

Holger Schünemann, MD, MSc, PhD, FRCP(C)

Chair and Professor

Department of Health Research Methods, Evidence and Impact



December 10, 2018 | NAS

Systematic Review: Best Practices and Special Considerations for Mechanistic Data

Strategies and Tools for Conducting Systematic
Reviews of Mechanistic Data to Support
Chemical Assessments

Disclosures



Guideline International Network - committees

No direct financial COI – reimbursement travel

Views expressed my own

Today

- Brief background
- Considerations about conflicts of interest
- Defining the purpose and review questions
- Inclusion and exclusion criteria
- Developing protocols
- Assessing the certainty of the evidence
- Assessing the credibility of a review

What is a Systematic Review?

Systematic reviews are scientific investigations in themselves, with pre-planned methods and an assembly of original studies as their "subjects." They synthesize the results of multiple primary investigations by using strategies that limit bias and random error*

*Mulrow CD. Ann Intern Med, 1987

A properly conducted systematic review summarizes the evidence from all **relevant** studies on the **topic of interest**, and it does so **concisely and transparently**

*Cook DJ, Sackett DL, Spitzer WO. J Clin Epidemiol 1995

Why do a systematic review?

Identify and rely on least biased data

Explore variability

Reduce random error

Inform decision-making

Inform future research

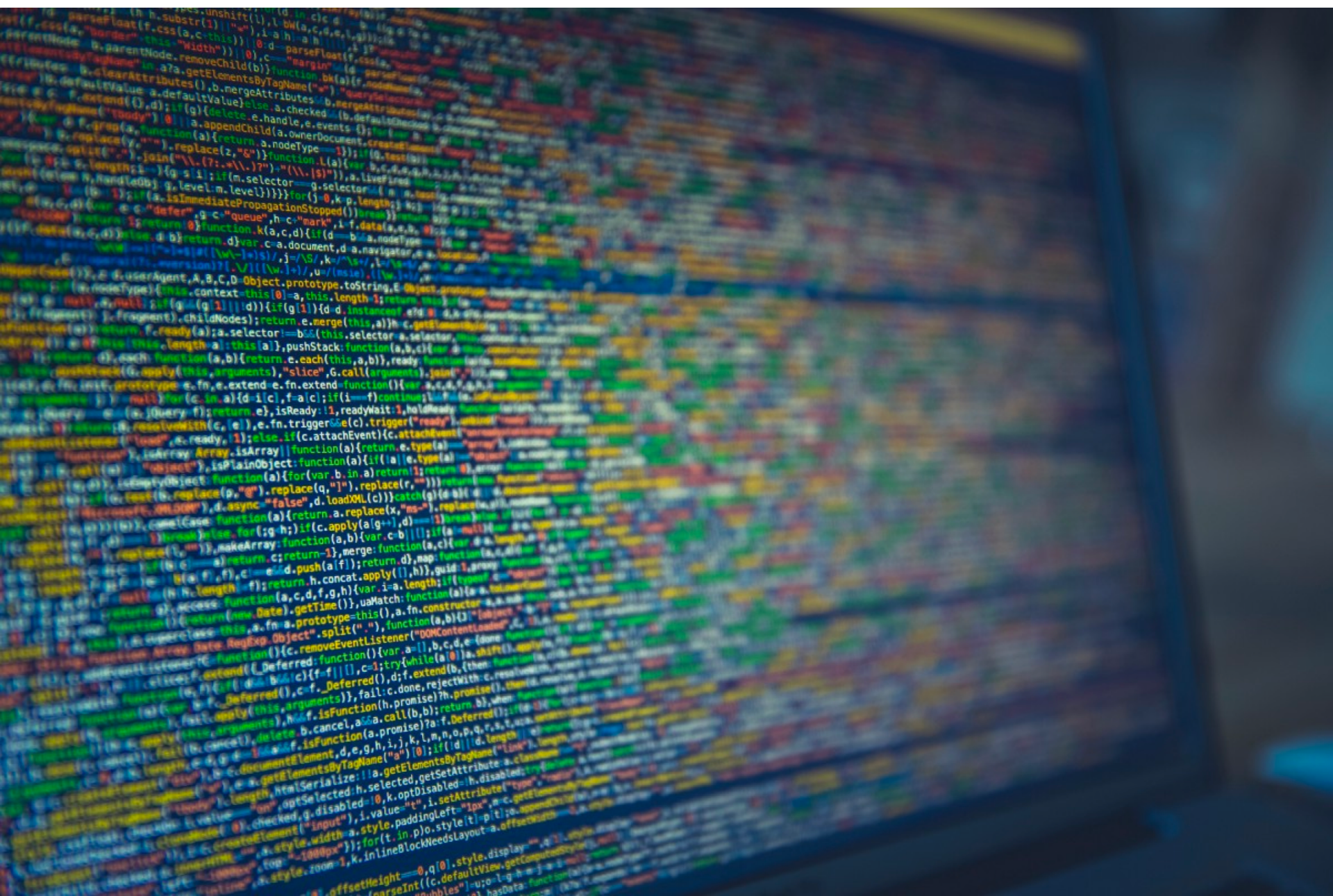
Why do we need systematic reviews?

Statistician Karl Pearson in 1904*

"Many of the groups ... are far too small to allow of any definite opinion being formed at all, having regard to the size of the probable error involved."

*Egger & Smith, BMJ, 1997

To much (irrelevant) Information

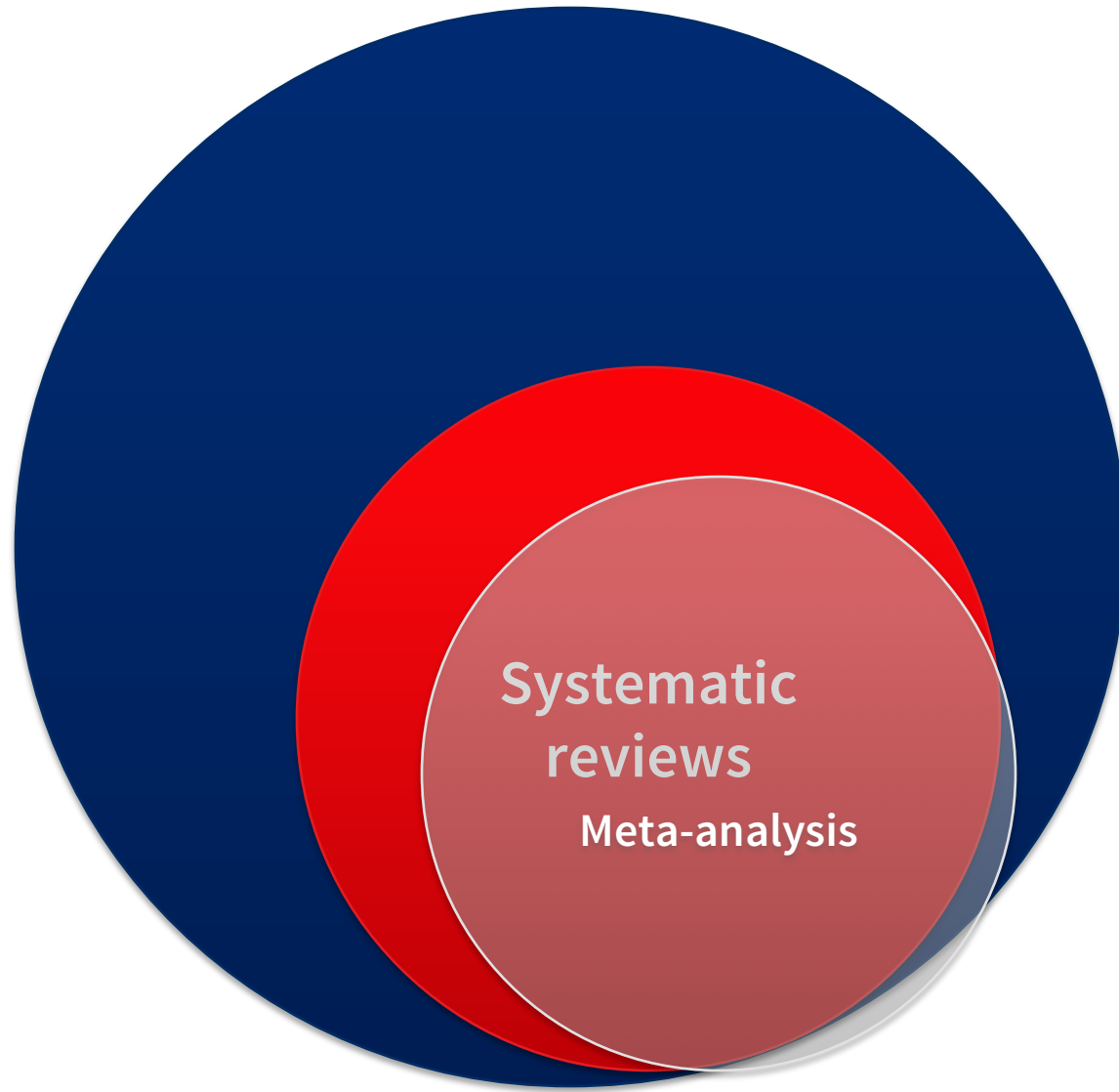


Narrative vs. Systematic Reviews

Feature	Narrative Review	Systematic Review
Question	Often broad in scope	Often a focused clinical question
Sources and search	Not usually specified, potentially biased	Comprehensive sources and explicit search strategy
Selection	Not usually specified, potentially biased	Criterion-based selection, uniformly applied
Appraisal	Variable	Rigorous critical appraisal
Synthesis	Often a qualitative summary	Quantitative summary*
Inferences	Sometimes evidence-based	Usually evidence-based

* A quantitative summary that includes a statistical synthesis is a meta-analysis.

Universe of Review Articles



METHODOLOGIC GUIDELINES FOR SYSTEMATIC
REVIEWS OF RANDOMIZED CONTROL TRIALS IN
HEALTH CARE FROM THE POTSDAM CONSULTATION
ON META-ANALYSIS

D. J. COOK,^{1*} D. L. SACKETT^{2†} and W. O. SPITZER^{3‡}

J Clin Epidemiol Vol. 48, No. 1, pp. 167–171, 1995

GUIDING PRINCIPLES

1. A systematic review must address a specific health care question. The question will determine which studies and data are relevant, and how they should be synthesized.
2. Methodology must reflect biology and the user and providers of health care. Therefore, a team with expertise in both the content area and methodology is ideally suited to conduct valid, useful systematic reviews.
3. A systematic review requires collaboration with the investigators who conducted the primary studies.

Systematic Reviews are subject to biases the way other research is.





Cochrane Handbook for Systematic Reviews of Interventions

[Español](#) [简体中文](#)

- ◆ Handbook for SRI
- ◆ Browse online
- ◆ Supplementary material
- ◆ Updates and corrections
- ◆ What's new?
- ◆ Version 6 (restricted)

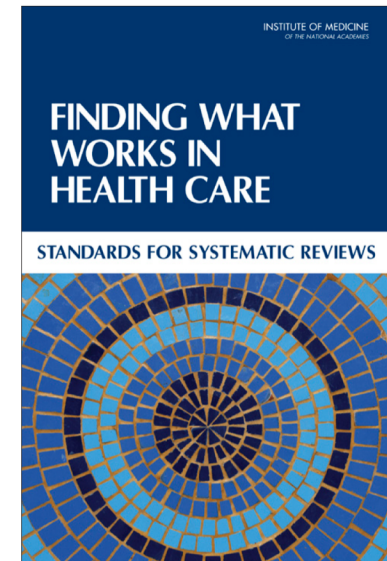
Update September 2018 - Version 6

We are pleased to announce good progress on the major update of the *Cochrane Handbook for Systematic Reviews of Interventions*, Version 6. A second edition of the book version of the Handbook will be published by Wiley in mid 2019. In addition, an online version will provide additional chapters and supplementary material. A number of chapters are either complete or nearly complete. We are making available draft PDF versions of these chapters to Cochrane members (Login required) to coincide with the Edinburgh Colloquium. A full list of Version 6 chapters and those available to view, are [here](#). Current version available below.

Finding What Works in Health Care

Standards for Systematic Reviews

These standards are for systematic reviews of comparative effectiveness research of therapeutic medical or surgical interventions





Health Assessment and Translation

OHAT Systematic Review

Completed Reports

Ongoing Evaluations

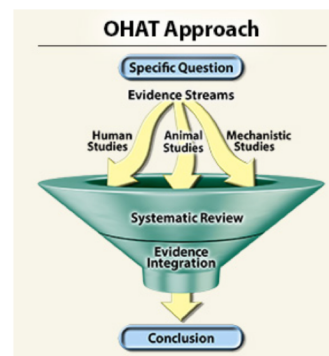
Topics Under Consideration

Nominate a Topic

Contact OHAT

OHAT Systematic Review

The NTP Office of Health Assessment and Translation (OHAT) develops literature-based evaluations to reach conclusions about potential human health hazards and to examine the state of the science. In 2012, OHAT began exploring and developing an approach for implementation of systematic review methodology to carry out these evaluations ([Birnbaum et al. 2013](#)). The output from an evaluation can vary including, but not limited to: NTP Monograph, state-of-the-science workshop report, or peer-reviewed journal publication. All evaluations follow a similar process, the OHAT Evaluation Process, with opportunities for external scientific, public, and interagency input.



SHARE THIS: [Email](#) [Facebook](#) [Twitter](#) [LinkedIn](#) [Print](#)

https://ntp.niehs.nih.gov/go/systematic_review

On This Page

- [Fact Sheet on Systematic Review](#)
- [Systematic Review Methods](#)
- [OHAT Systematic Review Publications](#)
- [Ongoing Methods Development Activities](#)
- [Past Methods Development Activities](#)

Fact Sheet on Systematic Review

See the [Fact Sheet on Systematic Review](#) to learn more about the systematic review process.

1. Formulate problem and develop protocol.
2. Search for and select studies for inclusion.
3. Extract data from studies.
4. Assess internal validity of individual studies.
5. Synthesize evidence and rate confidence in the body of evidence.
6. Translate confidence ratings into level of evidence for health effect.
7. Integrate evidence to develop hazard identification conclusions.



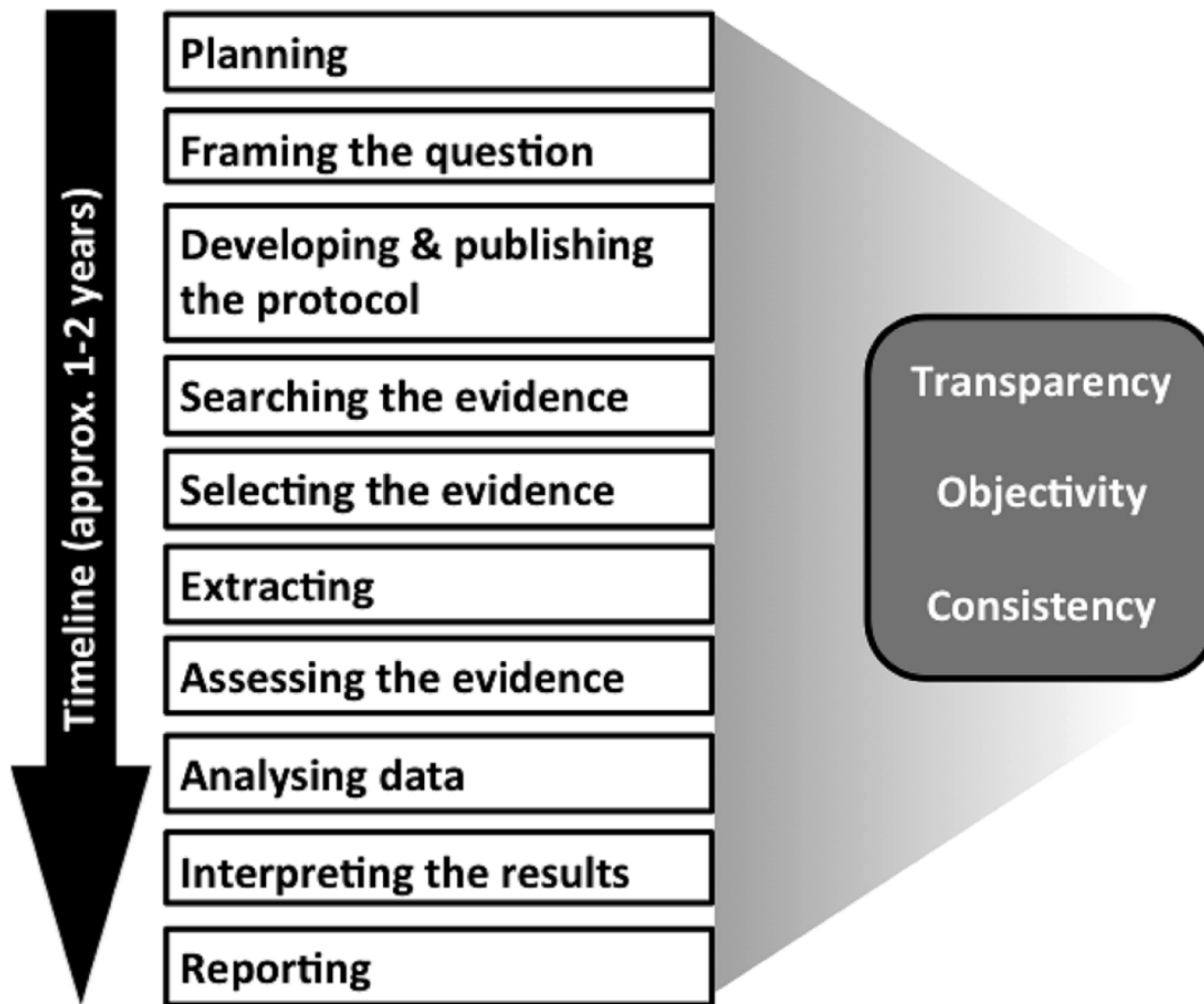


Fig. 1 Steps of a systematic review

Standards for Initiating a Systematic Review

STANDARD 2.1

Establish a team with appropriate expertise and experience to conduct the systematic review

- 2.1.1** Include expertise in the pertinent clinical content areas
- 2.1.2** Include expertise in systematic review methods
- 2.1.3** Include expertise in searching for relevant evidence
- 2.1.4** Include expertise in quantitative methods
- 2.1.5** Include other expertise as appropriate

STANDARD 2.2

Manage bias and conflict of interest (COI) of the team conducting the systematic review

- 2.2.1** Require each team member to disclose potential COI and professional or intellectual bias
- 2.2.2** Exclude individuals with a clear financial conflict
- 2.2.3** Exclude individuals whose professional or intellectual bias would diminish the credibility of the review in the eyes of the intended users

STANDARD 2.3

Ensure user and stakeholder input as the review is designed and conducted

- 2.3.1** Protect the independence of the review team to make the final decisions about the design, analysis, and reporting of the review

STANDARD 2.4

Manage bias and COI for individuals providing input into the systematic review

- 2.4.1** Require individuals to disclose potential COI and professional or intellectual bias
- 2.4.2** Exclude input from individuals whose COI or bias would diminish the credibility of the review in the eyes of the intended users

STANDARD 2.5

Formulate the topic for the systematic review

- 2.5.1** Confirm the need for a new review
- 2.5.2** Develop an analytic framework that clearly lays out the chain of logic that links the health intervention to the outcomes of interest and defines the key clinical questions to be addressed by the systematic review
- 2.5.3** Use a standard format to articulate each clinical question of interest
- 2.5.4** State the rationale for each clinical question
- 2.5.5** Refine each question based on user and stakeholder input

Defining the purpose and review questions



ELSEVIER

Contents lists available at [ScienceDirect](#)

Environment International

journal homepage: www.elsevier.com/locate/envint



Preface

Identifying the PECO: A framework for formulating good questions to explore the association of environmental and other exposures with health outcomes

Rebecca L. Morgan^a, Paul Whaley^b, Kristina A. Thayer^c, Holger J. Schünemann^{a,d,*}

^a Department of Health Research Methods, Evidence, and Impact (F), Canada Centre, McMaster University, Health Sciences Centre, Room
^b Lancaster Environment Centre, Lancaster University, Lancaster LA
^c Integrated Risk Information System (IRIS) Division, National Center for Environmental Health Protection Agency, Building B (Room 2111), Research Triangle Park
^d Department of Medicine, McMaster University, Health Sciences Centre

No guiding framework for operationalizing the PECO approach and the types of PECO questions researchers and decision-makers can answer

In environmental, public and occupational health research, specific challenges exist with identifying the exposure and comparator within the PECO

Five paradigmatic approaches and examples for identifying the exposure and comparator in systematic review and decision-making questions.

Developing protocols

STANDARD 2.6

Develop a systematic review protocol

- 2.6.1** Describe the context and rationale for the review from both a decision-making and research perspective
- 2.6.2** Describe the study screening and selection criteria (inclusion/exclusion criteria)
- 2.6.3** Describe precisely which outcome measures, time points, interventions, and comparison groups will be addressed
- 2.6.4** Describe the search strategy for identifying relevant evidence
- 2.6.5** Describe the procedures for study selection
- 2.6.6** Describe the data extraction strategy
- 2.6.7** Describe the process for identifying and resolving disagreement between researchers in study selection and data extraction decisions
- 2.6.8** Describe the approach to critically appraising individual studies

- 2.6.9** Describe the method for evaluating the body of evidence, including the quantitative and qualitative synthesis strategies
- 2.6.10** Describe and justify any planned analyses of differential treatment effects according to patient subgroups, how an intervention is delivered, or how an outcome is measured
- 2.6.11** Describe the proposed timetable for conducting the review

STANDARD 2.7

Submit the protocol for peer review

- 2.7.1** Provide a public comment period for the protocol and publicly report on disposition of comments

STANDARD 2.8

Make the final protocol publicly available, and add any amendments to the protocol in a timely fashion

... the time from search to submission was much longer for SRs with published protocols compared with [those without].

... while at the same time SRs with published protocols were better elaborated and reported.

Data extraction*

Two reviewers independently extracted data from all studies	58.8%	38.8%
One reviewer extracted data from all studies, and another reviewer checked/verified the extracted data for all (or a sample of) studies	22.5%	16.3%
Not reported	10.0%	38.8%
Other ^b	8.8%	6.3%
Risk of bias/quality assessed*	86.3%	60.0%

Risk of bias/quality assessment method

Two reviewers independently assessed all studies	58.0%	52.1%
Not reported	26.1%	43.8%
One reviewer assessed all studies, and another reviewer checked/verified the assessments for all (or a sample of) studies	10.1%	2.1%
Two reviewers assessed all studies, but authors did not state whether assessment was done independently	5.8%	2.1%

Mechanistic data

Mechanistic data come from a wide variety of studies and are generally not intended to identify a disease phenotype. This source of experimental data includes in vitro and in vivo laboratory studies directed at identifying the cellular, biochemical, and molecular mechanisms that are related to chemicals that produces particular adverse effects.

Another broad class of mechanistic data relates to the toxicokinetics of a chemical (NRC 2014a).



Historical criticism

- mixing apples and oranges

historical response

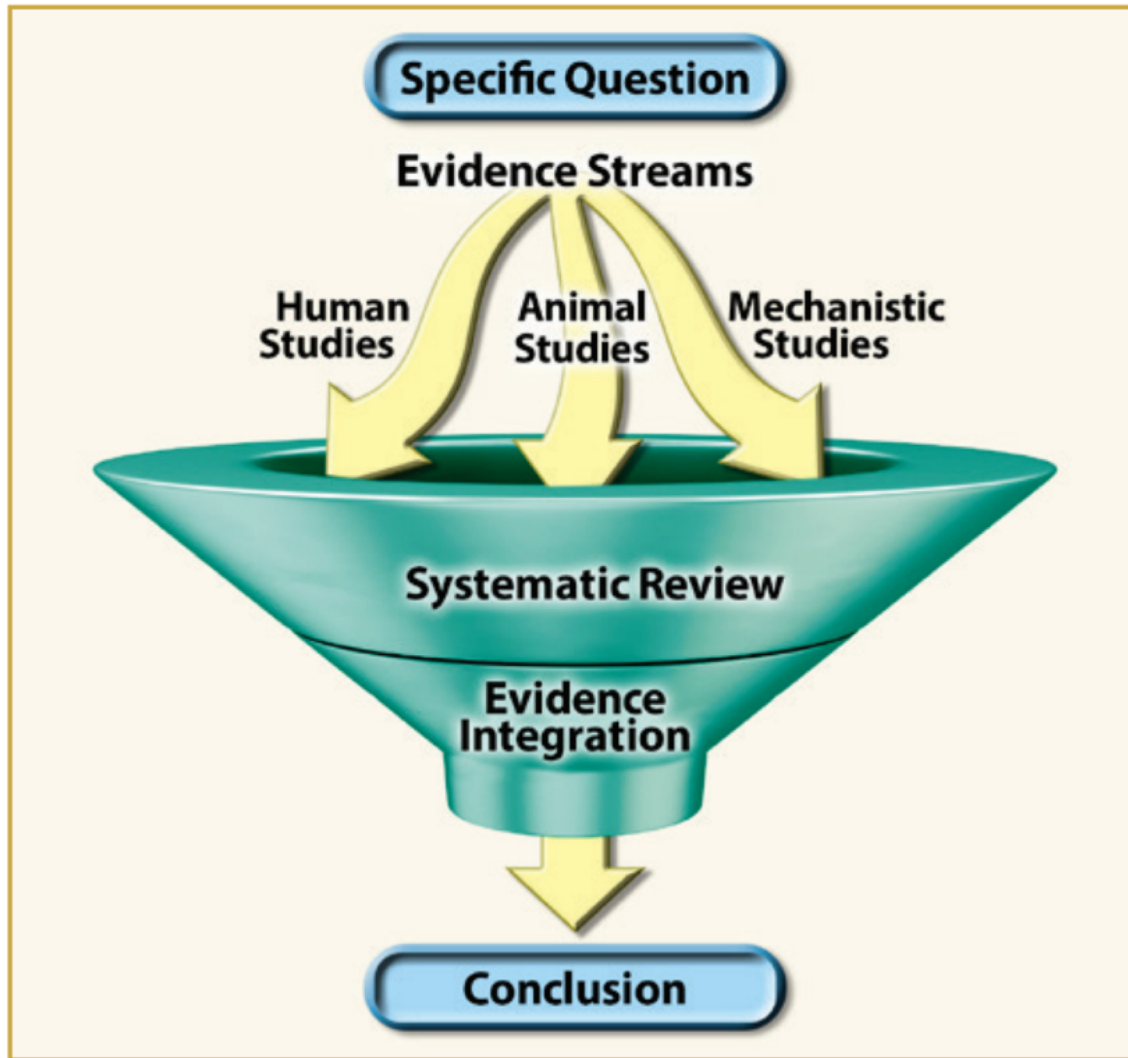
- interested in effect of fruit



But may be good to know through mechanistic data if effect is from Vitamin C and how it works



Inclusion and exclusion criteria



Confidence in the conclusion

Certainty of the evidence is also referred to as quality of the evidence, confidence in the effect estimate, and strength of the evidence. It is an assessment of the likelihood that the effect will not be substantially different from what the research found; i.e. different enough that it might affect a decision.

Confidence in the conclusion

GRADE

Certainty of evidence is best considered as the **certainty** that a true effect lies on one side of a specified threshold or within a chosen range.

Hultcrantz et al., JCE 2017



Certainty of evidence

How confident in the research?

Are the research studies well done? Risk of bias

Are the results consistent across studies ? Inconsistency

How directly do the results relate to our question? Indirectness

Is the effect size precise - due to random error? Imprecision

Are these all of the studies that have been conducted? Pub. Bias

Is there anything else that makes us particularly certain? Large effects, worst case scenario predictors still strong conclusions, exposure-effect relation

Bias in the design and execution of the studies only one of several domains – risk of bias

Systematic error or systematic deviation from the truth

Systematic review conclusions depend on risk of bias of included studies

- biased studies = misleading review results
- should we believe the results?

UCSF Navigation Guide: PFOA and Fetal Growth

REVIEW OCTOBER 2014 | VOLUME 122 | ISSUE 10

Environ Health Perspect; DOI:10.1289/ehp.1307893

The Navigation Guide—Evidence-Based Medicine Meets Environmental Health: Systematic Review of Human Evidence for PFOA Effects on Fetal Growth

Paula I. Johnson,¹ Patrice Sutton,¹ Dylan S. Atchley,¹ Erica Koustas,² Juleen Lam,² Saunak Sen,³ Karen A. Robinson,^{4,5,6} Daniel A. Axelrad,⁷ and Tracey J. Woodruff¹

in the direction of overall effect estimates.

Imprecision	0	We judged that the CI of the meta-analysis for birth weight is sufficiently narrow.
Publication bias	0	We found no reason to suspect publication bias. The search was comprehensive, and the studies were generally consistent among their findings, regardless of size or funding source.
Upgrade		
Large magnitude of effect	0	We did not consider the estimated effects large.
Dose response	0	Several studies in which association was modeled by categorized incremental exposure showed evidence of a dose-response relationship, but review authors agreed that the evidence was not compelling enough for an upgrade.
Confounding minimizes effect	0	We did not find evidence to suggest that possible residual confounders or biases would reduce effect estimate.
Overall quality of evidence (initial rating is "moderate")	Moderate	Moderate + (0) = moderate. (There were no upgrades or downgrades to change quality from the initial rating).
Summary of findings from meta-analysis	NA	We found decrements in fetal growth associated with PFOA exposure (see results from meta-analyses in Table 5).
Summary of qualitative findings	NA	Studies not included in the meta-analyses presented mixed results, mostly insignificant associations between PFOA and fetal growth (Figure 4; see also Supplemental Material, Figure S1).
Strength considerations		
Quality of body of evidence	NA	Moderate
Direction of effect estimate	NA	Birth weight decreased with increasing exposure to PFOA.
Confidence in effect estimate	NA	It is unlikely that a new study would have an effect estimate that would make the results of the meta-analysis null or insignificant.
Other compelling attributes of the data that may influence certainty	NA	None
Overall strength of evidence	Sufficient	Based on our analysis and interpretation of the evidence, we concluded that there is a positive association between exposure and outcome, and we believe with reasonable confidence that chance, bias, and confounding can be ruled out as an explanation.

NA, not applicable.
*See the Supplemental Material for more details on the process of downgrading and upgrading quality are presented in Table 2. "A" "0" quality rating indicates there were no upgrades or downgrades for each factor being evaluated across the body of evidence.

Human = moderate quality, sufficient strength

REVIEW OCTOBER 2014 | VOLUME 122 | ISSUE 10

Environ Health Perspect; DOI:10.1289/ehp.1307177

The Navigation Guide—Evidence-Based Medicine Meets Environmental Health: Systematic Review of Nonhuman Evidence for PFOA Effects on Fetal Growth

Erica Koustas,¹ Juleen Lam,¹ Patrice Sutton,² Paula I. Johnson,² Dylan S. Atchley,² Saunak Sen,³ Karen A. Robinson,^{4,5,6} Daniel A. Axelrad,⁷ and Tracey J. Woodruff²

and have low heterogeneity. Results are also consistent in magnitude and direction of effect estimates. Results of the meta-analysis do not appear to be strongly influenced by an individual study.

Indirectness	0	Mammalian studies did not have overlapping confidence bounds. Results are consistently in the same direction as human studies (Kimmel et al. 1984; U.S. EPA 2001).
Inconsistency	0	We found no reason to suspect publication bias. The search was comprehensive, and the studies were generally consistent among their findings, regardless of size or funding source.
Imprecision	0	We judged that the CI of the meta-analysis for birth weight is sufficiently narrow.
Publication bias	0	We found no reason to suspect publication bias. The search was comprehensive, and the studies were generally consistent among their findings, regardless of size or funding source.
Overall quality of evidence (initial rating is "high")	Moderate	Moderate + (0) = moderate. (There were no upgrades or downgrades to change quality from the initial rating).
Summary of findings from meta-analysis	NA	We found decrements in fetal growth associated with PFOA exposure (see results from meta-analyses in Table 5).
Summary of findings from qualitative analysis	NA	Studies not included in the meta-analyses presented mixed results, mostly insignificant associations between PFOA and fetal growth (Figure 4; see also Supplemental Material, Figure S1).
Overall strength of evidence	Sufficient	Based on our analysis and interpretation of the evidence, we concluded that there is a positive association between exposure and outcome, and we believe with reasonable confidence that chance, bias, and confounding can be ruled out as an explanation.

NA, not applicable. Ratings: -1, 1 level downgrade in quality; 0, no change in quality. Studies included in meta-analysis [source ID]: Abbott et al. 2007 [528], Hines et al. 2009 [260], Lau et al. 2006 [635] (birth weight data), White et al. 2007 [566], White et al. 2009 [312], White et al. 2011 [3862], and Wolf et al. 2007 [571] (cross-foster and windows of sensitivity data). Other studies [source ID]: Boberg 2008 et al. [3061], Fenton et al. 2009 [264], Hinderliter et al. 2005 [711], Hu et al. 2010 [68], Lau et al. 2006 [635] (fetal weight data), Onishchenko et al. 2011 [3610], Staples 1984 et al. [1871], Yabuta et al. 2010 [103], and York 2002 [5122].

Nonhuman animal = moderate quality, sufficient strength

REVIEW OCTOBER 2014 | VOLUME 122 | ISSUE 10

Environ Health Perspect; DOI:10.1289/ehp.1307923

The Navigation Guide—Evidence-Based Medicine Meets Environmental Health: Integration of Animal and Human Evidence for PFOA Effects on Fetal Growth

Juleen Lam,¹ Erica Koustas,¹ Patrice Sutton,² Paula I. Johnson,² Dylan S. Atchley,² Saunak Sen,³ Karen A. Robinson,^{4,5,6} Daniel A. Axelrad,⁷ and Tracey J. Woodruff²

Integration of human and animal evidence led to conclusion of "known" to be toxic to human reproduction and development

Integration of human and animal evidence led to conclusion of "known" to be toxic to human reproduction and development

GRADE and Rapid Response



ELSEVIER

Contents lists available at ScienceDirect

Environment International

journal homepage: www.elsevier.com/locate/en

Preface

Using GRADE to respond to health questions with different levels of evidence

Kristina A. Thayer^a, Holger J. Schünemann^{b,*}

^a Division of the National Toxicology Program, National Institute of Environmental Health Sciences, National Institutes of Health, Department of Health and Human Services, Mail Drop K2-02, Research Triangle Park, NC 27709, USA

^b Department of Clinical Epidemiology & Biostatistics, Department of Medicine, McMaster University, Health Sciences Centre, Room 2C

ARTICLE INFO

Article history:

Received 15 March 2016

Received in revised form 21 March 2016

Accepted 21 March 2016

Available online 26 April 2016

ABSTRACT

Increasing interest exists in applying the Grading of Recommendations (GRADE) approach to environmental health evidence. While individual studies and corresponding summary tables, such as evidence of risk, are typically presented, the methods that should be clearly described. In this article, we suggest that from narrative reviews, modelled (indirect) evidence, or evidence underlying judgments about the certainty in this evidence are presented transparently. Health questions that require assessing the certainty of answers may range from hours, to days or weeks, to a few years without short-term time pressures. Time frames of emergent, time-sensitive questions require relying on existing summaries or rapidly compiling the evidence without available full systematic reviews, expressing the certainty for users of the evidence and those who evaluate certainty in evidence between organizations tackling similar questions about the evidence. Narrative or other summaries of the evidence can be presented to

Table 1
Examples of GRADE applied across different time scenarios.

Type of response	Ultra-short emergency response: within one or more hours	Urgent response: one to two weeks	Rapid response: one to three months	Routine response: more than 3 months
Example	West Virginia Elk River spill Population: community exposed to the chemical spill. Intervention/exposure: chemicals in the spill that contaminated water supply. Comparison: no chemicals in the spill. Outcomes: genotoxicity, developmental or reproductive toxicity, liver toxicity and others.	Melamine in composite food products Population: healthy people Intervention/exposure: melamine from composition food products below 0.5 mg/kg body weight per day. Comparison: higher than 0.5 mg/kg body weight of melamine from composition food. Outcomes: renal insufficiency (assessed with renal clearance), urinary tract calculi, urinary tumors (used for this example of the certainty in the evidence).	Avian influenza Population: people with suspected avian influenza infection. Intervention/exposure: oseltamivir. Comparison: no oseltamivir. Outcomes: mortality, duration of hospitalization, incidence of lower respiratory tract complications (used for this example of the certainty assessment below), antiviral drug resistance existing before treatment, and serious adverse events.	PFOA and birth weight Population: women of reproductive age and fetuses (before and/or during pregnancy or development). Intervention/exposure: perfluorooctanoic acid (PFOA; CAS# 335-67-1) or its salts. Comparison: lower levels of PFOA. Outcomes: fetal growth, birth weight, other measures of fetal or newborn size.
Type of evidence	Available evidence: animal toxicology studies in rodents for two chemicals in the spill (a 28-day study and a teratology study) and SAR analyses for other chemicals in the spill with no toxicology data.	Available evidence: animal toxicology studies in rat and mice with exposures to various levels of melamine via feeding, including a control group. The utilized evidence should be supported by a literature search with transparent inclusion and exclusion criteria and a (narrative) summary of that evidence.	Available evidence: five randomized trials in patients with seasonal flu (summarized in systematic reviews), case studies of patients with avian influenza, in vitro and in vivo animal data.	Available evidence: a systematic review of 18 non-randomized (observational) studies (10 were included in a meta-analysis).

GRADE domains to assess certainty in the evidence: suggested approaches to making judgments or proposed judgments (note these are not necessarily reflecting judgments in the original scenarios).

Risk of bias	Animal studies: would be assessed by risk of bias (RoB) considerations for animal studies (e.g., randomization, blinding at outcome assessment, sufficient characterization of test compound, or whether all animals were accounted for). Ideally, RoB assessments would be available for individual studies and summarized across studies. In the Elk River example, the number of animal studies was small and could be assessed at the individual level within a short-time frame. A de novo risk of bias evaluation may not be feasible in cases where evidence is drawn from existing narrative risk assessments that summarize a large body of literature. Nevertheless, it may still be possible to assess risk of bias based on the uncertainties and evidence limitations described in the risk assessment. SAR: could be assessed using OECD model validation or similar guidance that recommends presentation of a defined domain of applicability for a defined endpoint supported by appropriate measures of goodness-of-fit (OECD, 2007). Could be assessed for both animal data and SAR (e.g., considering statistical or numerical uncertainty in model parameters).	Animal studies: would be assessed by risk of bias (RoB) considerations for animal studies (e.g., randomization, pathologists blinded in their assessments or all animals accounted for). In this case it appears that the animal studies did not report that it was randomized and, thus, may be at risk of bias.	Not serious	Serious based on some concern of risk of bias in the included studies (in the original report, the authors used an approach to rating certainty that accounted for risk of bias by lowering the certainty from high to moderate).
Imprecision	Could be assessed for both animal data and SAR (e.g., considering statistical or numerical uncertainty in model parameters).	While no summary estimates are available, an assessment could be guided by the availability of data from only 100 animals in different exposure groups which would result in wide confidence intervals.	Serious	Not serious
Inconsistency	Could be assessed for both animal data and SAR (e.g., assessing similarity of results based on applying different models).	Only one study was included and therefore no inconsistency is present (Guyatt et al., 2011d).	Not serious	Not serious
Publication bias	Could be assessed for both animal studies and SAR. A judgment of undetected might be reasonable if	Could be assessed using guidance for animal studies but a judgment of undetected might be reasonable if	Undetected	Undetected

Table 1

Examples of GRADE applied across different time scenarios.

Type of response	Ultra-short emergency response: within one or more hours	Urgent response: one to two weeks	Rapid response: one to three months	Routine response: more than 3 months
Example	West Virginia Elk River spill Population: community exposed to the chemical spill. Intervention/exposure: chemicals in the spill that contaminated water supply. Comparison: no chemicals in the spill. Outcomes: genotoxicity, developmental or reproductive toxicity, liver toxicity and others.	Melamine in composite food products Population: healthy people Intervention/exposure: melamine from composition food products below 0.5 mg/kg body weight per day. Comparison: higher than 0.5 mg/kg body weight of melamine from composition food. Outcomes: renal insufficiency (assessed with renal clearance), urinary tract calculi, urinary tumors (used for this example of the certainty in the evidence).	Avian influenza Population: people with suspected avian influenza infection. Intervention/exposure: oseltamivir. Comparison: no oseltamivir. Outcomes: mortality, duration of hospitalization, incidence of lower respiratory tract complications (used for certainty of antiviral before treatment adverse effects).	PFOA and birth weight Population: women of reproductive age and fetuses (before and/or during pregnancy or development). Intervention/exposure: perfluorooctanoic acid (PFOA; CAS# 335-67-1) or its salts. Comparison: lower levels of PFOA.
Type of evidence	Available evidence: animal toxicology studies in rodents for two chemicals in the spill (a 28-day study and a teratology study) and SAR analyses for other chemicals in the spill with no toxicology data.	Available evidence: animal toxicology studies in rat and mice with exposures to various levels of melamine via feeding, including a control group. The utilized evidence should be supported by a literature search with transparent inclusion and exclusion criteria and a (narrative) summary of that evidence.	Available trials in people (summary reviews), with avian in vivo and	
GRADE domains to assess certainty in the evidence: suggested approaches to making judgments or proposed judgments (original scenarios).				
Risk of bias	Animal studies: would be assessed by risk of bias (RoB) considerations for animal studies (e.g. randomization, blinding at outcome assessment, sufficient characterization of test compound, or whether all animals	Animal studies: would be assessed by risk of bias (RoB) considerations for animal studies (e.g. randomization, pathologists blinded in their assessments or all animals accounted for). In this case it	Not serio	

Rest of table summarizes:

- GRADE domains
 - risk of bias, imprecision, indirectness, inconsistency, publication bias, magnitude, etc.
- Certainty in evidence
- Possible summary statements

Risk of bias is not the same as

Imprecision

- random error due to sampling variation
- reflected in the confidence interval

Quality

- bias can occur in well-conducted studies
- not all methodological flaws introduce bias

Reporting

- good methods may have been used but not well reported

ORIGINAL ARTICLE

ROBIS: A new tool to assess risk of bias in systematic reviews was developed

Penny Whiting^{a,b,c,*}, Jelena Savović^{a,b}, Julian P.T. Higgins^{a,d}, Deborah M. Caldwell^a, Barnaby C. Reeves^e, Beverley Shea^f, Philippa Davies^{a,b}, Jos Kleijnen^{c,g}, Rachel Churchill^a, the ROBIS group

^a*School of Social and Community Medicine, University of Bristol, Canynge Hall, 39 Whatley Road, Bristol BS8 2PS, UK*

^b*The National Institute for Health Research Collaboration for Leadership in Applied Health Research and Care West at University Hospitals Bristol NHS Foundation Trust, 9th Floor, Whitefriars, Lewins Mead, Bristol BS1 2NT*

^c*Kleijnen Systematic Reviews Ltd, Unit 6, Escrick Business Park, Riccall Road, Escrick, York YO19 6FD, UK*

^d*Centre for Reviews and Dissemination, University of York, York YO10 5DD, UK*

^e*School of Clinical Sciences, University of Bristol, Bristol Royal Infirmary, Level Queen's Building, 69 St Michael's Hill, Bristol BS2 8DZ, UK*

^f*Community Information and Epidemiological Technologies Institute of Population Health, 1 Stewart Street, Room 319, Ottawa, Ontario, K1N 6N5, Canada*

^g*School for Public Health and Primary Care (CAPHRI), Maastricht University, PO Box 616, 6200 MD, Maastricht, The Netherlands*

Accepted 5 June 2015; Published online xxxx

Table 1. Summary of phase 2 ROBIS domains, phase 3, and signaling questions

	Phase 2				Phase 3
	1. Study eligibility criteria	2. Identification and selection of studies	3. Data collection and study appraisal	4. Synthesis and findings	Risk of bias in the review
Signaling questions	1.1 Did the review adhere to predefined objectives and eligibility criteria?	2.1 Did the search include an appropriate range of databases/ electronic sources for published and unpublished reports?	3.1. Were efforts made to minimize error in data collection?	4.1. Did the synthesis include all studies that it should?	A. Did the interpretation of findings address all of the concerns identified in domains 1 to 4? B. Was the relevance of identified studies to the review's research question appropriately considered? C. Did the reviewers avoid emphasizing results on the basis of their statistical significance?
	1.2 Were the eligibility criteria appropriate for the review question?	2.2 Were methods additional to database searching used to identify relevant reports?	3.2. Were sufficient study characteristics available for both review authors and readers to be able to interpret the results?	4.2. Were all predefined analyses reported or departures explained?	
	1.3 Were eligibility criteria unambiguous?	2.3 Were the terms and structure of the search strategy likely to retrieve as many eligible studies as possible?	3.3. Were all relevant study results collected for use in the synthesis?	4.3. Was the synthesis appropriate given the nature and similarity in the research questions, study designs, and outcomes across included studies?	
	1.4 Were all restrictions in eligibility criteria based on study characteristics appropriate?	2.4 Were restrictions based on date, publication format, or language appropriate?	3.4. Was risk of bias (or methodologic quality) formally assessed using appropriate criteria?	4.4. Was between-study variation minimal or addressed in the synthesis?	
	1.5 Were any restrictions in eligibility criteria based on sources of information appropriate?	2.5 Were efforts made to minimize error in selection of studies?	3.5. Were efforts made to minimize error in risk of bias assessment?	4.5. Were the findings robust, for example, as demonstrated through funnel plot or sensitivity analyses? 4.6. Were biases in primary studies minimal or addressed in the synthesis?	
Judgment	Concerns regarding specification of study eligibility criteria	Concerns regarding methods used to identify and/or select studies	Concerns regarding methods used to collect data and appraise studies	Concerns regarding the synthesis	Risk of bias in the review

Summary

Methods and standards for systematic reviews exist for various scientific fields

Overarching principles not different

Work on operationalization of the gaps

- E.g. how mechanistic data support statements about human health effects when the relevant mechanistic evidence has been systematically reviewed

Thank you

