Expectations for Conduct of Systematic Reviews in Chemical Risk Assessment (ECoSys-CRA): an interim code of practice

Discussion draft.

Paul Whaley, Lancaster Environment Centre | p.whaley@lancaster.ac.uk

> *"Readers often find that research reports fail to provide a clear and transparent account of the methods and adequate reporting of the results. If authors do not provide sufficient details of the conduct of their study, readers are left with an incomplete picture of what was done and found. Poorly reported research may result in misinterpretation and inappropriate application in clinical settings. New research projects may also be based on misleading evidence from poorly reported studies. As such, funds devoted to support research may not be used optimally." (Moher et al. 2014b)*
>
> *"Suboptimal systematic reviews and meta-analyses can be harmful given the major prestige and influence these types of studies have acquired." (Ioannidis 2016)*

Table of Contents

Context: What role could formal standards play in improving systematic reviews?

Quality management and standards: a brief introduction

Quality management consists of two interplaying processes of quality assurance and quality control. Quality assurance is the process for preventing mistakes or defects in the production of goods and their delivery to users; quality control specifically refers to the testing of products to uncover potential defects, with reporting mechanisms back to management to allow or deny product release (Manghani 2011).

Standards are 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 can contribute to both quality assurance and quality control: insofar as they describe practices which help ensure a product, service or practice is fit for purpose they offer guidance which can be expected to ensure a minimum level of quality is reached; when describing product performance, they can provide a set of quality benchmarks against which a product can be tested.[1]

The need for better quality management in systematic review

Quality management processes in academic publishing should ensure that only fit-for-purpose systematic reviews get published. In the broad context of quality assurance, "fit for purpose" means a product should be suitable for the intended purpose. In the specific context of systematic review, it means that SRs should be truthful (i.e. minimise risk of bias[2] in their results and conclusions); they should ask an important question; and they should include all the

[1] We can think about illustrating the QM/QA/QC discussion with examples from analytical chemistry, laboratory assays and GLP standards, to bring home how this might apply in the context of conducting and publishing fit-for-purpose SRs and secondary research.

[2] In our paper from our previous workshop (Whaley et al. 2016a) we described risk of bias in SRs as coming from three sources: "bias in the conduct of a review (e.g. because of inappropriate methods for identifying and selecting evidence for inclusion in the review); bias because the material available for the review is not representative of the evidence base as a whole (due to selective publication); and bias arising from flaws in the design, conduct, analysis and reporting of individual studies included in the review that can cause the effect of an intervention or exposure to be systematically under- or over-estimated".

information about methods and results such that a reader can judge the relevance and validity of their results and use their findings.

While quality management is fundamental to ensuring only high-quality systematic reviews (SRs) get published, the processes which researchers follow in planning, conducting and reporting SRs, and journals follow in deciding which SR manuscripts are worth publishing, receive inconsistent attention, and there is evidence that the largely informal approaches to quality assurance employed by journals are resulting in large numbers of low-quality systematic reviews being published. For example, Moher and colleagues estimated the number of stringently defined systematic reviews published in 2004 as approximately 2500, yet the PubMed search for the tag for "systematic review" yielded 8989 items for the same year (Moher et al. 2007). Ioannidis recently described the production of medical systematic reviews and meta-analyses as having reached "epidemic proportions" but estimates that only about 3% of manuscripts are "decent and clinically useful" (Ioannidis 2016).[3]

There is very little research into the quality of systematic reviews being published in the environmental health field. A screening of the literature by three of the authors (PW, GP, CH) shows inconsistent understanding of minimum reporting requirements for systematic reviews published in environmental health journals (Figure 1). The number of purported SRs published without protocols, which did not explicitly consider the generalisability of the evidence base, or which did not formally assess the internal validity of included evidence is cause for concern.[4]

Few environmental health journals appear to formally endorse any kind of publication standard for systematic reviews: only three of the journals which had published environmental health SRs at the time of the literature screening had either endorsed the PRISMA standard or recommended that authors follow it. At time of writing, only two dedicated environmental health journals seem to have formally endorsed PRISMA, and only one environmental health journal has an editor for systematic reviews (Whaley et al. 2016b).

The conclusion is that peer-review and current quality assurance practices by researchers and at journals, as they relate to ensuring proper conduct, reporting and publication of systematic reviews, are failing badly in medicine and could well be failing in environmental health.

[3] Participants with a particular interest in Day 1 proceedings will likely be struck by how many SRs are considered by Ioannidis to be doomed from the outset, by asking questions of too little research value.

[4] Note that this screening did not investigate the validity of methods used, only whether information was provided which might allow validity of methods to be determined.

A standard for systematic reviews in chemical risk research

The existing standards landscape

Standards for reporting and conduct of medical research have proliferated in the last two decades, with the EQUATOR Network's online Library for Health Research Reporting currently listing over 200 reporting guidelines. Although many of these guidelines are concerned with reporting of primary research, there are a number of guidelines for reporting of systematic reviews, such as the PRISMA checklist for systematic reviews of interventions (Moher et al. 2009) and the MOOSE reporting guidelines for systematic reviews of observational studies in medicine (Stroup et al. 2000). Guidelines which focus explicitly on conduct are much fewer but include the What Works in Health Care standards for systematic reviews (Eden et al. 2011) and the Cochrane Editorial Unit's "MECIR" expectations for conduct of systematic reviews of interventions (Chandler et al. 2013).

While in the environmental health and chemical risk assessment disciplines there is some published research into e.g. the key elements for judging the quality of a risk assessment (Fenner-Crisp, Dellarco 2016) which can be interpreted as a guideline for conduct, and there are also surveys of weight-of-evidence guidelines (Rhomberg et al. 2013; Agerstrand, Beronius 2016), there appear to be no formal standards for conduct of systematic reviews developed specifically for chemical risk assessment or its components. Given the potential value of a standard for conduct of systematic reviews in the chemical risk assessment disciplines, we believe one ought to be developed.[5]

The objectives of the ECoSys-CRA Standard

We are interested in developing a standard focusing explicitly on best practices in conducting SRs in CRA, deliberately going beyond encouraging the "constructive unease" generated by reporting standards (Schulz et al. 2014) which only indirectly guides authors in selecting methods for conducting SRs. ECoSys-CRA should describe a minimum set of

[5] Considered as an intervention, there is actually limited evidence of efficacy of guidelines on their own. There is some evidence that journals which endorse the CONSORT statement publish better-reporting clinical trials than those which do not, though the effect is modest (Altman, Simera 2014); however, a recent systematic review suggests limited to no efficacy of endorsement of standards as an intervention to improve the quality of published manuscripts (Page et al. 2016). Given the lack of systematic efforts to enforce standards, it is perhaps not surprising there is limited evidence of efficacy; furthermore, a minimum standard as a reference point against which to judge the quality of conduct of a systematic review is unlikely to be harmful. It should also be noted that improving publishing standards will require a complex intervention of which a standard is only part, though likely fundamental to distinguishing good practices from inadequate ones.

procedures for planning and conduct of a systematic review, such that if followed by a group of authors, would yield a SR which is fit for purpose.

Our other motive for proposing the standard is to bring into the open the disagreements and differing perceptions of what ought to be done in an SR, as are evidenced by the different approaches in the literature, as a first step in developing consensus among SR practitioners in the CRA disciplines of what needs to be done in order for a SR to be fit for purpose.

Methods for developing the ECoSys-CRA standard

High-quality, robust standards are typically based on three things: (a) *expectation* that the practices described in the standard contribute to being fit for purpose; (b) *hard evidence* that the practices contribute to fitness for purpose; and (c) *broad acceptance* of the practices among the practicing community. Standards should also be accompanied by an elucidation document which explains to the user the reasoning behind the inclusion of each clause in the standard.

These three components imply a certain process for developing a standard: (1) a systematic review of existing standards and guidelines to determine the need for a new standard; (2) a systematic review of the prevalence of current research practices; (3) critical appraisal of those practices for completeness, and face and construct validity; (4) a process to determine consensus on best practices and detailing the criteria for the standard.

Moher and colleagues recommend a four-step framework to accommodate these requirements, prior to activities to promote implementation of a standard (Moher et al. 2014a). We describe steps 1-4 as being the foundations on which the "three pillars" A-C of a robust standard are built (Figure 2).

These foundations and pillars of the ideal process put some restrictions on what we can realistically achieve at our workshop. While we arguably have good collective understanding of (a) for many of the steps of CRA, we have little of (b). Developing (b) via (1), (2) and (3) requires capacity not yet available to us. While we may be able to agree among ourselves for (c), the engagement and review processes we need to conduct to secure sufficiently broad consensus that we can describe ourselves as speaking for the SR community as a whole, are also beyond our reach at this immediate juncture.

While it would therefore be a misnomer to describe the outcome of our workshop as a "standard", it is accepted that standardisation is a process which can begin with the articulation of a general set of recommendations ("code of practice"), in lieu of resourcing the more rigorous

research and consensus-building processes which lead to formal standards.[6] We can therefore legitimately make a start with a placeholder standard in lieu of acquiring time and resource to conduct steps 1-4. Our outcome from the workshop will therefore be akin to a "code of practice".

Workshop plan

Objective / targeted outcome

Review each of the criteria in the draft ECoSys-CRA standard (see Excel table) for conduct of systematic reviews in chemical risk assessment and its related fields, aiming to achieve consensus on (a) which of the proposed criteria should be included, (b) how those criteria should be formulated, and (c) whether there are any additional criteria which should be included, and if so how they should be formulated. The outcome is a "code of practice" for conduct of SRs in the CRA disciplines.

Specific considerations for discussion

- For the purposes of discussion, we are **assuming that a systematic review is conducted in eight steps**: planning the SR; searching for evidence; selecting evidence for review; extracting data; appraisal of the validity of the evidence; synthesising the evidence; interpreting the evidence and summarising what it means for the review question; drawing conclusions (see Figure 3).
- Fundamental to standards is lack of ambiguity. Some guidelines require authors to e.g. "protect independence of the review team" but this is not adequate because it is ambiguous as to what level of protection is required, and therefore it cannot be determined when it has been achieved. In fact, protecting the independence of the review team is an objective; the purpose of a standard is to give guidance on how to achieve this. We should focus on articulating, as far as possible, **unambiguous standards** for conducting systematic reviews, bearing in mind that some judgement calls in the SR process resist unambiguous categorisation (e.g. decisions on whether data is too heterogeneous to permit meta-analysis).

[6] It is probably somewhat misleading for many existing guidelines to be called standards: while they might articulate a set of requirements, few are the outcome of an authentic standardisation process.

- Chemical risk assessment consists of **a number of sub-disciplines**, for which a range of SRs with different objectives and relevant types of evidence could be conducted. How much do we want to anticipate this in detailing our ECoSys-CRA standard?

- At this stage, the standard makes **no specific mention of quality control measures** at point of internal review, peer-review or publication which would prevent reviews which are not fit for purpose from being published. These should be discussed (see Figure 4).

- ISO and BSI standards stipulate **levels of requirement** for each clause in a standards document: "shall" (required, not optional); "should" (recommended but optional); "may" (permission to do); "can" (possibility or capability). Can we identify what actors "shall" and "should" do at this stage, or should we stick with a broad set of recommendations and leave deciding which are compulsory until a later stage in the process?

- While we have good detail on appraising limitations in design and conduct of included studies, there is no **guidance on assessing the generalisability of included studies**. Should this be a distinct step in the SR methodology for CRAs, and if so, how should it be conducted? (In medical SRs, it is only considered across the evidence base as a whole, but in CRA it seems a more fundamental component with a higher risk of biased interpretation.)

Note on consensus approaches

Consensus is defined by ISO 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 Guide 2:2004, definition 1.7]

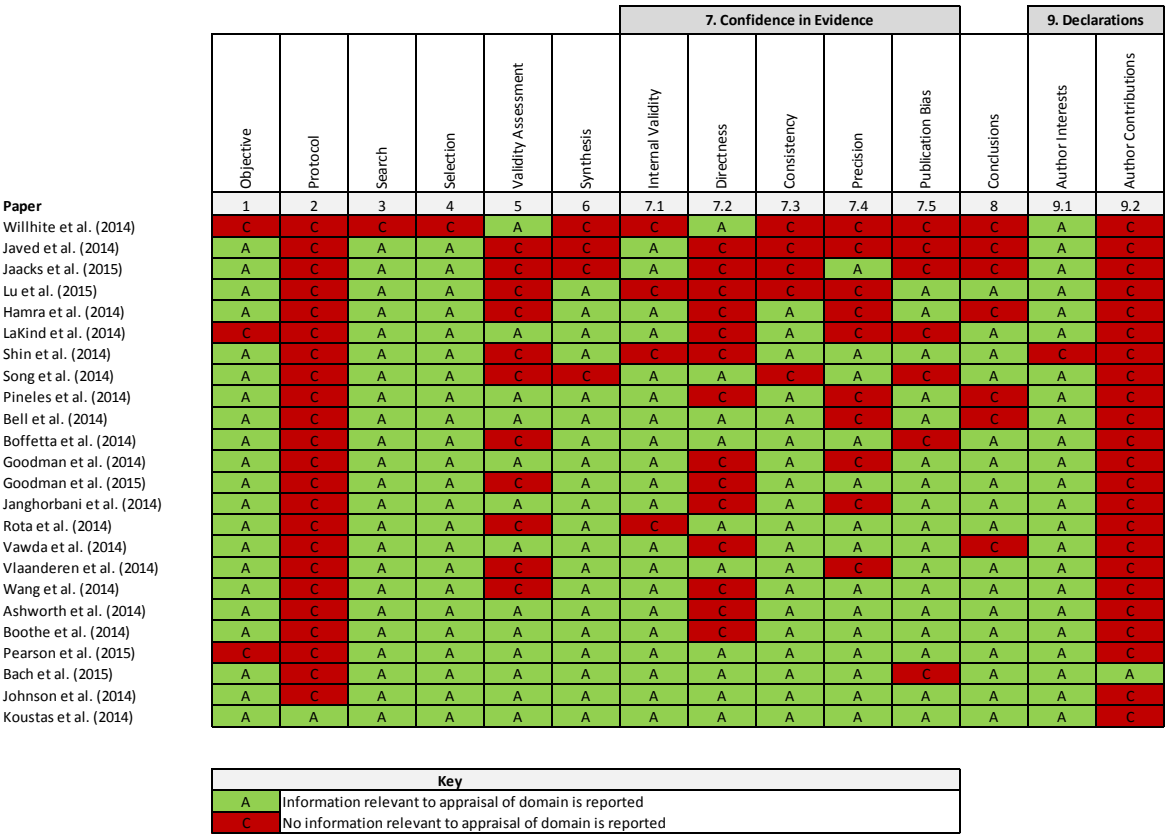
We will be following a consensus approach in determining the ECoSys-CRA criteria. This is the normal method used for defining standards and codes of practice at BSI and ISO level. Participants should note this may feel like a "race to the bottom" process for establishing a minimum set of criteria which ensure a product is fit for purpose, understood as a body of requirements which a skilled group of practitioners would not necessarily see as capturing best possible practice, but can live with as being good enough (i.e. a necessary tolerable minimum).

The purpose of defining the standard in this way is to produce a document of public record of minimum requirements for best practice which can be referred to, critiqued and improved over time. There will be some present who are highly committed to conflicting approaches in conducting a systematic review. In these circumstances, discussion should not become bogged down in the relative merits of these approaches, but try to capture consensus on what it is in

these approaches which can be expected to yield SRs which are fit for purpose, i.e. are truthful, useful, and contain enough information to judge relevance and validity (we are interested in what the standard ought to be, not how the standard might be met). Participants should aim for as high a standard as possible (opposition to a criterion ought always to be relevant and reasoned) but be aware that the desired outcome is a functional standard to which everyone can agree to adhere.

We do not expect to secure consensus on the day. In the event that consensus cannot be reached in a reasonable amount of time, the issues will be recorded and parked for resolution at a future date. We expect to organise teleconferences and email discussions in the coming weeks to identify and resolve contentious issues.

Figures



							7. Confidence in Evidence						9. Declarations	
	Objective	Protocol	Search	Selection	Validity Assessment	Synthesis	Internal Validity	Directness	Consistency	Precision	Publication Bias	Conclusions	Author Interests	Author Contributions
Paper	1	2	3	4	5	6	7.1	7.2	7.3	7.4	7.5	8	9.1	9.2
Willhite et al. (2014)	C	C	C	C	A	C	C	A	C	C	C	C	A	C
Javed et al. (2014)	A	C	A	A	C	C	A	C	C	C	C	C	A	C
Jaacks et al. (2015)	A	C	A	A	C	C	A	C	C	A	C	C	A	C
Lu et al. (2015)	A	C	A	A	C	A	C	C	C	C	A	A	A	C
Hamra et al. (2014)	A	C	A	A	C	A	A	C	A	C	A	C	A	C
LaKind et al. (2014)	C	C	A	A	A	A	A	C	A	C	C	A	A	C
Shin et al. (2014)	A	C	A	A	C	A	C	C	A	A	A	A	C	C
Song et al. (2014)	A	C	A	A	C	C	A	A	C	A	C	A	A	C
Pineles et al. (2014)	A	C	A	A	A	A	A	C	A	C	A	C	A	C
Bell et al. (2014)	A	C	A	A	A	A	A	A	A	C	A	C	A	C
Boffetta et al. (2014)	A	C	A	A	C	A	A	A	A	A	C	A	A	C
Goodman et al. (2014)	A	C	A	A	A	A	A	C	A	C	A	A	A	C
Goodman et al. (2015)	A	C	A	A	C	A	A	C	A	A	A	A	A	C
Janghorbani et al. (2014)	A	C	A	A	A	A	A	C	A	C	A	A	A	C
Rota et al. (2014)	A	C	A	A	C	A	C	A	A	A	A	A	A	C
Vawda et al. (2014)	A	C	A	A	A	A	A	C	A	A	A	C	A	C
Vlaanderen et al. (2014)	A	C	A	A	C	A	A	A	A	C	A	A	A	C
Wang et al. (2014)	A	C	A	A	C	A	A	C	A	A	A	A	A	C
Ashworth et al. (2014)	A	C	A	A	A	A	A	C	A	A	A	A	A	C
Boothe et al. (2014)	A	C	A	A	A	A	A	C	A	A	A	A	A	C
Pearson et al. (2015)	C	C	A	A	A	A	A	A	A	A	A	A	A	C
Bach et al. (2015)	A	C	A	A	A	A	A	A	A	A	C	A	A	A
Johnson et al. (2014)	A	C	A	A	A	A	A	A	A	A	A	A	A	C
Koustas et al. (2014)	A	A	A	A	A	A	A	A	A	A	A	A	A	C

Key	
A	Information relevant to appraisal of domain is reported
C	No information relevant to appraisal of domain is reported

Figure 1: Reporting of methods used in systematic reviews published in the top 20 environmental health journals between January 2014 and June 2015, as ranked by impact factor. See supplemental information for questionnaire and domain clarification, search and selection methods etc.

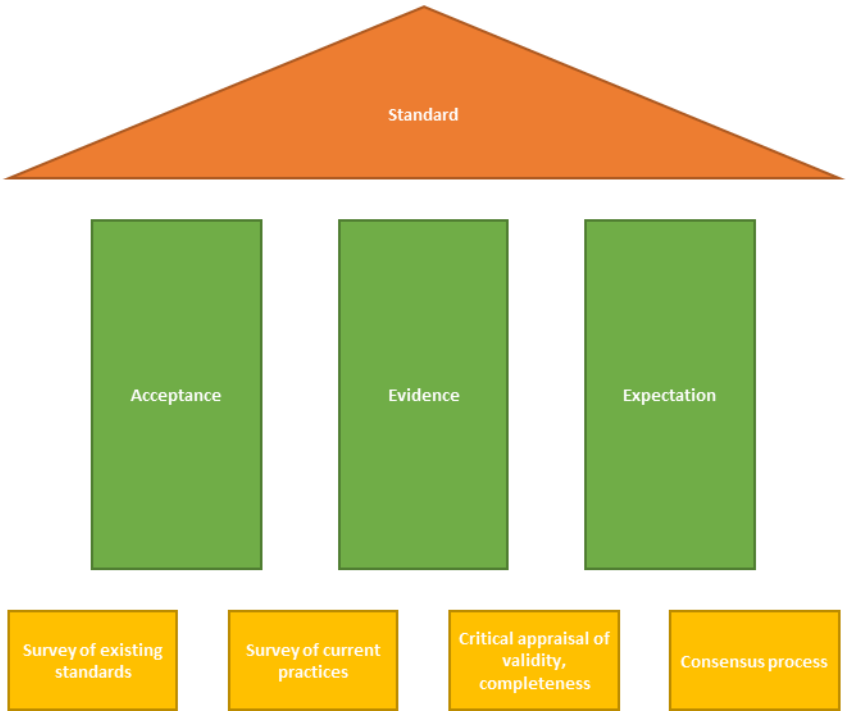


Figure 2: The foundations and "three pillars" of the standardisation process, whereby robust standards are founded on knowledge and critical appraisal of current practices, and consensus on what best practice would be, to generate acceptance for a standardised approach, and evidence and expectation that the standardised approach will be effective in generating fit-for-purpose systematic reviews.

216

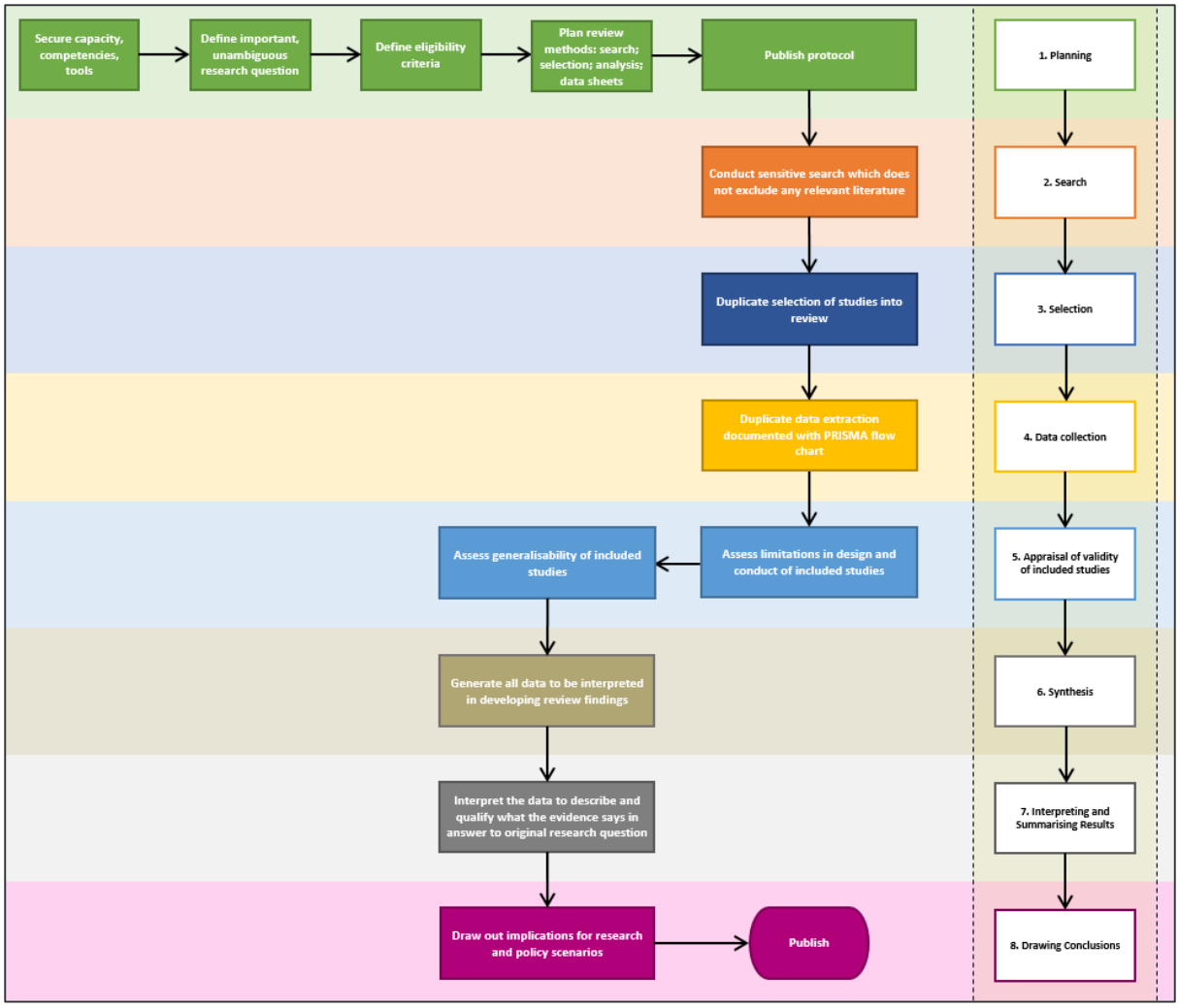


Figure 3: The steps of a conducting a SR, as assumed by the ECoSys-CRA framework

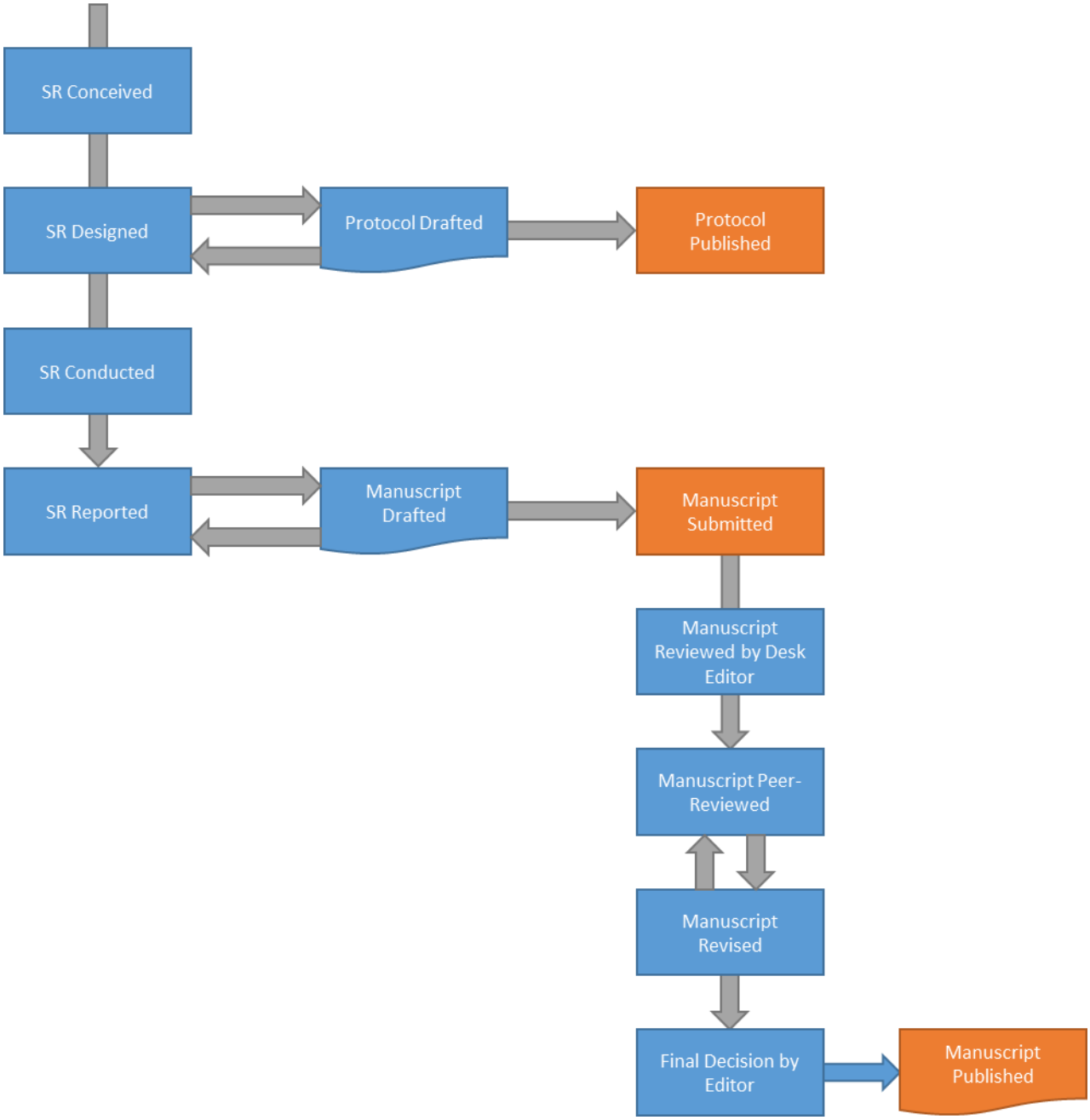


Figure 4: The steps of conducting, reporting and publishing a systematic review, for consideration of opportunities for quality control interventions.

Publication bibliography

Agerstrand, M.; Beronius, A. (2016): Weight of evidence evaluation and systematic review in EU chemical risk assessment: Foundation is laid but guidance is needed. *Environment international* 92-93, pp. 590–596. DOI: 10.1016/j.envint.2015.10.008.

Altman, D.G.; Simera, I. (2014): Using Reporting Guidelines Effectively to Ensure Good Reporting of Health Research. 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. 32–40.

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/, accessed 26/10/16.

Chandler, J.; Churchill, R.; Higgins, J.; Lasserson, T.; Tovey, D. (2013): Methodological standards for the conduct of new Cochrane Intervention Reviews. Cochrane Editorial Unit.

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.

Fenner-Crisp, P.A.; Dellarco, V.L. (2016): Key Elements for Judging the Quality of a Risk Assessment. *Environmental health perspectives* 124 (8), pp. 1127–1135. DOI: 10.1289/ehp.1510483.

Ioannidis, J.P.A. (2016): The Mass Production of Redundant, Misleading, and Conflicted Systematic Reviews and Meta-analyses. *The Milbank quarterly* 94 (3), pp. 485–514. DOI: 10.1111/1468-0009.12210.

Manghani, K. (2011): Quality assurance: Importance of systems and standard operating procedures. *Perspectives in clinical research* 2 (1), pp. 34–37. DOI: 10.4103/2229-3485.76288.

Moher, D.; Altman, D.G.; Schulz, K.F.; Simera, I. (2014a): 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, D.; Liberati, A.; Tetzlaff, J.; Altman, D.G. (2009): Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Annals of internal medicine* 151 (4), p. 264.

Moher, D.; Schulz, K.F.; Altman, D.G.; Hoey, J.; Grimshaw, J.; Miller, D. et al. (2014b): Characteristics of Available Reporting Guidelines. 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. 22–31.

Moher, D.; Tetzlaff, J.; Tricco, A.C.; Sampson, M.; Altman, D.G. (2007): Epidemiology and reporting characteristics of systematic reviews. *PLoS medicine* 4 (3), pp. e78. DOI: 10.1371/journal.pmed.0040078.

Page, M.J.; Shamseer, L.; Altman, D.G.; Tetzlaff, J.; Sampson, M.; Tricco, A.C. et al. (2016): Epidemiology and Reporting Characteristics of Systematic Reviews of Biomedical Research: A Cross-Sectional Study. *PLoS medicine* 13 (5), pp. e1002028. DOI: 10.1371/journal.pmed.1002028.

Rhomberg, L.R.; Goodman, J.E.; Bailey, L.A.; Prueitt, R.L.; Beck, N.B.; Bevan, C. et al. (2013): A survey of frameworks for best practices in weight-of-evidence analyses. *Critical reviews in toxicology* 43 (9), pp. 753–784. DOI: 10.3109/10408444.2013.832727.

Schulz, K.F.; Moher, D.; Altman, D.G. (2014): Ambiguities and Confusions Between Reporting and Conduct. 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. 41–47.

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. *JAMA* 283 (15), pp. 2008–2012.

Whaley, P.; Halsall, C.; Agerstrand, M.; Aiassa, E.; Benford, D.; Bilotta, G. et al. (2016a): Implementing systematic review techniques in chemical risk assessment: Challenges, opportunities and recommendations. *Environment international* 92-93, pp. 556–564. DOI: 10.1016/j.envint.2015.11.002.

Whaley, P.; Letcher, R.J.; Covaci, A.; Alcock, R. (2016b): Raising the standard of systematic reviews published in Environment International. *Environment international. DOI:* 10.1016/j.envint.2016.08.007.