italy); Kurt Straif (ISGlobal, Spain); Sebastian Hoffman (Evidence-based Toxicology Collaboration, Germany); Frank Pega (WHO staff member); Katya Tsaioun (Evidence-based Toxicology Collaboration, USA); and Laura Martino (European Food Safety Authority, Italy). The peer review comments were considered by the drafting group when developing the final manuscript.

The manuscript was edited prior to publication by John Dawson (Nairobi, Kenya). The development of this publication was coordinated by Richard Brown (Chemical Safety and Health Unit, Department of Environment, Climate Change and Health, WHO).



# Abbreviations

<b>COSTER</b>	Conduct of Systematic Reviews in Toxicology and Environmental Health Research
<b>EFSA</b>	European Food Safety Authority
<b>EPA</b>	Environmental Protection Agency
<b>GRADE</b>	Grading of Recommendations Assessment, Development and Evaluation
<b>HAWC</b>	Health Assessment Workspace Collaborative
<b>IARC</b>	International Agency for Research on Cancer
<b>IPCS</b>	International Programme on Chemical Safety
<b>IRIS</b>	Integrated Risk Information System
<b>NTP</b>	National Toxicology Program
<b>NTP OHAT</b>	Office of Health Assessment and Translation of the United States National Toxicology Program
<b>OECD</b>	Organisation for Economic Co-operation and Development
<b>PECO</b>	populations, exposures, comparators, and outcomes
<b>PECOTS</b>	populations, exposures, comparators, outcomes, timings, and settings of interest
<b>PFOA</b>	perfluorooctanoic acid
<b>PFOS</b>	perfluorooctane sulfonate
<b>PRISMA</b>	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
<b>SCiRAP</b>	Science in Risk Assessment and Policy
<b>SYRINA</b>	systematic review and integrated assessment
<b>WHO</b>	World Health Organization





# 1 Purpose

Systematic review is a methodology for identifying, selecting, appraising and synthesizing available evidence to answer a research question, in a way that maximizes transparency and minimizes bias and error in the review process.

Originally developed in the fields of clinical medicine and social sciences, interest in the application of systematic review principles for addressing environmental health issues is growing. Systematic review principles have the potential to advance the rigour and transparency of chemical risk assessments, in particular when the evidence appears to be contradictory, there is significant uncertainty, or different types of evidence need to be brought together effectively.

The overall purpose of this document is to provide guidance to chemical risk assessors who are not currently familiar with systematic review principles and processes via a high-level overview, without being prescriptive or endorsing any existing published systematic review, method or tool.

Knowledge of systematic review principles in institutions undertaking chemical risk assessments varies greatly. This framework is intended to describe the overall process and critical components of a systematic review and how they can be integrated in the chemical risk assessment process to increase rigour and transparency. It provides an introduction to the topic and describes the key steps in a systematic review, first in outline and then in detail, describing important considerations when applying systematic review principles to problems in chemical risk assessment. The framework also addresses the advantages and challenges of performing a systematic review and when it might, or might not, be appropriate to use systematic review in chemical risk assessment.

This framework will assist readers to understand assessments conducted by other institutions that have used systematic review principles. Also, for institutions that may be exploring the use of systematic review in their own assessments, this framework will assist with understanding the issues involved.

The framework was developed by a group of experts working in this field, but from a broad perspective that has been informed by a number of existing frameworks, without endorsing any particular existing framework as a gold standard method. The intention is to inform readers of the basic principles of systematic review rather than recommend the use of any particular framework, process or tool.



# 2 Introduction to systematic review and evidence integration

## 2.1 What is systematic review?

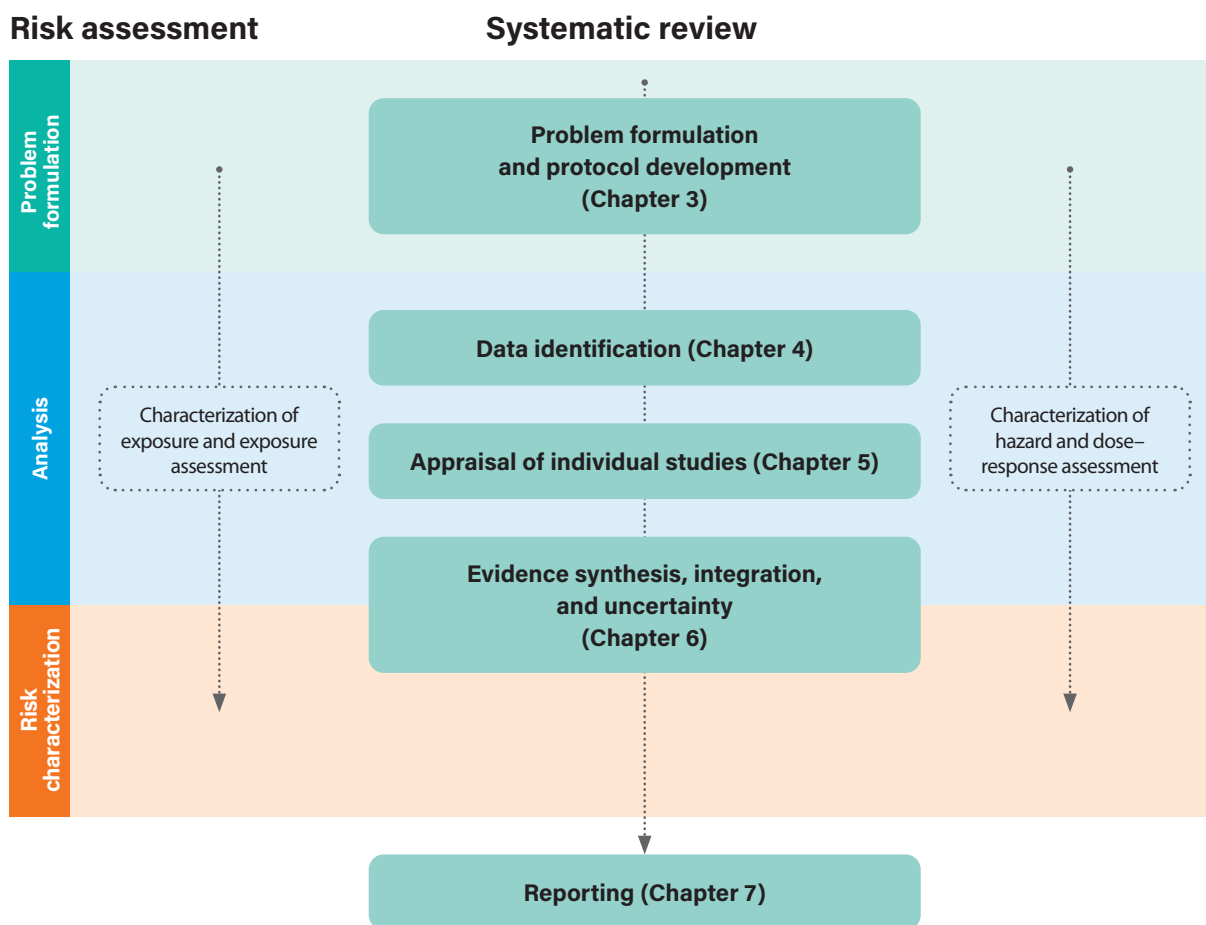
Systematic review is a methodology designed to minimize risk of bias and error and maximize transparency when answering a research question via a review of existing evidence. It is a particularly robust approach to evidence synthesis, that is, the identification, selection, appraisal and synthesis of evidence relevant to answering a research question. Because high-quality evidence synthesis has a fundamental role in chemical risk management, where there is a need to aggregate complex and sometimes contradictory bodies of evidence into statements of what is and is not known in relation to understanding and mitigating the potential health effects of exposure to a chemical substance, the use of systematic review methods is of increasing interest to chemical risk assessors.

In the context of chemical risk assessment (Figure 2.1), systematic review methods can increase the transparency and rigour of each step of the evidence evaluation workflow, supporting:

- problem formulation and protocol development
- answering individual subquestions using existing data
- interpreting results and drawing conclusions in the context of identified uncertainties
- documenting the reasons for each judgement made in the review process
- comprehensive reporting of the methods and results of the risk assessment.

**Systematic review is a scientifically robust approach to interpreting complex, often contradictory evidence into statements of what is already known in relation to a research question.**

**Figure 2.1 How systematic review methods integrate with the overall risk assessment process**



## 2.2 Origins of systematic review

The principles and methods of systematic review are well established in the field of evidence-based medicine (for example, the Cochrane Collaboration<sup>1</sup>) (1), the social sciences (for example, the Campbell Collaboration<sup>2</sup>), and the broader environmental sciences (for example, the Collaboration for Environmental Evidence<sup>3</sup>). A summarized history of the development of systematic review methods from the perspective of health care research is presented in *A brief history of research synthesis*, by Chalmers, Hedges and Cooper (2).

Historically, systematic review methods originated in meta-analysis (a statistical technique for pooling the results of multiple similar studies into a single meta-study). However, by the 1990s it had become apparent that without certain methodological controls, reviews using meta-analytical techniques can produce biased results.

Bias in synthesizing evidence comes from three sources:

- the evidence available for inclusion in the review not being representative of the evidence base as a whole (that is, publication bias);
- insufficient consideration of how limitations in design and conduct of studies included in a review can bias their results, leading to these biases being transmitted through to the final result of a review;
- limitations in design, conduct and reporting of the synthesis itself (for example, selective use of evidence or inappropriate methods for statistical analysis).

<sup>1</sup> <https://www.cochrane.org/>.

<sup>2</sup> <https://campbellcollaboration.org/>.

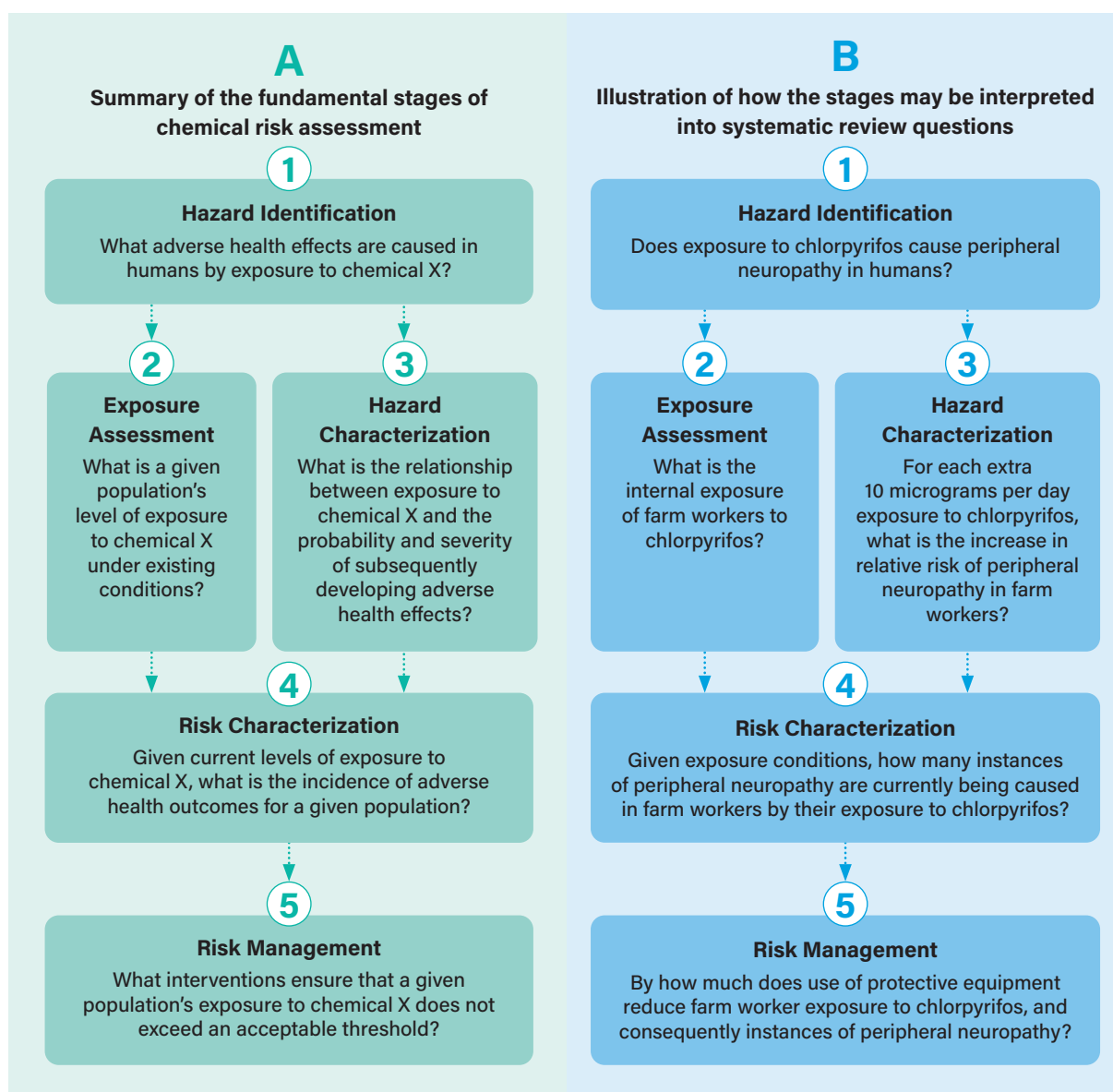
<sup>3</sup> <http://www.environmentalevidence.org/>.

Systematic review was born of the recognition that specific methods were needed to address all three issues and prevent the findings of meta-analyses from appearing more robust than are merited. This gradually led to the distinction being made between a meta-analysis and a "systematic review." Nowadays, the term "meta-analysis" usually refers to the use of statistical methods for pooling the results of individual studies, while "systematic review" refers to the broader processes of minimizing bias in an evidence review.

The methodological rigour associated with high-quality systematic reviews has resulted in their becoming a major scientific reference point across the fields of medicine and social sciences, used to assess the impact of health care interventions and inform the setting of health care and social policies.

The potential for replicating the success of implementing systematic review in medicine has seen an acceleration of interest in applying systematic review methods in the context of chemical risk assessment. The methods are being developed and applied to such issues as food and feed safety, air pollution, noise pollution, and occupational health in relation to chemical exposure. Systematic review questions can be asked at each stage of the risk assessment process (Figure 2.2).

**Figure 2.2 Stages of chemical risk assessment**



A = Archetypal research questions associated with each stage of the risk assessment process

B = Illustrative examples of questions that might be asked in systematic reviews being conducted in support of the risk assessment and risk management of a pesticide

Considerable effort is being invested to promote the development of systematic reviews and their adaptation to the fields of toxicology, chemical risk assessment, exposure sciences, and environmental and occupational epidemiology by research groups, scientific organizations, nongovernment organizations, government agencies and public institutions (for a partial summary, see Box 2.1).

### Box 2.1 Examples of initiatives to utilize systematic review methods in chemical risk assessment processes

The following are examples of initiatives to utilize systematic review methods in chemical risk assessment processes. Illustrative references are provided of relevant documents, publications or websites. Asterisks (\*) indicate initiatives that have published best-practice recommendations or handbooks for conduct of systematic reviews.

**European Food Safety Authority (EFSA).** EFSA is a European Union risk assessment agency with one of the longest histories of engagement with systematic review methods, first publishing on the topic in 2010 (3) and leading methods development across the systematic review workflow (4).

**United States Environmental Protection Agency (EPA).** The United States EPA Integrated Risk Information System (IRIS) programme is engaged in multiple applications of systematic review methods to the hazard assessment and evidence scoping process, with publications including a series of systematic reviews investigating the health effects of phthalate exposure (5).

**World Health Organization (WHO).** WHO has used systematic review methods to evaluate evidence relating to several environmental health challenges, including air pollution (6), noise pollution (7) and nanomaterials (8). Recently, WHO collaborated with the International Labour Organization to conduct the first protocol-based systematic reviews for estimating the global burden of disease attributable to exposure to occupational and environmental risk factors (9, 10).

**Navigation Guide.\*** A research initiative from the University of California San Francisco Program on Reproductive Health and the Environment, in 2014 the Navigation Guide was one of the first published frameworks for systematic review methods in environmental health (11).

**United States National Toxicology Program.\*** The Office of Health Assessment and Translation of the United States National Toxicology Program published its framework and handbook for conduct of systematic reviews in environmental health in 2014, and updated the handbook in 2019 (12).

**Texas Commission for Environmental Quality.** One of the first United States state-level initiatives to adapt systematic review methods to the chemical risk assessment context (13).

**Endocrine Disruption Exchange.** A research-focused environmental health nongovernmental organization that has adopted systematic methods for analysing health risks from chemical exposure (14, 15).

**Evidence-Based Toxicology Collaboration.** A cross-sectoral collaboration based at Johns Hopkins Bloomberg School of Public Health, the Evidence-Based Toxicology Collaboration advocates and develops systematic review methods for toxicological research and chemical risk assessment contexts (16).

**GRADE Working Group.** Initially a medical and public health-focused global network of researchers, GRADE has established a project group to develop systematic review methods for environmental health contexts, including application to chemical risk assessment (17).

**Other initiatives.** A number of cross-sectoral collaborative efforts have been undertaken, including outlining frameworks for the conduct of specific challenges in chemical risk assessment. These include assessment of the endocrine-disrupting potential of chemicals (18)\*, defining the value of systematic review methods specifically in the chemical risk assessment context (19), and recommendations for the conduct of systematic reviews in toxicology and environmental health research (COSTER)\* (20).

An important aspect of the evolution of systematic review methods in chemical risk assessment has been the focus on handling the broad information requirements and wide variety of types of evidence that are characteristic of the chemical risk assessment context. Systematic review methods in medicine have typically



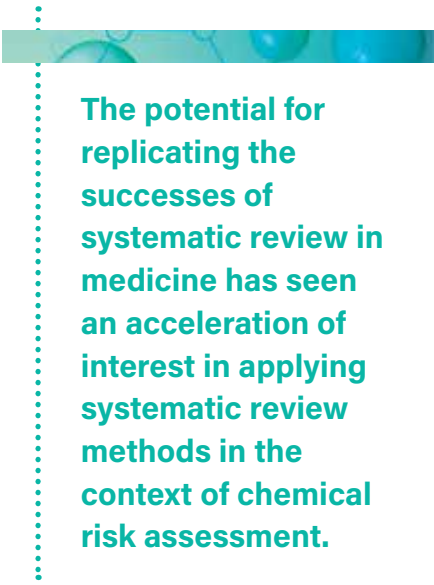
addressed narrow questions with relatively homogeneous evidence (exceptions to this include preclinical and public health research and guideline development). In risk assessment, systematic review methods have been adapted to accommodate all types of risk-relevant data, and particularly to study appraisal, synthesis and integration of evidence from animal experiments, human epidemiological studies, mechanistic studies and exposure data. Here, the systematic review principles of minimized bias and maximized transparency of methods are being applied to refine and improve the chemical risk assessment process.

## 2.3 Why is systematic review of importance in chemical risk assessment?

By its very nature, scientific evidence is heterogeneous, always complex and sometimes conflicting: studies are conducted according to different methods, each of varying relevance to the risk assessment issue in question. Their findings are often inconsistent and subject to varying degrees of bias.

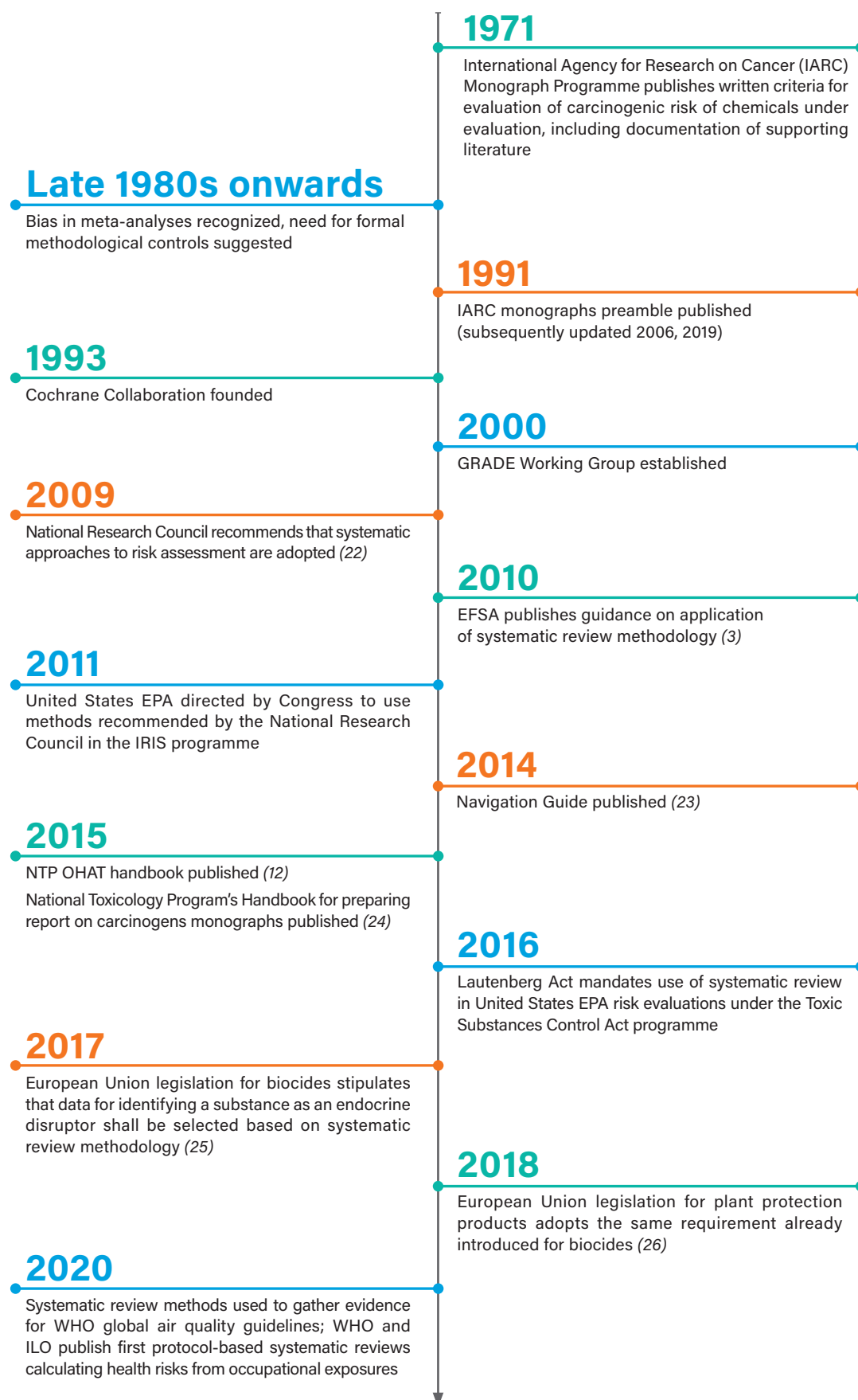
There is already a long history of development of robust methods for handling the complexity of evidence in chemical risk assessment (21, 22). Systematic review methods build on this history, providing a framework for evidence assessment that lends itself well to delivering the transparency, validity, stakeholder confidence in results, utility of outcome, resource efficiency and reproducibility of findings expected of chemical risk assessment (19). The high level of standardization of systematic review methods, a thorough documentation process, and assurance that all relevant primary and other types of information have been identified, selected and compiled mean that the use of systematic review methods constitutes a further advance in the objectivity, transparency and reproducibility of a chemical risk assessment process. A timeline of the development of systematic review methods with respect to chemical assessments is shown in Figure 2.3.

Systematic review outputs are considered as having high potential to support more transparent and robust conclusions in chemical risk assessment practice, to reduce (or at least better characterize) uncertainty, and to inform risk management, regulatory decision-making and policy. In turn, this may facilitate risk communication to all stakeholders, ensure transparency and maintain public trust in the risk analysis process.



**The potential for replicating the successes of systematic review in medicine has seen an acceleration of interest in applying systematic review methods in the context of chemical risk assessment.**

**Figure 2.3 Brief timeline of the development of systematic review methods with respect to chemical assessments**



## 2.4 Expertise and resources required for conducting a systematic review

Systematic reviews can be used in various situations: to assess the state of the science; to identify and describe uncertainties and inconsistencies in the evidence; to analyse and integrate evidence to inform conclusions about potential hazards or risks associated with environmental chemical exposures; to help resolve controversies; to identify knowledge gaps and research needs; and to support recommendations for risk management actions (3, 16, 19, 23, 27).

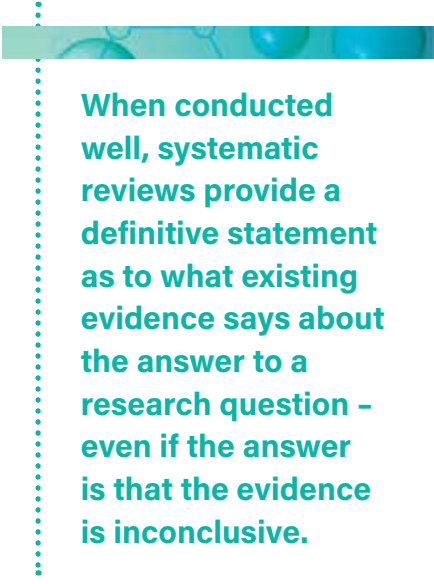
Compared to more classical, narrative toxicological reviews, for a research question of a given scope systematic reviews may require increased time, resources (capacity) and diversity of expertise (capability), covering both topical and methodological expertise. They are a collaborative effort, involving a multidisciplinary team including subject matter experts in toxicology, epidemiology, risk assessment and other scientific disciplines, as well as database and information specialists (who are particularly important for designing effective strategies for locating relevant evidence) and statisticians. While multidisciplinary teams are long familiar in chemical risk assessment, extra resources associated with conducting systematic reviews are potentially necessary because of the high level of methodological rigour and transparency that they entail.

When conducted well, systematic reviews provide a definitive account of what existing evidence says in answer to a research question – even if the definitive account is that the evidence is inconclusive. However, there is not always the time or resources, or even the need, to conduct a full systematic review for each question that may be relevant to a risk management decision. Inevitably, choices have to be made about the scope (how much evidence is to be reviewed?) and completeness (how comprehensive will the analysis be?) of the review in order to achieve useful results within the time and resource constraints imposed by the decision-making context.

A number of factors typically influence the scope and comprehensiveness of a review. They are often interrelated, requiring an iterative decision-making process to ensure each is accounted for. For example, resource availability in particular may be contingent on either or both the importance of the issue to stakeholders and the significance of the consequences of the decision that the review is supporting.

The following factors tend to increase the feasibility and value of a systematic review approach to a chemical risk assessment problem:

- sensitivity of a topic or its importance to stakeholders;
- significant economic, environmental, social, public health or individual consequences of the decisions that the review will inform;
- conflicting evidence or a large degree of uncertainty around a question;
- existence of plentiful resources and expertise for conducting a systematic review (it is worth noting that areas with well defined methods may require fewer resources than answering novel questions);
- sufficient time for purpose, for example when it might be more important to have an unbiased answer after a potentially lengthy period of research rather than an approximate and potentially biased answer in a more urgent context.



**When conducted well, systematic reviews provide a definitive statement as to what existing evidence says about the answer to a research question – even if the answer is that the evidence is inconclusive.**

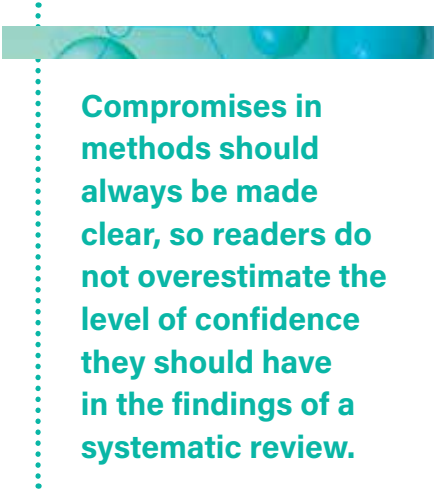
Compromises to make review processes faster and less resource intensive tend to focus on either narrowing the scope of a review, or using less rigorous methods, or both. Narrowing the scope to one or two questions that are critical to decision-making is an effective way of reducing resource requirements while preserving maximum rigour, though it may not be sufficient to accelerate a review process being undertaken within strict time constraints. A risk of narrowing the scope of a review is that it does not provide enough of the information upon which risk managers need to make an informed decision.

Methods for exploring evidence to help with prioritization processes, such as systematic evidence maps (28, 29), are being developed to provide evidence-based approaches to identifying narrow research questions that can support chemical risk management.

There are also rapid review techniques that have been proposed for time-constrained research scenarios (30). These include truncated search strategies, reduced data extraction, limited critical appraisal, and simplified methods for synthesizing results. These should be undertaken cautiously as they will reduce the accuracy of the result of the review and potentially introduce significant bias, for example if evidence is included partially or selectively (see Chapter 4), or critical appraisal processes fail to identify ways in which existing research might be producing misleading results (see Chapter 5).

Key to optimal use of resources is understanding and being transparent about any compromises in methods in conducting a systematic review, and informing users of the review about the consequences of those compromises for the results of the review and any uncertainties they might have introduced.

So as not to mislead readers about the rigour of the methods that have been employed, when review methods do not include a predefined protocol, comprehensive searching, selection process, critical appraisal of individual studies or certainty or strength of evidence assessment, the result should not be referred to as a "systematic review". Transparency should also always be preserved: in the event that compromises have understandably had to be made, these should be made clear to the reader alongside their potential impact on accuracy of results and certainty in conclusions.



**Compromises in methods should always be made clear, so readers do not overestimate the level of confidence they should have in the findings of a systematic review.**

## 2.5 Typical steps of a systematic review

The following are the typical steps taken in conducting a systematic review (the chapters addressing each step are in parentheses):

1. question formulation, taking into consideration resource availability and requirements and stakeholder needs in relation to the knowledge requirements of the situation (Chapter 3);
2. planning and protocol development and publication, in which the plan for answering the question in a transparent, unbiased fashion is developed (Chapter 3);
3. search for evidence, in which a comprehensive strategy for finding all relevant evidence is defined, reducing the risk of selective use of studies (Chapter 4);
4. screening the evidence, in which evidence that is actually relevant to the review question is selected from the search results, reducing risk of selective use of studies (Chapter 4);
5. data extraction, in which all data relevant to the review are abstracted from the included studies, further reducing risk of selective inclusion of data in the review (Chapter 4);

6. critical appraisal of the included evidence, in which the informativeness of a study, given the systematic review objectives and its potential to produce biased results, is evaluated (Chapter 5);
7. synthesis and integration of the evidence, in which findings from the included studies are aggregated into an overall answer to the research question (Chapter 6);
8. assessment of certainty in the findings, in which the strengths and limitations of the overall body of evidence are appraised to determine the credibility of the results of the systematic review (Chapter 6);
9. writing up the systematic review, in which the objectives, methods and results are reported in sufficient detail that the findings of the systematic review can be appraised and contextualized by its readers (Chapter 7).

## References

1. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ et al., editors. Cochrane handbook for systematic reviews of interventions, version 6.0 (updated July 2019). Cochrane; 2019 ([www.training.cochrane.org/handbook](http://www.training.cochrane.org/handbook), accessed 25 April 2021).
2. Chalmers I, Hedges LV, Cooper H. A brief history of research synthesis. *Evaluation and the Health Professions*. 2002;25(1):12–37.
3. European Food Safety Authority. Application of systematic review methodology to food and feed safety assessments to support decision making. *EFSA Journal*. 2010;8(6):1637 (<https://www.efsa.europa.eu/fr/efsajournal/pub/1637>, accessed 25 April 2021).
4. European Food Safety Authority. Scientific report on principles and process for dealing with data and evidence in scientific assessments. *EFSA Journal*. 2015;13(5):4121. doi:10.2903/j.efsa.2015.4121.
5. Radke EG, Yost EE, Roth N, Sathyanarayana S, Whaley P. Application of US EPA IRIS systematic review methods to the health effects of phthalates: lessons learned and path forward. *Environment International*. 2020;145:105820. doi:10.1016/j.envint.2020.105820.
6. WHO guidelines for indoor air quality: household fuel combustion. Geneva: World Health Organization; 2014 (<https://apps.who.int/iris/handle/10665/141496>, accessed 25 April 2021).
7. Jarosińska D, Héroux M-È, Wilkhu P, Creswick J, Verbeek J, Wothge J et al. Development of the WHO Environmental Noise Guidelines for the European Region: An Introduction. *International Journal of Environmental Research and Public Health*. 2018;15(4):813. doi:10.3390/ijerph15040813.
8. WHO guidelines on protecting workers from potential risks of manufactured nanomaterials. Geneva: World Health Organization; 2017 (<https://apps.who.int/iris/bitstream/handle/10665/259671/9789241550048-eng.pdf>, accessed 25 April 2021).
9. Pega F, Náfrádi B, Momen N, Ujita Y, Streicher K, Prüss-Üstün A et al. Global, regional, and national burdens of ischemic heart disease and stroke attributable to exposure to long working hours for 194 countries, 2000–2016: A systematic analysis from the WHO/ILO Joint Estimates of the Work-related Burden of Disease and Injury. *Environment International*. 2021;154:106595. doi:10.1016/j.envint.2021.106595.
10. Pega F, Momen N, Ujita Y, Driscoll T, Whaley P. Systematic reviews and meta-analyses for the WHO/ILO Joint Estimates of the Work-related Burden of Disease and Injury. *Environment International*. 2021;155:106605. doi:10.1016/j.envint.2021.106605.
11. Navigation Guide: an evidence-based medicine methodology to bridge the gap between clinical and environmental health sciences. University of California San Francisco, Program on Reproductive Health and the Environment (<https://prhe.ucsf.edu/navigation-guide>, accessed 25 April 2021).

12. Handbook for conducting systematic reviews for health effects evaluations. United States National Toxicology Program; 2019 (<https://ntp.niehs.nih.gov/go/ohathandbook>, accessed 25 April 2021).
13. Schaefer HR, Myers JL. Guidelines for performing systematic reviews in the development of toxicity factors. *Regulatory Toxicology and Pharmacology*. 2017;91:124–41.
14. Rochester JR, Boldena AL, Kwiatkowski CF. Prenatal exposure to bisphenol A and hyperactivity in children: a systematic review and meta-analysis. *Environment International*. 2018;114:343–56.
15. Pelch KE, Reade A, Wolffe TAM, Kwiatkowski CF. PFAS health effects database: protocol for a systematic evidence map. *Environment International*. 2019;130:104851.
16. Hoffmann S, de Vries RBM, Stephens ML, Beck NB, Dirven HAAM, Fowle JR 3rd et al. A primer on systematic reviews in toxicology. *Archives of Toxicology*. 2017;91:2551–75.
17. Morgan RL, Beverly B, Ghersi D, Schünemann HJ, Rooney AA, Whaley P et al. GRADE guidelines for environmental and occupational health: a new series of articles in *Environment International*. *Environment International*. 2019;128:11–12. doi:10.1016/j.envint.2019.04.016.
18. Vandenberg LN, Ågerstrand M, Beronius A, Beausoleil C, Bergman A, Bero LA et al. A proposed framework for the systematic review and integrated assessment (SYRINA) of endocrine disrupting chemicals. *Environmental Health*. 2016;15:74. doi:10.1186/s12940-016-0156-6.
19. 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. *Environment International*. 2016;92–93:556–64.
20. Whaley P, Aiassa E, Beausoleil C, Beronius A, Bilotta G, Boobis A et al. Recommendations for the conduct of systematic reviews in toxicology and environmental health research (COSTER). *Environment International*. 2020;143:105926.
21. National Research Council. Risk assessment in the federal government: managing the process. Washington (DC): National Academy Press; 1983.
22. National Research Council. Science and decisions: advancing risk assessment. Washington (DC): National Academy Press; 2009.
23. Woodruff TJ, Sutton P. The Navigation Guide systematic review methodology: a rigorous and transparent method for translating environmental health science into better health outcomes. *Environmental Health Perspectives*. 2014;122(10):1007–14. doi: 10.1289/ehp.1307175.
24. Handbook for preparing report on carcinogens monographs. United States Department of Health and Human Services, National Toxicology Program; 2015 (<https://ntp.niehs.nih.gov/whatwestudy/assessments/cancer/handbook/index.html>, accessed 25 April 2021).
25. Commission Delegated Regulation (EU) 2017/2100 of 4 September 2017 setting out scientific criteria for the determination of endocrine-disrupting properties pursuant to Regulation (EU) No 528/2012 of the European Parliament and Council. Brussels: European Commission; 2017.
26. Commission Regulation (EU) 2018/605 of 19 April 2018 amending Annex II to Regulation (EC) No 1107/2009 by setting out scientific criteria for the determination of endocrine disrupting properties. Brussels: European Commission; 2018.
27. Rooney AA, Boyles AL, Wolfe MS, Bucher JR, Thayer KA. Systematic review and evidence integration for literature-based environmental health science assessments. *Environmental Health Perspectives*. 2014;122:711–8.

28. Walker VR, Boyles AL, Pelch KE, Holmgren SD, Shapiro AJ, Blystone CR et al. Human and animal evidence of potential transgenerational inheritance of health effects: an evidence map and state-of-the-science evaluation. *Environment International*. 2018;115:48–69.
29. Wolffe TAM, Whaley P, Halsall C, Rooney AA, Walker VR. Systematic evidence maps as a novel tool to support evidence-based decision-making in chemicals policy and risk management. *Environment International*. 2019;130:104871.
30. Tsertsvadze A, Chen YF, Moher D, Sutcliffe P, McCarthy N. How to conduct systematic reviews more expeditiously? *Systematic Reviews*. 2015;4:160.







# 3 Problem formulation and protocol development

## 3.1 Introduction

Problem formulation is the first step in a systematic review process in which the purpose and scope of the review are explicitly defined through careful planning. Protocol development is the second step in the systematic review process and involves a priori documentation of the approach and methods to be used to search for, select, and appraise individual studies, synthesize and assess the body of evidence, and develop conclusions. Collectively, these processes establish the “what” and the “how” of the overall risk assessment and of the components of the risk assessment that will be evaluated using systematic review (see Figure 3.1 below).

## 3.2 Problem formulation

Problem formulation is recognized as the first step in both a systematic review and a chemical risk assessment (1). In the context of chemical risk assessment, problem formulation is described as the process by which the assessment is defined and the plan for characterizing risk is developed (2, 3) – a process that involves dialogue to clarify management goals, the purpose and scope of the assessment, and the resources available to conduct the assessment (1). When utilizing systematic review to help facilitate the chemical risk assessment, problem formulation is also the planning step that specifically involves characterizing which components of the risk assessment (for example, hazard identification, exposure assessment) will be conducted using systematic review, and how such methods can aid in characterization of uncertainty across a chemical risk assessment.

In the context of systematic review, problem formulation (also referred to as question formulation) involves characterizing and refining the purpose and goals based on considerations of background knowledge, the available evidence base, time frame, and resource availability. It starts with defining the goal or objectives of the systematic review, which in turn is dependent on the context of the intended application. During problem formulation, factors to be considered include purpose of the review, goals of the organization, volume and breadth of the underlying evidence base, established knowledge of the chemical in question, uncertainty or controversy around the topic, depth of assessment required to meet objectives, and resources and time available to conduct the review. The dialogue encountered may involve iterative consideration of the guiding questions in Box 3.1. In practice, this iterative dialogue may frequently involve limiting the goals and objectives to meet available resources by relying on existing authoritative or systematic assessments for some aspects, thus allowing the goal of the systematic review to be very focused on a specific topic.

### Box 3.1 Questions faced during problem formulation

The following questions may be faced during problem formulation.

- What is the problem and the specific context of the assessment?
- What is the overall risk management goal?
- What is the depth (scope) and breadth of analysis required to provide adequate information to address the risk management goal?
- What are the timelines and what resources are available?
- What is the most suitable assessment methodology (systematic review, scoping review, evidence map, mixed review, conduct primary research studies)?
- What are the anticipated uses of the review?

A rigorous, well planned problem formulation process is a major factor in ensuring that an assessment project yields a successful result. It helps ensure that a systematic review targets, in a resource-efficient way, the critical information needed for informing risk management decisions, and helps ensure that the results of the review process are credible. Moreover, it helps with ensuring, via suitable engagement with stakeholders in the planning process, that they accept the findings of the review process. It is recommended that problem formulation is carried out by a multidisciplinary team also involving risk managers to ensure that the scope of the review addresses management needs. Importantly, the problem formulation process should be transparent and documented (often included as the rationale or background in the protocol, as discussed below).

#### 3.2.1 Structured format for the research question

In systematic review, the outcome of question formulation is a defined research question, often in the form of a PECO (populations, exposures, comparators, and outcomes) question. In some situations, the research question may take the form of a PECOTS (populations, exposures, comparators, outcomes, timings, and settings of interest) question, or may be modified to only selected components to address the needs of the systematic review (for example, only population and exposure evaluated in a systematic review on exposure prevalence). In the case of chemical risk assessment, often multiple PECO questions are required to accommodate the various components of risk assessment, and comparators are generally inverse scenarios of the exposure (for example, exposure to chemical X compared to lack of exposure to chemical X or low exposure to chemical X). Subquestions can also be developed to facilitate evaluation of contextual topics that are important to decision-making related to key questions. Examples of PECO questions related to evaluation of hazard in environmental health are provided by Morgan et al. (4). Exposure and risk-based PECO questions (and similar) are demonstrated by EFSA (5).

#### PECO:

**P – populations**

**E – exposures**

**C – comparators**

**O – outcomes**

#### 3.2.2 Developing focused (narrow) research questions for systematic review

It is recommended that systematic review topics are as narrow and focused as possible. A primary component of problem formulation is thus identification of a specific topic in the context of other aspects identified in Box 3.1. Identifying a narrow and specific topic will help to limit the resources required to conduct the review. In practice, topics are often selected on the basis of mandate, nomination of a substance to an agency for

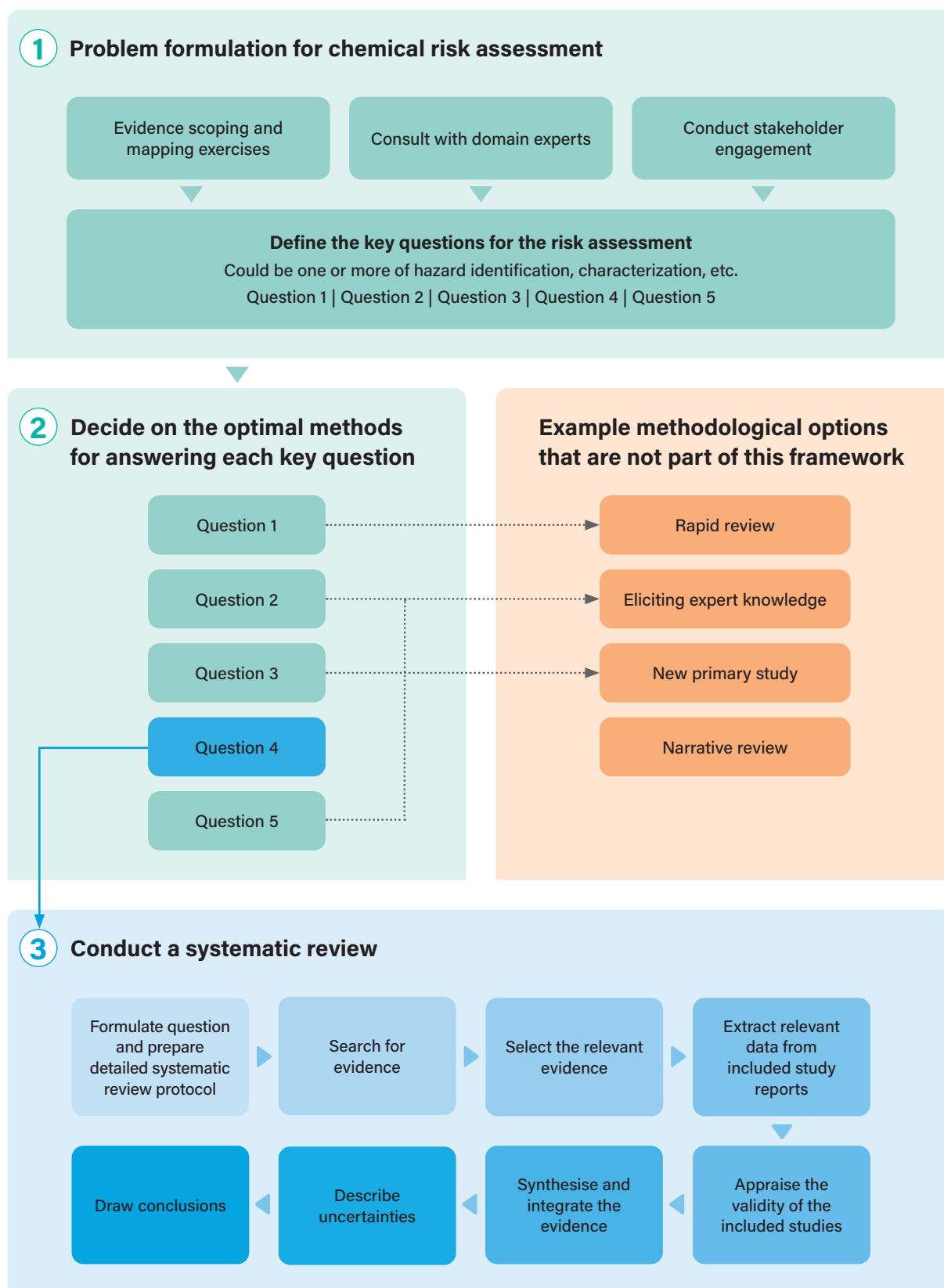
evaluation, or degree of controversy, though in all cases, refinement of the topic is important to developing targeted research questions for systematic review. As a first step in refining or prioritizing a topic, a broad search of the literature should be conducted as a means of characterizing the evidence base. Documents, such as reviews or regulatory assessments, should be surveyed to determine key issues and potential data gaps, and registries and repositories searched to identify potentially ongoing systematic reviews of similar topics. In establishing this background knowledge, experts in the topic should also be consulted. If prior systematic reviews have been conducted, the potential for an update to those reviews (pending availability of newer data) can also be considered. If feasible, systematic maps or scoping reviews (see Chapter 2) can be conducted prior to the systematic review to characterize the landscape of available literature; these exercises are regarded as a best-practice method for providing a comprehensive map of evidence and identifying data gaps, thus facilitating decision-making in determining which topics to carry forward to systematic review (6, 7). For an example of how mapping of evidence can be used for determining systematic review topics, see Lohner, Toews and Meerpohl (8). If systematic maps are not available or cannot be conducted, tabular summaries relevant to the possible evidence base, such as by evidence stream or end-points within an outcome, can help to facilitate the prioritization process. Various examples of how a research question can be focused or narrowed around specific elements of the PECO include the following (4).

- **Populations.** These could be narrowed to sensitive or target populations, such as pregnant women or children (and experimental models of such).
- **Exposures.** These could be narrowed by exposure scenario (for example, occupational), route (for example, inhalation), dose ranges (for example, relevance to environmental exposures) or timing (for example, developmental window).
- **Comparators.** These could be narrowed to only non-exposed or a specific level associated with low exposure.
- **Outcomes.** These could be narrowed to a subset of outcomes (for example, cancer, developmental, specific cancer types, specific developmental end-points).

Because it may not be necessary or feasible to conduct a systematic review for every health outcome or every exposure scenario in a chemical risk assessment – particularly for a data-rich chemical – the problem formulation process will often involve both identification and prioritization of topics. During this stage, it may also be determined that for some topics, approaches other than systematic reviews (such as rapid reviews or systematic maps, carrying out primary research, or eliciting expert knowledge) may be more optimal for some components of chemical risk assessment (Figure 3.1). Thus, it may be determined that not all available data associated with a chemical risk assessment will be reviewed in a systematic fashion, but rather only specific PECO combinations that are of greatest priority are selected to be included in a systematic review, and those that have sufficient data upon which to conduct a systematic review (as data-poor scenarios may not be suitable for systematic review). No prescriptive method or criteria exist for prioritizing, though reasonable considerations include quantitative (versus qualitative) analyses, biological relevance, exposure relevance, degree of controversy, organization goals, review purpose and resource considerations. Some criteria for prioritizing questions for systematic review are provided by EFSA (5).

Figure 3.1 provides a conceptual demonstration of how a general problem formulation conducted for risk assessment may involve multiple questions, some of which may be addressed with systematic review, and other questions using different approaches. Thus, systematic review may alone, or in combination with other approaches, be used to address the overarching risk assessment issue. In the figure, the steps of the systematic review process discussed in this guidance are also illustrated. Here, the systematic review approach is presented in response to a single question; in Chapter 6, a figure is presented in which the systematic review findings may be aggregated with findings from multiple analyses, some using other methods, into an overall risk assessment result.

**Figure 3.1 Conceptual demonstration of how a general problem formulation conducted for risk assessment may involve multiple questions**

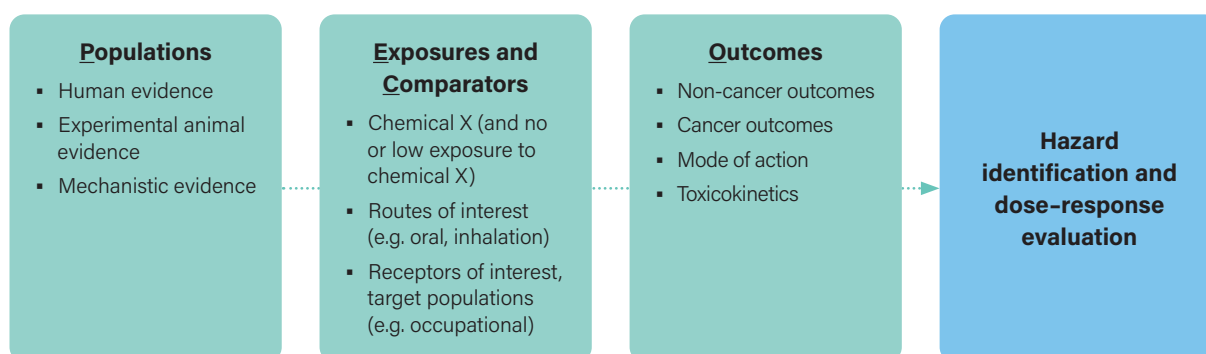


### 3.2.3 Use of conceptual models to inform planning, scoping and framing the research questions

Logic models and analytical frameworks (often referred to as conceptual models) have been used in systematic review practice for a more structured and transparent approach to problem formulation (9–13), as well as in risk assessment (14, 15). During the early stages of the review, conceptual models can help reduce the complexity of the problem at hand, identify conceptual boundaries, focus the analysis on specific questions of interest, and highlight knowledge gaps and areas of uncertainty; they can also inform on the degree of stakeholder engagement and on the type of expertise in the systematic review team that are necessary to answer the review questions.

Conceptual models can be particularly useful to support framing the review. Roth, Sandström and Wilks (13) developed a logic, pathway-oriented model to facilitate the breakdown of the various risk assessment dimensions (hazard, exposure and risk) for framing primary and secondary PECO questions. The PECO can be expressed as a visual figure that lays out the chain of logic, thereby clearly providing linkages among the populations of interest, exposures, and outcomes of interest, as well as contextual factors that may influence the linkages (13, 16, 17) (see example in Figure 3.2). Because of their versatility, conceptual models have been proposed as central decision, prioritization and communication tools throughout the whole systematic review process (10).

**Figure 3.2 Demonstrative example of a PECO-based analytical framework for evaluation of hazard in a risk assessment**




### 3.2.4 Important considerations for exposure during problem formulation

Regardless of the scope of the assessment, the concept of exposure should be addressed in problem formulation. Depending on the context of the assessment, the exposure component of risk assessment may be directly evaluated using systematic review or may be included contextually as part of evaluating or prioritizing hazard and dose–response data (18, 19). Exposure assessment typically requires consideration of all relevant exposure information, including sources, media, and routes of exposure; exposure duration and frequency; and external and internal exposure pathways (absorption, distribution, metabolism and excretion (ADME)/toxicokinetics). Exposure metrics may include real-life exposure data such as biomonitoring (for example, blood levels), food consumption and occurrence data, or modelled data (when, for example, real-life human exposure data are not available). There is a need to manage the overall complexity of exposure data in terms of feasibility and what priority such information should have for being included in the review. Conceptual models can inform which exposure information is critical to the review and should be prioritized in a chemical risk assessment context (13).

### 3.2.5 Challenges of applying systematic reviews to chemical risk assessments

It should be recognized that there are areas of chemical risk assessment that are less well-developed in the context of systematic review tools and processes and thus more challenging. These aspects warrant particular recognition with regard to feasibility, as in many cases methods or tools may need to be significantly refined in order to include them in the systematic review (20, 21). Examples of these aspects are the inclusion and use of mechanistic and pharmacokinetic data as part of hazard and dose–response; lack of widely accepted tools for appraising and integrating exposure data; and a general void of methods or applications for chemical risk assessment in ecological risk assessment.

Regarding mechanistic and pharmacokinetic data (two separate types of data), these are often described as being contextual or associated with subquestions. However, if these data are anticipated to be critical in making determinations of hazard or dose–response, such data should be included in the review and subjected to the same appraisal and structured evaluation process applied for other data types. Both mechanistic data and pharmacokinetic data can be generated from different study types (such as epidemiological studies or in vitro studies), and thus should not be confused with evaluation of evidence streams (human, animal or mechanistic). Chapter 6 contains additional discussion regarding mechanistic data, which should be considered at the problem formulation stage. Specifically, the approach for identification, organization, and assessment of mechanistic data needs to be considered in problem formulation. Given the large and growing number of publications with mechanistic information, it is very useful to consider these data stepwise, depending on the human or animal evidence identified (for example, if human evidence of nephrotoxicity is identified, then data on potential mechanisms for nephrotoxicity will be identified and assessed). As part of determining inclusion of mechanistic data, it is important to consider that critical appraisal tools and structured approaches for integrating mechanistic findings via systematic review in the context of chemical risk assessment are largely unavailable or underdeveloped for evaluating exposure information in a systematic review.



**Challenging areas of  
applying systematic  
review to chemical  
risk assessment:**

**Mechanistic data**

**Exposure data**

**Toxicokinetic data**

**Ecological data**

## 3.3 Protocol development

### 3.3.1 Developing and publishing a protocol

Once the review question is well defined, a specific a priori approach is formalized via development of a systematic review protocol. The protocol provides specific information as to how the systematic review will be carried out (that is, the methods for conducting the review). The protocol should be developed by a multidisciplinary team with expertise in the subject matter (including toxicology, epidemiology, statistical analysis, exposure sciences and risk assessment expertise) as well as a systematic review methodologist and an information specialist. Planning the methods upfront in a protocol safeguards against arbitrary decision-making during the assessment process and protects from cognitive biases, as the outcomes of the available studies are not yet known when the methods are defined (22). A protocol should be developed regardless of the plans for dissemination. However, as part of best practices, the protocol should be published or registered in a fashion that is viewable to the public. There are many benefits of making a protocol available prior to initiation of a review (Box 3.2).

### Box 3.2 Benefits of making a protocol publicly available

The benefits of making a protocol publicly available include:

- provides an opportunity for peer review and stakeholder comment to improve quality and increase confidence of end users (23);
- avoids unnecessary duplication of efforts;
- reduces the likelihood of changes made ad hoc that may bias the review outcomes (selective reporting) (24);
- promotes best systematic review practice by providing transparency in the review process and outcomes;
- promotes accountability, research integrity and consistency in conducting the work among the review team.

Registration can be achieved via open-access databases of systematic review protocols such as PROSPERO, and a variety of platforms exist for simply making the protocol available to the public (for example, Open Science Framework,<sup>4</sup> EQUATOR Network,<sup>5</sup> Zenodo,<sup>6</sup> EFSA Knowledge Junction<sup>7</sup>). Publishing protocols in a peer-reviewed journal is becoming more common in environmental and occupational health as well as toxicology. Protocol publication is useful for two main reasons: (a) by committing to a written plan before a research project begins, the risk that a group's expectations regarding the results of a review will affect their decisions about including and interpreting evidence is reduced; and (b) publication provides an opportunity for external peer review of proposed methods, which helps with early identification of issues that could undermine the credibility or value of a systematic review.

### 3.3.2 Elements of a systematic review protocol

Though there is flexibility in the depth of information provided in a protocol, certain common elements can be identified (5, 16, 19, 25). As summarized in Chapter 7, required elements broadly include:

- review question and associated context and rationale, including motivation to conduct the work;
- methods for searching for, selecting and extracting data;
- methods for appraising individual studies;
- methods for synthesizing and integrating the body of evidence;
- methods for assessing the reliability of the final result and for developing conclusions.

With respect to the methods for selecting evidence, the protocol should include the search strategy (including databases and search syntax), which should be developed with an information specialist (26); inclusion and exclusion criteria (including study screening and selection criteria); and the procedure for carrying out the selection of evidence. Protocols should specifically address how each data source will be handled, in particular unpublished study reports such as those submitted by applicants for regulatory compliance, as well as other grey literature and secondary literature sources such as reviews (see Chapter 4).

Similarly, the methods and process for extracting evidence should be well described, including development of extraction templates and specific methods to evaluate study validity (for example, the type of critical appraisal tool that will be used to assess risk of bias in the included studies, and methods to assess other aspects of validity). The methods for evaluating the body of evidence should be well characterized, including both

<sup>4</sup> OSF registries: part of the Open Science Framework <https://osf.io/registries>.

<sup>5</sup> EQUATOR Network: Enhancing the QUALity and Transparency Of health Research <http://www.equator-network.org/>.

<sup>6</sup> Zenodo: open-access resource funded by the European Commission <https://zenodo.org/>.

<sup>7</sup> EFSA Knowledge Junction <https://zenodo.org/communities/efsa-kj/>.



quantitative and qualitative synthesis strategies. In addition, the protocol should identify team members and roles, provide a process for identifying and resolving disagreements between reviewers, describe procedures for disclosing and managing conflict of interest, and provide a projected timeline. Changes to the protocol during the conduct of the review should be well documented; the rationale and potential impact on the review due to changes should be discussed in the reporting phase (see Chapter 7).

## References

1. International Programme on Chemical Safety. Human health risk assessment toolkit: chemical hazards. Harmonization Project Document No. 8. Geneva: World Health Organization; 2010 (<https://www.who.int/publications/i/item/9789241548076>, accessed 26 April 2021).
2. Framework for ecological risk assessment. EPA/630/R-92/001. Washington (DC): United States Environmental Protection Agency; 1992.
3. International Programme on Chemical Safety. Integrated risk assessment. Report prepared for the WHO/ UNEP/ILO International Programme on Chemical Safety. Chapter II: Framework for the integration of health and ecological risk assessment. Geneva: World Health Organization; 2001 ([https://www.who.int/ipcs/publications/en/ch\\_2.pdf?ua=1](https://www.who.int/ipcs/publications/en/ch_2.pdf?ua=1), accessed 26 April 2021).
4. Morgan RL, Whaley P, Thayer KA, Schunemann HJ. Identifying the PECO: a framework for formulating good questions to explore the association of environmental and other exposures with health outcomes. *Environment International*. 2018;121(Pt 1):1027–31.
5. European Food Safety Authority. Application of systematic review methodology to food and feed safety assessments to support decision making. *EFSA Journal*. 2010;8(6):1637 (<https://www.efsa.europa.eu/fr/efsajournal/pub/1637>, accessed 25 April 2021).
6. James KL, Randall NP, Haddaway NR. A methodology for systematic mapping in environmental sciences. *Environmental Evidence*. 2016;5(7):13.
7. Wolffe TAM, Whaley P, Halsall C, Rooney AA, Walker VR. Systematic evidence maps as a novel tool to support evidence-based decision-making in chemicals policy and risk management. *Environment International*. 2019;130:104871.
8. Lohner S, Toews I, Meerpohl JJ. Health outcomes of non-nutritive sweeteners: analysis of the research landscape. *Nutrition Journal*. 2017;16:55.
9. Anderson LM, Petticrew M, Rehfuss E, Armstrong R, Ueffing E, Baker P et al. Using logic models to capture complexity in systematic reviews. *Research Synthesis Methods*. 2011;2:33–42.
10. Kneale D, Thomas J, Harris K. Developing and optimising the use of logic models in systematic reviews: exploring practice and good practice in the use of programme theory in reviews. *PLoS One*. 2015;10:e0142187.
11. Harris JL, Booth A, Cargo M, Hannes K, Harden A, Flemming K et al. Cochrane qualitative and implementation methods group guidance series, paper 2: methods for question formulation, searching, and protocol development for qualitative evidence synthesis. *Journal of Clinical Epidemiology*. 2018;97:39–48.
12. Rehfuss EA, Booth A, Brereton L, Burns J, Gerhardus A, Mozygemba K et al. Towards a taxonomy of logic models in systematic reviews and health technology assessments: a priori, staged, and iterative approaches. *Research Synthesis Methods*. 2018;9:13–24.
13. Roth N, Sandström J, Wilks MF. A case study applying pathway-oriented thinking to problem formulation for planning a systematic review. *Environment International*. 2020;140:105768.



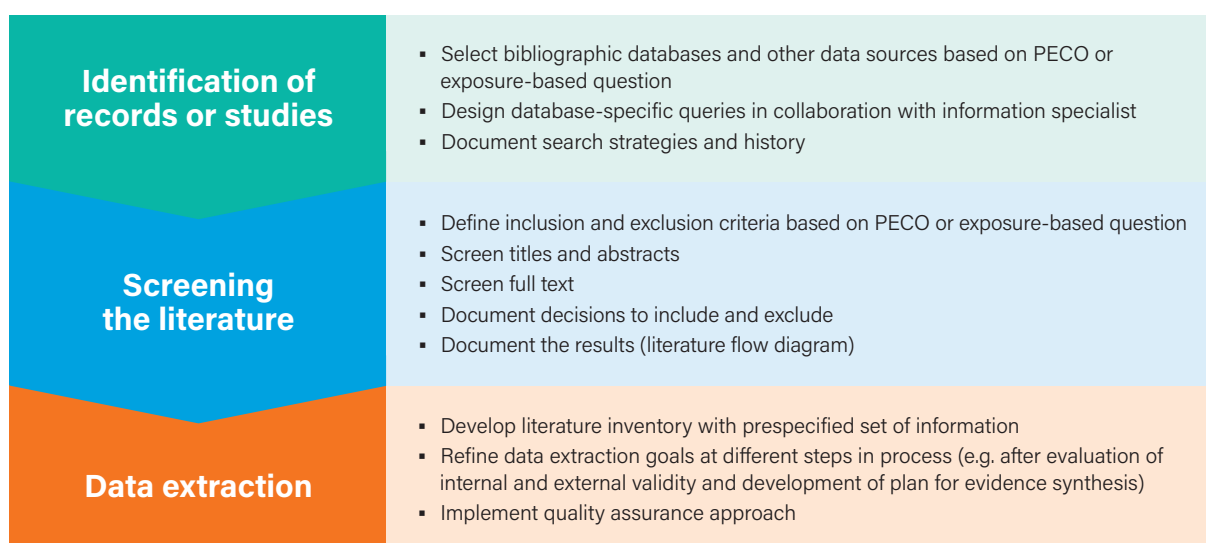
14. Rhomberg LR, Goodman JE, Bailey LA, Prueitt RL, Beck NB, Bevan C et al. A survey of frameworks for best practices in weight-of-evidence analyses. *Critical Reviews in Toxicology*. 2013;43(9):753–84.
15. Solomon KR, Wilks MF, Bachman A, Boobis A, Moretto A, Pastoor TP et al. Problem formulation for risk assessment of combined exposures to chemicals and other stressors in humans. *Critical Reviews in Toxicology*. 2016;46:835–44.
16. Institute of Medicine. Finding what works in healthcare: standards for systematic reviews. National Academies Press; 2011.
17. Aiassa E, Higgins JPT, Frampton GK, Greiner M, Afonso A, Amzal B et al. Applicability and feasibility of systematic review for performing evidence-based risk assessment in food and feed safety. *Critical Reviews in Food Science and Nutrition*. 2015;55(7):1026–34. doi:10.1080/10408398.2013.769933.
18. Procedure manual. Rockville, MD: United States Preventive Services Task Force, Program Office; 2015 (<https://www.uspreventiveservicestaskforce.org/Page/Name/procedure-manual>, accessed 26 April 2021).
19. Handbook for conducting a literature-based health assessment using OHAT approach for systematic review and evidence integration. United States Department of Health and Human Services, National Toxicology Program, Office of Health Assessment and Translation; 2019 update ([https://ntp.niehs.nih.gov/ntp/ohat/pubs/handbookmarch2019\\_508.pdf](https://ntp.niehs.nih.gov/ntp/ohat/pubs/handbookmarch2019_508.pdf), accessed 26 April 2021).
20. 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. *Environment International*. 2016;92–93:556–64.
21. Wikoff D, Urban JD, Harvey S, Haws LC. Role of risk of bias in systematic review for chemical risk assessment: A case study in understanding the relationship between congenital heart defects and exposures to trichloroethylene. *International Journal of Toxicology*. 2018;37(2):125–143. doi: 10.1177/1091581818754330.
22. European Food Safety Authority, Martino L, Aiassa E, Halldórsson TI, Koutsoumanis PK, Naegeli H et al. Draft framework for protocol development for EFSA's scientific assessments. EFSA supporting publication 2020:EN-1843. doi:10.2903/sp.efsa.2020.EN-1843.
23. Stewart L, Hoher D, Shekelle P. Why prospective registration of systematic reviews makes sense. *Systematic Reviews*. 2012;1:7.
24. Kirkham JJ, Altman DG, Williamson PR. Bias due to changes in specified outcomes during the systematic review process. *PLoS ONE*. 2010;5(3):e9810.
25. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews*. 2015;4:1.
26. National Research Council. Review of the Environmental Protection Agency's Integrated Risk Information System (IRIS) process. Washington (DC): National Academies Press; 2014.



# 4 Search, selection and extraction

This chapter describes the process for identifying, retrieving and selecting studies relevant to the research question and for extracting data from those studies (illustrated in Figure 4.1). The approach to this step is defined by the research question developed during the scoping and problem formulation steps and detailed in the PECO statement or other statement relevant to the topic (for example, PECOTS). The search strategy should be developed with the input of information specialists knowledgeable in the use of electronic bibliographic databases and the development and testing of search strings and strategies, in conjunction with review authors. The records that result from the literature search strings are then screened using the prespecified criteria. Subsequently, study information and data are extracted from the selected studies. Guidance for searching, selecting and extracting data for systematic reviews in the context of chemical risk assessment has been published and is generally similar across methods (1–6).

**Figure 4.1 Search, selection and data extraction process flow**



## 4.1 Identification of records or studies

The goal of the search and selection steps of a systematic review is to obtain as many potentially relevant records as possible, in a way that is not selective, and to minimize the inclusion of non-relevant records in a manner that is both manageable and reproducible. The efficiency of the process depends on optimizing the approaches used in initial searching and screening stages. Collaboration with information specialists (such as research librarians) is essential to striking the right balance. In some cases, the search strategy used in other published systematic reviews may be applied or adapted because the research questions are on the same or a similar topic. For example, a published systematic review of the health effects related to exposure to a chemical in the environment or workplace may be consulted to develop a search strategy for a different chemical and the same health effects. Because the goal of the review is to identify a comprehensive set

of relevant records, it is common that a large percentage of the retrieved records are ultimately excluded because they are not considered relevant.

The components in a literature search process involve selecting information sources; developing, testing and conducting the search strategy for each source (or supplementing the literature search if required); documenting the search strategy and the process, including revisions to or deviations from the process; and updating the literature search. Prespecified criteria are instrumental in the identification of information sources, in development of the search strategies for each source, and in the subsequent screening step.

#### 4.1.1 Criteria for identifying, retrieving and screening records

Criteria are developed based on the PECO statement specifying the populations, exposures, comparators, and outcomes relevant to the research question. Additional requirements may also be used on a case-by-case basis for the types of study designs, methodologies, or other study attributes determined relevant to the review. Inclusion and exclusion criteria may also involve whether to include publication dates, publications in non-native languages, studies from sources without peer review, grey literature, citations with no abstract, or other sources. Reasons for excluding records should be documented and records may be sorted into categories that reflect exclusion decisions. Abstract-only citations (that is, abstracts without a full-text publication, such as conference proceedings) may be useful to identify primary research articles that are in press or have not been published. Review articles and other relevant material may be binned in a category for background information.

#### 4.1.2 Sources of literature

The specific bibliographic databases and other data sources to use for the search will be determined by the focus of the review (for example, health effects, exposure, other). PubMed, Web of Science, Scopus, and Embase are commonly used sources for reviews related to chemical exposure and health (Table 4.1). These overlapping “academic” databases of journals focus on medical and life science, social science, and toxicology literature; the use of multiple sources increases the probability that as many relevant records as possible are identified. To a limited extent, these databases may identify some grey literature.

**Table 4.1 Examples of bibliographic databases relevant to systematic reviews on chemical risk assessment health topics**

Databases	Characteristics
PubMed (United States National Library of Medicine)	More than 27 million citations from the biomedical literature from MEDLINE, life science journals and online books
Web of Science (Clarivate Analytics)	Multidisciplinary research platform provides access to multiple databases containing citations for journals and conference proceedings in the sciences, social sciences, arts and humanities
Scopus (Elsevier)	Abstract and citation database of peer-reviewed literature, including scientific journals, books and conference proceedings in science, technology, medicine, social sciences, arts and humanities
Embase (Elsevier)	Biomedical database covering journal articles and conference abstracts from 1947 indexed using Embase Indexing and Emtree, Elsevier's Life Science Thesaurus
Chemical or toxicological databases	Examples include the Organisation for Economic Co-operation and Development (OECD) eChemPortal, INCHEM and PubChem
Grey literature sources	Examples include GreyNet, OpenGrey, the Grey Literature Report, and the interagency Science.gov Alliance

Grey literature (Box 4.1) includes document types that are found outside the scientific journals, such as governmental and nongovernmental technical reports (for example, risk assessments, scientific opinions), conference proceedings, book chapters or dissertations, as well as many other types of secondary literature.<sup>8</sup> Government documents have often been developed using a rigorous internal and external peer review process<sup>9</sup> and are likely to be important sources when conducting a systematic review in a chemical risk assessment context. Indeed, these information sources can constitute a crucial entry portal to the regulatory and scientific literature, for example at the problem formulation stage, as they may contribute information on technical or research activities and knowledge gaps for the topic of interest. Including grey literature searches can be challenging but can result in broadening the scope of the review, amongst other advantages (7). Some limitations that need to be considered and addressed on a case-by-case basis may include lack of or inadequately documented quality assurance methods for data collection, lack of peer review and potential conflicts of interest.

#### Box 4.1 Grey literature

Grey literature is a field in library and information science that deals with the production of, distribution of, and access to multiple document types produced on all levels of government, academics, business, and organization in electronic and print formats not controlled by commercial publishing, i.e. where publishing is not the primary activity of the producing body.

*Source:* GreyNet International.

### 4.1.3 Development and testing of search queries

Search terms can be identified using previously published reviews or assessments from government institutions, indexing terms used in bibliographic databases (such as PubMed MeSH terms), and text words from a set of relevant primary research publications identified during the problem formulation stage. The search terms are combined into strings using Boolean operators, for example AND and OR, that will be used to search titles and abstracts in the databases (see example in Table 4.2). Refinement and validation of the search queries is important. Unique search strings and strategies tailored to the search requirements specific to each database are needed. The development of search queries can be complicated; therefore, specialists in library and information science should always be involved in designing the search strategy and queries. The search queries can be tested in each database using a test set of primary research publications identified during the problem formulation stage. The search query should be able to retrieve most of this test set of publications. If the test query results in the identification of a large number of off-topic citations, this also provides an opportunity to change the search string to eliminate them. (For common search topics, such as air pollution, there may even be published search strategies available in the literature.)

The search queries in Table 4.2 were developed to answer two key questions regarding research on exposure to perfluorooctanoic acid (PFOA) or perfluorooctane sulfonate (PFOS) and effects on the immune system. The key questions were: (a) What is the hazard identification category for an association between exposure to PFOA or PFOS or their salts and immunotoxicity or immune-related health effects based on integrating levels of evidence from human and animal studies: Known, Presumed, Suspected, Not classifiable, or Not identified to be a hazard to humans? and (b) How does the evidence from other relevant studies (such as mechanistic studies) support or refute the biological plausibility of the association between exposure to PFOA or PFOS or their salts and immunotoxicity or immune-related health effects? In this example the search strings were constructed using terms in the title and abstract (tiab) and PubMed MeSH headings (mh).

<sup>8</sup> GreyNet International: <http://www.greynet.org/greysourceindex/documenttypes.html>.

<sup>9</sup> As in the case of the United States Agency for Toxic Substances and Disease Registry, the United States EPA, the European Chemicals Agency, and EFSA.

**Table 4.2 Example of search string used in PubMed to identify research on PFOA or PFOS and immune effects**

Database	Search terms
PubMed	<p>perfluoroalkyl*[tiab] OR perfluorocaprylic[tiab] OR perfluorocarbon*[tiab] OR perfluorocarboxyl*[tiab] OR perfluorochemical*[tiab] OR (perfluorinated[tiab] AND (C8[tiab] OR carboxylic[tiab] OR chemical*[tiab] OR compound*[tiab] OR octanoic[tiab])) OR PFAA*[tiab] OR "fluorinated polymer"[tiab] OR "fluorinated polymers"[tiab] OR (fluorinated[tiab] AND (polymer[tiab] OR polymers[tiab])) OR (fluorocarbon[tiab] AND (polymer[tiab] OR polymers[tiab])) OR Fluoropolymer*[tiab] OR (fluorinated[tiab] AND telomer*[tiab]) OR fluorotelomer*[tiab] OR fluoro-telomer*[tiab] OR fluorosurfactant*[tiab] OR "FC 143"[tiab] OR FC143[tiab] OR 335-67-1 [rn] OR Pentadecafluorooctanoate*[tiab] OR Pentadecafluorooctanoate*[tiab] OR pentadecafluorooctanoic[tiab] OR pentadecafluorooctanoic[tiab] OR "pentadecafluoro-1-octanoic"[tiab] OR "pentadecafluoro-n-octanoic"[tiab] OR "perfluoro-1-heptanecarboxylic"[tiab] OR perfluorocaprylic[tiab] OR perfluoroheptanecarboxylic[tiab] OR perfluorooctanoate[tiab] OR perfluorooctanoate[tiab] OR "perfluoro octanoate"[tiab] OR "perfluorooctanoic acid"[nm] OR perfluorooctanoic[tiab] OR perfluorooctanoic[tiab] OR "perfluoro octanoic"[tiab] OR "perfluoro-n-octanoic"[tiab] OR "perfluorooctanoyl chloride"[tiab] OR PFOA[tiab] OR APFO[tiab] OR 1763-23-1[rn] OR 307-35-7[rn] OR "1-octanesulfonic acid"[tiab] OR "1-perfluorooctanesulfonic"[tiab] OR "1-perfluorooctanesulfonic"[tiab] OR "heptadecafluoro-1-octanesulfonic"[tiab] OR "heptadecafluoro-1-octane sulfonic"[tiab] OR "heptadecafluorooctanesulfonic"[tiab] OR "heptadecafluorooctane sulfonic"[tiab] OR "heptadecafluorooctane sulfonic"[tiab] OR "perfluoroalkyl sulphonate"[tiab] OR perfluorooctanesulfonate[tiab] OR perfluorooctanesulfonate[tiab] OR "perfluorooctane sulfonate"[tiab] OR "perfluorooctane sulfonate"[tiab] OR "perfluoro-n-octanesulfonic"[tiab] OR perfluorooctanesulfonic[tiab] OR perfluorooctanesulfonic[tiab] OR "perfluorooctane sulfonic acid"[nm] OR "perfluorooctane sulfonic"[tiab] OR "perfluorooctane sulfonic"[tiab] OR perfluorooctanesulphonic[tiab] OR perfluorooctanesulphonic[tiab] OR "perfluorooctane sulphonic"[tiab] OR "perfluorooctane sulphonic"[tiab] OR perfluorooctylsulfonic[tiab] OR PFOS [tiab]</p> <p>AND</p> <p>immunology[sh] OR immune[tiab] OR immunocomp*[tiab] OR immunogen*[tiab] OR immunolog*[tiab] OR immunotox*[tiab] OR immunotoxins[mh] OR immunity[tiab] OR autoimmun*[tiab] OR "host resistance"[tiab] OR immunocompetence[mh] OR "immune system"[mh] OR spleen[tiab] OR splenic[tiab] OR splenocyt*[tiab] OR thymus[tiab] OR thymic[tiab] OR thymocyt*[tiab] OR leukocyt*[tiab] OR granulocyt*[tiab] OR basophil*[tiab] OR eosinophil*[tiab] OR neutrophil*[tiab] OR lymph[tiab] OR lymphoid*[tiab] OR lymphocyt*[tiab] OR "b-lymphocyte"[tiab] OR "b-lymphocytes"[tiab] OR "t-lymphocyte"[tiab] OR "t-lymphocytes"[tiab] OR "killer cell"[tiab] OR "killer cells"[tiab] OR "NK cell"[tiab] OR "NK-cell"[tiab] OR "NK-cells"[tiab] OR macrophag*[tiab] OR "mast cell"[tiab] OR "mast cells"[tiab] OR monocyt*[tiab] OR phagocyt*[tiab] OR dendrit*[tiab] OR "t-cell"[tiab] OR "t cell"[tiab] OR "t cells"[tiab] OR "t-cells"[tiab] OR "T helper"[tiab] OR "T-helper"[tiab] OR "b-cell"[tiab] OR "b cell"[tiab] OR "b cells"[tiab] OR "b-cells"[tiab] OR antibod*[tiab] OR histamine*[tiab] OR histocompatib*[tiab] OR</p> <p>immunoglobulins[mh] OR immunoglobulin*[tiab] OR "immunoglobulin A"[tiab] OR IgA[tiab] OR "immunoglobulin D"[tiab] OR IgD[tiab] OR "immunoglobulin E"[tiab] OR IgE[tiab] OR "immunoglobulin G"[tiab] OR IgG[tiab] OR "immunoglobulin M"[tiab] OR IgM[tiab] OR "antigens, CD"[mh] OR CD3 [tiab] OR CD4 [tiab] OR CD8 [tiab] OR CD25 [tiab] OR CD27 [tiab] OR CD28 [tiab] OR CD29 [tiab] OR CD45*[tiab] OR cytokines[mh] OR cytokine*[tiab] OR chemokine*[tiab] OR inteferon*[tiab] OR interleukin*[tiab] OR "IL-6"[tiab] OR "IL-8"[tiab] OR lymphokine*[tiab] OR monokine*[tiab] OR ("tumor necrosis"[tiab] AND (factor[tiab] OR factors[tiab])) OR "TNF alpha"[tiab] OR "TNFalpha"[tiab] OR "immune system diseases"[mh] OR autoimmun*[tiab] OR addison[tiab] OR rheumatoid[tiab] OR glomerulonephritis[tiab] OR diabetes[tiab] OR graves[tiab] OR lupus[tiab] OR thyroiditis[tiab] OR hypersensitiv*[tiab] OR sensitization OR hyperresponsiv*[tiab] OR allergy[mh] OR allerg*[tiab] OR atopy[tiab] OR atopic[tiab] OR dermatitis[tiab] OR eczema[tiab] OR otitis[tiab] OR "ear infection"[tiab] OR "ear inflammation"[tiab] OR Respiratory tract infections[mh] OR (respiratory[tiab] AND infection*[tiab]) OR asthma[tiab] OR bronchitis[tiab] OR pneumonia[tiab] OR bronchiolitis[tiab] OR rhinitis[tiab] OR sinusitis[tiab] OR wheez*[tiab] OR crackle*[tiab] OR cough[mh] OR cough*[tiab] OR dyspnea[tiab] OR gastroenteritis[tiab] OR inflammation[mh] OR inflammat*[tiab] OR pro-inflammat*[tiab] OR anti-inflamm*[tiab] OR "inflammation mediators"[mh] OR autacoid*[tiab] OR eicosanoid*[tiab] OR prostaglandin*[tiab] OR immunomodulation[mh] OR immunomodul*[tiab] OR immunotherap*[tiab] OR vaccin*[tiab] OR immuniz*[tiab] OR immunosuppress*[tiab] OR desensitiz*[tiab] OR immunoproteins[mh] OR immunoprotein*[tiab] OR "c-reactive protein"[tiab] OR CRP[tiab] OR "complement component" [tiab] OR (complement[tiab] AND (C1 OR C2 OR C3 OR C4 OR C5 OR C6 OR C7 OR C8 OR C9))</p>

Source: United States National Toxicology Program (8).

Once the results of the searches have been compiled, removal of duplicate references will be required, which can be a labour-intensive process. This may be accomplished through an automated process using reference management software, but will probably also require manual review and deletion of duplicates.

#### 4.1.4 Additional search strategies

Some informative publications may not be identified using the search process described above. Publications may not be indexed correctly in the bibliographic database, the database may not include those journals, or results presented in the paper may not be included in the abstract. Additional strategies include forward and backward searches. Forward searching identifies articles that cite the key study (this may be automated in specific databases), and backward searching identifies articles that were cited by the key study. Backward searches can be done manually by reviewing the cited references, typically in the introduction and discussion sections of a paper, including those cited in review articles. References cited in authoritative reports by government and other institutions are also useful sources. Abstracts for presentations and posters from some professional meetings may be published on the society's website or in a journal. In addition, it may be useful to manually search the table of contents for certain relevant journals. Unique strategies may be needed to identify grey literature. These may include websites of government institutions, portals that allow searching of government agency websites, for example the OECD eChemPortal or the INCHEM database, abstracts from professional meetings, and other compilations.

#### 4.1.5 Documentation of the literature search results

Accurate documentation of the search strategy is essential to the transparency of the systematic review process. Documentation of literature searches using bibliographic databases should include the databases and date range covered by the search, search terms and search strings used that were applied, and dates that the searches were performed. Other literature identification strategies should be documented, including the steps taken, even if they do not produce additional records. The results of the literature search is a compiled list of unique records with abstracts organized using a reference management software (such as EndNote, DistillerSR) – some examples are shown in Table 4.3. The use of two open-source applications by the National Toxicology Program Report on Carcinogens and the IARC has been described (9).

**Table 4.3 Example tools for data identification and data extraction**

Database	Characteristics
EndNote (Clarivate Analytics)	Reference management software
DistillerSR (Evidence Partners)	Designed for screening and documenting the process of data identification in systematic reviews and also other steps (data extraction, evidence appraisal)
Swift-Active Screener (Sciome)	Web-based, collaborative systematic review software. This tool uses statistical models to automatically prioritize records based on review feedback
Health Assessment Workspace Collaborative (HAWC)	A free, web-based, open-source content management system for human health assessments
Table Builder	A free, web-based, open-source content management system for human health assessments; source code available for reuse (9)



## 4.2 Screening the literature

Eligibility for inclusion in the review is determined using a set of predefined inclusion and exclusion criteria developed based on the PECO statement. Citations that meet the inclusion criteria are retained and categorized as appropriate.

### 4.2.1 Study selection process

Literature screening involves a sequential process of title and abstract screening and then full-text screening to identify relevant publications. The process, with criteria and plans for documentation, should be described in the systematic review protocol. Generally, two reviewers should conduct the screening of the database of records independently, documenting their decisions, and there should be a process for resolving any differences that occur. This reduces the risk that the screening process is inadvertently selective. At this point in the systematic review process neither the quality nor the results of the study are considered. It may be desirable to categorize the studies that are retained into groups useful for later organization.

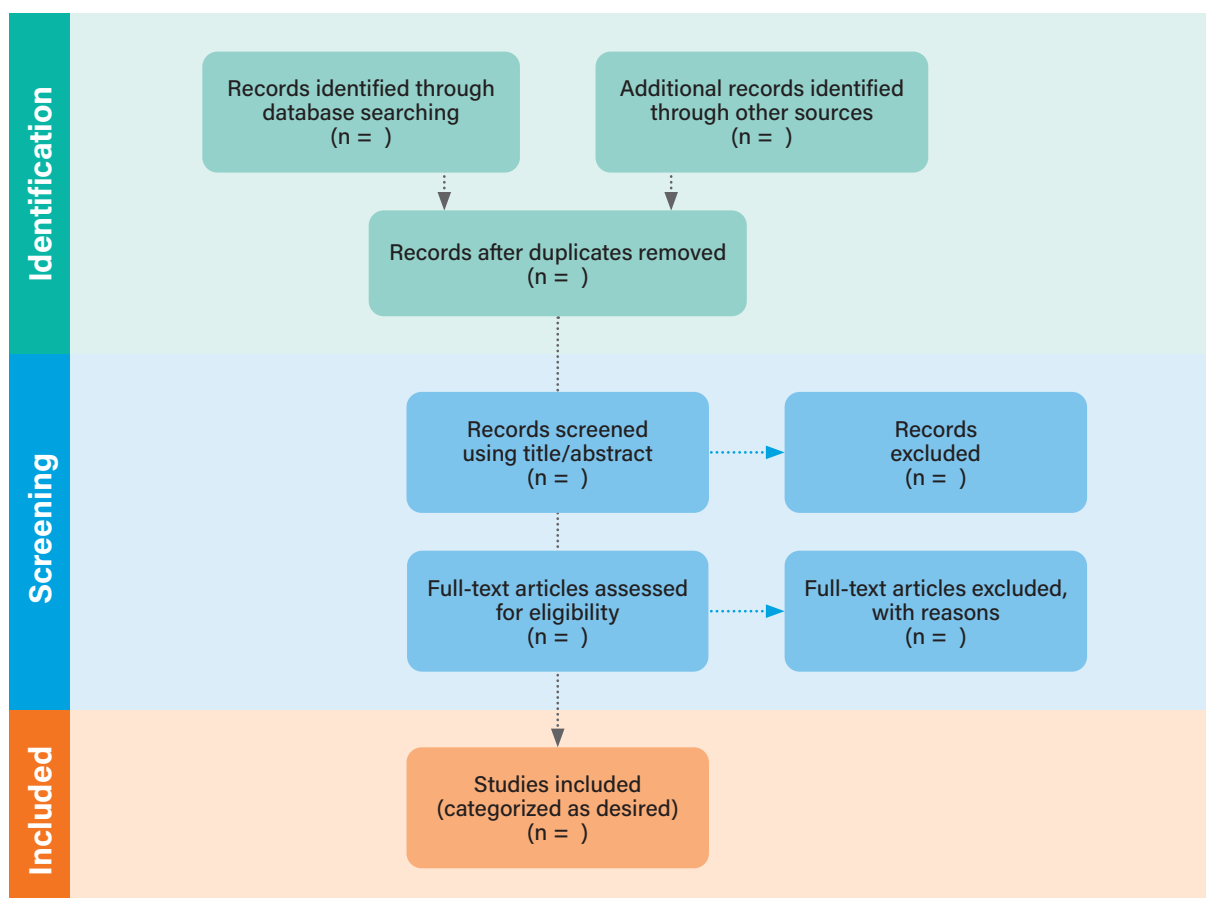
The screening process will depend on the number of retrieved citations and the complexity and scope of the research question. Some data sets may be easily screened using only titles and abstracts, with a few citations reserved for more in-depth review (full-text screening). Automated screening using machine learning approaches has made the process of screening large numbers of records more efficient, reducing the workload required (10).

### 4.2.2 Presentation of screening results

The presentation of the results of the screening process should document the numbers of records retrieved from each database, unique records after eliminating duplicates, records in each exclusion category, and the final numbers of included records categorized that make sense for the organization of the review. There are multiple ways to present the screening results, but often a literature flow diagram is used (Figure 4.2, adapted from the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) Statement). To increase the transparency of the screening process, the PRISMA authors state that the reasons for excluding articles during full-text screening should be provided. Sometimes the same study results may be reported in multiple publications, and the number of included studies and number of included publications and articles may need to be reported separately.



**Figure 4.2 Example of a literature flow diagram summarizing the results of the screening process**



Source: Adapted from the PRISMA Statement (11, 12). The figure is based on the flow diagram presented by Moher et al. (11) and can be obtained from the PRISMA website (<http://prisma-statement.org/>).

## 4.3 Data extraction

The extraction of data (that is, information about a study and the study's results) requires knowledge of the details in an individual study and what is common across sets of similar studies (for example, same end-points, study designs or exposure settings). The strategy for data extraction should be described in the systematic review protocol, which should describe the number of extractors, the training process and quality assurance procedures (for example, single extractor with quality assurance review or two extractors independently). The protocol should also describe the process for resolving differences between extractors. It may be helpful to develop a data extraction form to facilitate standardization and to pilot-test the form using a small subset of included studies. Often the form will be revised to address issues raised during the pilot test.

For the most part, data extraction is a manual process that is resource and time intensive. Because some studies may report multiple sets of results, efficiencies require that only the results that best inform the synthesis of findings for the review and its conclusions be extracted. However, decisions to extract data should not be based on the magnitude, direction or precision of the effect estimates. For some sets of studies, it may be appropriate not to extract the results of studies that are deemed less informative because of methodological limitations, low internal or external validity, reporting deficiencies, or other reasons described in the systematic review protocol.

Because some decisions about what findings to present cannot be made until all of the relevant studies are evaluated and the details of the findings are understood, extraction of data from individual studies is likely to be iterative. Initially, relevant types of data may be basic study information (authors, citation, descriptive information pertaining to study design, health end-points) and information that will be used to evaluate internal and external validity for study outcomes. If multiple results are presented for an outcome by a study, the choice of which data to extract will flow from the synthesis plan in the protocol.

The specific types of information to extract and the level of detail may be aided by the database tools and software available for storing the information. Some software tools available at this writing, including HAWC and DistillerSR, have been developed for storing and organizing study information and results. Such tools help to standardize terminology (that is, discipline-specific ontologies) and ensure that standards for data quality are met. In addition, these tools can facilitate the development of graphic and tabular means of data summarization (see Table 4.3). Other tools may become available in the future. Table 4.4 presents a list of databases and resources for systematic review, by category.

With regard to quality assurance for data extraction, an approach to minimize errors and maintain accuracy in data entry and presentation should be planned and described in the systematic review protocol. The quality assurance approach should detail the database management system to be used, and the process for data entry and verification. Any transformations or conversions of the data originally reported in a source document should be documented and retained in the database, as well as any calculations using the data. Automation of the process to the extent possible reduces the opportunity for errors in entry, manipulation and presentation of data.

**Table 4.4 Databases and resources, by category**

Category	Source
Academic databases	Embase <a href="https://www.embase.com/">https://www.embase.com/</a> PubMed <a href="https://www.ncbi.nlm.nih.gov/pubmed/">https://www.ncbi.nlm.nih.gov/pubmed/</a> Scopus <a href="https://www.scopus.com/">https://www.scopus.com/</a> Web of Science <a href="https://apps.webofknowledge.com/">https://apps.webofknowledge.com/</a>
Chemical and toxicological databases	INCHEM. Chemical Safety Information from Intergovernmental Organizations database <a href="http://www.inchem.org/">http://www.inchem.org/</a> OECD eChemPortal <a href="https://www.echemportal.org/echemportal/">https://www.echemportal.org/echemportal/</a> PubChem <a href="https://pubchem.ncbi.nlm.nih.gov/">https://pubchem.ncbi.nlm.nih.gov/</a>
Grey literature general resources	Grey Literature Report <a href="http://www.greylit.org/">http://www.greylit.org/</a> GreyNet <a href="http://www.greynet.org/">http://www.greynet.org/</a> OpenGrey <a href="http://www.opengrey.eu">http://www.opengrey.eu</a> U.S. Interagency Science.gov Alliance <a href="https://www.science.gov">https://www.science.gov</a>
Tools and software for data collection and management	Endnote <a href="https://endnote.com">https://endnote.com</a> DistillerSR <a href="https://www.evidencepartners.com">https://www.evidencepartners.com</a> Health Assessment Workspace Collaborative (HAWC) <a href="https://hawcproject.org">https://hawcproject.org</a> Swift-Active Screener <a href="https://www.sciome.com/swift-activescreener/">https://www.sciome.com/swift-activescreener/</a>

## References

1. Handbook for conducting a literature-based health assessment using OHAT approach for systematic review and evidence integration. United States Department of Health and Human Services, National Toxicology Program, Office of Health Assessment and Translation; 2019 update ([https://ntp.niehs.nih.gov/ntp/ohat/pubs/handbookmarch2019\\_508.pdf](https://ntp.niehs.nih.gov/ntp/ohat/pubs/handbookmarch2019_508.pdf), accessed 26 April 2021).
2. European Food Safety Authority. Application of systematic review methodology to food and feed safety assessments to support decision making. *EFSA Journal*. 2010;8(6):1637 (<https://www.efsa.europa.eu/fr/efsajournal/pub/1637>, accessed 25 April 2021).
3. European Food Safety Authority. Submission of scientific peer-reviewed open literature for the approval of pesticide active substances under Regulation (EC) No. 1107/2009. *EFSA Journal*. 2011;9(2):2092. doi:10.2903/j.efsa.2011.2092.
4. ECHA (European Chemicals Agency) and EFSA (European Food Safety Authority) with the technical support of the Joint Research Centre (JRC), Andersson N, Arena M, Auteri D, Barmaz S, Grignard E et al. Guidance for the identification of endocrine disruptors in the context of Regulations (EU) No. 528/2012 and (EC) No. 1107/2009. *EFSA Journal*. 2018;16(6):5311. doi:10.2903/j.efsa.2018.5311. ECHA-18-G-01-EN.
5. Guidance to JECFA experts on systematic literature searches. Geneva: World Health Organization; 2017 ([https://www.who.int/groups/joint-fao-who-expert-committee-on-food-additives-\(jecfa\)/procedural-guidance](https://www.who.int/groups/joint-fao-who-expert-committee-on-food-additives-(jecfa)/procedural-guidance), accessed 26 April 2021).
6. Leenaars M, Hooijmans CR, van Veggel N, ter Riet G, Leeflang M, Hooft L et al. A step-by-step guide to systematically identify all relevant animal studies. *Laboratory Animals*. 2012;46:24–31. doi:10.1258/la.2011.011087.
7. Mahood Q, van Eerd D, Irvin E. Searching for grey literature for systematic reviews: challenges and benefits. *Research Synthesis Methods*. 2014;5:221–34.
8. Immunotoxicity associated with exposure to perfluorooctanoic acid or perfluorooctane sulfonate. NTP Monograph. United States Department of Health and Human Services, National Toxicology Program; 2016.
9. Shapiro AJ, Antoni S, Guyton KZ, Lunn RM, Loomis D, Rusyn I et al. Software tools to facilitate systematic review used for cancer hazard identification. *Environmental Health Perspectives*. 2018;126(10):104501. doi:10.1289/EHP4224.
10. O'Mara-Eves A, Thomas J, McNaught J, Miwa M, Ananiado S. Using text mining for study identification in systematic reviews: a systematic review of current approaches. *Systematic Reviews*. 2015;4:1–22.
11. Moher D, Liberati A, Tetzlaff J, Altman DG, the PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA Statement. *PLoS Medicine*. 2009;6(7):e1000097. doi:10.1371/journal.pmed1000097.
12. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA et al. The PRISMA Statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *PLoS Medicine*. 2009;6(7):e1000100. doi:10.1371/journal.pmed.1000100.



# 5 Appraisal of individual studies

## 5.1 Multiple aspects of critical appraisal

Chemical risk assessment often includes procedures to critically assess study quality; however, assessment methods may be inconsistently applied and lack transparency. Research and regulatory activities to facilitate structured evaluation of study quality have resulted in development of different approaches, including methods based on systematic review methodology. Evidence appraisal in systematic review involves applying structured and standardized approaches to critically assessing and summarizing the strengths and weaknesses of each individual study included in the review. Appraisal is a critical step because the certainty of the evidence that is synthesized in a systematic review determines the trustworthiness of the result of the review.

A challenge when illustrating this process is to clarify what study characteristics are subject to evaluation, as terminology and approaches are not standardized and vary across fields and textbooks. Further, best practices for critically appraising the wide variety of study types commonly encountered in chemical risk assessment are still under development. Approaches and tools are more well developed for some study types (such as randomized controlled trials, a type of human controlled exposure studies, or experimental animal studies) relative to others (such as in vitro studies, human observational studies).



**Main critical appraisal areas:**  
Reporting quality  
Internal validity  
External validity  
Sensitivity

A first fundamental distinction is between aspects related to the design and conduct of a study and the reporting of the study. Studies that are poorly reported are problematic because they do not provide all the information about their methods and results that is necessary to conduct a study appraisal. Failure to report certain aspects of a study limits the reviewer's ability to identify problems in study design and conduct, or to distinguish them from poor reporting (1).

### 5.1.1 Internal validity, external validity and sensitivity

Another aspect that requires clarification is the relation between internal and external validity of the studies included in a review. Internal validity addresses the degree to which bias or "a systematic error, or deviation from the truth, in results" (2) is minimized in the studies of interest. Bias arises through problems in the design or conduct of a study; it can operate in either direction (resulting in systematic underestimation or overestimation of the parameter in question) and can vary in magnitude (2). External validity (comparable to relevance, directness, applicability, generalizability, and lack of external bias under different methods and frameworks) reflects the extent to which the study results provide direct evidence for addressing the research question (for example, applicability of the animal model or dosing route). External validity also reflects the utility of the study results as a correct basis for generalizations to other circumstances (2, 3).

Study sensitivity or informativeness (4) addresses the ability of a study to detect a true effect and is a third, and relatively recent, concept that some groups distinguish from internal and external validity (5). It involves evaluation of aspects of study design and conduct, such as the sensitivity of the animal model, dose or exposure level (for example, use of low or environmentally relevant dose), and administration of the test compound during sensitive windows of exposure (such as during gestation), and ensuring that measurements

are conducted at appropriate time points when sensitive effects can be expected to be observed. In the case of an insensitive study, a false conclusion of no effect may be drawn because of study design factors that lead to the failure to identify differences that truly exist.

Bias (and issues of sensitivity) can occur at different levels in a study. For instance, in studies aiming at assessing an exposure–outcome association, some biases apply to the whole study (for example, those arising from bias in assigning animals to experimental groups), some biases are specific to the outcome being measured, some biases apply mainly to the method used for the outcome measurement, some biases are relevant to the exposure and method for measuring the exposure, and other biases are specific to the results (such as bias in selection of the reported result) (6). Hence, in these studies risk of bias assessment must be specific to a particular result for a particular exposure–outcome pair (and time point, if applicable), and the assessment of the exposure–outcome dependent aspects should be applied for as many exposure–outcome pairs as needed, for each individual study included in the review.

### 5.1.2 Validity versus quality

Study validity differs from quality, which is a broad term that covers the concepts of internal validity, external validity, sensitivity, quality of reporting and compliance with standards (7). Therefore, it follows that an assessment of risk of bias (or of degree of internal validity) in a study is not equivalent to assessing overall study quality. A study can be conducted to the highest possible standards (for example, in compliance with regulatory requirements) but still have serious biases (7, 8).

Risk assessment has a decades-long history of developing standards for conducting studies (for example, on good laboratory practices); however, applying a consistent approach to evaluating the quality of published studies is a more recent development. The “Klimisch criteria” are an example of methods developed for assessing the extent to which a scientific study is of sufficient quality (or “reliability”) to be used to determine health risk or hazards from chemical exposure (9). The Klimisch criteria mix aspects of reporting quality and internal and external validity and use scoring systems for evaluation of “reliability” and “relevance” that are challenging to implement transparently or consistently, despite several attempts to refine these criteria (10, 11). Techniques for evaluating evidence in systematic review contexts have advanced significantly since 1997, when the criteria were first published, and more recent approaches that promote systematic and transparent evidence appraisal, such as those described below, should be used when applying systematic review methodology for chemical risk assessment.

Many frameworks focus the appraisal of individual studies on internal validity, and issues related to external validity are assessed separately when evaluating certainty in bodies of evidence selected for review (12, 13). Some tools are available for assessing sensitivity as well as both internal and external validity at individual study level (14–16). Whether assessed at the individual study level or the body of evidence level, or both, the approach to assessing internal validity, external validity, and sensitivity should be described in the protocol.

### 5.1.3 Validity versus precision

Internal and external validity differ from precision. Precision concerns the ability of a study to provide similar results when repeated under the same conditions and is related to random error. Imprecision is addressed by expressing the sampling estimates as confidence intervals at the individual study level, and for pooled estimates, and also typically in systematic review, it is not assessed at the individual study level. For questions aiming to test a hypothesis, lack of power is another (statistical) study limitation that must be accounted for. Imprecision and lack of power are fundamentally different, since their potential impact on the results is related respectively to bias (lack of power) and random error (imprecision). They may also differ in the direction of the effect, where lack of power is likely to bias a study towards the null and issues with internal validity may bias either towards or away from the null, depending on the biases involved.

## 5.2 Examples of threats to internal validity and sensitivity

### 5.2.1 Human observational studies

Risk of bias evaluations of human observational studies assess selection bias, information bias and confounding. While there are differences between the approaches that have been developed to evaluate health studies of chemical exposures, tools generally include analyses of confounding, selection of participants into the study, measurement and classification of exposures, attrition (or loss to follow-up), missing data, measurement of outcomes and selective reporting of the results (1). The consideration of confounding and its consequences for exposure misclassification are particularly challenging. The evaluation of measurement error is an important consideration that can affect data related to exposures, outcomes, confounders or modifiers (7). Characteristics of a study that may reduce sensitivity should also be considered, such as a narrow exposure range in the study population or inadequate timing of the outcome measurements (for example, not allowing enough time for the outcome or disease to develop) (15).

### 5.2.2 Experimental animal studies

Biases commonly considered in laboratory animal studies can arise from the randomization process, the experimental conditions across study groups (that is, whether they are identical or not), the blinding of research personnel and of outcome assessors, the completeness of outcome data, the methods for exposure characterization and outcome assessment, the presence of selective reporting, and the statistical methods. Some factors are included based on generally accepted use in toxicology, and for others there is a growing body of empirical evidence that elements shown to bias clinical trial (such as lack of blinding of outcome assessors) are also associated with changes in effect size in experimental animal studies, thereby representing sources of bias (17). The following aspects may decrease sensitivity of toxicity studies and bias results towards the null, and are important for evaluating study design and conduct: sensitivity of animal species or strains for the outcome of interest; the use of inadequate methods for measuring the outcomes; and timing of exposure and outcome measurements (for example, whether the study allowed sufficient time for the effect or disease to develop).

### 5.2.3 Mechanistic and modelled data studies

Mechanistic data include a broad category of end-points designed to inform the chemical, molecular, cellular or physiological mechanisms by which substances contribute to potential health effects. Principal sources of bias will differ substantially by study design. Several types of modelled or “in silico” data are also relevant for chemical risk assessment, such as quantitative structure–activity relationship (QSAR) data. At this point, few structured approaches have been developed for assessing modelled data, with the framework for systematic review and integrated assessment (SYRINA) of endocrine-disrupting chemicals having outlined one such approach (18). Mechanistic data reported from in vivo studies in humans or animals can be assessed for risk of bias with the same tools as for other end-points in experimental animal studies. However, studies with mechanistic data from cells or tissues with an in vitro exposure regime may require a modified approach. The United States National Toxicology Program developed a risk of bias tool to assess in vitro exposure studies that was used as a case example in a systematic review of PFOA and PFOS immunotoxicity (19). Based on that tool, EFSA is also developing a risk of bias assessment tool for in vitro studies on developmental neurotoxicity.<sup>10</sup> The Science in Risk Assessment and Policy (SciRAP) group has also developed a tool to assess

---

<sup>10</sup> Draft risk of bias assessment tool for in vitro studies on developmental neurotoxicity. European Food Safety Authority, in progress.



in vitro exposure studies.<sup>11</sup> The National Toxicology Program tool focuses on risk of bias, while the SciRAP tool also includes consideration of study sensitivity and separate modules for reporting and external validity (relevance). The impact of risk of bias relative to other aspects of quality will depend on how mechanistic studies are used in a scientific assessment. Particularly when mechanistic studies are used to interpret the biological plausibility of human or animal studies, it is critical to assess the external validity (or applicability) of the mechanistic data as well as the internal validity.

## 5.3 Process for evidence appraisal

As with other steps of the systematic review process, the method for evidence appraisal should be (a) predefined in the review protocol; (b) structured in a way to minimize bias, random error and subjectivity and maximize impartiality; (c) scientifically defensible; and (d) thoroughly documented and reported. The protocol should specify the appraisal tool, application of the tool to different study designs, criteria for reaching ratings on individual appraisal questions (as well as an overall study rating if assessed), and whether external validity is to be assessed at the individual study level or at the body of evidence level (see Chapter 6). It is recommended that modifications specific to the topic (which are often warranted) should be identified in the protocol. The protocol should also outline the workflow for study appraisal, including number of reviewers and process for resolving differences between reviewers, as well as procedures for pilot-testing the tool and revising and documenting changes.

Evidence appraisal requires multidisciplinary expertise (for example, domain expertise and expertise in the specific study design – epidemiology, toxicology and statistics). Each study included in the review should be appraised by two mutually independent assessors and a system should be in place for solving possible divergences (for example, by involving a third reviewer). Before being formally implemented in a systematic review, critical appraisal tools should be pilot-tested on a subset of relevant studies and refined as appropriate (for example, through discussion with subject matter experts and stakeholder engagement as necessary). The pilot-testing should include all assessors to promote consistency on the specific tool used.

Critical appraisal tools are fundamental instruments (20, 21). An appraisal tool contains a predefined list of study design-specific sources of bias and guidance for judging the risk of bias due to each of them. Ideally this judgement should attempt to quantify the impact of each source of bias on the likely direction and magnitude of the study estimated parameters and, for a scientifically defensible method, it should be based on empirical evidence (22). Tools usually provide a scale for expressing the judgement (for example, high risk of bias, low risk of bias, some concern) as well as a narrative description of the basis for that rating.

If the approach to evidence appraisal involves assessing external validity in addition to internal validity and sensitivity, the appraisal tool should also contain the list of possible threats to external validity and guidance for assessing them.

Appraisal tools are generally tailored to assess studies of a specific design type, and a number of critical appraisal tools have been used as part of chemical assessments. Several recent tools for risk of bias assessment for human observational studies and experimental animal studies have a domain-based structure in which different types of bias are considered in turn (1, 22–24). Some tools (16, 22, 25) include specific signalling questions or criteria that should be answered to help the judgement on the risk of bias, minimize subjectivity and enhance clarity. Whether the approach uses tiered signalling questions or is less structured, context-specific expertise is needed, and all tools require scientific judgement to ensure the appraisal reaches conclusions that can be scientifically justified.

---

<sup>11</sup> Science in Risk Assessment and Policy: [www.scirap.org](http://www.scirap.org).



Some approaches benchmark the assessment of studies or provide context by comparing studies to an “idealized version” of the original study. This can be viewed as a repeat of the study by using a design that eliminates all sources of bias and where the aspects related to external validity are specified as they are described in the original study (14, 15, 25, 26).

Different types of tools are available for different study designs and types, with varying degrees of coverage (for example, focusing on internal validity only or covering also elements related to external validity), details for guiding the assessment, supporting empirical evidence, development, testing and validation (16, 21, 23, 27, 28).

## **5.4 Use and reporting of critical appraisal of individual studies in the review**

### **5.4.1 Accounting for bias from individual studies in the synthesis**

There are principally two approaches to reach a judgement on the bias of the body of evidence, by outcome, accounting for the strengths and limitations of each individual study included in a systematic review. The first method groups studies by categories of risk of bias and assesses external validity across the group in a separate step, and the second method considers internal and external validity together at the study level. These approaches are briefly described below and addressed in Chapter 6, as the critical appraisal is then considered in evidence synthesis and evidence integration. Similarly, if the decision is made to assess external validity at the body of evidence, this would be done when making considerations on the body of evidence (see Chapter 6).

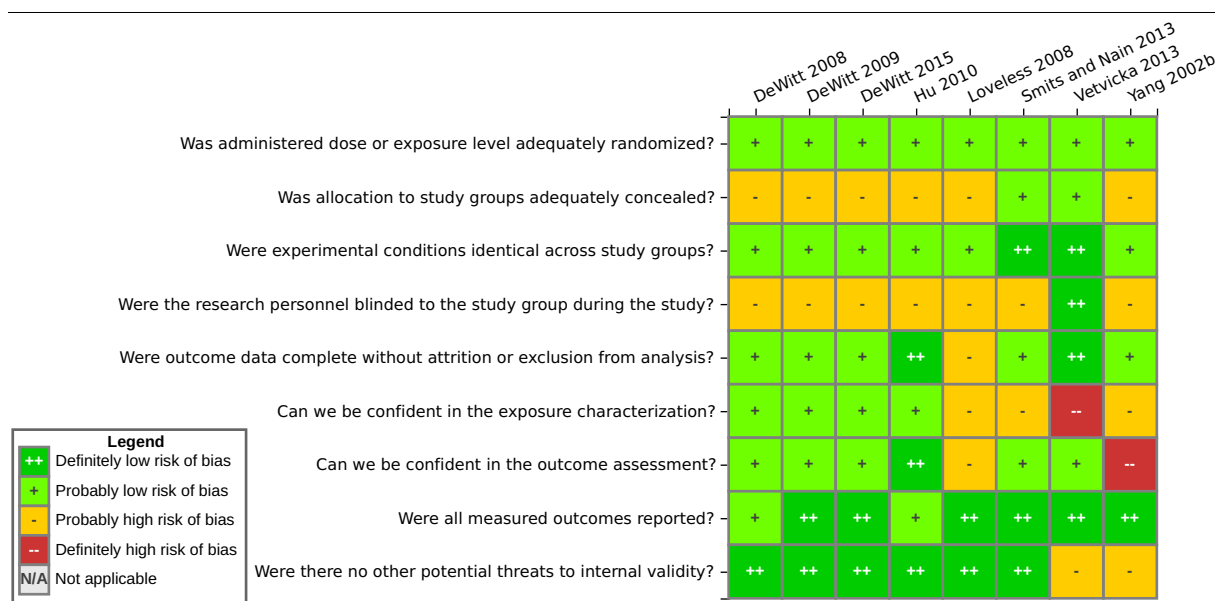
A widely applied method is to group the included studies according to summary categories of risk of bias (for example, high risk of bias, low risk of bias, some concern). Judgements on the sources of bias (or domains) are combined to categorize studies for different levels of bias using prespecified rules (for example, if one domain or question that is considered critical is rated high risk of bias, the study cannot be at low risk of bias) (21). Then, typically in these approaches, the result of study appraisal is addressed through sensitivity analyses or exploratory subgroup analyses. In most cases the studies with lower risk of bias are used to drive conclusions.

Another method involves adjusting for internal and external bias at the individual study level. Information about the biases may come from empirical evidence from an external collection of meta-analyses (29), eliciting expert knowledge (26), or a combination of the two (30). These methods allow the subsequent conduct of bias-adjusted meta-analysis.

### **5.4.2 Reporting evidence appraisal**

As for all other steps in systematic review, the methods, assumptions, process and results of evidence appraisal should be thoroughly documented and reported, including any possible deviation from the review protocol. This implies illustrating the finalized critical appraisal tools resulting from pilot-testing or use along with the process for their application (for example, number of reviewers involved or rationale for resolving conflicts) and reporting, for each outcome, the rating of the risk of bias domains at study level and across the body of evidence. At the individual study level the rationale for the rating should also be reported. An example of reporting the results of evidence appraisal is shown in Figure 5.1 and Table 5.1.

**Figure 5.1 Example of risk of bias heatmap for experimental animal studies in a systematic review**



Source: United States National Toxicology Program (37).

**Table 5.1 Example risk of bias rating and rationale for each rating for DeWitt 2008**

Risk of bias question	Risk of bias rating and rationale for rating
Was administered dose or exposure adequately randomized?	<b>Probably low risk of bias:</b> Study states animals were randomized to treatment group. Correspondence with the author indicated that animals were semi-randomized by weight. Animals were removed from the shipping container and placed in cages in order (either 1 to 8 or 1 to 16 cages). The animals were weighed and if the weights were not statistically different animals were left in the cage, but if they were different, they were adjusted until evenly distributed. Therefore, animals were distributed to minimize imbalance between groups on weight.
Was allocation to study groups adequately concealed?	<b>Probably high risk of bias:</b> Study did not report if allocation was concealed. Correspondence with the author indicated that there was no allocation concealment.
Were experimental conditions identical across study groups?	<b>Probably low risk of bias:</b> Controls received the same vehicle (deionized water) as the treated animals. Experimental conditions were provided, but it was not specified or documented that they were the same across study groups. It was stated that cage controls were included and were treated identically to all other mice.
Were the research personnel blinded to the study group during the study?	<b>Probably high risk of bias:</b> Not reported – study did not report if personnel were blinded to study group during the study. Correspondence with the author indicated personnel were not blinded in order to maintain appropriate dosing solution and drinking-water dose.
Were outcome data complete without attrition or exclusion from analysis?	<b>Probably low risk of bias:</b> Outcome data appear to be complete. Methods specify 8 animals/dose group and end-point; however, tables and figures do not specify N so it is not possible to further verify possibility of missing data.

Risk of bias question	Risk of bias rating and rationale for rating
Can we be confident in the exposure characterization?	<b>Probably low risk of bias:</b> PFOA was the ammonium salt and was stated to be $\geq 98\%$ pure, but was not independently confirmed. For the dose-response studies PFOA was administered via drinking-water (as opposed to gavage for the other studies). Average daily water consumption was reported and average daily intake was reported. Serum PFOA concentrations were obtained using appropriate methods. Average accuracy was stated to be 104% for low and 89.4% for high QC sera and 104% and 98% for low and high dosed sera. Average coefficient of variation for replicate analysis was 11.7%.
Can we be confident in the outcome assessment?	<b>Probably low risk of bias:</b> Antibody synthesis (SRBC test) was conducted using well established methods (++ for assessment methods), delayed-type hypersensitivity responses were tested using well established methods (++ for assessment methods), spleen and thymus weight were assessed with an acceptable method (+ for assessment methods). Study did not report if outcome assessors were blind to dose group. Correspondence with the author indicated outcome assessors were "blind to the study group" on day of sample collection and for running the antibody ELISA except for head of lab. "I was the only one who knew what animal belonged to what group, but all everyone else saw was a number that did not correspond to dose group." Final rating for assessment methods and blinding = ++ for antibody/DTH and + for organ weight and other end-points.
Were all measured outcomes reported?	<b>Probably low risk of bias:</b> All outcomes outlined in the abstract, introduction, and methods were reported. End-points that did not reach statistical significance are not shown in tables or figures.
Were there no other potential threats to internal validity?	<b>Definitely low risk of bias:</b> Appropriate statistical analyses were conducted. Study did not report if homogeneity of variance was confirmed before two-way ANOVA. Correspondence with the author indicated homogeneity was confirmed. The initial C57BL/6J mice had become genetically susceptible to ulcerative dermatitis so the animal model was switched to C57BL/6N for the dose-response. This does not appear to have impacted the results. No other threats to internal validity were identified.

*Note:* This study is depicted in the first column of the risk of bias heatmap in Figure 5.1.

*Source:* National Toxicology Program (31).


## References

1. Rooney, AA, Cooper GS, Jahnke GD, Lam J, Morgan RL, Boyles AL et al. How credible are the study results? Evaluating and applying internal validity tools to literature-based assessments of environmental health hazards. *Environment International*. 2016;92–93:617–29.
2. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ et al., editors. *Cochrane handbook for systematic reviews of interventions*, version 6.0 (updated July 2019). Cochrane; 2019 ([www.training.cochrane.org/handbook](http://www.training.cochrane.org/handbook), accessed 25 April 2021).
3. Reitsma JB, Rutjes AWS, Whiting P, Vlassov VV, Leeflang MMG, Deeks JJ. Chapter 9: Assessing methodological quality. In: Deeks JJ, Bossuyt PM, Gatsonis C, editors. *Cochrane handbook for systematic reviews of diagnostic test accuracy*, version 1.0.0. Cochrane Collaboration; 2009.

4. IARC monographs on the identification of carcinogenic hazards to humans: preamble. International Agency for Research on Cancer; 2019 (<https://monographs.iarc.fr/wp-content/uploads/2019/07/Preamble-2019.pdf>, accessed 27 April 2021).
5. Cooper GS, Lunn RM, Agerstrand M, Glenn BS, Kraft AD, Luke AM et al. Study sensitivity: evaluating the ability to detect effects in systematic reviews of chemical exposures. *Environment International*. 2016;92–93:605–10.
6. Higgins JP, Savović J, Page MJ, Sterne JA. Revised Cochrane risk of bias tool for randomized trials. 2016.
7. Review of EPA's Integrated Risk Information System (IRIS) process. Washington (DC): National Research Council; 2014 ([http://www.nap.edu/openbook.php?record\\_id=18764](http://www.nap.edu/openbook.php?record_id=18764), accessed 29 April 2021).
8. European Food Safety Authority. Submission of scientific peer-reviewed open literature for the approval of pesticide active substances under Regulation (EC) No. 1107/2009. *EFSA Journal*. 2011;9(2):2092. doi:10.2903/j.efsa.2011.2092.
9. Klimisch H-J, Andreae M, Tillmann U. A systematic approach for evaluating the quality of experimental toxicological and ecotoxicological data. *Regulatory Toxicology and Pharmacology*. 1997;25:1–5.
10. Schneider K, Schwarz M, Burkholder I, Kopp-Schneider A, Edler L, Kinsner-Ovaskainen A et al. "ToxRTool", a new tool to assess the reliability of toxicological data. *Toxicology Letters*. 2009;189(2):138–44.
11. Roth N, Cifroy P. A critical review of frameworks used for evaluating reliability and relevance of (eco) toxicity data: perspectives for an integrated eco-human decision-making framework. *Environment International*. 2016;95:16–29.
12. Handbook for conducting a literature-based health assessment using OHAT approach for systematic review and evidence integration. United States Department of Health and Human Services, National Toxicology Program, Office of Health Assessment and Translation; 2019 update ([https://ntp.niehs.nih.gov/ntp/ohat/pubs/handbookmarch2019\\_508.pdf](https://ntp.niehs.nih.gov/ntp/ohat/pubs/handbookmarch2019_508.pdf), accessed 26 April 2021).
13. Woodruff TJ, Sutton P. The Navigation Guide systematic review methodology: a rigorous and transparent method for translating environmental health science into better health outcomes. *Environmental Health Perspectives*. 2014;122(10):1007–14. doi:10.1289/ehp.1307175.
14. Handbook for preparing report on carcinogens monographs. United States Department of Health and Human Services, National Toxicology Program; 2015 (<https://ntp.niehs.nih.gov/whatwestudy/assessments/cancer/handbook/index.html>, accessed 25 April 2021).
15. Radke EG, Glenn B, Galizia A, Persad A, Nachman R, Batesonet T et al. Development of outcome-specific criteria for study evaluation in systematic reviews of epidemiology studies. *Environment International*. 2019;130:104884. doi:10.1016/j.envint.2019.05.078.
16. Beronius A, Molander L, Zilliacus J, Rudén C, Hanberg A. Testing and refining the Science in Risk Assessment and Policy (SciRAP) web-based platform for evaluating the reliability and relevance of in vivo toxicity studies. *Journal of Applied Toxicology*. 2018;38:1460–70.
17. Vesterinen HM, Egan K, Deister A, Schlattmann P, Macleod MR, Dirnagl U. Systematic survey of the design, statistical analysis, and reporting of studies published in the 2008 volume of the *Journal of Cerebral Blood Flow and Metabolism*. *Journal of Cerebral Blood Flow and Metabolism*. 2011;31(4):1064–72.
18. Vandenberg LN, Ågerstrand M, Beronius A, Beausoleil C, Bergman A, Bero LA et al. A proposed framework for the systematic review and integrated assessment (SYRINA) of endocrine disrupting chemicals. *Environmental Health*. 2016;15:74. doi:10.1186/s12940-016-0156-6.
19. Monograph on immunotoxicity associated with exposure to perfluorooctanoic acid (PFOA) and perfluorooctane sulfonate (PFOS). Research Triangle Park, NC: National Toxicology Program; 2016.

20. Risk of bias tools. riskofbias.info; 2021 (<https://www.riskofbias.info/welcome>, accessed 29 April 2021).
21. The National Toxicology Program Office of Health Assessment and Translation (OHAT) risk of bias rating tool for human and animal studies. Research Triangle Park, NC: National Toxicology Program; 2015 (<https://ntp.niehs.nih.gov/go/riskbias>, accessed 29 April 2021).
22. Sterne JA, Hernan MA, Reeves BC, Savovic J, Berkman ND, Viswanathan M et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ*. 2016;355:i4919.
23. Hooijmans CR, Rovers MM, de Vries RB, Leenaars M, Ritskes-Hoitinga M, Langendam MW. *BMC Medical Research Methodology*. 2014;14:1–9.
24. Pega F, Norris SL, Backes C, Bero LA, Descatha A, Gagliardi D et al. 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*. 2020;135:105039. doi:10.1016/j.envint.2019.105039.
25. Morgan RL, Thayer KA, Holloway AC, Santesso N, Blain R, Eftim S et al. Report of the 28th Annual Conference of the International Society for Environmental Epidemiology (ISEE), Rome, 2016.
26. Turner RM, Spiegelhalter DJ, Smith GC, Thompson SG. Bias modelling in evidence synthesis. *Journal of the Royal Statistical Society: Series A*. 2009;172:21–47.
27. Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Annals of Internal Medicine*. 2011;155:529–36.
28. Higgins JPT, Savović J, Page MJ, Elbers RG, Sterne JAC. Chapter 8: Assessing risk of bias in a randomized trial. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ et al., editors. *Cochrane handbook for systematic reviews of interventions*, version 6.0 (updated July 2019). Cochrane; 2019 ([www.training.cochrane.org/handbook](http://www.training.cochrane.org/handbook), accessed 25 April 2021).
29. Welton NJ, Ades AE, Carlin JB, Altman DG, Sterne JAC. Models for potentially biased evidence in meta-analysis using empirically based priors. *Journal of the Royal Statistical Society: Series A*. 2009;172(1):119–36. doi:10.1111/j.1467-985X.2008.00548.x.
30. EFSA Scientific Colloquium 23 – Joint European Food Safety Authority and Evidence-Based Toxicology Collaboration Colloquium. Evidence integration in risk assessment: the science of combining apples and oranges, 25–26 October 2017, Lisbon, Portugal. European Food Safety Authority; 2018. doi:10.2903/sp.efsa.2018.EN-1396.
31. Figure D13: RoB heatmap – PFOA antibody response in animals. In: Monograph on immunotoxicity associated with exposure to perfluorooctanoic acid (PFOA) and perfluorooctane sulfonate (PFOS). Research Triangle Park, NC: National Toxicology Program; 2016 (<https://hawcproject.org/summary/visual/104/>, accessed 29 April 2021).





# 6 Evidence synthesis and evidence integration

## 6.1 Introduction

Following critical appraisal of individual studies, the next step in conducting a systematic review involves synthesizing evidence to develop conclusions. For assessments involving multiple outcomes or populations – for example, hazard identification or exposure assessments of chemicals – an additional step after evidence synthesis is required, that is, evidence integration. Fundamental to these activities, which involve evaluation of the body of evidence as a whole, is the implementation of a prespecified and structured approach to reach a decision about a potential hazard, exposure or risk associated with chemicals. Though the specific approach will probably vary by assessment depending on the scope, objectives and types of data in the evidence base, the general steps require (a) evidence synthesis and confidence assessment, (b) evidence integration, and (c) overall uncertainty assessment. The resulting conclusion may vary depending on the purpose of the assessment, ranging from hazard identification (likely to or unlikely to present a hazard for an outcome) to hazard characterization (characterization of dose–response), exposure assessment (how prevalent and high is the exposure to the chemical) or risk characterization (derivation of a health-based benchmark). The following sections will outline some key considerations for synthesizing and integrating evidence for chemical risk assessment.

It is notable that various terminologies and definitions of such are utilized in practice, and thus defining evidence synthesis, data, bodies of evidence, evidence stream and evidence integration within the context of a specific systematic review is critical as a first step. Relevant terms are defined in Box 6.1.

### Box 6.1 Definitions of terms for evidence synthesis and evidence integration

**Body of evidence.** Collection of data across studies on the same or related end-points, outcomes or exposure parameter within an evidence stream.

**Confidence.** Measure of certainty that results presented in a body of evidence reflect the true relationship between the exposure and outcome. Results should be presented for each body of evidence (for example, human and animal studies separately). Note: other documents may refer to confidence as certainty or uncertainty in the evidence.

**Data.** Experimental or observational information obtained from individual studies that meet inclusion criteria.

**End-point.** Experimental or observational measurements that relate to an outcome, such as sperm production, organ weight or fertility.

**Evidence.** A collection (or body) of available data (information) that is deemed relevant for the specific context and objectives of the assessment.

**Evidence integration.** The process of integrating synthesized evidence across evidence streams to develop conclusions. This step may also involve qualitative and quantitative methods and allows for the characterization of uncertainty in the overall assessment.

**Evidence synthesis.** Combining the available evidence within a stream of evidence. The method by which evidence is combined may be qualitative or quantitative.

**Outcome.** A measurable change in the health status of an individual (for example, increased breast cancer risk, liver dysfunction) or a group of individuals (for example, increased mortality rate). Evaluation of an outcome (such as reproductive toxicity) may involve assessment of multiple end-points (such as sperm concentration, motility and morphology).

**Study design.** A protocol or plan for conducting a study by using a set of experimental methods or procedures to measure specified end-points or outcomes in a given test system (for example, population exposed during a specific time frame, in silico, in vitro, in vivo systems) to answer a defined research hypothesis or question.

**Streams of evidence.** Different types of evidence that are grouped around a population, similar toxicity end-point, or health or exposure outcome, based on predefined criteria (such as animal evidence, in vitro evidence, mechanistic evidence; or cohort, case-control, cross-sectional evidence; or evidence on consumption or occurrence). Note: other documents may refer to streams of evidence as lines of evidence or subsets.

**Uncertainty.** All types of limitations in the knowledge available to the assessors at the time an assessment is conducted and within the available time and resources (for example, any type of limitation in the available evidence, in its interpretation, and in the outcome or conclusions of the assessment). Uncertainty can be characterized using different methods, either qualitatively or quantitatively. The higher the uncertainty, the less confidence in the risk assessment outcome.

**Weight of evidence.** A term traditionally used to refer to the process of collecting, appraising, synthesizing and integrating evidence (within but also across streams of evidence) and accounting for the related uncertainty for scientific questions aiming at assessing causality. In this document, this term is replaced by "evidence synthesis" or "evidence integration", as appropriate.

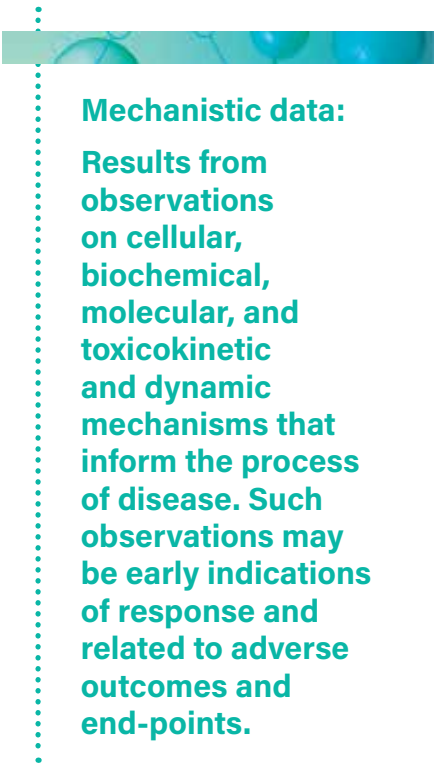


## 6.2 Evidence synthesis

As a first step in implementing the process, the underlying evidence available must be synthesized in a summary format. Early consideration (at the problem formulation step) of how evidence may be grouped and how potential patterns of data across evidence streams may be assessed can provide insight on where best to focus resources relative to the approach that will ultimately be used for evidence integration. Evidence synthesis typically involves developing a narrative based on evidence tables that presents a summary of the results in the context of the objectives of the assessment (for example, for hazard identification or a full risk assessment, the PE(C)O). Data are grouped on the same or similar study subjects,<sup>12</sup> end-points, outcomes or health effects across studies within each stream of evidence. The synthesis is generally more straightforward with animal and (to a certain extent) human studies, as the data are more likely to be similar across studies, and more complex with mechanistic data, which are often very heterogeneous. As such, some considerations when synthesizing mechanistic data for hazard identification are presented below.

### 6.2.1 Considerations when synthesizing mechanistic data

Mechanistic data are generally focused on end-points that precede the manifestation of health effects and as such can be biomarkers related to the health effects of interest. These data can be observed at various levels of biological complexity, including molecular or cellular responses upstream of or distant from the health effect or disease. Depending on the objectives of the assessment, mechanistic data may be utilized to characterize biological plausibility, characterize a mode of action, or address potential uncertainties within or across evidence streams (for example, species differences). Mechanistic data could be used as the evidence base on which to develop conclusions in the absence of human or animal studies, although resulting conclusions might have much less certainty because the data are on end-points upstream of clear health effects. Relevant mechanistic data provide valuable information for the assessment that may be identified from a wide variety of study designs, including in vivo, in vitro or in silico studies using human samples or data from animal models. A well defined conceptual model is essential to guide the approach for identification, assessment, synthesis and use of relevant mechanistic data needs, which has to be prespecified in the protocol. Given the large and growing number of publications with mechanistic information, it is very useful to consider these data stepwise depending on the human or animal evidence identified or use other constructs (such as adverse outcome pathways (AOPs) or key characteristics) to prioritize and focus resources (for example, if human evidence of nephrotoxicity is identified, then data on potential mechanisms for nephrotoxicity will be identified and assessed). To date, their greatest utility in the context of hazard-based systematic reviews is in characterizing biological plausibility for the health effect of interest. Mechanistic data can also inform species differences, dose-response, or characterize mode of action – elements that are important considerations in evidence integration analyses. Mechanistic data have also been used to identify and link biomarkers of exposure, effect and susceptibility. In environmental health evaluations, mechanistic data are often assessed separately (within their own evidence stream) and are subjected to critical appraisal of the internal validity (for example,



**Mechanistic data:**  
**Results from observations on cellular, biochemical, molecular, and toxicokinetic and dynamic mechanisms that inform the process of disease. Such observations may be early indications of response and related to adverse outcomes and end-points.**

<sup>12</sup> Note: for exposure, the study subjects can also be food items.

assessment of methodological quality) on an individual study or experiment basis, and external validity (generalizability or applicability of dose and model system to human and animal evidence), as part of developing conclusions. To facilitate synthesis of mechanistic data (or more broadly, patterns of evidence across evidence streams), organization using adverse outcome pathways, key characteristics or similar construct can be helpful. Generally, data are organized around the biology of health effects or diseases relevant to the review topic. In some cases, mechanistic data can be used to integrate similar toxicological observations across organ systems and species. If there is a well established a priori understanding of the biological significance underlying the high confidence human evidence, then further analysis of mechanistic data may not be necessary.

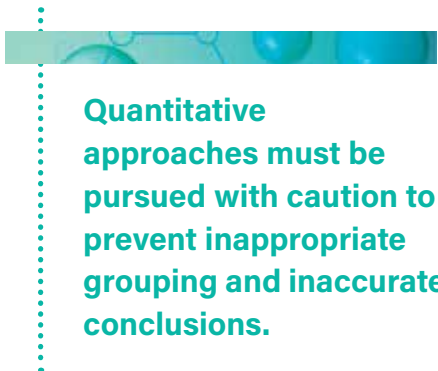
### 6.2.2 Qualitative or quantitative approaches to evidence synthesis

Generally, the synthesis presents the data by category, often in tabular format relevant to the research question, which may be organized by evidence stream (for example, human versus animal), by end-point within a health outcome (for example, heart rate versus cardiovascular morbidity as part of an overall cardiovascular health outcome), perhaps by study model or prevalence or dose (as it relates to environmental exposure), or by food item (for example, for consumption data). When possible, for hazard identification it is helpful to synthesize these elements using a parallel format for animal and human evidence on a particular outcome (such as direction of effect, magnitude and dose–response relationship of the findings) to prepare bodies of evidence for next steps in the process (assessing certainty and integrating evidence). An important goal is to organize and display the results to explore whether results are reasonably homogeneous or to explain the sources of heterogeneity in the results.

Evidence syntheses via quantitative or qualitative approaches will depend on the characteristics of the data within the evidence base. The degree of heterogeneity within the body of evidence will dictate whether a qualitative (narrative) or quantitative (for example, meta-analysis) synthesis of the data is developed. Given the diversity in exposure metrics, animal models, end-points and outcome assessment methods for studies considered in most chemical risk assessments, it is likely that many data sets will not lend themselves to formal meta-analyses, and thus a narrative synthesis of the data would be appropriate. Both toxicokinetic and toxicodynamic models can assist in the evaluation of evidence streams across assay platforms (for example, in vivo, in vitro and in silico). Specific toxicokinetic programmes allow for extrapolation across in vivo and in vitro situations (for example, in vitro to in vivo extrapolation, IVIVE), and this can support more quantitative approaches. It is of note that some researchers may make a distinction between qualitative and narrative synthesis, with the former referring to the synthesis of qualitative evidence and the latter describing the synthesis of quantitative evidence using non-statistical approaches (1). In the risk assessment community (and in this document), qualitative and narrative syntheses are used interchangeably and refer to the synthesis of any type of data in the evidence base using non-statistical approaches.

### 6.2.3 Quantitative approaches

If data are amenable, quantitative synthesis techniques are encouraged and may involve a variety of approaches depending on the type of evidence. Many advancing techniques allow for combined data analyses beyond standard meta-analyses of observational data in human studies. These can include categorical regression, Bayesian methods for combined evaluation of dose–response, multicriteria decision analysis, expanded meta-analytical techniques such as meta-regression, and machine learning techniques. Quantitative methods for evidence synthesis allow accounting for some sources of



**Quantitative approaches must be pursued with caution to prevent inappropriate grouping and inaccurate conclusions.**

uncertainty already in the synthesis process. For example, in meta-analyses, studies are weighted according to their precision, and sensitivity analyses and exploratory subgroup analyses can account for bias by restricting the analysis to studies of high validity. Bias-adjusted meta-analysis allows the quantification of some uncertainties arising from the individual studies (such as risk of bias and indirectness). In addition, subsets of data, each with different potential biases (for example, population-based versus hospital-based case-control studies), can be conducted to inform triangulation approaches. However, other sources of uncertainty, such as unexplained inconsistency of results and publication bias, cannot be accounted for in the meta-analysis and must be addressed in other ways (for example, see subsection 6.2.5 below). Quantitative approaches must be pursued with caution, as inappropriate grouping of end-points for analysis can result in inaccurate conclusions. Consultation with experts (such as statisticians or subject matter experts) to determine the appropriateness of grouping end-points will offer important insight when pursuing quantitative approaches for evidence synthesis.

### 6.2.4 Qualitative approaches

When data do not lend themselves to quantitative analysis, a qualitative synthesis should be pursued. Structured, criteria-based methods are preferred (for example, data grouped and plotted using graphic visuals such as forest plots), as they help interpret and understand the trend of the results. However, qualitative syntheses may have limited usefulness in quantifying risk, and therefore quantitative syntheses should be pursued whenever feasible.

The approach pursued for synthesis will depend both on the overall heterogeneity of the evidence base and the consistency of the findings. Many factors can contribute to heterogeneity within environmental health data used for chemical risk assessment, and consideration of these factors will determine the best approach for synthesizing evidence within an evidence stream.

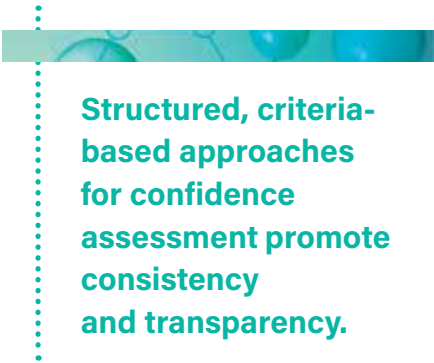
- **Population.** Considerations include occupational, population-based sample, children, adults, sex, species or strain, life stage.
- **Study design.** Data that are used to address environmental health questions comprise several types of evidence. Human studies can consist of a variety of study designs that include cohort, case-control and cross-sectional studies. Experimental animal data can also vary greatly depending on animal models used and the specific question being addressed for that study. These aspects are important to consider within and across the evidence base.
- **Exposure.** Factors related to exposure, such as the route of exposure, the timing (for example with regard to life stage) or duration of exposure, the dose level or dose range, may also vary across studies that assessed the same or similar end-points or outcomes. A related consideration is the age of the population at the time of the exposure or the time of the outcome assessment. Exposure measurements can also have an impact (for example, whether exposure was measured directly or via proxy; objectively measured or self-reported; or measured at the level of the individual or the environment).
- **Data analysis.** Data across relevant studies within an end-point group may be presented as continuous or dichotomous values, or may vary by route or metric (such as oral dose, dose in feed), and thus often need to be standardized or accommodated for.
- **Internal validity (when bias-adjusted meta-analysis is not possible).** Differences in the confidence in the relationship between the exposure and end-point due to internal validity (risk of bias) or other study quality features can vary across studies with an end-point group and can contribute to heterogeneity of findings.

The outcome of the evidence synthesis step, whether qualitative or quantitative, is a summary of the results within each stream, by end-point or cluster of end-points, or exposure outcome (such as consumption outcomes). Even when the data are sufficiently similar to support a quantitative analysis, inclusion of a

qualitative summary may be necessary to transparently describe findings. The combined use of qualitative and quantitative methods also more readily allows for sensitivity analyses and more readily facilitates assessing confidence in the evidence, as well as assessment of overall uncertainty in the assessment, an important component of risk characterization. The synthesis step also can facilitate further evaluation of some risk of bias domains (for example, confounding for epidemiological evidence) by checking whether expectations about impacts on the direction of bias were realized across a set of study results. If results of studies predicted to have more serious impacts from a specific bias are comparable to study results not expected to be affected, then the specific bias may not have been operating to the degree predicted.

### 6.2.5 Confidence assessment

Once the data are organized, a confidence assessment is conducted. Though various approaches exist, a structured, criteria-based approach is generally considered best practice because it promotes consistency and transparency. It is also important to emphasize that expert judgement is critical when applying structured approaches so that the structure does not become an algorithm that forces reviewers to reach conclusions that they do not consider are supported by the data. The GRADE (Grading of Recommendations Assessment, Development and Evaluation) framework (2) is a widely adopted approach that incorporates and expands on the Hill considerations on causality (3, 4). It is used in a modified form for hazard identification in environmental health by the Office of Health Assessment and Translation of the United States National Toxicology Program (5) and the Navigation Guide (6). Modified Hill considerations are also incorporated in other structured qualitative approaches (such as the WHO/International Programme on Chemical Safety (IPCS) framework on mode of action/species concordance analysis) aimed to increase transparency and consistency in integrating evidence for testing hypotheses of modes of action (or adverse outcome pathways) and, in turn, inform the risk assessment process (7). Confidence in bodies of evidence is assessed for the following characteristics.



**Structured, criteria-based approaches for confidence assessment promote consistency and transparency.**

- **Internal validity.** This addresses the degree to which bias or a systematic error or deviation of the truth in the results is minimized in the studies of interest; it should be considered when assessing confidence in the body of evidence (can also be referred to as risk of bias; see Chapter 5).
- **Consistency.** Different studies, populations, species, etc., showing similar findings would increase confidence.
- **Precision.** Evidence of a significant difference between exposed and unexposed groups (may be presented as statistical significance  $P < 0.05$  or biological significance) would increase confidence; a narrow confidence interval around a point estimate would also increase confidence.
- **Directness.** A measure of how well the end-points or outcomes address the prespecified PECO; data from more relevant end-points would increase confidence.
- **Publication bias.** Unequal representation of findings from the literature would decrease confidence.
- **Dose-response.** Simple (linear) or complex (non-linear) relationships increase confidence, while expected but missing dose-response relationships decrease confidence.
- **Strength of association.** Large or severe effects increase confidence (if there is also evidence of precision).

For hazard identification, confidence assessment should be conducted separately within evidence streams. In assessing the confidence, it is important to clarify the confidence as it relates to subsets of data (as appropriate). For example, in a toxicological evidence base, it is common to observe heterogeneous findings across end-points, doses, species, etc., and thus characterizations of confidence may be most meaningful when applied both to subsets of data and to larger groups of data.

The result of the confidence in the evidence step is a clear statement as to the confidence in the bodies of evidence in the context of the PECO. Conclusions about confidence should be described for each of the bodies of evidence included in the evidence synthesis (that is, on the same outcome or health effects basis) for the human and animal evidence.

## 6.3 Evidence integration

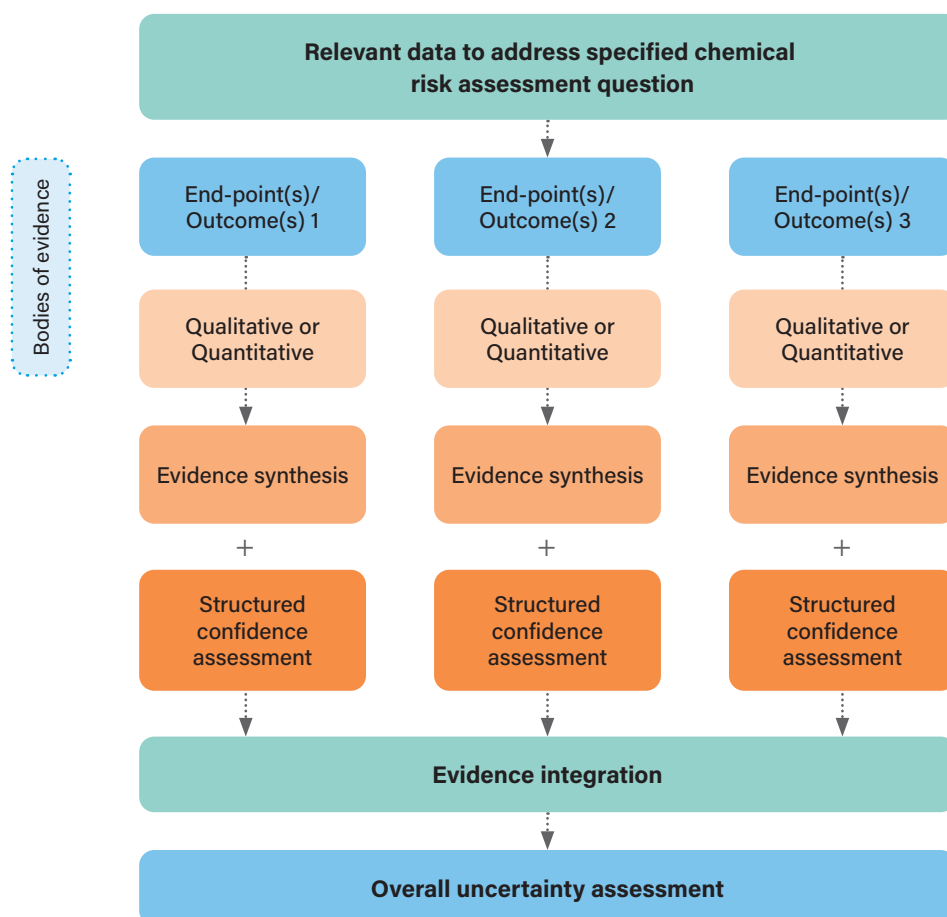
Integration is a concept more routinely used in systematic reviews in the context of hazard identification of chemicals as, unlike reviews in clinical medicine, the underlying toxicological evidence bases are composed of different types of evidence that may be reported in human (typically observational) studies, in experimental animal studies, and in mechanistic studies (often in vitro exposure studies, though also derived from in vivo studies in humans and experimental animals). The evidence integration process across evidence streams for hazard identification allows for the combination of all available evidence relevant to the review question to develop conclusions based on the totality of data. In some cases, the synthesis and confidence of evidence steps will have addressed outcomes in both human and animal evidence streams. In others, there may only be animal or mechanistic studies. In each case, the integration process considers the findings described in the qualitative or quantitative data synthesis and the certainty of the evidence for each outcome by evidence stream to determine conclusions that directly address the PECO. The specific approach for integration will be prespecified in the protocol but may involve a matrix-based approach (for example, a matrix describing how the confidence in animal and human bodies of evidence are combined to reach different hazard conclusions) or techniques for eliciting expert knowledge. In general, higher confidence in the human and animal bodies of evidence results in stronger conclusions. The use of mechanistic data is particularly notable; they are often used to support evaluation of biological plausibility associated with hazard conclusions or extrapolation approaches in dose–response assessments.

If the aim of the assessment, defined at problem formulation, is to estimate a parameter and a mathematical model is used, uncertainties related to each factor can be combined using a deterministic or probabilistic approach (see section 3.2 for information on how problem formation can outline methods that can aid in characterizing uncertainties for a given chemical risk assessment question).

Taken together, the grouping of data by outcome (either a health outcome or an exposure outcome) (and, for health outcomes, further by end-point when applicable), the synthesis of data within each outcome, and integration of outcome evidence across evidence streams provides a stepwise approach for synthesizing data and integrating evidence from diverse and complex data sets to address environmental health questions (Figure 6.1). This approach also provides a transparent process for reaching conclusions that can support hazard characterization when considered with dose–response information, or support risk characterization when considered with exposure information.

The interpretation of the results should transparently outline how and why such decisions were made, gaps in knowledge, and the potential impact of new or additional data on the conclusions reached. Limitations of the review should be described, such as decisions to restrict the evaluation to human studies only with the implications of excluding potential animal data. Limitations of the evidence base should also be addressed, such as few studies available, or heterogeneous bodies of evidence, or no published data on a key outcome. Finally, the exposure context of the conclusions should be described. If the objectives of the review are to complete a full risk assessment, exposure and dose–response will be considered; depending on the research question outlined in the protocol, an exposure context may still be included, even if the conclusions are for hazard characterization or other purposes.

**Figure 6.1 Workflow in evidence synthesis and evidence integration**



*Note:* End-points or outcomes in a body of literature relevant to addressing a chemical risk assessment question can include in vivo or in vitro data from human and animal studies, or relevant data from exposure studies. These groupings of end-points or outcomes (that is, bodies of evidence) undergo qualitative (narrative) or quantitative syntheses, depending on the degree of heterogeneity, and an assessment of confidence. The end-point or outcome bodies of evidence are then integrated to reach conclusions and overall uncertainty in the evidence is assessed.

## 6.4 Overall uncertainty assessment

Uncertainty analysis is an integral and fundamental component of the scientific assessment process (8–13). Uncertainty refers to “all types of limitations in the knowledge available to the risk assessors at the time an assessment is conducted and within the time and resources available for the assessment” (12, 13). Uncertainty analysis is the process of identifying limitations in scientific knowledge (that is, data gaps) and evaluating their implications for scientific conclusions, if possible, in terms of the possible range and probability of possible answers to the assessment question (which is often a quantitative estimate of risk or a health-based benchmark to assess potential risk). In systematic review, confidence in individual studies is assessed through risk of bias (or similar), and the body of evidence is typically assessed through structured analyses. The output from these types of approaches can also be used to inform the qualitative characterization of uncertainty in risk assessment. For example, if a candidate study used to develop a toxicity value had a low risk of bias and was part of a body of evidence that exhibited consistency, directness, evidence of a dose–response, and other qualities, the resulting uncertainty related to the risk assessment may be low, and vice versa. The systematic review approach also facilitates quantitative uncertainty characterization in



various capacities. These include identification of areas or parameters that have low confidence and warrant investigation of sensitivity in the risk assessment. Such approaches may include bias-adjusted analyses to assess the impact of a given risk of bias domain on the findings. If meta-analytical approaches have been utilized, the variability and heterogeneity assessments associated with such approaches also contribute to a quantitative uncertainty evaluation of the risk assessment.

## 6.5 Conclusions for hazard identification or risk assessment

Systematic review principles in chemical risk assessment are most commonly applied in the context of developing hazard conclusions and identification of end-points suitable for dose-response analyses. The decision to conduct a full systematic review in a chemical risk assessment is determined during the problem formulation process (see Chapter 3) and may be one of many analyses or the only analysis that contributes to the overall risk assessment findings (Figure 6.2). Pending the objectives of the overall assessment, the data syntheses and integration may be used to facilitate the risk assessment process by going beyond a hazard conclusion and characterizing dose-response relationships for the purposes of developing a health-based benchmark, such as an acceptable daily intake (ADI) or similar representative value of exposures without adverse effects. The comprehensive appraisal of individual studies, combined with the evidence synthesis and characterization of certainty findings within the bodies of evidence, can assist in selection of candidate studies for developing health-based benchmarks for each outcome. The appraisal of individual studies, which may serve as a basis for deriving a reference dose, is recommended by the National Research Council (14), as reference values should only be set on studies with sufficient quality. In addition to having confidence in the individual study validity, its representativeness for the given outcome relative to the overall evidence base (and hazard-based conclusions, where applicable) can be considered. The use of quantitative techniques, such as meta-regression, greatly facilitate this process by eliminating the need to select single studies, thereby utilizing more of the available data to support development of health-based benchmarks. Once pivotal studies are selected and the critical experimental dose (that is, the so-called point of departure or starting point, such as a no observed adverse effect level (NOAEL) or a benchmark dose (BMD)) is determined, the uncertainty or safety factor selection can also be informed by the systematic review. Occupational exposure limits are the equivalent in the occupational health context and could also be determined based on systematic review. Depending on the outcome and study model, the evidence from the review may also facilitate decisions related to extrapolation approaches based on mode of action, adverse outcome pathways or key characteristics.

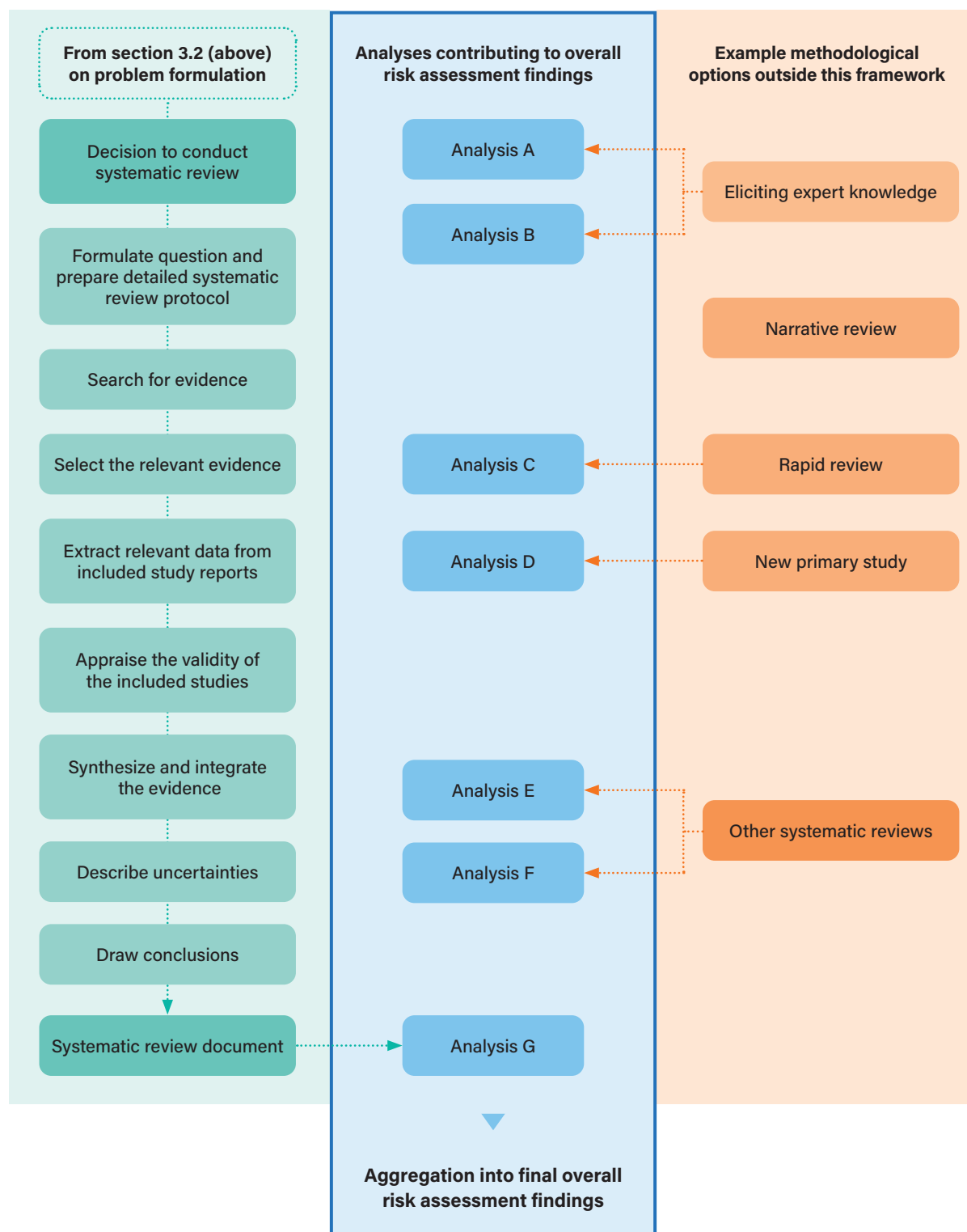


### Risk assessment

#### The four steps in risk assessment:

**hazard identification, hazard characterization, exposure assessment and risk characterization.**

**Figure 6.2 Schema outlining how systematic review analyses can contribute to overall risk assessment findings**



*Note:* The fit between conduct of a systematic review, as determined by the problem formulation process described in Chapter 3, and other analyses by different methods (including additional systematic reviews) in coming to an overall conclusion in a complete risk assessment process. In many cases, there may not be different types of analysis methods implemented in an assessment (for example, a systematic review may be the only analysis method implemented to facilitate risk assessment).



In summary, data can be synthesized using qualitative or quantitative approaches depending on how the data are organized and features of the bodies of evidence, with most data sets in environmental health supporting qualitative syntheses. Within each evidence stream, similar end-points or measures within a health or exposure outcome should be evaluated together and then integrated across evidence streams when appropriate. The synthesis of the available evidence that critically assesses, summarizes, and possibly quantifies the strengths and weaknesses of the studies included in the review represents a critical step in the systematic review process and provides a transparent, structured and standardized approach for decision-making in environmental health.

## References

1. Popay J, Roberts H, Sowden A, Petticrew M, Arai L, Rodgers M et al. Guidance on the conduct of narrative synthesis in systematic reviews: a product from the ESRC Methods Programme. Version 1: April 2006 (<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.178.3100&rep=rep1&type=pdf>, accessed 4 May 2021).
2. Morgan RL, Beverly B, Gherzi D, Schünemann HJ, Rooney AA, Whaley P et al. GRADE guidelines for environmental and occupational health: a new series of articles in *Environment International*. *Environment International*. 2019;128:11–12. doi:10.1016/j.envint.2019.04.016.
3. Hill AB. The environment and disease: association or causation? *Proceedings of the Royal Society of Medicine*. 1965;58:295–300.
4. Schünemann H, Hill S, Guyatt G, Akl EA, Ahmed F. The GRADE approach and Bradford Hill's criteria for causation. *Journal of Epidemiology and Community Health*. 2011;65:392–5.
5. Rooney AA, Boyles AL, Wolfe MS, Bucher JR, Thayer KA. Systematic review and evidence integration for literature-based environmental health science assessments. *Environmental Health Perspectives*. 2014;122:711–8.
6. Woodruff TJ, Sutton P. The Navigation Guide systematic review methodology: a rigorous and transparent method for translating environmental health science into better health outcomes. *Environmental Health Perspectives*. 2014;122(10):1007–14. doi:10.1289/ehp.1307175.
7. Meek ME, Boobis A, Cote I, Dellarco V, Fotakis G, Munn S et al. New developments in the evolution and application of the WHO/IPCS framework on mode of action/species concordance analysis. *Journal of Applied Toxicology*. 2014;34(1):1–18. doi:10.1002/jat.2949.
8. International Programme on Chemical Safety. Guidance document on evaluating and expressing uncertainty in hazard characterization, second edition. Harmonization Project Document No. 11. IPCS Project on the Harmonization of Approaches to the Assessment of Risk from Exposure to Chemicals. Geneva: World Health Organization (within the framework of the Inter-Organization Programme for the Sound Management of Chemicals); 2018 (<https://apps.who.int/iris/handle/10665/259858>, accessed 4 May 2021).
9. Group of Chief Scientific Advisors. Scientific advice to European policy in a complex world. Scientific Opinion No. 7, September 2019. Brussels: European Commission, Directorate-General for Research and Innovation; 2019 ([https://ec.europa.eu/info/sites/info/files/research\\_and\\_innovation/groups/sam/ec\\_rtd\\_sam-scientific-advice\\_092019.pdf](https://ec.europa.eu/info/sites/info/files/research_and_innovation/groups/sam/ec_rtd_sam-scientific-advice_092019.pdf), accessed 4 May 2021).
10. Making sense of science for policy under conditions of complexity and uncertainty. Berlin: Science Advice for Policy by European Academies; 2019 (<https://doi.org/10.26356/MASOS>, accessed 4 May 2021).

11. European Food Safety Authority. Approaches to assess and manage scientific uncertainty: examples from EU Agency Network for Scientific Advice agencies. Brussels: European Union; 2018 (<https://publications.europa.eu/en/publication-detail/-/publication/9880c8bc-83eb-11e8-ac6a-01aa75ed71a1>, accessed 4 May 2021).
12. EFSA Scientific Committee, Benford D, Halldorsson T, Jeger MJ, Knutsen HK, More S, Naegeli H et al. Guidance on uncertainty analysis in scientific assessments. *EFSA Journal*. 2018;16(1):39.
13. Benford D, Halldorsson T, Jeger MJ, Knutsen HK, More S, Naegeli H et al. Scientific opinion on the principles and methods behind EFSA's guidance on uncertainty analysis in scientific assessment. *EFSA Journal*. 2018;16(1):e05122.
14. National Research Council. Review of the Environmental Protection Agency's Integrated Risk Information System (IRIS) process. Washington (DC): National Academies Press; 2014.



# 7 Expectations for the reporting of systematic reviews

Systematic reviews, as is the case for all scientific publications, should be comprehensively reported to render their methods and findings as transparent as possible to the reader.

Detailed reporting allows a reader to understand the basis of the scientific judgements made in the course of conducting a systematic review (such as the reasons for rating studies during critical appraisal, or evaluating the overall certainty of the evidence), determine their own confidence in the conclusions that the authors have drawn from the evidence they have synthesized, and assess the applicability of those conclusions to their own decision-making situation.

The highest levels of transparency are not possible unless, prior to conduct, the methods for a systematic review are recorded as a protocol and published in a public database, and that protocol is then followed with any deviations from planned methods being appropriately justified. Expectations for reporting therefore cover both the planned methods for a systematic review and the final systematic review itself.

Since systematic reviews are complex projects consisting of many steps, reporting standards are similarly complex and lengthy. However, because reporting standards follow the structure of a systematic review project, they do help simplify the planning and reporting process, effectively functioning as a structured checklist of tasks that need to be completed for comprehensive final reporting of a systematic review. Reporting standards also function as templates for writing up a systematic review – if followed, they make the complex job of reporting interconnected methods much easier to accomplish.

A complication is that a chemical risk assessment will often have to be reported using a format mandated by legislative or regulatory requirements. A chemical risk assessment that has used systematic approaches will then have to apply the following reporting expectations within the required reporting format.

The reporting expectations that are summarized here can be considered the minimum elements necessary to ensure that requirements for transparency and confidence in the conclusions of the systematic review are met. The table is followed by useful resources that facilitate comprehensive, detailed reporting of systematic reviews, as would be required for publication in scientific journals.

**Table 7.1 Systematic review: reporting expectations and explanations**

Protocol and methods

Final systematic review manuscript

Expectation	Explanation
<b>Authorship</b> ● ●	
<ul style="list-style-type: none"> <li>List all contributors</li> <li>Describe sources of support</li> <li>Declare interests of both</li> </ul>	Readers of a systematic review need to know who contributed to the project to be able to put the research project in its full context. Systematic review authors should therefore list and describe the roles of everyone involved in the planning, conduct and decision-making of a systematic review project. Contributors and sources of support (both financial and in kind), and their interests in the review, should be listed for both transparency and to allow them to be credited for their work.
<b>Introductory matter (Chapter 3)</b> ● ●	
<ul style="list-style-type: none"> <li>Identify the document as a systematic review</li> <li>State rationale and context</li> <li>Present objectives (PECO or appropriate)</li> </ul>	Readers need to be able to quickly identify whether a given systematic review is relevant to their decision-making context. A systematic review should therefore clearly identify in its title what it is investigating and that it is a systematic review. The registration number and location of the pre-published protocol should be provided, the reason for conducting the review, and clearly formulated objectives should also be presented.
<b>Search strategy (Chapter 4)</b> ● ●	
<ul style="list-style-type: none"> <li>Describe information sources (databases, contact with study authors, search engines, etc.)</li> <li>Present search strings used in databases</li> </ul>	The reader of a systematic review must be able to reproduce the search strategy that the authors of the systematic review employed to find information of potential relevance to their research objective. Systematic review authors should therefore report which information sources they searched and the methods for searching them, including full search strings for all databases.
<b>Selection (Chapter 4)</b> ● ●	
<ul style="list-style-type: none"> <li>State eligibility (inclusion and exclusion) criteria</li> <li>Describe method for screening</li> </ul>	Because a search strategy yields evidence of only potential relevance to the review, systematic review authors must then describe how they selected for inclusion in their systematic review all the evidence of actual relevance to their research objectives. The authors should state their inclusion criteria and describe how they screened the search results.

Expectation	Explanation
<b>Data collection and extraction (Chapter 4)</b> ● ●	
<ul style="list-style-type: none"> <li>List all data items</li> <li>Describe method of extraction</li> <li>Describe data storage software</li> </ul>	For readers to understand how relevant data from included studies is appropriately transferred into the systematic review for synthesis, systematic review authors should define all the variables for which data were sought, describe how data extraction was conducted, and explain how data from the included studies were captured and summarized for analysis. This would include presenting the data extraction forms, details about which authors extracted data, how extraction was quality controlled, and the software used for storing the extracted data.
<b>Appraisal of individual studies (Chapter 5)</b> ● ●	
<ul style="list-style-type: none"> <li>Describe methods for assessing validity of individual included studies</li> <li>Describe how validity assessment will be incorporated into any data synthesis</li> </ul>	Critical appraisal of the evidence is a fundamental element of communicating to the reader how certain they can be about the overall findings of an evidence synthesis. Authors should describe the criteria used for distinguishing studies of greater and lesser validity. When applying the criteria, the authors should explain how they came to their judgements and quote relevant text from the study they are appraising. The procedures for appraising study validity (for example, duplicate assessment) should also be made clear. Traffic-light diagrams can be useful for presenting the results of the assessment. Finally, how the results of the assessment of the validity are incorporated into the data synthesis should be described (for example, by sensitivity analysis, assessment of certainty in results).
<b>Synthesis and integration (Chapter 6)</b> ● ●	
<ul style="list-style-type: none"> <li>Present the principal summary measures</li> <li>Describe the statistical and qualitative techniques for combining studies</li> <li>Describe methods for assessment of characteristics of cumulative evidence relevant to interpreting results (certainty or confidence assessment)</li> <li>If conducted, describe the methods for integrating multiple streams of evidence</li> </ul>	Authors should describe the criteria they will use to decide when study results can legitimately be pooled, the measures that will summarize such pooled results (for example, standardized mean difference, risk ratios), and the statistical methods used to generate them. For results that cannot legitimately be pooled, any qualitative assessment methods should be described. Methods for characterizing features of the cumulative body of evidence should also be described, such as appropriate tests for heterogeneity or publication bias. The methods by which multiple streams of evidence are to be appraised in terms of the extent to which they reinforce or contradict each other should, if conducted, also be presented.

Expectation	Explanation
<b>Results (Chapter 6)</b> ●	
<ul style="list-style-type: none"> <li>Describe the results of the study selection process, including a flowchart and list of studies excluded at full-text screening</li> <li>Present a table of the key characteristics of each included study</li> <li>Present the pooled summary results and results of assessment of cumulative characteristics of the evidence</li> <li>Describe, using a structured framework, the results of any qualitative analysis</li> <li>Present the results of the integration of the evidence streams</li> </ul>	<p>The results of each step of the systematic review process following the methods planned in the above sections should be carefully reported. To make the selection process transparent, authors should present a flowchart showing how much irrelevant evidence was excluded at each stage of the screening process (number of studies in search results, excluded in title or abstract screening, excluded at full-text screening). For studies excluded during full-text screening, the reason for exclusion should be reported. The characteristics of each of the included studies should be summarized in enough detail to allow the reader to follow the evidence synthesis process, and pooled results should be clearly presented using (for example) forest plots and other appropriate visual aids.</p>
<b>Interpretation (Chapter 6)</b> ●	
<ul style="list-style-type: none"> <li>Summarize the main findings</li> <li>Present the certainty of the evidence for main findings</li> <li>Describe limitations in the systematic review approach</li> </ul>	<p>One of the most important functions of a systematic review is explaining to the reader how certain they can be in its summary results. To achieve transparency, authors should first present the main results of the review accompanied by a description of how the cumulative characteristics of the evidence (such as heterogeneity and overall risk of bias) should affect the reader's confidence in those summary results. This should be supported by a summary of results table to facilitate interpretation of what can be a complex body of narrative text. Second, limitations in the review methods that could affect its results and the authors' interpretation of them should also be made clear.</p>

## Resources to help with reporting systematic reviews

There is plentiful guidance on how to report systematic reviews of medical research. One of the most cited is the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA). To date, there is little specific guidance on reporting of systematic reviews for chemical risk assessment and environmental health research. The following may nonetheless be useful.

- **PRISMA reports adapted for environmental health research.** In 2017, the journal *Environment International* made minor adaptations to the PRISMA reporting standard to fit its requirements for systematic review submissions that address environmental risks and hazards, including potential effects of exposure to chemical substances on human health. The reports can be downloaded from the journal website<sup>13</sup>. The original set of PRISMA reporting standards can be downloaded from the PRISMA website<sup>14</sup>.
- **ROSES reports for environmental research.** In 2018, the Collaboration for Environment Evidence endorsed the *RepOrting standards for Systematic Evidence Syntheses in environmental research* (ROSES) reports as a template for reporting systematic reviews. While these are not tailored specifically to chemical risk assessment and toxicological research, they cover a broader range of approaches to conduct a systematic review and may therefore be preferable to some authors. They are available from the ROSES website<sup>15</sup>.
- **NTP OHAT templates for reporting components of a systematic review.** The updated 2019 Office of Health Assessment and Translation of the United States National Toxicology Program (NTP OHAT) *Handbook for conducting systematic reviews for health effects evaluations* includes a series of exemplar reporting templates for search strategies, risk of bias assessments, data extraction summary tables, and others. The handbook is available from the NTP OHAT website<sup>16</sup>.
- **For search strategies.** The PRESS *Peer Review of Electronic Search Strategies* guideline provides a detailed checklist for reporting search strategies, and is available from the journal website<sup>17</sup>. There is also an elucidation document<sup>18</sup> that gives the reasons for each of the checklist items.
- **For comprehensive guidance on conduct and reporting of systematic reviews, with links to useful resources.** The Australian Government National Health and Medical Research Council maintains a website for health guideline developers, which includes detailed modular guidance on how to conduct a systematic review. This is in the “Develop” section of the website<sup>19</sup>.

---

<sup>13</sup> <https://www.elsevier.com/journals/environment-international/0160-4120/guidance-notes>

<sup>14</sup> <http://www.prisma-statement.org/>

<sup>15</sup> <https://www.roses-reporting.com/>

<sup>16</sup> <https://ntp.niehs.nih.gov/whatwestudy/assessments/noncancer/handbook/index.html>

<sup>17</sup> [https://www.jclinepi.com/article/S0895-4356\(16\)00058-5/fulltext](https://www.jclinepi.com/article/S0895-4356(16)00058-5/fulltext)

<sup>18</sup> <https://www.cadth.ca/press-2015-guideline-explanation-and-elaboration>

<sup>19</sup> <https://www.nhmrc.gov.au/guidelinesforguidelines/develop>





# 8 Future directions

## 8.1 Conducting systematic reviews with limited resources and the automation of systematic review

Systematic reviews are resource-intensive projects that currently take research teams of three to four people about 12–18 months to complete, even when addressing a tightly focused question and including a relatively small number of primary studies. This raises concerns about the feasibility of the methodology for contexts where resources are limited, such as for risk assessors working in low- and middle-income countries.

The issue of resource constraints is long familiar in systematic review. Most projects have to be conducted with small teams, and even relatively well resourced agencies such as the United States Environmental Protection Agency have to complete large projects using small groups of researchers (1). The focusing of questions is a familiar strategy for reducing the amount of evidence that has to be handled. However, it has been acknowledged that this strategy may not be practically available in risk assessment contexts where information requirements are often broad (2). It must also be acknowledged that even small systematic review projects may not be feasible in low- and middle-income country contexts.

Because high-quality systematic reviews are transparent and should produce summaries of risk-relevant research at minimum risk of bias, systematic reviews produced in one region should provide usable information for another. However, different regions have different specific issues, and risk estimates will often need to be recalculated for local conditions. Reducing the resources required to conduct systematic reviews is therefore a priority issue, to address both the volume of evidence that needs to be analysed, and the issues of equitable conduct of risk assessment between higher- and lower-income countries and regions.

While the work being undertaken to reduce the resource requirements for conducting systematic reviews will not bring about immediate change, it consists of three broad programmatic components that, over time, will make conducting systematic reviews more accessible across the globe.

1. **Creation of evidence maps and literature inventories.** Comprehensive understanding of what evidence is available as it relates to a research topic is critical for effective use of resources when planning research. The creation of inventories of risk-relevant scientific evidence reduces the resources required for a broad programme of systematic review-supported risk assessment by allowing resources to be reliably allocated to critical questions where systematic review will make a real difference to policy decisions.
2. **Automated systematic review.** Tools that at least semi-automate the screening step of a systematic review are already available, reducing some of the resources required for conducting a systematic review. As these improve and machine learning techniques are extended to other steps of the systematic review process, such as automated data extraction and risk of bias assessment, the resources required to conduct systematic reviews will reduce. Development of these as open-source tools will help make them affordable in low- and middle-income country contexts.
3. **Machine-readable research.** Most research is still published in electronic documents in the form of manuscripts and study reports. These are not machine readable: either a person has to spend time manually extracting data from these documents, or computers have to be taught how to read them, using laborious and often quite limited linguistic processing algorithms. Tools that facilitate complete, accurate, machine-readable reporting of research will significantly lower the amount of resources required to conduct systematic reviews.

One major remaining obstacle to making risk-relevant evidence widely available and readily reviewable is how so much research is still kept behind paywalls and not published in open access format. For systematic review to become truly accessible, and the literature inventory and automation tools to provide complete summaries of research, requires wholesale adoption of open publishing practices worldwide.

## 8.2 Evolving applications of systematic review in risk assessment

During development of this guidance, the landscape for applying systematic review to risk assessment evolved substantially owing largely to increased education and training on evidence-based techniques as well as an increase in practical applications by risk assessors. In addition to the tools that will help facilitate systematic review described above, a number of technical aspects specific to toxicology and risk assessment have been recognized as areas of interest for furthering the application of systematic review. These include the following.

- **Application of systematic review to interrogate mechanistic pathways.** Systematic review methods are being employed to identify and organize mechanistic data via pathway-based analyses. These approaches include traditional risk assessment techniques, such as mode of action (that is, chemical-specific evaluations of mechanistic pathways), as well as non-targeted organizational approaches, such as key characteristics that can be used to identify potential modes of action. Significant efforts are also being made to apply systematic methods to develop adverse outcome pathways (AOPs), highlighted by the ongoing development of guidance from an OECD working group on the matter. The application of systematic mapping and systematic review allows the risk assessor to objectively and systematically identify and characterize key events and key event relationships in mechanistic pathways that aid in assessing the risk from individual chemical stressors and specific adverse outcomes. Elucidation of mechanistic pathways via systematic methods can help to characterize the biological plausibility of a response, characterize the generalizability of non-human studies in human risk assessment (to assess the external validity), and provide important information needed to characterize and extrapolate a dose–response relationship observed in experimental animal studies to humans in risk assessment. There is significant interest in tools to help address challenges in terminology via ontologies, as well as tools to critically appraise these heterogeneous data types.
- **Application of systematic review to exposure data.** To date, most systematic review efforts have focused on application to hazard and dose–response. Recognizing, however, that the assessment of exposure is equally important to chemical risk assessment, there has been an increased emphasis on development of tools and practical experience in applying systematic review to exposure data. Authoritative agencies, including EFSA and the United States EPA (under the Toxic Substances Control Act), have conducted the most assessments to date applying these evidence-based tools to exposure data. Continued development of critical appraisal tools for these evidence types remains an intense area of interest.

## References

1. Radke EG, Yost EE, Roth N, Sathyanarayana S, Whaley P. Application of US EPA IRIS systematic review methods to the health effects of phthalates: lessons learned and path forward. *Environment International*. 2020;145:105820. doi:10.1016/j.envint.2020.105820.
2. Whaley P, Edwards S, Kraft A, Nyhan K, Shapiro A, Watford S et al. Knowledge organization systems for systematic chemical assessments. *Environmental Health Perspectives*. 2020;128(12):125001–11. doi:10.1289/EHP6994.



# Annex. WHO Chemical Risk Assessment Network

The WHO Chemical Risk Assessment Network is a voluntary collaborative initiative whose overall goal is to improve chemical risk assessment globally through facilitating sustainable interaction between institutions on chemical risk assessment issues and activities. The Network was established to enhance global efforts to assess risks to human health from exposure to chemicals. The activities of the Network promote the objectives of the Strategic Approach to International Chemicals Management (SAICM).

The Network was established in 2013, and at the end of 2020 consisted of 92 institutions involved with chemical risk assessment activities, from 52 countries.

## Network objectives

- provide a forum for scientific and technical exchange;
- facilitate and contribute to capacity-building;
- promote best practices and the harmonization of methodologies;
- assist in the identification of research needs and promote the application of new science in risk assessment practice;
- assist in the identification of emerging risks to human health from chemicals;
- share information about work programmes to avoid duplication of effort;
- upon request, assist WHO in the development of training and other materials in support of the above.

In order to inform the development of this publication, prior to drafting the text a questionnaire was sent to Network institutions to assess the level of knowledge and interest in systematic approaches within institutions undertaking chemical risk assessments. Responses were received from 15 institutions. Participant responses indicated that most institutions have an understanding of systematic review approaches and are interested in learning more, but only a small proportion of institutions are actively utilizing systematic approaches. Participants' interest was reported to be specific to research topics (for example, endocrine disruptors, food contaminants and pesticide exposure). Participants noted an interest in learning how institutions overcome challenges such as cost and time burden, as well as methods for assessing data quality. Responses to the questionnaire also indicated the need for improved methodology for systematic review approaches, standardized protocols, and training to enhance the quality and reproducibility of systematic reviews in order to implement the use of these approaches in risk assessment.





**World Health  
Organization**

**Environment, Climate Change and Health (ECH)**

World Health Organization

20 Avenue Appia

CH-1211 Geneva 27

Switzerland

[ipcsmail@who.int](mailto:ipcsmail@who.int)

9789240034488



9 789240 034488