**Title:** A code of practice for the Conduct Of Systematic reviews in Toxicology and Environmental health Research (COSTER)

## Author List

*Corresponding and first author*: **Paul Whaley**, Lancaster Environment Centre, Lancaster University, Lancaster, LA1 4YQ, UK | p.whaley@lancaster.ac.uk

*Other authors*:
**Elisa Aiassa**, European Food Safety Authority (EFSA), Assessment and Methodological Support unit. Via Carlo Magno 1/A, 43126 Parma, Italy | elisa.aiassa@efsa.europa.eu
**Claire Beausoleil,** ANSES (French Agency for Food, Environmental and Occupational Health Safety), Risk Assessment Department, Chemical Substances Assessment Unit, F-94700 Maisons-Alfort, France | claire.beausoleil@anses.fr
**Anna Beronius**, Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden | anna.beronius@ki.se
**Gary Bilotta**, School of Environment and Technology, University of Brighton, Brighton, UK | g.s.bilotta@gmail.com
**Alan Boobis** National Heart & Lung Institute, Imperial College London, London, UK | a.boobis@imperial.ac.uk
**Rob de Vries,** SYRCLE, Department for Health Evidence, Radboud Institute for Health Sciences, Radboudumc, Nijmegen, The Netherlands | rob.devries@radboudumc.nl
**Annika Hanberg**, Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden | annika.hanberg@ki.se
**Sebastian Hoffmann,** Evidence-based Toxicology Collaboration at Johns Hopkins Bloomberg School of Public Health, Paderborn, Germany | sebastian.hoffmann@seh-cs.com
**Neil Hunt**, Yordas Group, Lancaster Environment Centre, Lancaster University, Lancaster, LA1 4YQ, UK | n.hunt@yordasgroup.com
**Carol F. Kwiatkowski**, The Endocrine Disruption Exchange. P.O. Box 54, Eckert, CO, 81418, USA | kwiatkowskicf@gmail.com
**Juleen Lam**, University of California, San Francisco and California State University, East Bay. 28500 Carlos Bee Blvd Room 502, Hayward, CA 94542, USA | Juleen.Lam@csueb.edu
**Steven Lipworth,** Royal Society of Chemistry, Burlington House, Piccadilly, London, W1J 0BA, UK
**Olwenn Martin**, Institute for the Environment, Health and Societies, Brunel University London, Uxbridge, UK | olwenn.martin@brunel.ac.uk
**Nicola Randall,** Harper Adams University, Newport, Shropshire, UK | nrandall@harper-adams.ac.uk
**Lorenz Rhomberg PhD ATS**, Gradient, 20 University Road, Cambridge, MA 02138, USA lrhomberg@gradientcorp.com
**Andrew A. Rooney**, Division of the National Toxicology Program, National Institute of Environmental Health Sciences, North Carolina, USA | andrew.rooney@nih.gov
**Holger J. Schünemann**, McGRADE Centre and Michael G De Groote Cochrane Canada Centre, Dept. of Health Research Methods, Evidence and Impact, McMaster University, 1280 Main Street West, Hamilton, ON, Canada | schuneh@mcmaster.ca
**Daniele Wikoff,** ToxStrategies, 31 College Place, Suite B118B, Asheville, NC 28801, USA | dwikoff@toxstrategies.com
**Taylor Wolffe**, Lancaster Environment Centre, Lancaster University, Lancaster, LA1 4YQ, UK | t.wolffe@lancaster.ac.uk
**Crispin Halsall**, Lancaster Environment Centre, Lancaster University, Lancaster, LA1 4YQ, UK | c.halsall@lancaster.ac.uk

## Declaration of Interests

Due to the objective of the project being to establish, across a wide range of stakeholders, a consensus view on sound and good practice in the conduct of environmental health systematic reviews, participants in the process were selected because of their varying interests in the conduct of environmental health research. Funding was provided by Lancaster University to support travel costs of authors who would otherwise be unable to attend (PW, CH, LR, JL, AR) and Dr Jennifer McPartland (non-authoring workshop participant, see acknowledgements). With regard to the development of COSTER, the authors declare they have no apparent competing financial interests, and certify that their freedom to design, conduct, interpret, and publish the research was not compromised by any controlling sponsor. These forms are included as supplemental information. PW, as organiser of the meeting and lead author of the manuscript, declares personal fees from Elsevier Ltd (Environment International), the Cancer Prevention and Education Society, the Evidence Based Toxicology Collaboration and Yordas Group, and grants from Lancaster University, which are outside the submitted work but relate to the development and promotion of systematic review and other evidence-based methods in environmental health research, delivering training around these methods, and providing editorial services. Each author has declared their interests using the International Committee of Medical Journal Editors Form for Disclosure of Potential Conflicts of Interest; these are available as Supplemental Materials.

## Highlights (long version)

- COSTER is a research standard which describes a minimum set of recommended practices for ensuring an environmental health systematic review is credible, transparent and useful.
- COSTER is intended to aid in the planning and conduct of environmental health SRs, and to function as a benchmark against which SRs and related SR standards, guidance and frameworks can be assessed.
- Anticipated users of COSTER include researchers, research commissioners, journal editors, research quality managers, and any other stakeholder with an interest in the quality of conduct of a systematic review.
- Agreement on COSTER was reached via a consensus process. The final consensus group consisted of 21 systematic review practitioners and related experts from industry, non-government organisations, government agencies and academia.

- COSTER is a first step, not the final word, in defining an authoritative, comprehensive standard for conduct of systematic reviews in the environmental health sciences.


## Abstract

**Background**: There are several standards which make explicit a consensus view on sound practice in systematic reviews (SRs) for the medical sciences. Until now, no equivalent standard has been published for SRs which focus on human health risks posed by exposure to environmental challenges, chemical or otherwise.

**Objectives**: To develop an expert, cross-sector consensus view on a key set of recommended practices which can function as a standard for the planning and conduct of SRs in the environmental health sciences.

**Methods**: A draft set of practices was derived from two existing standards for SRs in biomedicine and discussed at an international workshop of 33 participants from government, industry, non-government organisations, and academia. The guidance was revised over six follow-up webinars, multiple rounds of email feedback, and bilateral phone calls, until there was group consensus that a comprehensive framework for the planning and conduct of high-quality environmental health SRs had been articulated.

**Results**: The Conduct of Systematic Reviews in Toxicology and Environmental Health Research (COSTER) standard is a code of practice consisting of 70 performance elements across eight performance domains, representing the consensus view of a diverse group of experts as to what constitutes “sound and good” practice in the conduct of environmental health SRs.

**Discussion**: COSTER provides a set of practices which, if followed, should facilitate the production of credible, high-value SRs of environmental health evidence. COSTER clarifies sound and good practice in a number of controversial aspects of SR conduct, including the management of conflicts of interest, inclusion of grey literature, and protocol registration and publication. Not all of the practices are yet commonplace but environmental health SRs would benefit from their use.

## 1 Introduction

In the fields of toxicology, epidemiology, environmental health and chemical risk assessment (henceforth abbreviated as "environmental health (EH) research"), systematic reviews (SRs) are increasingly being conducted and used by academics, non-government organisations, industry and regulators to characterise health hazards and risks posed by exposure to environmental challenges (Whaley et al. 2016a). One of the drivers of this growing interest is increasing recognition of how systematic methods offer a potential new benchmark in best practice for aggregating and summarising evidence in support of policy decisions (EFSA 2010; Rooney et al. 2014; NAS 2017, 2014; Stephens et al. 2016).

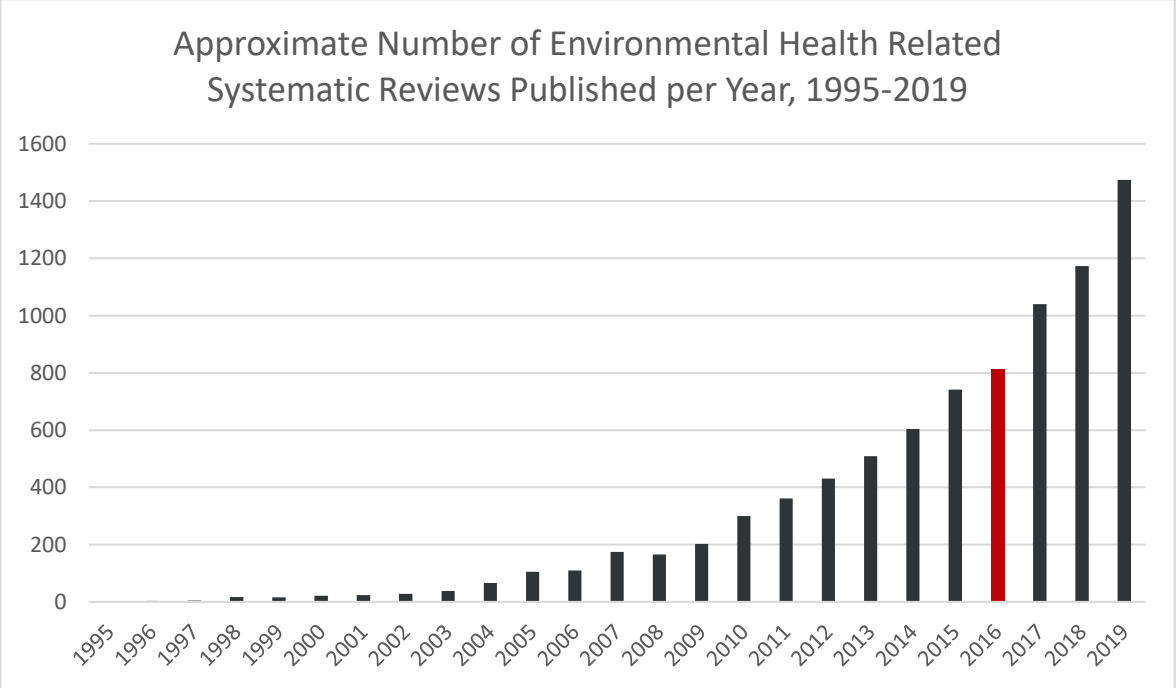


***Figure 1***: Chart showing annual increase in number of publications on topics related to EH research with the term "systematic review" in the title, indexed in Web of Science. The total number of publications has approximately doubled since 2016. **Search**: TITLE: ("systematic review"), **Refined by**: WEB OF SCIENCE CATEGORIES: ( PUBLIC ENVIRONMENTAL OCCUPATIONAL HEALTH OR TOXICOLOGY ) AND [excluding] WEB OF SCIENCE CATEGORIES: ( PHARMACOLOGY PHARMACY ), **Timespan**: All years (1995-2019 shown). **Indexes**: SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, BKCI-S, BKCI-SSH, ESCI, CCR-EXPANDED, IC. **Date of search**: 4 February 2020

In service of this interest, there is a burgeoning number of documents which purport to provide varying types of guidance for conducting SRs in EH research. These include, for example: a handbook by the US National Toxicology Program Office of Health Assessment and Translation, first published in 2015 and updated in 2019 (NTP OHAT 2019); guidance

documents by the Texas Commission on Environmental Quality (Schaefer and Myers 2017), the European Food Safety Authority (EFSA 2015, 2010), and the US Environmental Protection Agency application of SR methods in Toxic Substances Control Act (TSCA) risk evaluations (EPA 2018); the 2019 updates to the Preamble to the International Agency for Research on Cancer Monographs and Instructions to Authors (IARC 2019a, 2019b); the SYRINA framework describing systematic methods for the identification of endocrine disruptors (Vandenberg et al. 2016); and the Navigation Guide framework for environmental health SRs (Woodruff and Sutton 2014).

The challenge for the reader is in how SR guidance documents vary in their levels of comprehensiveness and detail, domains of applicability, the extent to which they have been tested and validated, and what they define (either implicitly or explicitly) as being essential SR methodology. For example, the NTP OHAT handbook is for SRs conducted in support of hazard assessment within a US regulatory framework, whereas the Navigation Guide is intended for a more general research context. While the Navigation Guide and NTP OHAT approaches are largely similar in principal methods including use of protocol, comprehensive search strategies, employment of a Cochrane-derived risk of bias approach to appraising study quality, and use of a GRADE-based approach to assessing confidence in a body of evidence, there are some differences between the two, and larger ones with other approaches. These include how SYRINA lays out a wider range of options which an SR team can choose between, how the draft TSCA approach scores study quality rather than implementing Cochrane guidance on risk of bias assessment (Singla et al. 2019), and how the IARC Monographs apply systematic principles to several but not all stages of the evidence review process. Some approaches apply to the entire chemical risk assessment process while others focus on a particular stage of it, and many SR guidance documents have been developed for the specific purposes unique to a particular entity. Overall, these documents do not therefore provide a collectively consistent, general standard for good practice in the planning and conduct of an EH SR.

Importantly, none of these guidance documents has been explicitly developed as a research standard. Standards are distinguished from guidelines and handbooks in that, in the form of a list of requirements, they “provide a set of agreed principles or criteria for a product, service or practice, such that users of those products can make reliable assumptions about their performance, safety, compatibility and/or other features as specified in the standard” (British Standards Institution 2016b). Standards vary in detail and prescriptiveness according to the function they perform, from “specifications” which set out detailed, absolute requirements, to flexible “codes of practice” which recommend “sound and good practice as currently undertaken by competent and conscientious practitioners” (British Standards Institution 2016a). Standards

158 set a benchmark against which the quality of a product can be evaluated – be it the safety of
159 electrical goods, compatibility between devices or, in the present case, the quality of a SR or SR
160 guidance document. The US Institute of Medicine considers the development and promulgation
161 of standards which provide clear, expert guidance on good practice to be an important
162 contributor to ensuring the quality of SRs (Eden et al. 2011).

163 The situation in EH research sits in contrast to the biomedical sciences, where standards for
164 both conducting and reporting SRs have been proliferating rapidly over the last three decades.
165 The EQUATOR Network's online *Library for Health Research Reporting* currently lists over 400
166 standards for reporting health research (EQUATOR Network 2020). Although many of the
167 standards in the library are concerned with reporting of primary research, there are also
168 numerous standards for reporting of SRs, such as the PRISMA checklist for systematic reviews
169 of interventions (Moher et al. 2009) and the MOOSE reporting guidelines for SRs of
170 observational studies in medicine (Stroup et al. 2000). Reporting standards are relevant for
171 informing the conduct of SRs because they imply a set of practices which need to be reported in
172 a scientific manuscript; sometimes these practices are even made explicit in an explanation
173 document, such as for the PRISMA Statement (Liberati et al. 2009). Standards which focus
174 directly on the conduct rather than reporting of SRs include the US Institute of Medicine (IOM)
175 *Finding What Works in Health Care: Standards for Systematic Reviews* (Eden et al. 2011) and the
176 Cochrane Editorial Unit's *Methodological Expectations for Conduct of Intervention Reviews*
177 (MECIR) standard (Chandler et al. 2013), recently updated to version 1.07 in November 2018
178 (Higgins et al. 2018).

179 While the universal nature of the fundamentals of SR methods should result in broad overlap
180 in sound SR practices between biomedical and EH research, the potential for cultural and
181 research-specific differences between the domains mean that direct applicability of biomedical
182 SR standards to EH research cannot be assumed (Haddaway et al. 2018a). These differences
183 include:

184 - the types of evidence being summarised, with a focus in EH on observational human,
185 experimental animal and *in vitro* study designs intended to elucidate disease
186 aetiology and identify health risks, as opposed to a prevalence of methods for
187 identifying effective treatments for disease using a body of evidence in which
188 randomised controlled trials in humans tends to be more readily available;

189 - the types of decision potentially being supported by EH SRs, such as defining the
190 conditions for acceptable use of chemical substances rather than informing
191 healthcare intervention decisions);

- and specific methodological challenges in evidence synthesis, with the need in EH research to integrate evidence from human, animal, *in vitro*, and *in silico* studies.

These differences mean that standards developed in biomedicine need to be methodically assessed, and potentially adapted and added to, by EH research practitioners. This was recognised in a 2014 expert workshop on EH SRs which, among other strategic proposals, recommended "development of a recognised 'gold standard' for SRs in toxicology and risk assessment [...] to address the growing number of purported SRs of unclear validity which are increasingly prevalent in the environmental health literature" (Whaley et al. 2016a).

A broad cross-section of relevant expert stakeholders was therefore convened with the objective of developing a standard for conduct of SRs in EH research. This was to be delivered in the form of a single authoritative consensus document which identifies a set of minimum recommended practices for EH SRs, and by so doing addresses inconsistencies, omissions and errors in current guidance.

## 2 Methods

A workshop was held 2 December 2016, attended by 31 expert participants selected to cover academic, policy, regulatory, non-government and industry sectors (see Supplemental Information 01). An initial longlist of 62 potential participants was drawn up by PW and CH. Participants were prioritised for invitation to the workshop based on a mixture of having a publishing history demonstrating at least some experience in systematic review or the principles thereof, professional reputation, economic sector, and word-of-mouth recommendation. An overall balance of expertise in SR methods, weight-of-evidence methods, chemical risk assessment, toxicology, environmental health research and chemicals policy was sought across the final group of participants, along with balanced representation from each sector including a target of at least two NGO participants. £5000 was made available by Lancaster University to facilitate balanced participation, covering travel costs for participants who would not otherwise be able to attend the workshop.

The new standard was developed using a consensus methodology, with "consensus" defined following International Organization for Standardization (ISO) terminology as "general agreement, characterized by the absence of sustained opposition to substantial issues by any important part of the concerned interests and by a process that involves seeking to take into account the views of all parties concerned and to reconcile any conflicting arguments" (ISO/IEC 2004).

The consensus process was seeded by two discussion documents (see Supplements 02 and 03) drafted by PW. A draft standard (Supplement 03), initially given the working title of "ECOSYS-CRA" before being renamed "COSTER", was created by combining version 2.3 of the Cochrane *MECIR* standards (Chandler et al. 2013) with the US Institute of Medicine *WWHC: Standards for Systematic Reviews* (Eden et al. 2011), henceforth referred to as MECIR and IOM respectively. These standards were taken to already represent a high degree of consensus and expectation of effectiveness of sound-practice requirements relating to general SR methods in biomedicine, thereby providing a solid basis for a standard for EH SRs. The discussion documents also outlined for participants the potential role of standards in quality management of research, explained how standards are developed, and described how the workshop and subsequent follow-up activities would be structured to facilitate consensus on a standard for conduct of SRs in EH research.

The ECOSYS-CRA draft standard was discussed element-by-element at the workshop by two break-out groups working in parallel, chaired by PW and JL. Feedback was solicited on the following: (a) which of the proposed criteria should be included as "sound and good practice" in a code of practice for SRs in toxicology and chemical risk assessment; (b) if and how the included criteria should be reformulated; (c) whether there were any additional criteria which should be included, and if so, how they should be formulated; and (d) questions for clarification and follow-up. Further detail on the assumptions, methodological decisions, and structure of the consensus process behind COSTER is provided in Supplement 02. The final workshop participant list is the aggregate of the authors of COSTER plus the non-authoring participants of the workshop listed in the Acknowledgements.

GB and CH took notes of the discussion in each group. Comments were collated into a redrafted document and cross-checked by PW against the Campbell Collaboration *MEC2IR* standard (Campbell Collaboration 2014). This was to check for any further possible performance elements which might be included in COSTER, as suggested in discussion at the workshop. The redrafted COSTER standard was then discussed in a series of six one-hour webinars held between January and June 2017, chaired by PW and attended on average by six participants (EA, ABe, RdV, KG, AH, NH, SH, CK, JL, OM, LR, AR, HS, KS, DW, CH, TW participated in at least one). The webinars were followed by email exchanges and bilateral phone calls between PW and various authors to finalise wording and agree that consensus had been reached.

The consensus process was closed by PW on 24 January 2018; participating authors confirmed agreement with the consensus by signing off as co-authors of this manuscript. Participants in the process who contributed to the workshop and related discussions but either

left the consensus process or did not sign off on the manuscript have been listed in the Acknowledgements. In total, 21 of the 31 workshop participants signed off as a manuscript author. Participants were not asked to provide their reasons for leaving the discussions. The majority of those leaving the process did so shortly after the workshop; when reasons were volunteered, it was due to personal capacity or governance issues relating to their professional position (e.g. wanting to avoid any implication that an employing organisation might be endorsing a specific standard). Only one participant who was closely involved in the development of the manuscript itself ultimately felt they could not sign off as an author, citing organisational governance issues. None of the participants opposed publication of COSTER.

## 3 Results

COSTER presents, in the form of 70 performance elements across 8 domains, the consensus view of 21 EH practitioners of a minimum set of recommended practices for the conduct of EH SRs. Since the consensus process results in a description of what the authors, as a group of competent and conscientious practitioners, agreed to be "sound and good practice" in the planning and conduct of a EH SR, COSTER should be interpreted as a standard around the level of a "code of practice" as defined by the British Standards Institution (British Standards Institution 2016a). "Sound and good practice" is interpreted by the authors as being a key set of actions which, if followed, should result in a EH SR having the following three essential characteristics:

1. being **useful**, addressing an important research question and advancing community understanding of an environmental health issue via a methodology of synthesising existing research;

2. being **transparent**, encouraging comprehensive consideration of the assumptions and methods employed in a SR such that, if they are adequately reported, a reader is able to appraise the validity of the SR's findings and assess their relevance to a given decision-making context;

3. being **credible**, minimising the risk that a SR's findings are biased either by limitations in the evidence base itself or in the processes used to locate and synthesise that evidence.

The COSTER domains cover the following methodological elements of the SR process: planning the SR; searching for evidence; selecting evidence for review; extracting data; critically appraising each individual included study; synthesising the evidence; interpreting the evidence and summarising what it means for the review question; and drawing conclusions (see Figure

291 2). The performance elements within each domain are listed in Table 1. Explanation of key
292 performance elements and guidance on how to use COSTER is provided in Table 2 and the
293 Discussion section of this manuscript.


294

295 *Figure 2. Conceptual structure of COSTER with objectives for each stage of the SR process*

| **COSTER v1.0.0: Performance elements for sound and good practice in the planning and conduct of environmental health systematic reviews** |
|---|
| **1. Planning the Review and Preparing the Protocol** |
| **1.1 Securing capacity, competencies and tools** |
| 1.1.1 Ensure the review team has sufficient combined competence to conduct the systematic review, including relevant expertise in: information science (for e.g. search strategies); evidence appraisal; statistical methods; domain or subject expertise; systematic review methods. |
| 1.1.2 Identify information management practices for each stage of the review, including reference and knowledge management tools, systematic review software, and statistics packages. |
| 1.1.3 Exclude people or organisations with apparent conflicts of interest relating to the findings of the review from analysis and decision-making roles in the review process. |
| 1.1.4 Disclose the roles and all potential conflicts of interest of all people and organisations involved in planning and conducting the review, including all providers of financial and in-kind support. |
| **1.2 Setting the research question to inform the scope of the review ("problem formulation")** |
| 1.2.1 Demonstrate the need for a new review in the context of the scientific value of the question, the importance to stakeholders of the question being asked, and the findings of any pre-existing primary research and/or evidence syntheses. |
| 1.2.2 Articulate the scientific rationale for each question via development of a theoretical framework which connects e.g. the exposure to the outcomes of interest (or otherwise as appropriate given the objectives of the review). |
| 1.2.3 For each research question to be answered by the review, prospectively define a statement of the research objective in terms of one or more of the following components, selected as appropriate:<br>• Population (objects of investigation, i.e. the entities to which exposures or interventions happen) |

| |
|---|
| • Exposure or Intervention (the administered change in conditions of the objects of investigation, to include timing, duration and dose)<br>• Comparator (the group to which the intervention or exposure groups are being compared)<br>• Outcome (the change being measured in the intervention or exposure group)<br>• Study design (specific design features of relevant research)<br>• Target condition (the object of a test method for diagnosis or detection) |
| **1.3 Defining eligibility criteria** |
| 1.3.1 Define and justify unambiguous and appropriate eligibility criteria for each component of the objective statement. |
| 1.3.2 Define the points at which screening for eligibility will take place (e.g. pre-screening based on title/abstract, full text screening, or both) |
| 1.3.3 For interventions, exposures and comparators: define as relevant to review objectives the eligible types of interventions and/or exposures, methods for measuring exposures, the timing of the interventions/exposures, and the interventions/exposures against which these are to be compared. |
| 1.3.4 For outcomes: define as relevant to review objectives the primary and secondary outcomes of interest (including defining which are apical and which are intermediate), what will be acceptable outcome measures (e.g. diagnostic criteria, scales) and the timing of the outcome measurement. |
| 1.3.5 For study designs: define eligible study designs per design features rather than design labels. |
| 1.3.6 Include all relevant, publicly-available evidence, except for research for which there is insufficient methodological information to allow appraisal of internal validity. |
| 1.3.7 Include evidence which is relevant to review objectives irrespective of whether its results are in a usable form. |
| 1.3.8 Include relevant evidence irrespective of language. |
| 1.3.9 Exclude evidence which is not publicly available. |
| **1.4 Planning the review methods at protocol stage** |
| 1.4.1 Design sufficiently sensitive search criteria, so that studies which meet the eligibility criteria of the review are not inadvertently excluded. |
| 1.4.2 Design “characteristics of included studies” table. |
| 1.4.3 Define the risk of bias assessment methods to be used for evaluating the internal validity of the included research. If observational studies are included, this should cover identification of plausible confounders. |
| 1.4.4 Design the methods for synthesising the included studies, to cover: qualitative and quantitative methods (with full consideration given to synthesis methods to be used when meta-analysis is not possible); assessment of heterogeneity; choice of effect measure (e.g. RR, OR etc.); methods for meta-analysis and other quantitative synthesis; pre-defined, appropriate effect modifiers for sub-group analyses. |
| 1.4.5 Define the methods for determining how, given strengths and limitations of the overall body of evidence, confidence in the results of the synthesis of the evidence for each outcome is to be captured and expressed. (For reviews which include multiple streams of evidence, this may need to be defined for each stream.) |
| 1.4.6 For reviews which include multiple streams of evidence (e.g. animal and human studies), define the methods for integrating the individual streams into an overall result. This should include a description of the relative relevance of populations (e.g. species, age, comorbidities etc.), |

| |
|---|
| exposures (e.g. timing, dose), and outcomes (direct or surrogate, acute or chronic model of disease, etc.), as appropriate, per which inferences about predicted effects in target populations can be made from observed effects in study populations. |
| 1.4.7 Pilot-test all components of the review process in which reviewer performance could affect review outcomes. This includes the design and usability of the data extraction form/s, and the conduct of the risk of bias assessment. |
| **1.5 Publishing the protocol** |
| 1.5.1 Create a permanent public record of intent to conduct the review (e.g. by registering the protocol in an appropriate registry) prior to conducting the literature search. |
| 1.5.2 As appropriate for review planning and question formulation, secure peer-review and public feedback on a draft version of the protocol, incorporating comments into the final version of the protocol. |
| 1.5.3 Publish the final version of the protocol in a public archive, prior to screening studies for inclusion in the review. |
| 1.5.4 Clearly indicate in the protocol and review report any changes in methods made after testing or conduct of any steps of the review. |
| **2. Searching for Evidence** |
| 2.1 Search all the key scientific databases for the topic, including national, regional and subject-specific databases. |
| 2.2 Define reproducible strategies for identifying and searching sources of grey literature (databases, websites etc.). |
| 2.3 Structure search strategies for each database, electronic and other source, using appropriate controlled vocabulary, free-text terms and logical operators in a manner which prioritises sensitivity. |
| 2.4 Search within the reference lists of included studies and other reviews relevant to the topic ("hand-searching") and consider searching in the reference lists of documents which have cited included studies. |
| 2.5 Search by contacting relevant individuals and organisations. |
| 2.6 Document the search methods and results in sufficient detail to render them transparent and reproducible. |
| 2.7 Re-run all searches and screen the results for potentially eligible studies within 12 months prior to publication of the review (screening at least at the level of title plus abstract). In deciding whether to incorporate new studies in the review, the importance of a possible change in results should be weighed against any delay in publication. Potentially eligible studies which have not been incorporated should be listed as "awaiting classification". |
| **3. Screening Evidence for Inclusion** |
| 3.1 Screening of each piece of evidence for inclusion to be conducted by at least two people working independently, with an appropriate process (e.g. third-party arbitration) for identifying and settling disputes. |
| 3.2 Document decisions in enough detail to allow presentation of the results of the screening process in a PRISMA flow chart. |
| 3.3 Studies which are excluded after assessment of full text should be listed in a table of excluded studies along with the reason for their exclusion (one reason is sufficient). |

| |
|---|
| 3.4 Do not exclude multiple reports of the same research (e.g. multiple publications, conference abstracts etc.); instead collate the methodological information from each of the reports as part of the data extraction process for each unit of evidence. |
| **4. Extracting Relevant Data from Included Study Reports** |
| 4.1 Collect characteristics of the included studies in sufficient detail to populate the planned "characteristics of included studies" table. |
| 4.2 Extraction of study characteristics and outcome data to be conducted by at least two people working independently with an appropriate process (e.g. third-party arbitration) for identifying and settling disputes. |
| 4.3 Assessment of risk of bias to be conducted separately from data extraction. Ideally, and where appropriate, risk of bias assessment should be conducted between extraction of study characteristics and extraction of outcome data (study results). |
| 4.4 Correct for errors and omissions in data reported in included studies by: (1) collecting the most detailed numeric data possible; (2) examining relevant retraction statements and errata for information; (3) obtaining where possible relevant unpublished data which is missing from reports and studies. |
| 4.5 Check accuracy of the numeric data in the meta-analysis utilising an appropriate process (e.g. third-party control). |
| **5. Appraising the Internal Validity of Included Studies** |
| 5.1 Appraise internal validity of each included study via the risk of bias assessment methodology specified in the protocol. |
| 5.2 Assess risk of bias per outcome or outcome-exposure pair (as appropriate) rather than per study. |
| 5.3 Risk of bias assessment is to be conducted by at least two people working independently, with an appropriate process (e.g. third-party arbitration) for identifying and settling disputes. |
| 5.4 Apply the risk of bias assessment tool thoroughly and consistently to each included study, recording each risk of bias judgement made by each reviewer, and any disagreements and how they were resolved. |
| 5.5 If there is empirical evidence which supports a judgement, comment but do not guess on likely direction and (if possible) magnitude of effect of bias. |
| 5.6 Provide appropriate explanation for judgement of risk of bias, making reference to decision processes described in the protocol, and using supporting quotes from study reports or noting if information was not available. |
| **6. Synthesising the Evidence / Deriving Summary Results** |
| 6.1 Undertake (or display) meta-analyses only when studies are sufficiently comparable as to render the combined result meaningful. |
| 6.2 Transform all scales (where appropriate) into common measures of outcome, explaining how each scale has been reinterpreted in the review. |
| 6.3 Use appropriate methods to assess the presence and extent of between-study variation (statistical heterogeneity) when undertaking a meta-analysis. |
| 6.4 If important statistical heterogeneity is observed, explain how this is accommodated in developing appropriate summary results for the review (e.g. by not pooling at all, by conducting subgroup analyses etc.) |

| |
|---|
| 6.5 Assess the potential for publication bias in the data (i.e. systematic differences between the evidence which was accessible to the review, and the evidence which was not). |
| 6.6 Assess potential impact of risk of bias in the synthesis, based on the results of the appraisal of risk of bias in the included studies (e.g. sub-group analysis excluding studies at high risk of bias; appropriate qualitative or quantitative approaches). |
| 6.7 Test the robustness of the results using sensitivity analyses (such as the impact of notable assumptions, imputed data, borderline decisions and studies at high risk of bias). |
| 6.8 If subgroup analyses are conducted, follow the subgroup analysis plan specified in the protocol, avoiding over-interpretation of any particular findings; sensible post-hoc analyses may also be carried out. |
| **7. Interpreting Results** |
| 7.1 Interpret the internal validity of the overall body of evidence by considering results of the appraisal of internal validity (risk of bias) of each included study. The review should describe the potential for biased summary results due to limitations in study design and conduct (e.g. extent of randomisation, blinding, confounding etc.) and the implications of these limitations for drawing conclusions based on the overall body of evidence. |
| 7.2 Interpret the consistency of the overall body of evidence, accounting for explainable and unexplainable variation between studies. If a meta-analysis has been conducted, consider statistical heterogeneity. Where appropriate, conduct sub-group and sensitivity analyses. |
| 7.3 Interpret any subgroup analyses without selective reporting of results or placing undue emphasis on specific findings. |
| 7.4 Interpret the precision of the results of any syntheses, taking care to interpret statistically non-significant results as findings of uncertainty rather than no effect, unless the confidence intervals are sufficiently narrow to rule out an important magnitude of effect. |
| 7.5 Interpret the magnitude of the observed effect. |
| 7.6 Interpret the dose-response relationship in the observed results. |
| 7.7 Interpret the potential effects of reporting and publication biases (e.g. unreported outcome data, unpublished studies etc.) on the observed results. |
| 7.8 Interpret the external validity of the overall body of evidence. Any inferences or predictions about effects in target populations which are made based on effects observed in the populations in the included studies should accord with the considerations defined in the protocol about the relative relevance of populations (e.g. species, age, comorbidities etc.), exposures (e.g. timing, dose), and outcomes (direct or surrogate, acute or chronic model of disease, etc.), as appropriate. Deviations from these considerations must be explained and justified. |
| 7.9 Include the “summary of findings” table. |
| 7.10 Summarise the quality of the overall body of evidence into an appropriate overall statement of confidence in the results of the synthesis. |
| **8: Drawing Conclusions** |
| 8.1 Draw out implications based only on findings from the synthesis of studies included in the review. |
| 8.2 Describe implications for research based on Population-Exposure-Comparator-Outcome or other appropriate formula consistent with that specified in the research objective. |

| |
|---|
| 8.3 Avoid describing policy implications in terms of specific actions authors feel that decision-makers should take. If authors feel it is necessary to describe policy implications, articulate them in terms of hypothetical scenarios rather than making specific policy recommendations. |

*Table 1: The full list of COSTER performance elements for sound and good practice in the planning and conduct of environmental health systematic reviews. The list of performance elements should be read alongside Table 2, which serves as an explanation and elucidation of the key recommendations of COSTER.*

# 4 Discussion

## 4.1 *How to use COSTER*

### *4.1.1 The role of COSTER in managing the quality of EH SR projects*

COSTER should be applied as one of a total of three types of mutually-supportive document which facilitate a comprehensive quality management process for the production of EH SRs. The three types of document are as follows:

a. **standards for conduct of research**, which describe key requirements for carrying out a sound and good piece of research;

b. **standards for reporting of research**, which describe the key information which needs to be presented to a reader in order that the quality of a piece of research can be evaluated;

c. **critical appraisal tools**, which help a reader analyse project documentation in order to determine the quality of a piece of research.

As a standard for conduct of EH SRs, COSTER fulfils (a) above, presenting a key set of recommended practices which, if followed, should result in a useful, transparent and credible SR product. It is a tool to be used when planning and conducting robust systematic reviews, and as a benchmark against which other SR guidance and frameworks can be assessed.

While COSTER could be used to inform the development of reporting standards (b) and critical appraisal tools (c), it has not been developed or tested for effectiveness in helping researchers report their SR, nor readers appraise a SR, and therefore should not be used for either purpose without appropriate adaptation. Examples of reporting standards being used for EH SRs include PRISMA (Moher et al. 2009) and ROSES (Haddaway et al. 2018b) . An example of a critical appraisal tool specific to EH SRs is CREST_Triage, which has been developed for facilitating consistent editorial screening decisions for EH SR submissions (Whaley and Isalski 2019).

COSTER is complementary to EH SR handbooks and guidance frameworks, describing in detail what authors of SRs should aim to achieve without providing instruction on how they should do it. This is important for allowing users of COSTER flexibility in the methods they employ, while providing clear guidance on the standard those methods need to meet. COSTER should be used alongside SR handbooks and frameworks to help determine whether the instructions and advice therein constitute good practice and to help identify potential omissions in their guidance.

#### *4.1.2 The target audience of COSTER*

COSTER is intended to be usable by any entity or practitioner responsible for or interested in conducting an EH SR project, who needs a benchmark against which they can evaluate the many different approaches they might be faced with, not all of which they can be familiar with in detail. Such entities include: independent scientists; journal editors receiving SR submissions; research teams wishing to conduct a SR; research commissioners seeking confidence that a contractor will conduct a successful SR project; quality assurance units in research-associated organisations seeking to implement consistent, sound and good SR practices; or regulatory authorities and scientific agencies seeking to demonstrate compliance with an agreed set of practices for conduct of research in a regulatory or other formal setting.

#### *4.1.3 Managing the number of performance elements in COSTER*

SRs are complex, multi-disciplinary projects which typically take 12-36 months to conduct (Borah et al. 2017; Haddaway and Westgate 2019). While 70 may seem like a large number of steps for a research team to complete, COSTER is comparable in size to IOM, which consists of 82 requirements across 4 domains, and MECIR 1.07, which consists of 75 requirements across 10 domains.

COSTER is designed to be used in parallel to the development, conduct, and reporting of a systematic review in an iterative manner which mirrors many of the considerations that should naturally arise for research teams undertaking each of these steps. Therefore, the fulfilment of COSTER's performance elements is anticipated to be already addressed or incorporated in a well-designed and well-conducted SR and would not constitute an additional burden in these scenarios. In other scenarios, COSTER should help identify oversights and limitations in methods which might threaten the integrity of a SR project. Correction of such issues should increase the quality of the resulting SR and be worth the additional effort.

#### *4.1.4 How should compliance with COSTER be described?*

When research teams report the use of COSTER in planning and conducting a SR, they are encouraged to avoid broad summary statements such as "the COSTER code of practice was

followed". Although prevalent in the literature, such self-reported statements are usually only partly true and may therefore mislead the reader about the exact methods used (Page and Moher 2017). Instead, authors should report that COSTER was used to inform the planning and conduct of a SR, and transparently describe whether and how they were able to fulfil each performance element. The elements are numbered to facilitate this process. Where people elect to depart from COSTER, justification should be provided. We recommend that COSTER-specific SR reporting standards be developed. In the interim, reporting may be facilitated by the use of standards such as PRISMA (Moher et al. 2009), ROSES (Haddaway et al. 2018b), or MOOSE (Stroup et al. 2000).

We also recommend that readers be very cautious in making any assumptions about the quality of a SR which uses or claims to have complied with COSTER. While COSTER is intended to help authors make good decisions about their EH SR methods, as a written document it has little power on its own to ensure they have been successful in making them. As is the case for any standard, claims of compliance with COSTER are open to potential abuse, either deliberate or inadvertent, as a mechanism for artificially elevating a reader's perception of the quality of a piece of research. A SR should therefore always be appraised using a valid, contextually appropriate tool before coming to any judgments about its quality.

### 4.2 *Comparing COSTER to other SR standards*

Because SR methods are relatively universal and independent of topic, there is substantial overlap between COSTER and other SR standards, including MECIR and IOM. However, COSTER is the first explicit effort by EH research practitioners to validate for their particular cultural and research context SR standards which are being applied in biomedicine. By doing this, COSTER contributes to resolving the question of which standards for the conduct of biomedical SRs can be applied to EH research. COSTER also provides a platform on which SR standards for environmental health research can be further developed, particularly in areas where the COSTER process has identified methodological guidance as being needed but immediate consensus on sound and good practice is elusive. In particular, this applies to assessing the external validity of included studies, which was presented for discussion in the initial draft of COSTER but not included in the final standard (see Supplemental Materials 3, Page 9, Performance Element 5.2).

Table 2 highlights key explanatory points for COSTER according to themes we believe are either unique to the context of EH research, address aspects of conduct of a systematic review for which it has historically been difficult in any field to achieve consensus on good practice, or we believe are a novel contribution to progressing SR standards in general. Where COSTER

closely follows the conventions of IOM and MECIR, we instead refer the reader to (Eden et al. 2011) and (Higgins et al. 2019) for explanation as to why the performance elements are considered sound and good practice in SR.

Table 2: Explanation and elucidation of key performance elements of COSTER

| | |
|---|---|
| **Performance elements** | 1.1.1 through 1.5.4 |
| **Theme** | Project planning |
| **Contribution of COSTER** | Emphasis on importance of practices in biomedical SRs for environmental health research |
| **Explanation:** It is not yet common practice for EH SRs to be conducted according to pre-published protocols, though has been changing since the date of the workshop – see e.g. (Mandrioli et al. 2018; Matta et al. 2019; Hansen et al. 2019). Protocol publication has value for reducing risk that changes in methods mid-project will bias the results of a SR, while also providing an opportunity for external peer-review of proposed approach and subsequent early identification of errors which, if left unresolved, could seriously undermine the validity of a resource-intensive project which can take years to conduct (Munafò et al. 2017). COSTER follows MECIR and IOM in providing detailed guidance on conduct of the planning and protocol phase of a SR, to help research teams avoid potentially costly errors and maximise the value of their project outcomes. | |
| **Performance elements** | 1.1.3, 1.1.4 |
| **Theme** | Disclosure and management of interests |
| **Contribution of COSTER** | Distinction between potential and apparent conflicts of interest to rationale for team selection in SRs |
| **Explanation:** COSTER adopts Columbia University's "Responsible Conduct of Research" definition of a conflict of interest (COI) as "a situation in which financial or other personal considerations would be considered by a reasonable person to have the potential to compromise or bias professional judgment and objectivity" (Columbia University 2004). In the Columbia University framework, "apparent" conflicts of interest are defined as situations "in which a reasonable person would think that the professional's judgment is likely to be compromised", while "potential" conflicts of interest are situations "that may develop into an apparent conflict of interest" (the reader should note that the framework provides a number of useful illustrative examples).<br>We believe the Columbia University COI policy offers a way of operationalising the describing and handling of risks to the integrity of a SR project from conflicts of interest. The | |

term “potential” can be applied to indicate an interest has been identified as being a conflict, but the conflict has been evaluated as not presenting a risk to the integrity of the project. When a conflict of interest is described as “apparent”, it means an interest has been evaluated and determined to present excess risk to the integrity of the project. Our view is this is a useful way of denoting interests: all identified conflicts are described; those which have been assessed for the possibility of being apparent but have been deemed not to pose excess risk to the integrity of the SR are described as potential; potential interests which have been deemed to pose excess risk are identified as apparent and are excluded.

COSTER allows for interests to be financial and/or non-financial. Similar to IOM, COSTER recognises that any potential COI can, in the right circumstances, become an apparent COI and that all potential COIs should be declared and managed. COSTER distinguishes itself from the IOM approach to COIs by emphasising that individuals with apparent conflicts of interest need only be excluded from analysis and decision-making roles in the review process. This leaves open the possibility of their involvement as individuals with specialist knowledge on which review teams can draw, while insulating the review process from risk of bias by prohibiting their involvement in decision-making. This allows EH SRs to utilise the full range of expertise of a field in which a large body of knowledge is contributed by special interest groups, and therefore many practitioners will likely have apparent COIs.

The intent of these performance elements is not to limit participation by excluding people with affiliation to broad sectors (e.g. academic grant holders, industry, or NGOs), but rather to make such associations transparent. The fundamental objective is to make interests clear and limit participation in decision-making roles when relevant interests pose a risk to the scientific integrity of a SR. In lieu of purpose-built declaration of interest forms for environmental health research, SR authors could consider using forms such as those published by the International Committee of Medical Journal Editors (International Committee of Medical Journal Editors 2013).

| | |
|---|---|
| **Performance elements** | 1.2.2, 1.4.6, 7.8 |
| **Theme** | Interpreting external validity of the evidence, and integrating multiple evidence streams |
| **Contribution of COSTER** | Adaptation of biomedical SR standards to specific context of EH research |

**Explanation:** Operationalising the interpretation of the value of non-human and *in vitro* evidence for understanding potential human health risks from environmental exposures remains a fundamental challenge in adapting SR methods to environmental health. For

healthcare interventions, IOM specifies the use of an "analytical framework which clearly lays out the chain of logic that links the health intervention to the outcomes of interest". COSTER applies this concept to the assessment of the external validity of evidence, to account for the importance in environmental health research of consistent, unbiased interpretation of an evidence base which is often indirect. Environmental health researchers are increasingly interested in how indirect mechanistic evidence can be organised in predictive networks (Villeneuve et al. 2014a, 2014b) or Key Characteristics frameworks (Smith et al. 2016; Arzuaga et al. 2019; Luderer et al. 2019) to help anticipate whether an environmental challenge will cause an adverse health outcome. In anticipation of the development of systematic approaches to developing and assessing the plausibility of such networks or framework analyses, in requirement 1.2.2 COSTER requires that authors offer the basic elements of a theoretical framework for interpreting the external validity of included studies as part of the protocol. The framework should describe why and to what extent different populations (e.g. species, developmental stage), exposures (e.g. timing, dose, similarity of substance / read-across) and outcomes (e.g. apical, intermediate) will be considered by the reviewers to be comparable to the target populations, exposures and outcomes of interest. Performance element 7.8 specifies that interpretation of the results of synthesis are made in accordance with this pre-specified framework.

While such inferential frameworks may currently be limited in scope, and there should be caution about overly-prescriptive use which can lead to spurious rejection of true hypotheses as much as spurious acceptance of false ones, the authors believe that the use of such frameworks is important in discouraging the sort of ad-hoc analysis of evidence which is vulnerable to expectation bias. COSTER takes an initial step in requiring the application of such frameworks for environmental health SRs.

| | |
|---|---|
| **Performance elements** | 1.2.3, 1.3.3, 1.3.4, 1.3.5, 1.3.9 |
| **Theme** | Formulation of research objectives |
| **Contribution of COSTER** | Formal clarification of use of PECO-style statements in formulating SR objectives in EH research |

**Explanation**: COSTER requires that SR objectives be formulated in a structured format using appropriate elements of the PECOTS (Population-Exposure/Intervention-Comparator-Outcome-Target Condition-Study Design) mnemonic. While questions around effects of chemical exposures are more common, some environmental health SRs investigate interventions (such as amelioration of the effects of exposures) and this is expressly allowed for in COSTER. COSTER also specifies in detail the specific aspects of the PECOTS elements

which should be considered in establishing the objectives of a EH SR, with elements such as timing of exposure being recognised as a potentially critical issue in reliably identifying health risks of chemical exposure, and an injunction that these be considered and defined as necessary. More specific guidance on good practice in the formulation of PECO statements has been developed since COSTER was finalised, to which prospective authors may wish to refer (Morgan et al. 2018).

| | |
|---|---|
| **Performance elements** | 1.3.6, 1.3.9, 3.4 |
| **Theme** | Including informally published or previously unpublished literature, regardless of usability in the planned analysis |
| **Contribution of COSTER** | Provides unambiguous rationale for exclusion of study reports due to insufficient information content |

**Explanation:** The consensus view of the authors is that grey literature (i.e. studies which have not been published in peer-reviewed journals) should be included in systematic reviews. This is because the relevance of evidence is determined by the SR objectives, not by the publication status of that evidence nor its compatibility with the analyses planned by the reviewers. The inclusion of grey literature also acts as a safeguard against the influence of publication bias; however, researchers should never assume that the grey literature which can be located will be representative of the grey literature overall. The authors also acknowledge that inclusion of grey literature can be daunting. Therefore, COSTER provides an explicit rationale for where researchers can draw the line on including study reports in a SR, as follows.

Firstly, in keeping with the SR principle of transparency, COSTER mandates that only publicly available information about a study is eligible for inclusion (requirement 1.3.9). A SR which brings into the public domain previously inaccessible information can be the mechanism by which such data becomes publicly accessible and therefore eligible for inclusion. This has happened with SRs from WHO (Mandrioli et al. 2018) and Cochrane (Jefferson et al. 2014).

Secondly, to prevent the inclusion in a SR of evidence which is potentially misleading but cannot be identified as such by the reviewers, COSTER mandates exclusion of studies for which there is insufficient information for risk of bias to be evaluated (performance element 1.3.6).

Thirdly, COSTER defines the included study itself, not documents describing the study, as the unit of evidence (performance element 3.4). Therefore, all publicly accessible study documents including conference abstracts etc. should be gathered and assessed for

information content as a whole, before a decision is made to exclude a study in accordance with performance element 1.3.6.

Fourthly, documents should be included in a SR regardless of whether their data fit the analysis plan of the reviewers. This is to ensure that study documents which may contain information of potential relevance to the SR's research objectives are not excluded from the data extraction step of the SR.

Many studies – especially epidemiological studies – cannot release detailed information on individual participants owing to privacy concerns and legal mandates. The intent of this requirement in COSTER is not to avoid such studies, but rather to ensure that the uses of study-specific findings within the larger analysis should be supported by those aspects of the underlying data that are available for public scrutiny. Authors should also be cautious about the potential for double-counting populations when combining multiple study reports, particularly when there is partial overlap between multiple documents.

| | |
|---|---|
| **Performance elements** | 1.5.1, 1.5.2, 1.5.3 |
| **Theme** | Protocol publication |
| **Contribution of COSTER** | Differentiates between protocol registration and publication as distinct steps of the methods development process |

**Explanation**: Protocol registries such as PROSPERO (Centre for Reviews and Dissemination) and preprint repositories such as Zenodo (CERN) and the Open Science Framework (OSF) allow authors to register their methods in advance of conducting a SR. In theory, such third-party version control of the registered protocol allows changes in methods to be audited, discouraging bias which can be introduced by ad-hoc decision-making. However, there are no protocol registries which currently require authors to submit sufficient information about methods that a registered protocol can be assumed to be a complete plan for conducting a SR. Nor do such registries have capacity to peer-review protocols for soundness of the proposed methods, at most performing only basic quality control checks. This leads to a situation in which the value of registration for ensuring the comprehensiveness and validity of methods for a given protocol is unclear. Therefore, it is the view of the authors that the current value of a registered protocol is primarily as a record of intent to conduct a SR, rather than serving as a guarantee of comprehensive documentation of methods prior to conduct of a SR.

COSTER addresses this ambiguity, of the status of registered protocol vs. comprehensive documentation of proposed methods, by specifying that authors of SRs take a two-step approach to protocol publication. As the first step, an outline of the proposed SR with the minimum of necessary information to characterise objectives and approach should be posted

on an appropriate public registry or functional equivalent thereof, over which the authors have no direct control (requirement 1.5.1). This first draft is the permanent public record of intent to conduct a systematic review, functioning to communicate research aims and help other review teams avoid planning duplicate SRs. As the second step, this draft can then be developed in further detail as a full protocol, which is submitted to external peer-review or other appropriate quality management process (requirement 1.5.2), and then published either in a scientific journal or a preprint repository (requirement 1.5.3). An example of journal publication of a protocol is provided by (Mandrioli et al. 2018), and in a public repository by (Martin et al. 2018). A general example of a two-stage peer-review process is provided by the Registered Reports model of publication (Chambers 2019).

| | |
|---|---|
| **Performance elements** | 1.4.3, 5 |
| **Theme** | Internal validity assessment |
| **Contribution of COSTER** | Explicit specification of risk of bias methods for assessing internal validity of included studies |

**Explanation**: To prevent systematic errors in included studies being transmitted through to the findings of a SR, it is necessary that each individual included study be assessed for internal validity, i.e. its potential to have biased results. Hence, COSTER explicitly requires each individual included study to be assessed for risk of bias. While anticipating direction and magnitude of bias is desirable, it is often not possible. This is due to a general lack of strong empirical evidence as to the magnitude and direction of bias for a given study design (one reason for this is that studies attempt to measure bias are very difficult to power, see e.g.(Giraudeau et al. 2016)). Judgements in a SR should be based on evidence not guess-work. Evidence-based assessments of internal validity which successfully quantify bias are consistent with COSTER.

COSTER does not state which instruments should be used by authors to assess risk of bias, leaving it to SR authors to determine which assessment methods are most suited to their research objectives (except, the tool should specifically target risk of bias). COSTER does, however, present a number of performance elements around the process of risk of bias assessment to ensure successful implementation of the risk of bias tool, whatever tool is selected. This includes assessing risk of bias per outcome (requirement 5.2) and making sure each judgement is transparent and grounded in the reviewed text (requirement 5.6). Risk of bias assessment should be sensitive to differences in study designs and use suitable assessment instruments accordingly. The assessment process should balance being sufficiently constrained in approach that it is conducted against a clear standard, but not be

so over-constrained that it ends up mischaracterizing limitations in research through comparison of a study to inappropriately rigid criteria. Various systematic reviews and evaluations of risk of bias assessment tools have been conducted (Wang et al. 2019; Krauth et al. 2013; Rooney et al. 2016) and a user of COSTER may wish to refer to such in deciding which tools to apply in a SR, in compliance with the relevant COSTER performance elements.

| | |
|---|---|
| **Performance elements** | 1.4.5, 7.1, 7.2, 7.4, 7.5, 7.6, 7.7, 7.8, 7.10 |
| **Theme** | Assessment of confidence in the overall body of evidence |
| **Contribution of COSTER** | Emphasis on evaluation of quality of evidence against pre-specified criteria known to be of importance when assessing confidence in the results of a SR |

**Explanation**: COSTER presents eight characteristics of a body of evidence which should be systematically evaluated in the course of determining how certain are the results of a SR. These apply to interpreting the overall strength of the evidence base, considered as a whole. While the characteristics are derived from those utilised in the GRADE framework (Guyatt et al. 2008; Guyatt et al. 2011), there are no specifications in COSTER regarding how they ought to be interpreted, except that the approach should be described in the protocol. The authors note there is ongoing work by the GRADE Working Group to further develop the GRADE methodology for the environmental health context (Morgan et al. 2016; Morgan et al. 2019), and that the US NTP OHAT (Rooney et al. 2014) and the Navigation Guide (Woodruff and Sutton 2014) both employ a close interpretation of the GRADE framework in their approaches to conducting SRs. A systematic approach to assessing confidence in a body evidence is important, because readers of a SR need a trustworthy analysis of the trustworthiness of the evidence. A high-quality review of low-quality evidence is still a trustworthy review – even if the review process has shown that the reader cannot put much trust in the evidence itself.

| | |
|---|---|
| **Performance elements** | 8.3 |
| **Theme** | Making policy recommendations |
| **Contribution of COSTER** | Emphasises that recommendations about interventions are often beyond the scope of a SR of health effects from environmental exposures |

**Explanation**: The development of environmental health policy requires accounting for a wide range of issues relating to evidence of health risks, due political process, and the values and preferences of stakeholders affected by the policy. Systematic reviews ask focused questions

which typically respond to only one or two of the full set of issues which may need to be accounted for by a decision-maker when developing policy. This is especially true for SRs of health effects of environmental exposures: while they address potential causes of adverse health outcomes (are etiological), they would not normally also investigate evidence for the effectiveness of interventions aimed at mitigating those adverse outcomes. While identifying threshold limits, which then inform policy decisions, is of course often the core business of this type of SR, COSTER adheres to the principle that the conclusions of a SR should not reach beyond the evidence which was included within it. COSTER therefore recommends authors resist answering questions about how best to mitigate the effects of an exposure or achieve a risk threshold when the evidence relating to this has not been addressed by the SR.

The authors recognise, however, that SRs characterising adverse outcomes from environmental exposures are often conducted to support policy decisions. COSTER therefore requires that, when they do form part of the conclusions of a SR, any policy implications be presented as hypothetical frameworks, whereby authors state that if certain conditions obtain, then a given intervention may be effective for mitigating harm. Any assumptions the authors make about values, other evidence and potential consequences of a decision should be made explicit when describing potential interventions to address an environmental exposure or mitigate health risks arising therefrom.

### 4.3 *Strengths and limitations of COSTER*

**Consensus process.** In developing COSTER, a deliberate attempt was made to emulate formal standards processes such as those followed by the British Standards Institution. We are not aware of other research standards which have sought to do this. Partial success was achieved: a group of 21 experts from various stakeholder groups agreed on a comprehensive set of recommendations which exceed current practice in many contemporary EH SRs with no objections to publication from non-author participants. However, a number of aspects of the consensus process could be improved upon. These include: securing greater capacity to organise and participate in more face-to-face meetings; a longer consensus process to accommodate the more challenging discussions, thereby potentially allowing a wider consensus covering more elements of the SR process; and implementation of more formal minute-taking and communication structures as would be provided by organisations such as BSI, for making the consensus process more auditable, improve transparency, and facilitate communication between participants in the consensus process.

Communicating the technicalities of methods for a formal consensus process at the same time as COSTER itself was also challenging. Standardisation processes have their own jargon and procedures which, while well-established in bodies such as ISO and BSI, are not always intuitive to understand. One example of this is the lack of fit between the language of standardisation, which employs terms such as "requirements", and the nature of codes of practice such as COSTER, which are sets of expert recommendations. We believe a detailed explication of the processes and jargon of standardisation would be useful in the current environment of largely informal approaches being taken to developing research standards. This could help make the authority and applicability of any putative standard much clearer. This would be of value for COSTER; however, as a manuscript it would need to be developed separately.

**Impact of conflicts of interest**: In order to secure cross-sector consensus, we invited participants because they had varied interests in developing a standard for conduct of EH SRs. We did not attempt to directly manage the interests of participants, as they were seen as desirable, but we did seek balance across stakeholder groups and domains of expertise. We believe involvement of a broad cross-section of stakeholder groups strengthens COSTER's generalisability and broadens its acceptability, while reducing the risk that any individual interest group has had excess influence on the consensus outcome. Given that COSTER does not seem to negate the principles of IOM or MECIR, we do not believe that the interests of those involved have not undermined the integrity of COSTER as a code of practice.

In terms of future work in this space, it could be beneficial to define what constitutes a COI in the context of development of a general SR standard, as it could facilitate proactive identification and management of COIs of which we were not aware. Defining a conflict of interest for SR development is not something we were able to do, and there was no guidance we were aware of which would support us in doing this. Cochrane guidance perhaps came closest, but we could not see how a party could have a direct financial stake in the outcome of COSTER. In relation to declarations of interest (DOIs), these are self-reported by authors; given the lack of definition of what constitutes a COI in this project, and the usual limitations of self-reporting interests, the DOI forms are varied in terms of what is reported. This could be improved upon in future.

**No systematic review of existing standards**. Rather than conduct a SR of existing standards and guidance of potential relevance to the development of COSTER, we relied on participants' tacit knowledge of these. We secured participation of stakeholders with experience developing the following SR frameworks: the Navigation Guide (Woodruff and Sutton 2014), the National Toxicology Program Office of Health Assessment and Translation (Rooney et al. 2014);

SYRINA (Vandenberg et al. 2016); the European Food Safety Authority (EFSA 2010); Cochrane's MECIR standards and the Cochrane Handbook (Higgins et al. 2011); GRADE (Morgan et al. 2016); the IARC Monographs Program (IARC 2015); and SYRCLE (de Vries, Rob B. M. et al. 2015). MECIR and IOM, as seed standards for COSTER, were selected as authoritative standards developed using robust methods, likely to be comprehensive and not misleading in either what they include or omit. These two existing standards provided 80 seed criteria (see Supplemental Materials 03). While a SR of existing standards and guidelines could have extended this list, it would have been a considerable task to undertake without obvious proportional benefit to a project which sought to define expert consensus on basic expected practices in EH SR. This is an element of the COSTER development methodology which could certainly be improved in future; a detailed discussion of this follows in Section 4.4 below.

**Potential for misuse of COSTER**. Terms such as "systematic review" and inappropriate reference to standards such as PRISMA are widespread in the literature. The value of all SRs is diminished by misuse of the term "systematic" and the publication of poor-quality SR manuscripts. With COSTER, we are seeking to avert this situation by giving authors, reviewers, editors and other stakeholders one important element of a SR quality management process which, when combined with appropriate reporting standards and critical appraisal tools, will give them confidence in their ability to distinguish good SRs from poor quality ones. While we hope COSTER will discourage the cherry-picking of convenient practices from among the current array of guidance documents, there is also a risk that COSTER will be misused. However, this is a problem common to all standards. At the very least, by providing an unambiguous set of criteria against which a SR can be compared, we would hope that it will be easier to identify when phrases such as "COSTER-compliant" and "employed systematic review methods" are being misused.

### 4.4 *Future development of COSTER*

As a code of practice, COSTER represents the first step in a broader research and consensus-building process which it is hoped will eventually yield a robust, international standard for conduct of systematic reviews in environmental health research. Formal standards are typically based on both expectation and empirical evidence that the practices described in the standard contribute to a product or process being fit for purpose, combined with broad acceptance of the practices among the community that is expected to adopt the standard (British Standards Institution 2016b). Since SR methods are still relatively new in environmental health research, it follows that while expectations for what should work can be captured, and the consensus view of small groups of experienced practitioners be secured, evidence for what is effective practice is not universally available. This is particularly true for areas in which SR methods are

not readily portable from social science and medical contexts to environmental health, or where environmental health researchers face challenges not encountered in other fields. Broad community consensus is also an unrealistic goal when only a small, albeit growing, part of the community is employing SR methods in conducting reviews of evidence. It also needs to be acknowledged that while COSTER represents the consensus view of the authors, other expert groups may disagree with some of the performance elements of COSTER. Such disagreement is healthy: by making explicit a set of key recommended practices for SR, COSTER serves as a focal point for advancing consensus across groups.

As community experience in environmental health SR develops over the next period, the authors suggest that future development of COSTER adapt the framework for development of reporting guidelines for health research presented in (Moher et al. 2014). This framework emphasises four steps:

1. a systematic review of existing standards and guidelines;
2. a systematic review of the prevalence of current research practices;
3. critical appraisal of existing guidelines and current research practices for completeness, face validity, and construct validity;
4. a process to determine community consensus on best practices and the criteria for a guideline.

Steps 1 and 2 would result in a larger seed-set of potential performance elements than was provided by selecting the MECIR and IOM standards as the basis for the current consensus. However, such a SR could be a significant undertaking, as it requires a decision as to what is relevant (e.g. should nutrition and public health standards be included?), and potentially interpreting the implied standards in several large handbooks, a large number of reporting standards and potentially even individual SR study reports as well. This is a major challenge for qualitative analysis.

Another benefit of Step 2 would be in providing evidence of what community practices actually are. This would be useful information for strategic implementation of COSTER, ongoing education about COSTER as a new standard, and identification of COSTER's relative advantages or otherwise over other standards and guidance.

Steps 1 and 2 provide data for Step 3, as they describe the extent to which current practices are aligned with what might be considered by relevant stakeholders as "best" practices, providing further empirical evidence for a formal standard. Step 3, when informed by step 2,

also allows recent or rare practices to be identified and considered for the standard, thereby extending the seed criteria beyond issues of common discussion.

Examples of recent methodological innovations in EH SR which could be considered for inclusion in future versions of COSTER include:

- more detailed performance elements for handling of specific types of evidence, including mechanistic and *in vitro* study designs, observational studies and controlled trials in humans;
- the handling of evidence of the efficacy of EH interventions, such as health benefits from introducing low-smoke cookstoves (Quansah et al. 2017);
- more advanced evidence integration techniques such as triangulation (Lawlor et al. 2016) and meta-regression (Phung et al. 2017);
- more detailed performance elements for critically appraising the external validity of included studies.

Step 4, as a broad consensus process, would provide a community view of where current practices fall short of expectation or need, or where specific processes might exceed what the community views as strictly necessary for conduct of a robust EH SR. For future versions of COSTER, this should be extended beyond the 21 people involved here; however, we note that while this will broaden the potential number of topics for discussion and the authority and general acceptance of COSTER, involving more people in the process will extend the time required for discussions as consensus will have to accommodate a greater diversity of opinion. Nor will this necessarily lead to a more demanding standard, if the current performance elements of COSTER are considered by a future-convened consensus group to be excessive. Care will need to be taken to maintain stakeholder balance as numbers of participants are increased.

We recommend COSTER be re-assessed according to the above methodology, with a view to an updated standard being published around 2025.

## 5 Conclusion

COSTER presents the consensus view of a group of expert practitioners as to a minimum set of key performance elements for planning and conducting a sound and good systematic review. The lack of current guidelines for conduct of high quality environmental health SRs, coupled with exponential growth in publication of SRs (Whaley et al. 2016b), justifies the introduction of COSTER as authoritative but intermediate guidance which authors and publishers can use to

immediately improve the quality of SRs. If followed, COSTER should significantly increase the likelihood of success and stakeholder acceptance of an environmental health SR project. As a first step in establishing a formal, community-wide standard, it is intended that COSTER be critiqued and improved over time, as part of a wider process which will ultimately yield a definitive description of minimum requirements for conduct of SRs in environmental health research.

**Acknowledgements**

We would like to thank Kate Jones and the Royal Society of Chemistry for hosting the workshop, and Lancaster University Faculty of Science and Technology and Lancaster Environment Centre for providing funding to run the workshop. Funding was also provided by the UK's Economic & Social Research Council (ESRC) "Radical Futures" programme and the Engineering & Physical Science Research Council (EPSRC) "Impact Acceleration Award" EP/K50421X/1 for developing systematic review methodology for environmental health.

The authors declare no competing financial interests. Further details on potential COIs have been provided in the DOI forms (see supplemental materials). The views expressed in this paper are those of the authors and do not necessarily reflect the views or policies of their respective employers or organisations. Previous submitted versions of the manuscript have been archived at https://doi.org/10.5281/zenodo.3560978.

We would also like to thank the following for their contribution to the workshop discussions: Sarah Bull (Royal Society of Chemistry); Richard Brown (World Health Organization); Kurt Straif (ret.) and Kathryn Guyton (International Agency for Research on Cancer); Julian Higgins (University of Bristol); Toby Lasserson (Cochrane Editorial Unit); Jennifer McPartland (Environmental Defense Fund); Sharon Munn (EU Joint Research Centre); Angelika Tritscher (World Health Organization); Christopher Weiss (US National Institute of Environmental Health Sciences). TW did not participate in the workshop but contributed to the consensus development calls and the manuscript.

## Publication bibliography

Arzuaga, Xabier; Smith, Martyn T.; Gibbons, Catherine F.; Skakkebæk, Niels E.; Yost, Erin E.; Beverly, Brandiese E. J. et al. (2019): Proposed Key Characteristics of Male Reproductive Toxicants as an Approach for Organizing and Evaluating Mechanistic Evidence in Human Health Hazard Assessments. In *Environmental health perspectives* 127 (6), p. 65001. DOI: 10.1289/EHP5045.

Borah, Rohit; Brown, Andrew W.; Capers, Patrice L.; Kaiser, Kathryn A. (2017): Analysis of the time and workers needed to conduct systematic reviews of medical interventions using data from the PROSPERO registry. In *BMJ open* 7 (2), e012545. DOI: 10.1136/bmjopen-2016-012545.

British Standards Institution (2016a): Different types of standards. Available online at http://www.bsigroup.com/en-GB/standards/Information-about-standards/different-types-of-standards/, updated on 26/10/16.

British Standards Institution (2016b): What is a standard and what does it do? Available online at http://www.bsigroup.com/en-GB/standards/Information-about-standards/what-is-a-standard/, checked on 26/10/16.

Campbell Collaboration (2014): Methodological Expectations of Campbell Collaboration Intervention Reviews (MEC2IR), Conduct Standards, version 1.0. Available online at http://www.campbellcollaboration.org/news_/Campbell_adopts_MEC2IR_guidelines.php, checked on 3/8/2017.

Centre for Reviews and Dissemination: PROSPERO: International prospective register of systematic reviews. University of York. Available online at https://www.crd.york.ac.uk/prospero/, checked on 7/21/2019.

CERN: Zenodo. European Organization for Nuclear Research. Available online at https://zenodo.org/, checked on 7/21/2019.

Chambers, Chris (2019): What's next for Registered Reports? In *Nature* 573 (7773), pp. 187–189. DOI: 10.1038/d41586-019-02674-6.

Chandler, Jackie; Churchill, Rachel; Higgins, Julian; Lasserson, Toby; Tovey, David (2013): Methodological standards for the conduct of new Cochrane Intervention Reviews. Cochrane Editorial Unit.

Columbia University (2004): Responsible Conduct of Research: Conflicts of Interest. Available online at http://ccnmtl.columbia.edu/projects/rcr/rcr_conflicts/foundation/index.html, checked on 1/3/2018.

de Vries, Rob B. M.; Hooijmans, Carlijn R.; Langendam, Miranda W.; van Luijk, Judith; Leenaars, Marlies; Ritskes-Hoitinga, Merel; Wever, Kimberley E. (2015): A protocol format for the preparation, registration and publication of systematic reviews of animal intervention studies. In *Evidence-based Preclinical Medicine* 2 (1), e00007. DOI: 10.1002/ebm2.7.

Eden, Jill; Levit, Laura A.; Berg, Alfred O.; Morton, Sally C. (2011): Finding what works in health care. Standards for systematic reviews. Washington, D.C.: National Academies Press.

EFSA (2010): Application of systematic review methodology to food and feed safety assessments to support decision making. In *EFSA Journal* 8 (6), p. 1637. DOI: 10.2903/j.efsa.2010.1637.

EFSA (2015): Principles and process for dealing with data and evidence in scientific assessments. In *EFSA Journal* 13 (6), p. 4121. DOI: 10.2903/j.efsa.2015.4121.

EPA (2018): Application of Systematic Review in TSCA Risk Evaluations. US EPA Office of Chemical Safety and Pollution Prevention (EPA Document # 740-P1-8001). Available online at https://www.epa.gov/assessing-and-managing-chemicals-under-tsca/application-systematic-review-tsca-risk-evaluations, checked on 5/8/2019.

EQUATOR Network (2020): Library for health research reporting. Available online at https://www.equator-network.org/library/, checked on 2/20/2020.

Giraudeau, Bruno; Higgins, Julian P. T.; Tavernier, Elsa; Trinquart, Ludovic (2016): Sample size calculation for meta-epidemiological studies. In *Statistics in medicine* 35 (2), pp. 239–250. DOI: 10.1002/sim.6627.

Guyatt, Gordon H.; Oxman, Andrew D.; Schünemann, Holger J.; Tugwell, Peter; Knottnerus, Andre (2011): GRADE guidelines: a new series of articles in the Journal of Clinical Epidemiology. In *Journal of clinical epidemiology* 64 (4), pp. 380–382. DOI: 10.1016/j.jclinepi.2010.09.011.

Guyatt, Gordon H.; Oxman, Andrew D.; Vist, Gunn E.; Kunz, Regina; Falck-Ytter, Yngve; Alonso-Coello, Pablo; Schünemann, Holger J. (2008): GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. In *BMJ (Clinical research ed.)* 336 (7650), pp. 924–926. DOI: 10.1136/bmj.39489.470347.AD.

Haddaway, Neal R.; Macura, Biljana; Whaley, Paul; Pullin, Andrew S. (2018a): Response to "Every ROSE has its thorns". In *Environ Evid* 7 (1), p. 20. DOI: 10.1186/s13750-018-0133-3.

Haddaway, Neal R.; Macura, Biljana; Whaley, Paul; Pullin, Andrew S. (2018b): ROSES RepOrting standards for Systematic Evidence Syntheses: pro forma, flow-diagram and descriptive summary of the plan and conduct of environmental systematic reviews and systematic maps. In *Environ Evid* 7 (1), p. 409. DOI: 10.1186/s13750-018-0121-7.

Haddaway, Neal R.; Westgate, Martin J. (2019): Predicting the time needed for environmental systematic reviews and systematic maps. In *Conservation biology : the journal of the Society for Conservation Biology* 33 (2), pp. 434–443. DOI: 10.1111/cobi.13231.

Hansen, Martin Rune Hassan; Jørs, Erik; Sandbæk, Annelli; Kolstad, Henrik Albert; Schullehner, Jörg; Schlünssen, Vivi (2019): Exposure to neuroactive non-organochlorine insecticides, and diabetes mellitus and related metabolic disturbances: Protocol for a systematic review and meta-analysis. In *Environment international* 127, pp. 664–670. DOI: 10.1016/j.envint.2019.02.074.

Higgins, J.P.T.; Thomas, J.; Chandler, J.; Cumpston, M.; Li, T.; Page, M. J.; Welch, V. A. (Eds.) (2019): Cochrane Handbook for Systematic Reviews of Interventions version 6.0 (updated July 2019). Cochrane. Available online at www.training.cochrane.org/handbook.

Higgins, Julian P. T.; Altman, Douglas G.; Gøtzsche, Peter C.; Jüni, Peter; Moher, David; Oxman, Andrew D. et al. (2011): The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. In *BMJ (Clinical research ed.)* 343, d5928.

Higgins, Julian P. T.; Lasserson, Toby; Chandler, Jackie; Tovey, David; Churchill, Rachel (2018): Methodological Expectations of Cochrane Intervention Reviews (MECIR). Version 1.07. Cochrane. Available online at https://community.cochrane.org/mecir-manual, updated on November 2018.

IARC (2015): IARC Monographs on the Evaluation of Carcinogenic Risks to Humans: Preamble. World Health Organization International Agency for Research on Cancer. Lyon, France.

IARC (2019a): IARC Monographs on the Identification of Carcinogenic Hazards to Humans: Preamble. IARC. Lyon, France.

IARC (2019b): Instructions for Authors for the Preparation of Drafts for IARC Monographs, updated on 6/17/2019.

International Committee of Medical Journal Editors (2013): ICMJE Form for Disclosure of Potential Conflicts of Interest. Available online at http://www.icmje.org/conflicts-of-interest/, checked on 1/3/2018.

ISO/IEC (2004): ISO/IEC 2:2004 Standardization and related activities -- General vocabulary. ISO/IEC. Switzerland. Available online at https://www.iso.org/standard/39976.html, checked on 7/20/2017.

Jefferson, Tom; Jones, Mark A.; Doshi, Peter; Del Mar, Chris B.; Hama, Rokuro; Thompson, Matthew J. et al. (2014): Neuraminidase inhibitors for preventing and treating influenza in

healthy adults and children. In *The Cochrane database of systematic reviews* (4), CD008965. DOI: 10.1002/14651858.CD008965.pub4.

Krauth, David; Woodruff, Tracey J.; Bero, Lisa (2013): Instruments for assessing risk of bias and other methodological criteria of published animal studies: a systematic review. In *Environmental health perspectives* 121 (9), pp. 985–992. DOI: 10.1289/ehp.1206389.

Lawlor, Debbie A.; Tilling, Kate; Davey Smith, George (2016): Triangulation in aetiological epidemiology. In *International journal of epidemiology* 45 (6), pp. 1866–1886. DOI: 10.1093/ije/dyw314.

Liberati, Alessandro; Altman, Douglas G.; Tetzlaff, Jennifer; Mulrow, Cynthia; Gøtzsche, Peter C.; Ioannidis, John P. A. et al. (2009): The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. In *BMJ (Clinical research ed.)* 339, b2700. DOI: 10.1136/bmj.b2700.

Luderer, Ulrike; Eskenazi, Brenda; Hauser, Russ; Korach, Kenneth S.; McHale, Cliona M.; Moran, Francisco et al. (2019): Proposed Key Characteristics of Female Reproductive Toxicants as an Approach for Organizing and Evaluating Mechanistic Data in Hazard Assessment. In *Environmental health perspectives* 127 (7), p. 75001. DOI: 10.1289/EHP4971.

Mandrioli, Daniele; Schlünssen, Vivi; Ádám, Balázs; Cohen, Robert A.; Colosio, Claudio; Chen, Weihong et al. (2018): WHO/ILO work-related burden of disease and injury: Protocol for systematic reviews of occupational exposure to dusts and/or fibres and of the effect of occupational exposure to dusts and/or fibres on pneumoconiosis. In *Environment international* 119, pp. 174–185. DOI: 10.1016/j.envint.2018.06.005.

Martin, O. V.; Bopp, S.; Ermler, S.; Kienzler, A.; McPhie, J.; Paini, A. et al. (2018): Protocol For A Systematic Review Of Ten Years Of Research On Interactions In Chemical Mixtures Of Environmental Pollutants.

Matta, Komodo; Ploteau, Stéphane; Coumoul, Xavier; Koual, Meriem; Le Bizec, Bruno; Antignac, Jean-Philippe; Cano-Sancho, German (2019): Associations between exposure to organochlorine chemicals and endometriosis in experimental studies: A systematic review protocol. In *Environment international* 124, pp. 400–407. DOI: 10.1016/j.envint.2018.12.063.

Moher, David; Altman, Douglas G.; Schulz, Kenneth F.; Simera, Iveta (2014): How to Develop a Reporting Guideline. In David Moher, Douglas G. Altman, Kenneth F. Schulz, Iveta Simera, Elizabeth Wager (Eds.): Guidelines for Reporting Health Research: A User's Manual. Oxford, UK: John Wiley & Sons, Ltd, pp. 14–21.

Moher, David; Liberati, Alessandro; Tetzlaff, Jennifer; Altman, Douglas G. (2009): Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. In *Annals of internal medicine* 151 (4), p. 264.

Morgan, Rebecca L.; Beverly, Brandy; Ghersi, Davina; Schünemann, Holger J.; Rooney, Andrew A.; Whaley, Paul et al. (2019): GRADE guidelines for environmental and occupational health: A new series of articles in Environment International. In *Environment international* 128, pp. 11–12. DOI: 10.1016/j.envint.2019.04.016.

Morgan, Rebecca L.; Thayer, Kristina A.; Bero, Lisa; Bruce, Nigel; Falck-Ytter, Yngve; Ghersi, Davina et al. (2016): GRADE: Assessing the quality of evidence in environmental and occupational health. In *Environment international* 92-93, pp. 611–616. DOI: 10.1016/j.envint.2016.01.004.

Morgan, Rebecca L.; Whaley, Paul; Thayer, Kristina A.; Schünemann, Holger J. (2018): Identifying the PECO: A framework for formulating good questions to explore the association of environmental and other exposures with health outcomes. In *Environment international* 121 (Pt 1), pp. 1027–1031. DOI: 10.1016/j.envint.2018.07.015.

Munafò, Marcus R.; Nosek, Brian A.; Bishop, Dorothy V. M.; Button, Katherine S.; Chambers, Christopher D.; Du Percie Sert, Nathalie et al. (2017): A manifesto for reproducible science. In *Nat Hum Behav* 1 (1), e124. DOI: 10.1038/s41562-016-0021.

NAS (2014): Review of EPA's Integrated Risk Information System (IRIS) Process. Washington (DC).

NAS (2017): Application of Systematic Review Methods in an Overall Strategy for Evaluating Low-Dose Toxicity from Endocrine Active Chemicals. Washington (DC).

NTP OHAT (2019): Handbook for Conducting a Literature-Based Health Assessment Using OHAT Approach for Systematic Review and Evidence Integration. US National Toxicology Program Office of Health Assessment and Translation. Available online at https://ntp.niehs.nih.gov/pubhealth/hat/review/index-2.html, checked on 5/8/2019.

Page, Matthew J.; Moher, David (2017): Evaluations of the uptake and impact of the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) Statement and extensions. A scoping review. In *Systematic reviews* 6 (1), p. 263. DOI: 10.1186/s13643-017-0663-8.

Phung, Dung; Des Connell; Rutherford, Shannon; Chu, Cordia (2017): Cardiovascular risk from water arsenic exposure in Vietnam: Application of systematic review and meta-regression analysis in chemical health risk assessment. In *Chemosphere* 177, pp. 167–175. DOI: 10.1016/j.chemosphere.2017.03.012.

Quansah, Reginald; Semple, Sean; Ochieng, Caroline A.; Juvekar, Sanjar; Armah, Frederick Ato; Luginaah, Isaac; Emina, Jacques (2017): Effectiveness of interventions to reduce household air pollution and/or improve health in homes using solid fuel in low-and-middle income countries: A systematic review and meta-analysis. In *Environment international* 103, pp. 73–90. DOI: 10.1016/j.envint.2017.03.010.

Rooney, Andrew A.; Boyles, Abee L.; Wolfe, Mary S.; Bucher, John R.; Thayer, Kristina A. (2014): Systematic review and evidence integration for literature-based environmental health science assessments. In *Environmental health perspectives* 122 (7), pp. 711–718. DOI: 10.1289/ehp.1307972.

Rooney, Andrew A.; Cooper, Glinda S.; Jahnke, Gloria D.; Lam, Juleen; Morgan, Rebecca L.; Boyles, Abee L. et al. (2016): How credible are the study results? Evaluating and applying internal validity tools to literature-based assessments of environmental health hazards. In *Environment international* 92-93, pp. 617–629. DOI: 10.1016/j.envint.2016.01.005.

Schaefer, Heather R.; Myers, Jessica L. (2017): Guidelines for performing systematic reviews in the development of toxicity factors. In *Regulatory toxicology and pharmacology : RTP* 91, pp. 124–141. DOI: 10.1016/j.yrtph.2017.10.008.

Singla, Veena I.; Sutton, Patrice M.; Woodruff, Tracey J. (2019): The Environmental Protection Agency Toxic Substances Control Act Systematic Review Method May Curtail Science Used to Inform Policies, With Profound Implications for Public Health. In *American journal of public health* 109 (7), pp. 982–984. DOI: 10.2105/AJPH.2019.305068.

Smith, Martyn T.; Guyton, Kathryn Z.; Gibbons, Catherine F.; Fritz, Jason M.; Portier, Christopher J.; Rusyn, Ivan et al. (2016): Key Characteristics of Carcinogens as a Basis for Organizing Data on Mechanisms of Carcinogenesis. In *Environmental health perspectives* 124 (6), pp. 713–721. DOI: 10.1289/ehp.1509912.

Stephens, Martin L.; Betts, Kellyn; Beck, Nancy B.; Cogliano, Vincent; Dickersin, Kay; Fitzpatrick, Suzanne et al. (2016): The Emergence of Systematic Review in Toxicology. In *Toxicological Sciences* 152 (1), pp. 10–16. DOI: 10.1093/toxsci/kfw059.

Stroup, D. F.; Berlin, J. A.; Morton, S. C.; Olkin, I.; Williamson, G. D.; Rennie, D. et al. (2000): Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. In *JAMA* 283 (15), pp. 2008–2012.

Vandenberg, Laura N.; Ågerstrand, Marlene; Beronius, Anna; Beausoleil, Claire; Bergman, Åke; Bero, Lisa A. et al. (2016): A proposed framework for the systematic review and integrated assessment (SYRINA) of endocrine disrupting chemicals. In *Environmental health : a global access science source* 15 (1), p. 74. DOI: 10.1186/s12940-016-0156-6.

Villeneuve, Daniel L.; Crump, Doug; Garcia-Reyero, Natàlia; Hecker, Markus; Hutchinson, Thomas H.; LaLone, Carlie A. et al. (2014a): Adverse outcome pathway (AOP) development I: strategies and principles. In *Toxicological sciences : an official journal of the Society of Toxicology* 142 (2), pp. 312–320. DOI: 10.1093/toxsci/kfu199.

Villeneuve, Daniel L.; Crump, Doug; Garcia-Reyero, Natàlia; Hecker, Markus; Hutchinson, Thomas H.; LaLone, Carlie A. et al. (2014b): Adverse outcome pathway development II: best practices. In *Toxicological sciences : an official journal of the Society of Toxicology* 142 (2), pp. 321–330. DOI: 10.1093/toxsci/kfu200.

Wang, Zhicheng; Taylor, Kyla; Allman-Farinelli, Margaret; Armstrong, Bruce; Askie, Lisa; Ghersi, Davina et al. (2019): A systematic review: Tools for assessing methodological quality of human observational studies.

Whaley, Paul; Halsall, Crispin; Agerstrand, Marlene; Aiassa, Elisa; Benford, Diane; Bilotta, Gary et al. (2016a): Implementing systematic review techniques in chemical risk assessment: Challenges, opportunities and recommendations. In *Environment international* 92-93, pp. 556–564. DOI: 10.1016/j.envint.2015.11.002.

Whaley, Paul; Isalski, Marek (2019): CREST_Triage. Available online at https://crest-tools.site/, checked on 2/20/2020.

Whaley, Paul; Letcher, Robert J.; Covaci, Adrian; Alcock, Ruth (2016b): Raising the standard of systematic reviews published in Environment International. In *Environment international*. DOI: 10.1016/j.envint.2016.08.007.

Woodruff, Tracey J.; Sutton, Patrice (2014): The Navigation Guide systematic review methodology: a rigorous and transparent method for translating environmental health science into better health outcomes. In *Environmental health perspectives* 122 (10), pp. 1007–1014. DOI: 10.1289/ehp.1307175.

791