

**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](mailto: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](mailto: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](mailto:claire.beausoleil@anses.fr)

**Anna Beronius**, Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden | [anna.beronius@ki.se](mailto:anna.beronius@ki.se)

**Gary Bilotta**, School of Environment and Technology, University of Brighton, Brighton, UK | [g.s.bilotta@gmail.com](mailto:g.s.bilotta@gmail.com)

**Alan Boobis** National Heart & Lung Institute, Imperial College London, London, UK | [a.boobis@imperial.ac.uk](mailto: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](mailto:rob.devries@radboudumc.nl)

**Annika Hanberg**, Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden | [annika.hanberg@ki.se](mailto: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](mailto:sebastian.hoffmann@seh-cs.com)

**Neil Hunt**, Yordas Group, Lancaster Environment Centre, Lancaster University, Lancaster, LA1 4YQ, UK | [n.hunt@yordasgroup.com](mailto:n.hunt@yordasgroup.com)

**Carol Kwiatkowski**, The Endocrine Disruption Exchange. P.O. Box 54, Eckert, CO, 81418, USA | [carolkw@tedx.org](mailto:carolkw@tedx.org)

**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](mailto:Juleen.Lam@csueb.edu)

**Steven Lipworth**, Royal Society of Chemistry, Burlington House, Piccadilly, London, W1J 0BA, UK | [s.lipworth@srl-consulting.com](mailto:s.lipworth@srl-consulting.com)

**Olwenn Martin**, Institute for the Environment, Health and Societies, Brunel University London, Uxbridge, UK | [olwenn.martin@brunel.ac.uk](mailto:olwenn.martin@brunel.ac.uk)

**Nicola Randall**, Harper Adams University, Newport, Shropshire, UK | [nrandall@harper-adams.ac.uk](mailto:nrandall@harper-adams.ac.uk)

**Lorenz Rhomberg PhD ATS**, Gradient, 20 University Road, Cambridge, MA 02138, USA | [lrhomberg@gradientcorp.com](mailto: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](mailto: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](mailto:schuneh@mcmaster.ca)

**Daniele Wikoff**, ToxStrategies, 31 College Place, Suite B118B, Asheville, NC 28801, USA | [dwikoff@toxstrategies.com](mailto:dwikoff@toxstrategies.com)

**Taylor Wolffe**, Lancaster Environment Centre, Lancaster University, Lancaster, LA1 4YQ, UK | [t.wolffe@lancaster.ac.uk](mailto:t.wolffe@lancaster.ac.uk)

**Crispin Halsall**, Lancaster Environment Centre, Lancaster University, Lancaster, LA1 4YQ, UK | [c.halsall@lancaster.ac.uk](mailto:c.halsall@lancaster.ac.uk)

50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## **Declaration of Interests**

Due to the objective of the project being to establish, across a broad selection 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. With regard to the development of COSTER, the authors declare they have no actual or potential competing financial interests, and certify that their freedom to design, conduct, interpret, and publish the research was not compromised by any controlling sponsor. For transparency in interests, each author has completed an International Committee of Medical Journal Editors Form for Disclosure of Potential Conflicts of Interest. These forms are included as supplemental information.

## 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 on a core set of requirements for sound practice in planning and conducting a SR in the environmental health sciences.

**Methods:** A draft set of requirements 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 and several rounds of email feedback, 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 requirements 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 sound-practice requirements 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, providing requirements relating to 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 introduction. Some aspects of SR, such as assessment of external validity at the level of individual study, are not yet sufficiently developed for consensus on sound practice to be achieved.

## 1 Introduction

In the fields of toxicology, 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 (Whaley et al. 2016a) to characterise health hazards and risks posed by exposure to environmental challenges. 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).

In service of this interest, there is a growing number of documents which provide varying types of guidance for conducting SRs in EH research. These include: 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), 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).

While these documents offer valuable guidance on conduct of SRs, they differ in their levels of comprehensiveness and detail, domains of applicability, and the extent to which the various practices they describe are either mandatory or optional. For example, the 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 OHAT approaches both employ a Cochrane-derived risk of bias approach to appraising study quality (Higgins et al. 2011), SYRINA lays out a wider range of options which an SR team can choose between, and the application of SR methods in TSCA does not follow the NTP-implemented Cochrane guidance to eschew scoring of study quality in assessing risk of bias. None of these documents therefore provide a collectively consistent standard for good practice in the planning and conduct of an EH SR.

The situation in EH research sits in contrast to the biomedical sciences, where standards for conducting and reporting SRs have been proliferating rapidly over the last three decades. The EQUATOR Network’s online *Library for Health Research Reporting* currently lists over 400 standards for reporting health research (<https://www.equator-network.org/library/>). Although

122 many of the listed standards are concerned with reporting of primary research, there are also  
123 numerous standards for reporting of SRs, such as the PRISMA checklist for systematic reviews  
124 of interventions (Moher et al. 2009) and the MOOSE reporting guidelines for SRs of  
125 observational studies in medicine (Stroup et al. 2000). Standards which focus explicitly on the  
126 conduct rather than reporting of SRs include the US Institute of Medicine (IOM) *Finding What  
127 Works in Health Care (WWHC): Standards for Systematic Reviews* (Eden et al. 2011) and the  
128 Cochrane Editorial Unit's *Methodological Expectations for Conduct of Intervention Reviews*  
129 (MECIR) standard (Chandler et al. 2013) recently updated to version 1.07 in November 2018  
130 (Higgins et al. 2018).

131 Standards are distinguished from guidelines and handbooks in that, in the form of a list of  
132 requirements, they “provide a set of agreed principles or criteria for a product, service or  
133 practice, such that users of those products can make reliable assumptions about their  
134 performance, safety, compatibility and/or other features as specified in the standard” (British  
135 Standards Institution 2016b). Standards vary in detail and prescriptiveness according to the  
136 function they perform, from “specifications” which set out detailed, absolute requirements, to  
137 flexible “codes of practice” which recommend “sound and good practice as currently undertaken  
138 by competent and conscientious practitioners” (British Standards Institution 2016a). Standards  
139 are always voluntary (except when laws and regulations refer to them and make them  
140 compulsory) but allow a benchmark to be set, against which the quality of a product can be  
141 evaluated. The development and promulgation of standards which provide clear, expert  
142 guidance on good practice are considered to be an important contributor to ensuring the quality  
143 of SRs (Eden et al. 2011).

144 While the universal nature of the fundamentals of SR methods should result in broad overlap  
145 in sound SR practices between biomedical and EH research, the potential for cultural and  
146 research-specific differences between the domains mean that direct applicability of biomedical  
147 SR standards to EH research cannot be assumed (Haddaway et al. 2018a). These differences  
148 include the types of evidence being summarised (with a focus in EH on observational human,  
149 experimental animal and *in vitro* study designs intended to elucidate disease aetiology and  
150 identify health risks, as opposed to a prevalence of methods for identifying effective treatments  
151 for disease using a body of evidence in which randomised controlled trials in humans tends to  
152 be more readily available), the types of decision potentially being supported by SRs (e.g.  
153 defining the conditions for acceptable use of chemical substances rather than informing  
154 healthcare intervention decisions), and specific methodological challenges in evidence synthesis  
155 (e.g. the need in EH research to integrate evidence from human, animal, *in vitro*, and *in silico*

156 studies). These differences mean that standards developed in biomedicine need to be  
157 methodically assessed, and potentially adapted and added to, by EH research practitioners.

158 A broad cross-section of EH expert stakeholders was therefore convened with the objective  
159 of adapting biomedical SR standards to the EH domain. The aim was to establish participating  
160 stakeholder consensus on a core set of requirements which describe the sound and good  
161 practices in the conduct of SRs in EH, which, if adhered to, will improve the quality of SR  
162 projects.

## 163 **2 Methods**

164 A workshop was held 2 December 2016, attended by 33 expert participants selected to cover  
165 academic, policy, regulatory, non-government and industry sectors (see Supplemental  
166 Information 01). Participants all had at least some experience in evidence synthesis, with an  
167 overall balance of expertise in SR methods, weight-of-evidence methods, chemical risk  
168 assessment, toxicology, environmental health research and chemicals policy being sought  
169 across the group. “Consensus” was defined following International Organization for  
170 Standardization (ISO) terminology as “general agreement, characterized by the absence of  
171 sustained opposition to substantial issues by any important part of the concerned interests and  
172 by a process that involves seeking to take into account the views of all parties concerned and to  
173 reconcile any conflicting arguments” (ISO/IEC 2004).

174 Development of the standard was seeded by two discussion draft documents (see  
175 Supplemental Information 02 and 03) developed by PW, presenting a draft standard derived  
176 from version 2.3 of the Cochrane *MECIR* standards (Chandler et al. 2013) and the US Institute of  
177 Medicine *WWHC: Standards for Systematic Reviews* (Eden et al. 2011). These standards were  
178 taken to already represent a high degree of consensus and expectation of effectiveness of sound-  
179 practice requirements relating to general SR methods in biomedicine, providing a solid basis for  
180 a standard for SRs in EH research. The discussion draft also outlined for participants the  
181 potential role of standards in quality management of research, explained how standards are  
182 developed, and described how the workshop and subsequent follow-up activities would be  
183 structured to facilitate consensus on a standard for conduct of SRs in EH research.

184 The draft standard was discussed requirement-by-requirement at the workshop by two  
185 break-out groups working in parallel, chaired by PW and JL. Input was solicited on the  
186 following: (a) which of the proposed criteria should be included in a code of practice for SRs in  
187 toxicology and chemical risk assessment; (b) if and how the included criteria should be

188 reformulated; (c) whether there were any additional criteria which should be included, and if  
189 so, how they should be formulated; and (d) questions for clarification and follow-up.

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

199 The consensus process was closed by PW on 24 January 2018; participating authors  
200 confirmed agreement with the consensus by signing off as co-authors of this manuscript.  
201 Participants in the process who contributed to the workshop and related discussions but were  
202 not able to sign off on the manuscript have been listed in the Acknowledgements.

### 203 **3 Results**

204 COSTER is a standard for design and conduct of SRs in environmental health research. It  
205 consists of 70 requirements divided across 8 domains. The domains cover: planning the SR;  
206 searching for evidence; selecting evidence for review; extracting data; critically appraising each  
207 individual included study; synthesising the evidence; interpreting the evidence and  
208 summarising what it means for the review question; and drawing conclusions (see Figure 1).

209 The number of requirements in each domain is a function of the number of decisions which  
210 need to be made at each stage of planning and conducting a SR, combined with whether or not  
211 consensus among the authors on sound practice relating to each decision-point was attainable.  
212 The requirements are listed in Table 1. Guidance on how to use COSTER is provided in the  
213 Discussion section below.

214 There was no consensus view in the group on the extent to which any individual requirement  
215 should be described as compulsory, desirable, or merely discretionary: while all requirements  
216 are considered by the authors to constitute sound practice, the authors did not believe it was  
217 possible to make general claims about the relative necessity of any single requirement across  
218 the wide variety of contexts in which EH SRs are conducted.

219 Readers should note that the term “requirement” is used in its technical sense as a provision  
 220 within a standard, not in the broader sense of something being compulsory.

221

<b>COSTER v1.0.0: Requirements for sound and good practice in the planning and conduct of environmental health systematic reviews</b>
<b>1. Planning the Review and Preparing the Protocol</b>
<b>1.1 Securing capacity, competencies and tools</b>
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.
<b>1.2 Setting the research question to inform the scope of the review (“problem formulation”)</b>
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: <ul style="list-style-type: none"> <li>• Population (objects of investigation, i.e. the entities to which exposures or interventions happen)</li> <li>• Exposure or Intervention (the administered change in conditions of the objects of investigation, to include timing, duration and dose)</li> <li>• Comparator (the group to which the intervention or exposure groups are being compared)</li> <li>• Outcome (the change being measured in the intervention or exposure group)</li> <li>• Study design (specific design features of relevant research)</li> <li>• Target condition (the object of a test method for diagnosis or detection)</li> </ul>
<b>1.3 Defining eligibility criteria</b>
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.
<b>1.4 Planning the review methods at protocol stage</b>
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.
<b>1.5 Publishing the protocol</b>
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.
<b>2. Searching for Evidence</b>
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".
<b>3. Screening Evidence for Inclusion</b>
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.
<b>4. Extracting Relevant Data from Included Study Reports</b>
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).
<b>5. Appraising the Internal Validity of Included Studies</b>
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.
<b>6. Synthesising the Evidence / Deriving Summary Results</b>
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.
<b>7. Interpreting Results</b>

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.
<b>8: Drawing Conclusions</b>
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: COSTER requirements for sound practice in the planning and conduct of environmental health systematic reviews. Table 1 should be read alongside Table 2, which serves as an explanation and elucidation of a number of the requirements of COSTER.*

## 4 Discussion

### 4.1 How to use COSTER

#### 4.1.1 The role of COSTER in ensuring high-quality EH SRs

COSTER is intended to help research teams ensure that a SR has the following three characteristics:

1. is useful, addressing an important research question, and advancing community understanding of an environmental health issue via a methodology of synthesising existing research;
2. is 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. is credible, minimising the risk that its findings are biased either by limitations in the evidence base itself or in the processes used to locate and synthesise that evidence.

The application of COSTER is best considered in the context of a broader quality management process that is facilitated by reference to three types of document:

- a. standards for conduct of research, which describe the core 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 its quality can be evaluated;
- c. critical appraisal tools, which help a reader analyse project documentation to determine the quality of a reported piece of research.

COSTER is intended to contribute to (a) above, offering guidance on best practice in conduct of SRs in the form of a list of requirements. COSTER has been designed neither as a reporting standard nor as a critical appraisal tool; rather, it is conceived as a tool or recipe for helping researchers plan and conduct robust systematic reviews, by making explicit a set of practices which a group of experts have agreed on as being sound and good. While it identifies requirements for sound practice in conduct of a EH SR, and could therefore inform the development of reporting standards and critical appraisal tools, it has not been developed or tested for effectiveness in helping researchers report their SR, nor readers appraise a SR, and should not be used for either purpose without appropriate adaptation.

253 COSTER is intended to be usable by any entity with a stake in the quality of conduct of an EH  
254 SR project. This could include: journals which might require SR submissions to comply with  
255 some or all of the COSTER requirements; research teams wishing to conduct a SR; research  
256 commissioners seeking confidence that a contractor will conduct a sound and good SR project;  
257 quality assurance units in research-associated organisations seeking to implement consistent,  
258 sound and good SR practices; or government agencies and scientists seeking to demonstrate  
259 compliance with an agreed set of practices for conduct of research in a regulatory setting. In  
260 conjunction with appropriate appraisal tools and reporting guidelines, thoughtful use of  
261 COSTER should increase the likelihood that a SR project is successful.

#### 262 *4.1.2 Managing the number of requirements in COSTER*

263 While 70 may seem like a large number of requirements for a research team to fulfil, SRs are  
264 complex, multi-disciplinary projects which typically take 12-36 months to complete (Borah et al.  
265 2017; Haddaway and Westgate 2019). Although numbers are not directly comparable because  
266 of differences in how the requirements of each stage of a review are presented, COSTER is  
267 comparable in size to WWHC, which consists of 82 requirements across 4 domains, and MECIR  
268 1.07, which consists of 75 requirements across 10 domains (Higgins et al. 2018).

269 The COSTER requirements are designed to be addressed in parallel to the development,  
270 conduct, and reporting of a systematic review in an iterative manner which mirrors many of the  
271 considerations that should naturally arise for research teams undertaking each of these steps.  
272 Therefore, the fulfilment of these requirements is anticipated to be already addressed or  
273 incorporated in a well-designed and well-conducted SR and would not constitute an additional  
274 burden in these scenarios. On the other hand, these requirements should help identify  
275 oversights and limitations in design and conduct that could undermine the integrity of a SR  
276 project which, if corrected, should increase the quality of the resulting product and ideally be  
277 worth the additional effort.

#### 278 *4.1.3 How should compliance with COSTER be described?*

279 When research teams report the use of COSTER in planning and conducting a SR, they are  
280 encouraged to avoid broad summary statements such as “the COSTER code of practice was  
281 followed”. Although prevalent in the literature, such self-reported statements are usually only  
282 partly true and may therefore mislead the reader about the exact methods used (Page and  
283 Moher 2017). Instead, authors should report that COSTER was used to inform the planning and  
284 conduct of a SR, and transparently describe whether and how they were able to fulfil each  
285 requirement. Since COSTER is not itself a reporting standard, this process may be facilitated by  
286 use of SR reporting standards such as PRISMA (Moher et al. 2009), ROSES (Haddaway et al.  
287 2018b), or MOOSE (Stroup et al. 2000). We recommend that COSTER-specific extensions of such

288 tools be developed, to address any potential inconsistencies between the reporting standards  
289 and the requirements of COSTER.

#### 290 4.1.4 *To what extent are the COSTER requirements compulsory?*

291 Because the COSTER consensus process captures the collective views of a relatively small  
292 group of expert (though representative) practitioners, and eschews language which would  
293 distinguish between requirements which are compulsory and those which are optional, it is  
294 roughly equivalent to the “code of practice” sub-category of standard. A code of practice is  
295 defined by the British Standards Institution (BSI) as a comprehensive set of requirements which  
296 “reflects current good practice ... as employed by competent and conscientious practitioners”  
297 (British Standards Institution 2016a). As such, COSTER is intended to provide an authoritative  
298 and comprehensive set of instructions on how to conduct SRs in EH research, while  
299 acknowledging that a broader consensus-building and best-practice research process should  
300 lead to further development of the standard in future (see section 4.3 for further discussion).

301 What it means to be compliant with COSTER requires some clarification, as the role which  
302 standards have in defining good practice is not necessarily intuitive. Five related considerations  
303 inform the issue of COSTER compliance.

304 The first consideration is that COSTER does not specify requirements for every step of a SR  
305 process. This is because COSTER is only the set of requirements on which the authors could  
306 agree, and not the full set of requirements for a complete SR process. One important reason for  
307 this discrepancy is that, while some steps of SR are believed to be important, it may not be the  
308 case that debate on appropriate methods for those steps is sufficiently settled for a consensus  
309 view of best practice to emerge. This was the case for assessment of external validity of  
310 individual included studies: while it was proposed as a domain in the pre-workshop discussion  
311 draft of COSTER (see Supplemental Materials), no consensus could be reached on requirements  
312 for this element of SR. Where COSTER specifies no requirements, as a standard it can obviously  
313 be exceeded.

314 The second consideration is that COSTER can be exceeded even where it does specify  
315 requirements. This is because the consensus process captures what participants can agree on as  
316 being “sound and good” practice, not what counts as exceptional practice. In that sense, COSTER  
317 presents a minimum set of requirements which, if a SR were to comply with them, would be  
318 likely to provide a satisfactory result. It does not mean that COSTER cannot or should not be  
319 exceeded – not only in areas not covered by COSTER (such as assessment of external validity),  
320 but also in areas where availability of resource may allow e.g. extremely sensitive searches over  
321 multiple databases to be conducted, large numbers of experts to be engaged in refining tools for

322 assessing risk of bias, etc. The more that SR authors are able to exceed the requirements of  
323 COSTER, the closer the final SR will be to being exceptional rather than simply sound and good.

324 The third consideration is that, while COSTER describes what needs to be done in order for a  
325 EH SR to be COSTER compliant, it does not describe the conditions under which COSTER should  
326 be used in a SR project. Nor does it describe how a requirement should be met. The intent of  
327 standards is to generate a shared understanding of basic good practice and provide a  
328 benchmark against which the quality of a process or product can be assured. Standards are  
329 voluntary (although laws and regulations may make reference to standards, and therefore make  
330 compliance with them compulsory). While COSTER, as a standard, provides such a benchmark  
331 for good practice, the issue of when any given research project, group, agency, journal or other  
332 party should adopt COSTER as their standard of practice is a matter for those parties  
333 themselves, not one which can be made for them by COSTER or its authors.

334 The fourth consideration is that highly-developed, formal standards utilise language such as  
335 “must”, “should” and “may” to communicate differences in importance of a requirement (such as  
336 when a requirement is fundamental to a process vs. when it might differentiate sound practice  
337 from exceptional practice). COSTER eschews this language as the consensus process did not  
338 extend to making such differentiations. As such, all the requirements of COSTER have equal  
339 standing – even though each requirement may not be of equal importance to the final quality of  
340 a EH SR. This is a limitation which can only be addressed by further development over time of  
341 the consensus view of sound and good practice in EH SRs. This also renders important the  
342 careful reporting of the extent to which a given SR has complied with the requirements of  
343 COSTER, as discussed in section 4.3 above.

344 Fifthly, given the above considerations, it can be anticipated that there will be circumstances  
345 in which a user may wish to be only partially compliant with COSTER. These could include  
346 situations where, although a project can still usefully be informed by the COSTER requirements,  
347 there might be e.g. severe time constraints limiting the methodological rigour achievable by a  
348 team of researchers; or when users disagree with the consensus view of COSTER, or find it non-  
349 applicable given specific research goals or project context. In such cases, users of COSTER  
350 should be explicit about which requirements were not followed and the reasons for so doing,  
351 thereby employing COSTER as a benchmark for making transparent the necessary compromises  
352 in a particular EH SR project. While deliberate non-compliance should be approached with  
353 caution, due to COSTER representing a consensus view of sound and good practices in the  
354 conduct of a SR, using COSTER as a framework for making transparent and justifying necessary  
355 deviations in methods would be a constructive use of the standard.



356 It may be that full compliance with COSTER is rare and its use as a planning tool and practice  
357 benchmark becomes the predominant application of COSTER.

358 To summarise:

- 359 • COSTER does not cover every step of the SR process, only the steps on which  
360 consensus about sound and good practice could be agreed.
- 361 • COSTER can be exceeded: compliance is about meeting or exceeding the base  
362 requirements. Decisions about how to comply with a COSTER requirement, or  
363 whether any given processes are equivalent to so doing, are made at the discretion of  
364 the user.
- 365 • A decision to comply with COSTER is voluntary. The full set of requirements of  
366 COSTER are compulsory only insofar as full compliance with COSTER is the objective  
367 of a SR team.
- 368 • There may be good reasons for selective compliance with the COSTER requirements  
369 but this should be approached cautiously and explained to the reader.
- 370 • EH SR projects which are selectively compliant with COSTER, or use alternative  
371 methods which the user believes are equivalent to a requirement, are recommended  
372 to use COSTER as a benchmark for making transparent their methods.

#### 373 **4.2 Comparing COSTER to other SR standards**

374 Because SR practices are relatively universal and independent of topic, there is substantial  
375 overlap between COSTER and other standards, including MECIR and WWHC. However, COSTER  
376 is the first explicit effort by EH research practitioners to validate for their particular cultural and  
377 research context SR standards which are being applied in biomedicine, bringing together  
378 multiple stakeholders with differing views of good practice in EH SR to establish a common  
379 consensus view which can be expressed in a set of requirements. By doing this, COSTER  
380 contributes to resolving the question of which standards in the conduct of biomedical SRs can  
381 be applied to EH research. COSTER also provides a platform on which SR standards for  
382 environmental health research can be further developed, particularly in areas where the  
383 COSTER process has identified methodological guidance as being needed but immediate  
384 consensus on sound and good practice is elusive. In particular, this applies to assessing the  
385 external validity of included studies.

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

Table 2: Key contributions and differences between COSTER and other SR standards, and explanation of significant COSTER requirements	
<b>Requirements</b>	1.1.1 through 1.5.4
<b>Theme</b>	Project planning
<b>Contribution of COSTER</b>	Emphasis on importance of practices in biomedical SRs for environmental health research
<p><b>Explanation:</b> 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 and 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 WWHC 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 the project outcomes.</p>	
<b>Requirements</b>	1.1.3, 1.1.4
<b>Theme</b>	Disclosure and management of interests
<b>Contribution of COSTER</b>	Distinction between potential and apparent conflicts of interest to rationale for team selection in SRs
<p><b>Explanation:</b> COSTER follows 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 (2004) 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 actual conflict of interest” (the reader should note that the framework provides a number of useful illustrative examples). These may be financial and/or non-</p>	

financial. Similar to WWHC, 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 WWHC 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 special knowledge on which review teams can draw, while insulating the review process from risk of bias by prohibiting their involvement in decision-making. This is to allow environmental health 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 have apparent COIs.

The intent of these provisions is not to limit participation by excluding participants with affiliation to broad sectors (academic grant holders, industry, or NGOs), but rather to make such associations transparent while focusing the limitations on decision-making roles by those with direct, topic-specific conflicts. 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).

<b>Requirements</b>	1.2.2, 1.4.6, 7.8
<b>Theme</b>	Interpreting external validity of the evidence, and integrating multiple evidence streams
<b>Contribution of COSTER</b>	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, WWHC 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. Provision 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 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.

<b>Requirements</b>	1.2.3, 1.3.3, 1.3.4, 1.3.5, 1.3.9
<b>Theme</b>	Formulation of research objectives
<b>Contribution of COSTER</b>	Formal clarification of use of PECO-style statements in formulating SR objectives in EH research
<p><b>Explanation:</b> COSTER requires that SR objectives be formulated in an appropriate 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 research SR, with elements such as timing of exposure being recognised as a potentially critical issue in reliably identifying health risks of chemical exposure, and a requirement 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 (Morgan et al. 2018b).</p>	
<b>Requirements</b>	1.3.6, 1.3.9, 3.4
<b>Theme</b>	Including informally-published or previously unpublished literature

<b>Contribution of COSTER</b>	Provides unambiguous rationale for exclusion of study reports due to insufficient information content
<p><b>Explanation:</b> The consensus view of the authors is that grey literature 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. The inclusion of grey literature also acts as one 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. Finally, the authors 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.</p> <p>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 evidence which does not provide sufficient information for risk of bias to be evaluated (requirement 1.3.6). Thirdly, COSTER defines the included study itself, not documents describing the study, as the unit of evidence (provision 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 requirement 1.3.6. 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.</p> <p>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.</p>	
<b>Requirements</b>	1.5.1, 1.5.2, 1.5.3
<b>Theme</b>	Protocol publication
<b>Contribution of COSTER</b>	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 pre-print repositories such as Zenodo (CERN) allow authors to register their methods in advance of conducting a SR. In theory, this 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 primary value of a registered protocol is 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 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).

<b>Requirements</b>	1.4.3, 5
<b>Theme</b>	Internal validity assessment
<b>Contribution of COSTER</b>	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. 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 requirements 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).

COSTER's requirements for risk of bias assessment (Domain 5) are strongly influenced by the following approaches to study appraisal: the Cochrane approach to risk of bias as articulated by Higgins et al. (2011); the adaption of the Cochrane approach to environmental health SRs by the National Toxicology Program Office of Health Assessment and Translation (Rooney et al. 2014) and the Navigation Guide (Woodruff and Sutton 2014); and the development process (though not the final tool, as COSTER predates the relevant publications) for the Risk Of Bias Instrument for Non-randomized Studies of Exposures (ROBINS-E) approach to assessing risk of bias of non-randomised studies of environmental health research has also been proposed (Morgan et al. 2019b; Morgan et al. 2018a). COSTER is not an endorsement of any particular approach to risk of bias assessment, and it should also be noted that COSTER predates the publication of the ROB2 approach to risk of bias assessment (Sterne et al. 2019).

<b>Requirements</b>	1.4.5, 7.1, 7.2, 7.4, 7.5, 7.6, 7.7, 7.8, 7.10
<b>Theme</b>	Assessment of quality of the overall body of evidence
<b>Contribution of COSTER</b>	Emphasis on evaluation of quality of evidence against pre-specified criteria known to be of importance when assessing certainty 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. 2019a), 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 the quality of the evidence is important, because readers of a SR need a trustworthy analysis of how much trust they can put in 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.

<b>Requirements</b>	8.3
<b>Theme</b>	Making policy recommendations
<b>Contribution of COSTER</b>	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 aetiological), they would not normally also investigate evidence for the effectiveness of interventions to mitigate 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 conclusions of a SR should not reach beyond the evidence which was included within it. COSTER therefore requires authors to resist answering questions about how best to mitigate the effects of an exposure or achieve a risk threshold when the evidence relating 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 policy implications be presented as hypothetical frameworks, whereby authors can state that if certain conditions obtain, then a given intervention may be effective for mitigating harm. Assumptions 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.

390

391

### 4.3 *Future development of COSTER*

392

As a code of practice, COSTER represents the first step in a broader research and consensus-building process which it is hoped will yield a robust, international standard for conduct of

393



394 systematic reviews in environmental health research. Formal standards are typically based on  
395 both expectation and empirical evidence that the practices described in the standard contribute  
396 to a product or process being fit for purpose, combined with broad acceptance of the practices  
397 among the community that is expected to adopt the standard. Since SR methods are still  
398 relatively new in environmental health research, it follows that while expectations for what  
399 should work can be captured, and the consensus view of small groups of experienced  
400 practitioners can be secured, evidence for what is effective practice is not yet available. This is  
401 particularly true for areas in which SR methods are not readily portable from the social science  
402 and medical contexts to environmental health, or where environmental health researchers face  
403 challenges not encountered in other fields.

404 Broad community consensus is also an unrealistic goal when only a small, albeit growing,  
405 part of the community is employing SR methods in conducting reviews of evidence. It also needs  
406 to be acknowledged that while COSTER represents the consensus view of the authors, other  
407 expert groups may disagree with some of the requirements of COSTER. Such disagreement is  
408 healthy; in that regard, by making explicit a set of requirements for SR, COSTER serves as a focal  
409 point for advancing consensus across groups.

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

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

420 Step 1 would result in a larger seed-set of potential requirements than was provided by  
421 selecting the MECIR and IOM standards as the basis for the current consensus. However, such a  
422 SR would be a significant undertaking, as it requires interpreting the implied standards in  
423 several large handbooks, a large number of reporting standards and potentially even individual  
424 SR study reports. This is a major challenge for qualitative analysis. Step 2 would provide  
425 evidence of what community practices actually are. Steps 1 and 2 provide data for Step 3, being  
426 a description of the extent to which current practices are aligned with what are considered

427 “best” practices, providing further empirical evidence for a formal standard. Step 4, as a broad  
428 consensus process, would provide a community view of where current practices fall short of  
429 expectation or need, or where specific processes might exceed what the community views as  
430 strictly necessary for conduct of a robust SR.

## 431 **5 Conclusion**

432 COSTER presents the consensus view of a group of expert practitioners as to a set of  
433 requirements for planning and conducting a sound and good systematic review. The lack of  
434 current guidelines for conduct of high quality environmental health SRs, coupled with  
435 exponential growth in publication of SRs (Whaley et al. 2016b), justifies the introduction of  
436 COSTER as authoritative but intermediate guidance which authors and publishers can use to  
437 immediately improve the quality of SRs. If followed, COSTER should significantly increase the  
438 likelihood of success and stakeholder acceptance of an environmental health SR project. As a  
439 first step in establishing a formal, community-wide standard, it is intended that COSTER be  
440 critiqued and improved over time, as part of a wider process which will ultimately yield a  
441 definitive description of the requirements for conduct of SRs in environmental health research.

442 *4100 words body text + 5000 words for Standard and explanatory tables*

## 443 **Acknowledgements**

444 We would like to thank Kate Jones and the Royal Society of Chemistry for hosting the  
445 workshop, and Lancaster University Faculty of Science and Technology and Lancaster  
446 Environment Centre for providing funding to run the workshop.

447 The authors declare no competing financial interests. Further details on potential COIs have  
448 been provided in the DOI forms (see supplemental materials). The views expressed in this paper  
449 are those of the authors and do not necessarily reflect the views or policies of their respective  
450 employers or organisations.

451 We would also like to thank the following for their contribution to the workshop discussions:  
452 Sarah Bull (Royal Society of Chemistry); Richard Brown (World Health Organization); Kurt  
453 Straif (ret.) and Kathryn Guyton (International Agency for Research on Cancer); Julian Higgins  
454 (University of Bristol); Toby Lasserson (Cochrane Editorial Unit); Jennifer McPartland  
455 (Environmental Defense Fund); Sharon Munn (EU Joint Research Centre); Angelika Tritscher  
456 (World Health Organization); Christopher Weiss (US National Institute of Environmental Health  
457 Sciences). TW did not participate in the workshop but contributed to the consensus  
458 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](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/prospéro/>, checked on 7/21/2019.

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

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](http://ccnmtl.columbia.edu/projects/rcr/rcr_conflicts/foundation/index.html), checked on 1/3/2018.

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.

493 EFSA (2010): Application of systematic review methodology to food and feed safety  
494 assessments to support decision making. In *EFSA Journal* 8 (6), p. 1637. DOI:  
495 10.2903/j.efsa.2010.1637.

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

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

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

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

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

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

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

518 Hansen, Martin Rune Hassan; Jørs, Erik; Sandbæk, Anneli; Kolstad, Henrik Albert;  
519 Schullehner, Jörg; Schlünssen, Vivi (2019): Exposure to neuroactive non-organochlorine  
520 insecticides, and diabetes mellitus and related metabolic disturbances: Protocol for a systematic  
521 review and meta-analysis. In *Environment international* 127, pp. 664–670. DOI:  
522 10.1016/j.envint.2019.02.074.

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

526 Higgins, Julian P. T.; Lasserson, Toby; Chandler, Jackie; Tovey, David; Churchill, Rachel  
527 (2018): Methodological Expectations of Cochrane Intervention Reviews (MECIR). Version 1.07.  
528 Cochrane. Available online at <https://community.cochrane.org/mecir-manual>, updated on  
529 November 2018.

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

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

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

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

540 Jefferson, Tom; Jones, Mark A.; Doshi, Peter; Del Mar, Chris B.; Hama, Rokuro; Thompson,  
541 Matthew J. et al. (2014): Neuraminidase inhibitors for preventing and treating influenza in  
542 healthy adults and children. In *The Cochrane database of systematic reviews* (4), CD008965. DOI:  
543 10.1002/14651858.CD008965.pub4.

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

548 Mandrioli, Daniele; Schlünssen, Vivi; Ádám, Balázs; Cohen, Robert A.; Colosio, Claudio; Chen,  
549 Weihong et al. (2018): WHO/ILO work-related burden of disease and injury: Protocol for  
550 systematic reviews of occupational exposure to dusts and/or fibres and of the effect of  
551 occupational exposure to dusts and/or fibres on pneumoconiosis. In *Environment international*  
552 119, pp. 174–185. DOI: 10.1016/j.envint.2018.06.005.

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

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

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

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

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

571 Morgan, Rebecca L.; Thayer, Kristina A.; Bero, Lisa; Bruce, Nigel; Falck-Ytter, Yngve; Gherzi,  
572 Davina et al. (2016): GRADE: Assessing the quality of evidence in environmental and  
573 occupational health. In *Environment international* 92-93, pp. 611–616. DOI:  
574 10.1016/j.envint.2016.01.004.

575 Morgan, Rebecca L.; Thayer, Kristina A.; Santesso, Nancy; Holloway, Alison C.; Blain, Robyn;  
576 Eftim, Sorina E. et al. (2018a): Evaluation of the risk of bias in non-randomized studies of  
577 interventions (ROBINS-I) and the 'target experiment' concept in studies of exposures: Rationale  
578 and preliminary instrument development. In *Environment international* 120, pp. 382–387. DOI:  
579 10.1016/j.envint.2018.08.018.

580 Morgan, Rebecca L.; Thayer, Kristina A.; Santesso, Nancy; Holloway, Alison C.; Blain, Robyn;  
581 Eftim, Sorina E. et al. (2019b): A risk of bias instrument for non-randomized studies of  
582 exposures: A users' guide to its application in the context of GRADE. In *Environment*  
583 *international* 122, pp. 168–184. DOI: 10.1016/j.envint.2018.11.004.

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

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

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

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

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

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

602 Rooney, Andrew A.; Boyles, Abee L.; Wolfe, Mary S.; Bucher, John R.; Thayer, Kristina A.  
603 (2014): Systematic review and evidence integration for literature-based environmental health  
604 science assessments. In *Environmental health perspectives* 122 (7), pp. 711–718. DOI:  
605 10.1289/ehp.1307972.

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

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

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

616 Sterne, Jonathan A. C.; Savović, Jelena; Page, Matthew J.; Elbers, Roy G.; Blencowe, Natalie S.;  
617 Boutron, Isabelle et al. (2019): RoB 2: a revised tool for assessing risk of bias in randomised  
618 trials. In *BMJ (Clinical research ed.)* 366, 14898. DOI: 10.1136/bmj.14898.

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

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



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

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

633 Woodruff, Tracey J.; Sutton, Patrice (2014): The Navigation Guide systematic review  
634 methodology: a rigorous and transparent method for translating environmental health science  
635 into better health outcomes. In *Environmental health perspectives* 122 (10), pp. 1007–1014. DOI:  
636 10.1289/ehp.1307175.

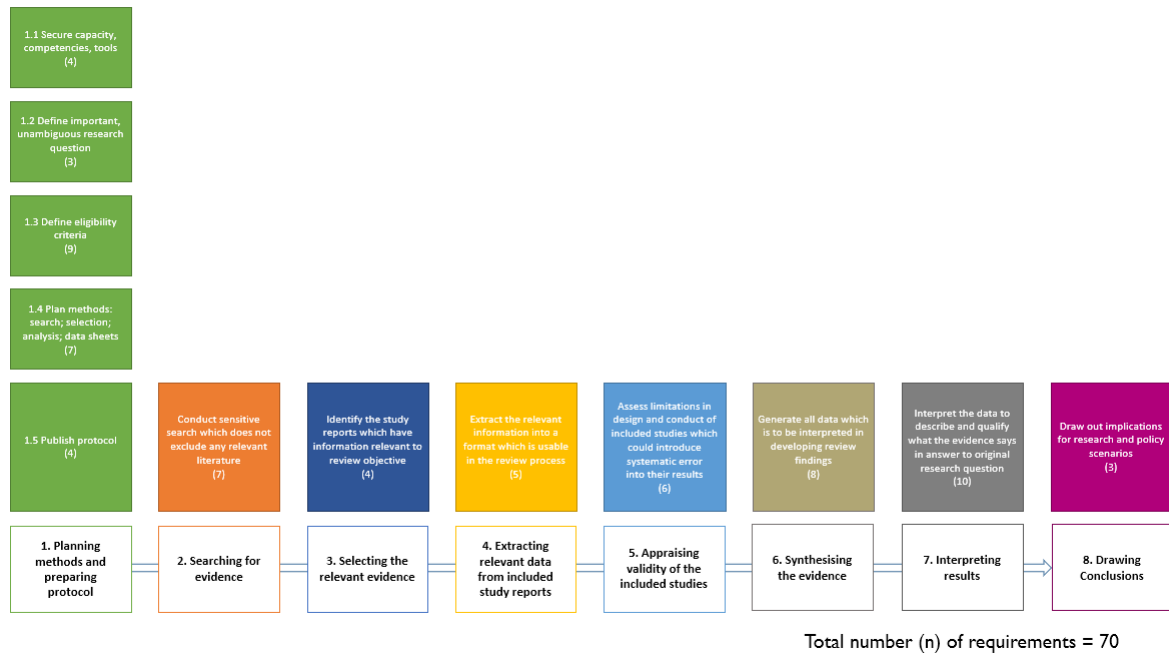
637

638

639

## 6 Figures

640



641

**Figure 1: Domains and conceptual structure of COSTER**