Title
Reviews in environmental health: How systematic are they?

Authors
Sutton, Patrice
Chartres, Nicholas
Rayasam, Swati DG
et al.

Publication Date
2021-07-01

DOI
10.1016/j.envint.2021.106473

Peer reviewed

HHS Public Access
Author manuscript
Environ Int. Author manuscript; available in PMC 2022 July 01.

Published in final edited form as:
Environ Int. 2021 July ; 152: 106473. doi:10.1016/j.envint.2021.106473.

Reviews in Environmental Health: How Systematic Are They?

Patrice Sutton[a,*], **Nicholas Chartres**[a,*], **Swati D.G. Rayasam**[a], **Natalyn Daniels**[a], **Juleen Lam**[b], **Eman Maghrbi**[b], **Tracey J. Woodruff**[a,**]

[a]UCSF Program on Reproductive Health and the Environment, Department of Obstetrics, Gynecology & Reproductive Sciences, Mailstop 0132, 550 16th Street, 7th Floor, San Francisco, CA, United States 94143

[b]Department of Health Sciences, California State University East Bay, SF 533, 25800 Carlos Bee Blvd, Hayward, CA, United States 94542

Abstract

Background: Synthesizing environmental health science is crucial to taking action to protect public health. Procedures for evidence evaluation and integration are transitioning from "expert-based narrative" to "systematic" review methods. However, little is known about the methodology being utilized for either type of review.

Objectives: To appraise the methodological strengths and weaknesses of a sample of "expert-based narrative" and "systematic reviews" in environmental health.

Methods: We conducted a comprehensive search of multiple databases and identified relevant reviews using pre-specified eligibility criteria. We applied a modified version of the Literature Review Appraisal Toolkit (LRAT) to three environmental health topics that assessed the utility, validity and transparency of reviews.

Results: We identified 29 reviews published between 2003–2019, of which 13 (45%) were self-identified as systematic reviews. Across every LRAT domain, systematic reviews received a higher

[*]First Authors patrice.sutton@ucsf.edu, nicholas.chartres@ucsf.edu. [**] Corresponding Author tracey.woodruff@ucsf.edu.
Conceptualization; TW, PS, JL
Data curation; N/A
Formal analysis; PS, JL, TW, SR, NC
Funding acquisition; TW, PS
Investigation; PS, SR, ND, JL, EM
Methodology; PS, JL, TW, NC
Project administration; PS, JL, TW
Resources; TW
Software; N/A
Supervision; PS, JL, TW
Validation; N/A
Visualization; SR, PS, TW, NC
Roles/Writing - original draft; PS, NC, SR, ND, JL, EM, TW
Writing - review & editing. PS, NC, SR, ND, JL, EM, TW

Competing Financial Interests statement: The authors declare they have no actual or potential competing financial interests.

Data Statement: All data used in the development of this study are made available in the tables presented in the manuscript

percentage of "satisfactory" ratings compared to non-systematic reviews. In eight of these domains, there was a statistically significant difference observed between the two types of reviews and "satisfactory" ratings. Non-systematic reviews performed poorly with the majority receiving an "unsatisfactory" or "unclear" rating in 11 of the 12 domains. Systematic reviews performed poorly in five of the 12 domains; 10 (77%) did not state the reviews objectives or develop a protocol; eight (62%) did not state the roles and contribution of the authors, or evaluate the internal validity of the included evidence consistently using a valid method; and only seven (54%) stated a pre-defined definition of the evidence bar on which their conclusions were based, or had an author disclosure of interest statement.

Discussion: Systematic reviews produced more useful, valid, and transparent conclusions compared to non-systematic reviews, but poorly conducted systematic reviews were prevalent. Ongoing development and implementation of empirically based systematic review methods are required in environmental health to ensure transparent and timely decision making to protect the public's health.

Keywords

Systematic Review; Methods; Bias; Environmental Health; Hazard Identification; Risk Assessment

1. INTRODUCTION

Scientific research linking the environment to beneficial and adverse health outcomes is rapidly unfolding. Evidence-based policy actions have produced major gains in health and reaped associated cost savings, as exemplified by tobacco control (Glantz and Gonzalez 2012, Lightwood, Dinno, and Glantz 2008), lead poisoning prevention (Tsai and Hatfield 2011), and programs to reduce air pollution implemented pursuant to the 1990 Clean Air Act (U.S. Environmental Protection Agency 2011). Conversely, failing to take timely action on scientific discoveries squanders opportunities to prevent harm, as demonstrated in the European Environment Agency's compilation of 34 case studies in *Late Lessons of Early Warnings* (European Environment Agency 2001, 2013).

Robust methods to synthesize what is known about the environmental drivers of health are crucial to making science actionable. Structured "systematic review" methods have been developed and empirically validated in the clinical setting over the past 30 years to support evidence-based decision-making (Fox 2017). By definition, systematic reviews " … identify, appraise and synthesize all the empirical evidence that meets pre-specified eligibility criteria to answer a specific research question … [using] explicit, systematic methods that are selected with a view aimed at minimizing bias, to produce more reliable findings to inform decision making" (Cochrane Library 2019). In contrast, environmental health has historically relied on "expert-based narrative review" methods, which do not follow pre-specified, consistently applied, and transparent rules (Woodruff and Sutton 2014, Whaley, Halsall, et al. 2016). Over the past decade, interdisciplinary collaborative efforts to apply robust clinical methods to the environmental health evidence base and decision context have led to the development and application of systematic review methods for environmental health (Woodruff, Sutton, and Group 2011, National Toxicology Program 2015, Johnson et al. 2014, Koustas et al. 2014, Lam et al. 2014, Vesterinen et al. 2015, Johnson et al. 2016, Lam et al. 2016, Lam et al. 2017, Morgan et al. 2016, National Academies of Sciences Engineering and Medicine 2018, Whaley, Halsall, et al. 2016, Lam et al. 2019, Paulo et al. 2019, Vandenberg et al. 2016).

As reviews summarizing the scientific body of evidence are critical for evidence-based decisions, it is important to understand the strengths and weaknesses of current methodological approaches. The uptake of systematic reviews in environmental health is advancing, but the standards by which the methods are applied have been somewhat variable (Whaley, Halsall, et al. 2016). Thus, new questions are arising such as “how systematic are reviews in environmental health?” and “do systematic reviews in environmental health result in more transparent and reliable reviews than traditional expert-based, narrative reviews?” Such unanswered questions are critical to demonstrate the benefits and challenges of using systematic review methods in environmental health (Whaley, Halsall, et al. 2016).

Therefore, our study aimed to assess the methods of reviews currently being implemented in environmental health; and establish if systematic review methods result in more transparent and methodologically sound conclusions than non-systematic review methods. To do this, we selected three environmental health exposures and associated health outcomes, and identified reviews (systematic or non-systematic) on each topic. We evaluated included reviews by applying a modified version of the Literature Review Appraisal Toolkit (LRAT) developed at the University of Lancaster to assess the utility, validity, and transparency of each by rating them across the tools different domains (Literature Review Appraisal Toolkit 2014). The LRAT was “derived from a number toolkits and appraisals of the methodological quality of literature reviews conducted in the medical sciences”, (Literature Review Appraisal Toolkit 2014) including the Cochrane Handbook for Systematic Reviews of Interventions (Higgins, Green, and Cochrane 2008), AMSTAR (A Measurement Tool to Assess the Methodological Quality of Systematic Reviews) (Shea et al. 2007) and PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) (Moher et al. 2009). Based on our results, we make recommendations for authors, peer-reviewers and journals to strengthen the methodology of environmental health reviews.

2. METHODS

2.1 Selection of review topics

We identified three topics (on a specific exposure and health outcome) in environmental health that had been assessed utilizing the Navigation Guide systematic review method. The Navigation Guide systematic review method was developed in 2009 (Woodruff, Sutton, and Group 2011) by an interdisciplinary group of experts lead by the University of California, San Francisco and including investigators of this current study (PS, TJW). It has since been endorsed and applied by the National Academy of Sciences (National Research Council 2014a, b, National Academies of Sciences Engineering and Medicine 2017), the World Health Organization (Paulo et al. 2019, Rugulies et al. 2019, Tenkate et al. 2019, Teixeira et al. 2019, Hulshof et al. 2019, Li et al. 2018, Godderis et al. 2018, Descatha et al. 2018, Mandrioli et al. 2018), and demonstrated in six proof-of-concept case studies (Johnson et al. 2014, Koustas et al. 2014, Lam et al. 2014, Vesterinen et al. 2015, Johnson et al. 2016, Lam et al. 2016, Lam et al. 2017, Lam et al. 2019). Thus, we wanted to compare this validated systematic review approach to other reviews that had assessed similar topics to the Navigation Guide case studies. Three of these proof of concept case studies were chosen as the review topics for this study as they had been peer-reviewed, published and had a direct relevance to environmental health. The three cases studies selected examined the relationship between: 1) air pollution and Autism Spectrum Disorder (ASD) (Lam et al. 2016); 2) polybrominated diphenyl ethers (PBDEs) and Intelligence Quotient (IQ) and/or Attention Deficit Hyperactivity Disorder (ADHD) (Lam et al. 2017), and; 3) formaldehyde and asthma (Lam et al. 2019).

2.2 Literature Review Appraisal Toolkit (LRAT)

The first version of the LRAT was published in 2013, to analyze the European Food Safety Authority (EFSA) evidence reviews for risk assessment against the standards of systematic review (Whaley 2013) and it underwent several developments thereafter. We selected LRAT as *"the point of the toolkit is to help users navigate the credibility of a synthesis of evidence (such as a literature review or an expert opinion), to come to a more informed opinion as to the extent to which they should believe its conclusions",* and therefore allows for the evaluation of both systematic and non-systematic reviews, offering greater flexibility in the breath of reviews that can be evaluated. (Literature Review Appraisal Toolkit 2014) While it was derived from several toolkits and appraisals of the methodological quality of reviews including the Cochrane Handbook for systematic reviews of interventions (Higgins, Green, and Cochrane 2008), AMSTAR (Shea et al. 2007) and PRISMA (Moher et al. 2009), it is suitable for evaluating any literature or evidence review, systematic, narrative, or otherwise, in the peer-reviewed literature, "*so long as it is hypothesis-driven*"(Literature Review Appraisal Toolkit 2014).

Thus, we determined that this was the most appropriate framework to use in our analysis as it allowed the evaluation of domains that were important to conducting reviews (systematic and non-systematic) in environmental health, to the highest standards.

2.3 Data sources

The database search that had been conducted for each Navigation Guide systematic review case study was used to identify eligible reviews for inclusion in this current study. The details of each of these systematic searches have previously been described in the individual case studies (Lam et al. 2016, Lam et al. 2017, Lam et al. 2019) and are available in Supplemental Materials Appendix I.

2.4 Eligibility criteria

We included reviews that: 1) had a study question that was identical, similar, or related to one of the three case study questions; 2) did not include any original data, other than meta-analyses of the included primary studies and; 3) had a publication date during or after 2011 for formaldehyde and asthma reviews only (this inclusion criteria was added for this case study due to a large number of potentially eligible reviews identified). We included reviews in any language. Reviews were included whether they were systematic or not.

2.5 Definition of systematic and non-systematic review

For included reviews, we defined a systematic review as one that was self-identified by the review's authors as "systematic". We use the term "non-systematic" to define any reviews that were not classified as "systematic" by the review's authors.

2.6 Selection of reviews

Three investigators (ND, PS, JL, working in pairs) independently screened the reviews against the inclusion criteria. Agreement was reached by consensus. If agreement could not be reached, a third investigator adjudicated the outcome.

2.7 Data collection and analysis

We used a modified version of the Literature Review Appraisal Toolkit (LRAT) to evaluate the utility, validity and transparency of the included literature reviews (Literature Review Appraisal Toolkit 2014). Each investigator received two to three hours of training, including pilot-testing the published LRAT instructions on one of the reviews. Following the pilot-testing, we made revisions to improve the clarity of the LRAT instructions and ensure consistency in the extraction and subsequent ratings (Table 1). Furthermore, we modified domains three, eight and nine by adding additional questions to explicitly capture important information relevant to that particular domain. We therefore ended up with a total of 12 domains.

Five investigators (HA, KD, ND, HT & PS), which included public health students, research assistants and a senior environmental health scientist with expertise on systematic reviews, independently extracted data in pairs according to the domains assessed in the LRAT. All extracted data from the reviews was tabulated and coded in MS Excel (Microsoft, Redmond WA, USA, 2016 MSO). The same investigators then rated each study domain independently in duplicate. For each LRAT domain, there were three possible appraisal ratings: (1) Satisfactory: conducted according to a clear, valid, and consistent procedure; (2) Unclear: insufficient documentation to allow evaluation; or (3) Unsatisfactory: positive evidence of invalid or inconsistent procedure. For the three domains (three, eight and nine) where we adapted the tool there were two possible appraisal ratings for the 'b' question, 'Yes' or 'No.' Discrepancies in the ratings were resolved by consensus, with the consensus rationale recorded for each decision. If agreement could not be reached, a senior scientist (PS) decided the rating. To assure quality assurance and quality control two seniors investigators (PS, SR) compared the final ratings and rationales to the published review to ensure accuracy and consistency in the final coding. Additionally, investigators did not evaluate reviews that they had co-authored.

When synthesizing the final appraisal ratings for the adapted domains (three, eight and nine) we classified the domains rated as 'Yes' as "satisfactory" and those as 'No' as 'unsatisfactory'. We reported and visually presented the percentage of "satisfactory" appraisal ratings for each LRAT domain over all reviews, as well as stratified by systematic versus non-systematic reviews. For consistency, when reporting the results we refer to the three additional questions 3b, 8 b and 9 b as domains and therefore report on 12 domains. Cross-tabulations were performed for evaluating possible associations between systematic reviews and non-systematic reviews and adherence of each LRAT Domain (classification of a 'satisfactory' rating) using the Fisher's Exact Test. All analyses were performed using the statistical software package R version 4.0.3

3. RESULTS

Over the three topics, we identified a total of 8,177 total records for screening (Figure 1). For the topic of Air Pollution and ASD we identified 1,155 records for screening, from which 6 reviews met our inclusion criteria (de Cock, Maas, and van de Bor 2012, Guxens and Sunyer 2012, Kalkbrenner, Schmidt, and Penlesky 2014, Rossignol, Genuis, and Frye 2014, Suades-González et al. 2015, Lam et al. 2016); for PBDEs and IQ/ADHD we identified 2,540 records for screening, from which 10 reviews met our inclusion criteria (Chao et al. 2014, Kim et al. 2014, Roth and Wilks 2014, Berghuis et al. 2015, Pinson, Bourguignon, and Parent 2016, Vrijheid et al. 2016, de Cock, Maas, and van de Bor 2012, Muir 2003, Brandt 2012, Lam et al. 2017); and for formaldehyde and asthma we identified 4,482 from which 13 reviews met our inclusion criteria (Golden 2011, Heinrich 2011, McGwin, Lienert, and Kennedy 2011, Baur, Bakehe, and Vellguth 2012, Hulin et al. 2012, Nielsen, Larsen, and Wolkoff 2013, Rumchev, Spickett, and Graham 2013, Tagiyeva and Sheikh 2014, Kundu, De, and Mitra 2015, Nurmatov et al. 2015, Patelarou, Tzanakis, and Kelly 2015, Yao et al. 2015, Lam et al. 2019) (Figure 1). Studies were excluded at the full text screening stage based on study question, review duplication, or (for the topic of formaldehyde and asthma) publication year (Figure 1). Consensus was reached on every included review without the need for a third investigator to adjudicate any decision on a review's eligibility and when rating each study domain with the tool.

All included reviews were published between 2003 and 2019. See Supplemental Materials Appendix II for details of every included review.

3.1 LRAT appraisal results overall

Thirteen of the 29 reviews (45%) were systematic reviews. Systematic reviews received a higher percentage of "satisfactory" LRAT ratings compared to non-systematic reviews across every domain (Figure 2). In eight of these domains, there was a statistically significant difference observed between the two types of reviews, and "satisfactory" ratings: 'Search Strategy' (Domain 4) (P=<0.001); 'Selection Process' (Domain 5) (P=0.003); 'Appraisal of directness of evidence' (Domain 6) (P=0.02); 'Appraisal of methodological quality of evidence' (Domain 7) (P=0.01); 'Synthesis of Evidence' (Domain 8a) (P=0.03); 'Meta-analysis' (Domain 8b) (P=0.01); 'Summation of Findings' (Domain 9a) (P=<0.001); and 'a priori definition of "sufficiency" or Evidence bar' (Domain 9b) (P=0.01).The frequency and percentage of all systematic and non-systematic reviews receiving a "satisfactory" rating for each LRAT domain across all topics and for each individual case study, and the results of theFisher's Exact Test are presented in Supplemental Materials Appendix III.

3.1.1 Non-Systematic Reviews—Every non-systematic review received an "unsatisfactory" or "unclear" rating in four of the 12 domains, including: 'The objective of the review' (Domain 1); 'The use of a protocol' (Domain 2); 'Appraisal of methodological quality of evidence' (Domain 7); and 'Meta-analysis' (Domain 8b) (Figure 2 & Supplemental Materials Appendix III). Only one domain, (Domain 6), 'Appraisal of directness of evidence' had more than 50% of non-systematic reviews receive a "satisfactory" rating.

3.1.2 Systematic reviews—

Across the systematic reviews, more than 50% received an "unsatisfactory" or "unclear" rating in 5 of the 12 domains: Ten (77%) for 'The objective of the review' (Domain 1) or 'The use of a protocol' (Domain 2); and eight (62%) for 'Roles and contributions' (Domain 3b), 'Appraisal of methodological quality of evidence' (Domain 7), and 'Meta-analysis' (Domain 8b) (Figure 2 & Supplemental Materials Appendix III). While only just over half of the systematic reviews (seven (54%)) received a "satisfactory" rating for 'Disclosures of interests' (Domain 3a) and 'a priori definition of "sufficiency" or Evidence bar' (Domain 9b). 'Appraisal of directness of evidence' (Domain 6) was the only domain in which all systematic reviews received a "satisfactory" rating. (Supplemental Materials Appendix II).

3.2 LRAT appraisal results by case study topic

3.2.1 Air pollution and ASD—

Three of the six reviews (50%) were systematic reviews (Figure 3). Across every LRAT domain, systematic reviews received a higher percentage of "satisfactory" LRAT ratings compared to non-systematic reviews. No statistically significant differences were observed between the two types of reviews and "satisfactory" ratings across any of the LRAT domains (Figure 3 & Supplemental Materials Appendix III). Every non-systematic review received an "unsatisfactory" or "unclear" rating in eight of the 12 domains and only one domain (Domain 6), 'Appraisal of directness of evidence' had more than 50% of non-systematic reviews receive a "satisfactory" rating. Across the systematic reviews, more than 50% received an "unsatisfactory" or "unclear" rating in five domains, including: 'The objective of the review' (Domain 1); 'The use of a protocol' (Domain 2); 'Roles and contributions' (Domain 3b); 'Appraisal of methodological quality of evidence' (Domain 7); and 'Meta-analysis' (Domain 8b). Every Systematic review, received a "satisfactory" rating for the domains 'Search strategy' (Domain 4) and 'Appraisal for the directness of the evidence' (Domain 6).

3.2.2 PBDEs and IQ/ ADHD—

Three of the ten reviews (30%) were systematic reviews (Figure 4). Across every LRAT domain, systematic reviews again received a higher percentage of "satisfactory" LRAT ratings compared to non-systematic reviews. No statistically significant differences were observed between the two types of reviews and "satisfactory" ratings across any of the LRAT domains (Figure 4 & Supplemental Materials Appendix III). Every non-systematic review received an "unsatisfactory" or "unclear" rating in seven of the 12 domains and more than 50% received an "unsatisfactory" or "unclear" rating in the other five domains. Across the systematic reviews, more than 50% once again received an "unsatisfactory" or "unclear" rating in five domains, including: 'The objective of the review' (Domain 1); 'The use of a protocol' (Domain 2); 'Roles and contributions' (Domain 3b); 'Appraisal of methodological quality of evidence' (Domain 7); and 'Meta-analysis' (Domain 8b). Every Systematic review again received a "satisfactory" rating for the domains ‘Search strategy’ (Domain 4) and ‘Appraisal for the directness of the evidence’ (Domain 6).

3.2.3 Formaldehyde and asthma—Seven of the 13 reviews (54%) were systematic reviews (Figure 5). Once again, across every LRAT domain, systematic reviews received a higher percentage of “satisfactory” LRAT ratings compared to non-systematic reviews. In one of these domains, there was a statistically significant difference observed between the two types of reviews, and “satisfactory” ratings: ‘Search Strategy’ (Domain 4) (P=0.02) (Figure 5 & Supplemental Materials Appendix III). Every non-systematic review received an “unsatisfactory” or “unclear” rating in six of the 12 domains and only one domain, ‘Appraisal for the directness of the evidence’ (Domain 6) had more than 50% of non-systematic reviews receive a “satisfactory” rating. Across the systematic reviews, more than 50% received an “unsatisfactory” or “unclear” rating in seven domains, including: ‘The objective of the review’ (Domain 1); ‘The use of a protocol’ (Domain 2); ‘Disclosure of interests’ (Domain 3a); ‘Roles and contributions’ (Domain 3b); the ‘Appraisal of methodological quality of evidence’ (Domain 7); ‘Meta-analysis’ (Domain 8b); and ‘a priori definition of “sufficiency” or Evidence bar’ (Domain 9b). Every systematic review received a “satisfactory” rating for ‘Appraisal for the directness of the evidence’ (Domain 6).

4. DISCUSSION

Our analysis of 29 reviews across three environmental health topics demonstrated that systematic reviews were consistently rated as “satisfactory” in the LRAT domains and met the criteria for a well-conducted review more often than non-systematic reviews. In only one domain did more than 50 percent of non-systematic reviews receive a “satisfactory” rating. Thus, the systematic reviews in this study have been demonstrated to have greater utility, validity, and transparency than non-systematic reviews when evaluating the harms of hazardous environmental exposures.

At the same time, however, almost all systematic reviews performed poorly across the LRAT domains, for example by failing to: state the review’s objective; develop a protocol; report the conflicts of interest (COI) of the review authors; evaluate the quality of the included evidence using valid methods; use meta-analysis to summarize study results; and by not stating a pre-defined definition of the evidence bar on which their conclusions were based. The domains in LRAT reflect the critical elements of a high quality systematic review, which have been developed to create a less biased evaluation of the existing literature (Higgins, Green, and Cochrane 2008, Shea et al. 2007, Moher et al. 2009). Therefore, the application of the LRAT tool to this sample of systematic reviews highlights the need for ongoing development, dissemination and implementation of empirically based systematic review methods to ensure that all systematic reviews implement these critical elements. Reviews that intentionally implemented these elements achieved “satisfactory” ratings for all the LRAT domains (Lam et al. 2016, Lam et al. 2017, Lam et al. 2019).

Possible reasons for the inconsistencies in the methods applied across the systematic reviews, however, may be due to the ongoing evolution and development of systematic review methods in environmental health. The systematic reviews included in this study were published between 2003–2019, however, the Navigation Guide methodology was not published until 2011 (Woodruff, Sutton, and Group 2011), and the first case study was not published until 2014 (Johnson et al. 2014, Koustas et al. 2014). Thus, the systematic reviews published prior to this time frame would not be expected to conform necessarily with these best practices, since they had not yet been established in the field of environmental health. However, poorly conducted systematic reviews have critical real-world implications for public policy. For example, the U.S. Environmental Protection Agency's (EPA's) current methods within the Office of Chemical Safety and Pollution Prevention (OCSPP) for implementing "systematic review" within the Toxic Substances Control Act (TSCA) fall so far short of internationally recognized standards that they will lead to biased evaluations of the science used for decision-making to the detriment of public health (Singla, Sutton, and Woodruff 2019). Therefore, in order for systematic reviews to produce reliable and valid answers to environmental health questions, they need to be based on empirically based standards, backed by authoritative bodies, for reducing bias and ensuring transparency.

To this end, several organizations have recommended and applied the use of PECO statements (Population, Exposure, Comparator and Outcome) in conducting environmental health assessments (Rooney et al. 2014, Woodruff, Sutton, and Group 2011), as these statements help authors systematize the objectives of the review (Whaley, Halsall, et al. 2016). Further, the use of protocols minimizes bias and ensures transparency in a review's process by pre-defining how the questions will be formulated, searches conducted, studies evaluated, and evidence synthesized. (Institute of Medicine 2011)

The use of accepted and transparent methods to assess the quality of the primary studies included in the systematic reviews is also essential to ensuring the integrity of the evaluations of environmental health hazards (Rooney et al. 2016). It is therefore critical to promote the implementation of this step in the systematic review process, along with further development of empirically based tools to assess the quality of the various types of evidence included in environmental health systematic reviews (Mandrioli and Silbergeld 2016, Bero et al. 2018).

Lack of disclosures identified in this study is seen consistently throughout other areas of research (Forsyth et al. 2014, Roseman et al. 2011, Baethge 2013). Across several fields of research, industry sponsorship and author COI have been found to be associated with outcomes that favor the industry sponsor (Lundh et al. 2017, Huss et al. 2007, Barnes and Bero 1998). Therefore, an essential step in quantifying this potential bias on environmental health research is being able to identify who the funders of the research are, and whether the authors of the study have a financial conflict of interest. The enforcement of penalties for non-compliance to journal COI policies could deter authors from failing to disclose such conflicts (Institute of Medicine 2009).

Meta-analyses produce quantitative estimates of risk that are critical to regulatory decision-making, can decrease or increase confidence in the body of evidence, and are very useful in environmental health as they can help increase the statistical power of environmental epidemiological studies (Higgins, Green, and Cochrane 2008). Less than half of the systematic reviews used a meta-analysis to analyze the results, which may in part be due to the limitations of the design and reporting of primary research that makes combining data from different studies extremely challenging. Thus, efforts that promote consistency in study reporting are needed to support meta-analyses opportunities when synthesizing study results. Further, not all meta-analyses are equal and there is a need to promote how to implement them appropriately, just like any other aspect of science or systematic reviews.

To minimize bias, it is also essential to pre-define the evidence needed for a specific finding, i.e., a definition of “sufficient” or other nomenclature describing the strength of the association between exposure and health outcome. For example, there were two systematic reviews of PBDEs and IQ/ADHD, (Kim et al. 2014, Roth and Wilks 2014) in addition to the Navigation Guide 2017 review (Lam et al. 2017). Neither the Kim nor the Roth and Wilks reviews defined the meaning of its strength of evidence terms. The Roth and Wilks review concluded that while the evidence is “suggestive”, it does not substantiate a “causal” relationship, however neither “suggestive” or “causal” is defined (Roth and Wilks 2014). The Kim review evaluated “evidence for causality” between exposure and health outcomes using the Bradford-Hill framework, and concluded there is a “possible relationship between BFR [brominated flame retardant] exposure and serious health consequences, namely … neurobehavioral and developmental outcomes in children… ”, with “possible” not defined (Kim et al. 2014). Without definitions to the evidence terms it is difficult to assess whether these findings are different from each other and thus are left open to interpretation. Therefore, it is difficult to interpret the evidence evaluation for researchers, policy makers, or the public, which further undermines the ultimate goal of timely, health protective decision-making.

4.1 Study Limitations

While we conducted a comprehensive search and followed explicit and well-defined inclusion and exclusion criteria for the included reviews, our study only assessed three environmental health topics. Therefore, an analysis across additional topics may present different findings. Further, the tool used in this study was an early version of an appraisal tool and required additional interpretation before its application. Despite this limitation, however, the LRAT domains were based on toolkits and appraisals of the methodological quality of literature reviews conducted in the medical sciences that appraise essential features of a review that reduce bias and increase transparency. We therefore believe it was the most appropriate available tool to conduct this analysis. Finally, while our review extends to 2019, none of the included reviews have been published since 2017. Thus, due to recent developments in the field of systematic review, the relevance of the findings may be limited, with an underestimation of the number of systematic reviews that comply with each LRAT domain.

While the application of the LRAT to this sample of reviews has demonstrated that systematic reviews have greater utility, validity, and transparency than non-systematic, and are thus superior in informing decisions around the harms of hazardous environmental exposures, efforts must be made to continue to improve their standards. Systematic review methods are new to environmental health and our results are consistent with studies that have documented a high prevalence of poorly conducted and reported systematic reviews in the clinical literature (Page et al. 2016, Ioannidis 2016). The lack of consistently applying robust methods to synthesize the available data identified in our research may be prevalent in the field. Pilot data from *Environment International* found "serious omissions in reporting [in] 19 of 25 SRs [systematic reviews] published in the top environmental health journals through 2014–2015" (Whaley, Letcher, et al. 2016). This same journal has now appointed an associate editor for systematic reviews and has made compulsory the Preferred Reporting Items for Systematic Reviews (PRISMA) guideline (Moher et al. 2009). A necessary first step to evaluating the methodological rigor of systematic reviews is complete and accurate reporting, therefore these reporting guidelines should be mandatory throughout all journals as recommended by *Environment International* (Bero 2017). Further, to realize the benefits of systematic reviews over expert-based narrative reviews in informing environmental health decision-making, it is essential for practitioners, journal editors, and policy makers, advocates, journalists, and other end-users, to become rapidly educated and competent at scrutinizing what constitutes reliable, reproducible, and transparent systematic review methods. Therefore, enforcing standards that systematic reviews must meet in order for them to be published will be equally critical in improving the quality of systematic reviews in environmental health.

5. CONCLUSION

Robust methods to synthesize what is known about the environmental drivers of health are crucial to making science actionable. Systematic reviews produced more transparent and valid conclusions than narrative, non-systematic reviews but poorly conducted systematic reviews were prevalent. Ongoing development, dissemination and implementation of empirically based systematic review methods are therefore required to ensure that all systematic reviews meet the domains in the LRAT that reflect the critical elements of a high quality systematic review, which have been developed to create a less biased and more transparent evaluation of the existing literature. Further, standardized application of tools to evaluate the quality of systematic reviews during the journal peer-review process would provide crucial training and feedback to scientists, journal editors, and policy makers, and help ensure only high-quality systematic reviews are produced.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements:

Hazel Tesoro, Veena Singla, Hannah Archibald, Max Aung and Kristen Davis.

This work was supported by the Broadreach Foundation, San Francisco, CA [132805A-2019]; the Clarence E. Heller Charitable Foundation, San Francisco, CA [131176A 2019]; the Passport Foundation, San Francisco, CA [A132316 2019]; the National Institute of Environmental Health Sciences, Research Triangle Park, NC [P01ES022841], and the US Environmental Protection Agency, Washington, DC [RD83543301].

6. REFERENCES

Baethge C 2013. "The effect of a conflict of interest disclosure form using closed questions on the number of positive conflicts of interest declared - a controlled study." PeerJ 1:e128. doi: 10.7717/peerj.128. [PubMed: 24024081]

Barnes DE, and Bero LA. 1998. "Why review articles on the health effects of passive smoking reach different conclusions." Jama 279 (19):1566–70. doi: 10.1001/jama.279.19.1566. [PubMed: 9605902]

Baur X, Bakehe P, and Vellguth H. 2012. "Bronchial asthma and COPD due to irritants in the workplace - an evidence-based approach." J Occup Med Toxicol 7 (1):19. doi: 10.1186/1745-6673-7-19. [PubMed: 23013890]

Berghuis SA, Bos AF, Sauer PJ, and Roze E. 2015. "Developmental neurotoxicity of persistent organic pollutants: an update on childhood outcome." Arch Toxicol 89 (5):687–709. doi: 10.1007/s00204-015-1463-3. [PubMed: 25618547]

Bero L 2017. "Systematic Review: A Method at Risk for Being Corrupted." Am J Public Health 107 (1):93–96. doi: 10.2105/AJPH.2016.303518. [PubMed: 27854519]

Bero L, Chartres N, Diong J, Fabbri A, Ghersi D, Lam J, Lau A, McDonald S, Mintzes B, Sutton P, Turton JL, and Woodruff TJ. 2018. "The risk of bias in observational studies of exposures (ROBINS-E) tool: concerns arising from application to observational studies of exposures." Syst Rev 7 (1):242. doi: 10.1186/s13643-018-0915-2. [PubMed: 30577874]

Brandt B 2012. "Household Environmental Toxins and Neurodevelopment in Children." accessed June 20. https://scholarworks.montana.edu/xmlui/bitstream/handle/1/966/BrandtB0512.pdf.

Chao HR, Huang HL, Hsu YC, Lin CW, Lin DY, Gou YY, and Chen KC. 2014. "Impact of Brominated POPs on the Neurodevelopment and Thyroid Hormones of Young Children in an Indoor Environment—A Review." Aerosol Air Qual. Res. 14:1320–1332. doi: 10.4209/aaqr.2013.05.0156.

Cochrane Library. 2019. "About Cochrane Reviews." accessed May 11. https://www.cochranelibrary.com/about/about-cochrane-reviews.

de Cock M, Maas YG, and van de Bor M. 2012. "Does perinatal exposure to endocrine disruptors induce autism spectrum and attention deficit hyperactivity disorders? Review." Acta Paediatr 101 (8):811–8. doi: 10.1111/j.1651-2227.2012.02693.x. [PubMed: 22458970]

Descatha A, Sembajwe G, Baer M, Boccuni F, Di Tecco C, Duret C, Evanoff BA, Gagliardi D, Ivanov ID, Leppink N, Marinaccio A, Magnusson Hanson LL, Ozguler A, Pega F, Pell J, Pico F, Prüss-Üstün A, Ronchetti M, Roquelaure Y, Sabbath E, Stevens GA, Tsutsumi A, Ujita Y, and Iavicoli S. 2018. "WHO/ILO work-related burden of disease and injury: Protocol for systematic reviews of exposure to long working hours and of the effect of exposure to long working hours on stroke." Environ Int 119:366–378. doi: 10.1016/j.envint.2018.06.016. [PubMed: 30005185]

European Environment Agency. 2001. Late lessons from early warnings: the precautionary principle 1896–2000.

European Environment Agency. 2013. Late lessons from early warnings: science, precaution, innovation.

Forsyth SR, Odierna DH, Krauth D, and Bero LA. 2014. "Conflicts of interest and critiques of the use of systematic reviews in policymaking: an analysis of opinion articles." Syst Rev 3:122. doi: 10.1186/2046-4053-3-122. [PubMed: 25417178]

Fox DM 2017. "Evidence and Health Policy: Using and Regulating Systematic Reviews." Am J Public Health 107 (1):88–92. doi: 10.2105/AJPH.2016.303485. [PubMed: 27854522]

Glantz S, and Gonzalez M. 2012. "Effective tobacco control is key to rapid progress in reduction of non-communicable diseases." Lancet 379 (9822):1269–71. doi: 10.1016/S0140-6736(11)60615-6. [PubMed: 21963004]

Godderis L, Boonen E, Cabrera Martimbianco AL, Delvaux E, Ivanov ID, Lambrechts MC, Latorraca COC, Leppink N, Pega F, Prüss-Ustün AM, Riera R, Ujita Y, and Pachito DV. 2018. "WHO/ILO work-related burden of disease and injury: Protocol for systematic reviews of exposure to long working hours and of the effect of exposure to long working hours on alcohol consumption and alcohol use disorders." Environ Int 120:22–33. doi: 10.1016/j.envint.2018.07.025. [PubMed: 30055358]

Golden R 2011. “Identifying an indoor air exposure limit for formaldehyde considering both irritation and cancer hazards.” Crit Rev Toxicol 41 (8):672–721. doi: 10.3109/10408444.2011.573467. [PubMed: 21635194]

Guxens M, and Sunyer J. 2012. “A review of epidemiological studies on neuropsychological effects of air pollution.” Swiss Med Wkly 141:w13322. doi: 10.4414/smw.2011.13322. [PubMed: 22252905]

Heinrich J 2011. “Influence of indoor factors in dwellings on the development of childhood asthma.” Int J Hyg Environ Health 214 (1):1–25. doi: 10.1016/j.ijheh.2010.08.009. [PubMed: 20851050]

Higgins JPT, Green S, and Cochrane C. 2008. Cochrane Handbook for Systematic Reviews of Interbentions. Chichestter, England Hoboken, NJ Wiley-Blackwell.

Hulin M, Simoni M, Viegi G, and Annesi-Maesano I. 2012. “Respiratory health and indoor air pollutants based on quantitative exposure assessments.” Eur Respir J 40 (4):1033–45. doi: 10.1183/09031936.00159011. [PubMed: 22790916]

Hulshof CTJ, Colosio C, Daams JG, Ivanov ID, Prakash KC, Kuijer PPFM, Leppink N, Mandic-Rajcevic S, Masci F, van der Molen HF, Neupane S, Nygård CH, Oakman J, Pega F, Proper K, Prüss-Üstün AM, Ujita Y, and Frings-Dresen MHW. 2019. “WHO/ILO work-related burden of disease and injury: Protocol for systematic reviews of exposure to occupational ergonomic risk factors and of the effect of exposure to occupational ergonomic risk factors on osteoarthritis of hip or knee and selected other musculoskeletal diseases.” Environ Int 125:554–566. doi: 10.1016/j.envint.2018.09.053. [PubMed: 30583853]

Huss A, Egger M, Hug K, Huwiler-Müntener K, and Röösli M. 2007. “Source of funding and results of studies of health effects of mobile phone use: systematic review of experimental studies.” Environ Health Perspect115 (1):1–4. doi: 10.1289/ehp.9149 [PubMed: 17366811]

Institute of Medicine. 2009. “Conflict of Interest in Medical Research, Education, and Practice.” In.

Institute of Medicine. 2011. “Finding What Works in Health Care: Standards for Systematic Reviews.” In.

Ioannidis JP 2016. “The Mass Production of Redundant, Misleading, and Conflicted Systematic Reviews and Meta-analyses.” Milbank Q 94 (3):485–514. doi: 10.1111/1468-0009.12210. [PubMed: 27620683]

Johnson PI, Koustas E, Vesterinen HM, Sutton P, Atchley DS, Kim AN, Campbell M, Donald JM, Sen S, Bero L, Zeise L, and Woodruff TJ. 2016. “Application of the Navigation Guide systematic review methodology to the evidence for developmental and reproductive toxicity of triclosan.” Environ Int 92–93:716–28. doi: 10.1016/j.envint.2016.03.009.

Johnson PI, Sutton P, Atchley DS, Koustas E, Lam J, Sen S, Robinson KA, Axelrad DA, and Woodruff TJ. 2014. “The Navigation Guide - evidence-based medicine meets environmental health: systematic review of human evidence for PFOA effects on fetal growth.” Environ Health Perspect 122 (10):1028–39. doi: 10.1289/ehp.1307893. [PubMed: 24968388]

Kalkbrenner AE, Schmidt RJ, and Penlesky AC. 2014. “Environmental chemical exposures and autism spectrum disorders: a review of the epidemiological evidence.” Curr Probl Pediatr Adolesc Health Care 44 (10):277–318. doi: 10.1016/j.cppeds.2014.06.001. [PubMed: 25199954]

Kim YR, Harden FA, Toms LM, and Norman RE. 2014. “Health consequences of exposure to brominated flame retardants: a systematic review.” Chemosphere 106:1–19. doi: 10.1016/j.chemosphere.2013.12.064. [PubMed: 24529398]

Koustas E, Lam J, Sutton P, Johnson PI, Atchley DS, Sen S, Robinson KA, Axelrad DA, and Woodruff TJ. 2014. “The Navigation Guide - evidence-based medicine meets environmental health: systematic review of nonhuman evidence for PFOA effects on fetal growth.” Environ Health Perspect 122 (10):1015–27. doi: 10.1289/ehp.1307177. [PubMed: 24968374]

Kundu S, De A, and Mitra S. 2015. “Formaldehyde: Fact sheet reflecting uses, exposure hazards, health effects and toxicological profile-A broad overview for medical professionals and embalmers.” JCER 3 (2):183–196. doi: 10.5455/jcer.201521.

Lam J, Koustas E, Sutton P, Cabana M, Whitaker E, Padula A, H Vesterinen, Daniels N, and Woodruff TJ. 2019. “Applying the Navigation Guide Systematic Review MethodologyCase Study #6.” https://prhe.ucsf.edu/sites/g/files/tkssra341/f/Formaldehyde%20protocol%20FINAL%20UPLOADED%20TO%20PROSPERO%202016-05-03.pdf.

Lam J, Koustas E, Sutton P, Johnson PI, Atchley DS, Sen S, Robinson KA, Axelrad DA, and Woodruff TJ. 2014. “The Navigation Guide - evidence-based medicine meets environmental health: integration of animal and human evidence for PFOA effects on fetal growth.” Environ Health Perspect 122 (10):1040–51. doi: 10.1289/ehp.1307923. [PubMed: 24968389]

Lam J, Lanphear BP, Bellinger D, Axelrad DA, McPartland J, Sutton P, Davidson L, Daniels N, Sen S, and Woodruff TJ. 2017. “Developmental PBDE Exposure and IQ/ADHD in Childhood: A Systematic Review and Meta-analysis.” Environ Health Perspect 125 (8):086001. doi: 10.1289/ehp1632. [PubMed: 28799918]

Lam J, Sutton P, Kalkbrenner A, Windham G, Halladay A, Koustas E, Lawler C, Davidson L, Daniels N, Newschaffer C, and Woodruff T. 2016. “A Systematic Review and Meta-Analysis of Multiple Airborne Pollutants and Autism Spectrum Disorder.” PLoS One 11 (9):e0161851. doi: 10.1371/journal.pone.0161851. [PubMed: 27653281]

Li J, Brisson C, Clays E, Ferrario MM, Ivanov ID, Landsbergis P, Leppink N, Pega F, Pikhart H, Prüss-Üstün A, Rugulies R, Schnall PL, Stevens G, Tsutsumi A, Ujita Y, and Siegrist J. 2018. “WHO/ILO work-related burden of disease and injury: Protocol for systematic reviews of exposure to long working hours and of the effect of exposure to long working hours on ischaemic heart disease.” Environ Int 119:558–569. doi: 10.1016/j.envint.2018.06.022. [PubMed: 30125833]

Lightwood JM, Dinno A, and Glantz SA. 2008. “Effect of the California tobacco control program on personal health care expenditures.” PLoS Med 5 (8):e178. doi: 10.1371/journal.pmed.0050178. [PubMed: 18752344]

Literature Review Appraisal Toolkit. 2014. “Literature Review Appraisal Toolkit (LRAT).” accessed May 14. http://policyfromscience.com/lrat/about-the-lra-toolkit/.

Lundh A, Lexchin J, Mintzes B, Schroll JB, and Bero L. 2017. “Industry sponsorship and research outcome.” Cochrane Database Syst Rev 2:Mr000033. doi: 10.1002/14651858.MR000033.pub3. [PubMed: 28207928]

Mandrioli D, Schlunssen V, Adam B, Cohen RA, Colosio C, Chen W, Fischer A, Godderis L, Goen T, Ivanov ID, Leppink N, Mandic-Rajcevic S, Masci F, Nemery B, Pega F, Pruss-Ustun A, Sgargi D, Ujita Y, van der Mierden S, Zungu M, and Scheepers PTJ. 2018. “WHO/ILO work-related burden of disease and injury: Protocol for systematic reviews of occupational exposure to dusts and/or fibres and of the effect of occupational exposure to dusts and/or fibres on pneumoconiosis.” Environ Int 119:174–185. doi: 10.1016/j.envint.2018.06.005. [PubMed: 29958118]

Mandrioli D, and Silbergeld EK. 2016. “Evidence from Toxicology: The Most Essential Science for Prevention.” Environ Health Perspect 124 (1):6–11. doi: 10.1289/ehp.1509880. [PubMed: 26091173]

McGwin G, Lienert J, and Kennedy JI. 2011. “Formaldehyde exposure and asthma in children: a systematic review.” Cien Saude Colet 16 (9):3845–52. doi: 10.1590/s1413-81232011001000020. [PubMed: 21987327]

Moher D, Liberati A, Tetzlaff J, Altman DG, and PRISMA Group. 2009. “Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement.” J Clin Epidemiol 62 (10):1006–12. doi: 10.1016/j.jclinepi.2009.06.005. [PubMed: 19631508]

Morgan RL, Thayer KA, Bero L, Bruce N, Falck-Ytter Y, Ghersi D, Guyatt G, Hooijmans C, Langendam M, Mandrioli D, Mustafa RA, Rehfuess EA, Rooney AA, Shea B, Silbergeld EK, Sutton P, Wolfe MS, Woodruff TJ, Verbeek JH, Holloway AC, Santesso N, and Schünemann HJ. 2016. “GRADE: Assessing the quality of evidence in environmental and occupational health.” Environ Int 92–93:611–6. doi: 10.1016/j.envint.2016.01.004.

Muir T 2003. “Is there a relationship between the rise in thyroid and neurodevelopmental health effects in North America and the rise in concentrations of PBDEs in the environment?” Organohalogen Compounds (61):195–198.

National Academies of Sciences Engineering and Medicine. 2017. “Application of Systematic Review Methods in an Overall Strategy for Evaluating Low-Dose Toxicity from Endocrine Active Chemicals.” In.

National Academies of Sciences Engineering and Medicine. 2018. Progress Toward Transforming the Integrated Risk Information System (IRIS) Program: A 2018 Evaluation. Washington, DC: National Academies Press.

National Research Council. 2014a. "Review of EPA's Integrated Risk Information System (IRIS) Process." In.

National Research Council. 2014b. Review of the Environmental Protection Agency's State-of-the-Science Evaluation of Nonmonotonic Dose-Response Relationships as they Apply to Endocrine Disruptors. Washington, DC.

National Toxicology Program. 2015. Handbook for Conducting Systematic Reviews for Health Effects Evaluations. National Institute of Environmental Health Sciences.

Nielsen GD, Larsen ST, and Wolkoff P. 2013. "Recent trend in risk assessment of formaldehyde exposures from indoor air." Arch Toxicol 87 (1):73–98. doi: 10.1007/s00204-012-0975-3. [PubMed: 23179754]

Nurmatov UB, Tagiyeva N, Semple S, Devereux G, and Sheikh A. 2015. "Volatile organic compounds and risk of asthma and allergy: a systematic review." Eur Respir Rev 24 (135):92–101. doi: 10.1183/09059180.00000714. [PubMed: 25726560]

Page MJ, Shamseer L, Altman DG, Tetzlaff J, Sampson M, Tricco AC, Catala-Lopez F, Li L, Reid EK, Sarkis-Onofre R, and Moher D. 2016. "Epidemiology and Reporting Characteristics of Systematic Reviews of Biomedical Research: A Cross-Sectional Study." PLoS Med 13 (5):e1002028. doi: 10.1371/journal.pmed.1002028. [PubMed: 27218655]

Patelarou E, Tzanakis N, and Kelly FJ. 2015. "Exposure to indoor pollutants and Wheeze and asthma development during early childhood." Int J Environ Res Public Health 12 (4):3993–4017. doi: 10.3390/ijerph120403993. [PubMed: 25872014]

Paulo MS, Adam B, Akagwu C, Akparibo I, Al-Rifai RH, Bazrafshan S, Gobba F, Green AC, Ivanov I, Kezic S, Leppink N, Loney T, Modenese A, Pega F, Peters CE, Prüss-Üstün AM, Tenkate T, Ujita Y, Wittlich M, and John SM. 2019. "WHO/ILO work-related burden of disease and injury: Protocol for systematic reviews of occupational exposure to solar ultraviolet radiation and of the effect of occupational exposure to solar ultraviolet radiation on melanoma and non-melanoma skin cancer." Environ Int 126:804–815. doi: 10.1016/j.envint.2018.09.039. [PubMed: 30792021]

Pinson A, Bourguignon JP, and Parent AS. 2016. "Exposure to endocrine disrupting chemicals and neurodevelopmental alterations." Andrology 4 (4):706–22. doi: 10.1111/andr.12211. [PubMed: 27285165]

Rooney AA, Boyles AL, Wolfe MS, Bucher JR, and Thayer KA. 2014. "Systematic review and evidence integration for literature-based environmental health science assessments." Environ Health Perspect 122 (7):711–8. doi: 10.1289/ehp.1307972. [PubMed: 24755067]

Rooney AA, Cooper GS, Jahnke GD, Lam J, Morgan RL, Boyles AL, Ratcliffe JM, Kraft AD, Schunemann HJ, Schwingl P, Walker TD, Thayer KA, and Lunn RM. 2016. "How credible are the study results? Evaluating and applying internal validity tools to literature-based assessments of environmental health hazards." Environ Int 92–93:617–29. doi: 10.1016/j.envint.2016.01.005.

Roseman M, Milette K, Bero LA, Coyne JC, Lexchin J, Turner EH, and Thombs BD. 2011. "Reporting of conflicts of interest in meta-analyses of trials of pharmacological treatments." Jama 305 (10):1008–17. doi: 10.1001/jama.2011.257. [PubMed: 21386079]

Rossignol DA, Genuis SJ, and Frye RE. 2014. "Environmental toxicants and autism spectrum disorders: a systematic review." Transl Psychiatry 4:e360. doi: 10.1038/tp.2014.4. [PubMed: 24518398]

Roth N, and Wilks MF. 2014. "Neurodevelopmental and neurobehavioural effects of polybrominated and perfluorinated chemicals: a systematic review of the epidemiological literature using a quality assessment scheme." Toxicol Lett 230 (2):271–81. doi: 10.1016/j.toxlet.2014.02.015. [PubMed: 24583043]

Rugulies R, Ando E, Ayuso-Mateos JL, Bonafede M, Cabello M, Di Tecco C, Dragano N, Durand-Moreau Q, Eguchi H, Gao J, Garde AH, Iavicoli S, Ivanov ID, Leppink N, Madsen IEH, Pega F, Prüss-Üstün AM, Rondinone BM, Sørensen K, Tsuno K, Ujita Y, and Zadow A. 2019. "WHO/ILO work-related burden of disease and injury: Protocol for systematic reviews of exposure to long working hours and of the effect of exposure to long working hours on depression." Environ Int 125:515–528. doi: 10.1016/j.envint.2018.11.011. [PubMed: 30737040]

Rumchev K, Spickett J, and Graham J. 2013. "Exposures to Organic Pollutants and Respiratory Illnesses in Adults and Children." In Effects of Persistent and Bioactive Organic Pollutants on Human Health. John Wiley & Sons, Inc.

Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, Porter AC, Tugwell P, Moher D, and Bouter LM. 2007. "Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews." BMC Med Res Methodol 7:10. doi: 10.1186/1471-2288-7-10. [PubMed: 17302989]

Singla VI, Sutton PM, and Woodruff TJ. 2019. "The Environmental Protection Agency Toxic Substances Control Act Systematic Review Method May Curtail Science Used to Inform Policies, With Profound Implications for Public Health." Am J Public Health 109 (7):982–984. doi: 10.2105/ajph.2019.305068. [PubMed: 31166745]

Suades-González E, Gascon M, Guxens M, and Sunyer J. 2015. "Air Pollution and Neuropsychological Development: A Review of the Latest Evidence." Endocrinology 156 (10):3473–82. doi: 10.1210/en.2015-1403. [PubMed: 26241071]

Tagiyeva N, and Sheikh A. 2014. "Domestic exposure to volatile organic compounds in relation to asthma and allergy in children and adults." Expert Rev Clin Immunol 10 (12):1611–39. doi: 10.1586/1744666X.2014.972943. [PubMed: 25399826]

Teixeira LR, Azevedo TM, Bortkiewicz A, Corrêa da Silva DT, de Abreu W, de Almeida MS, de Araujo MAN, Gadzicka E, Ivanov ID, Leppink N, Macedo MRV, de S Maciel EMG, Pawlaczyk-Łuszczyńska M, Pega F, Prüss-Üstün AM, Siedlecka J, Stevens GA, Ujita Y, and Braga JU. 2019. "WHO/ILO work-related burden of disease and injury: Protocol for systematic reviews of exposure to occupational noise and of the effect of exposure to occupational noise on cardiovascular disease." Environ Int 125:567–578. doi: 10.1016/j.envint.2018.09.040. [PubMed: 30683322]

Tenkate T, Adam B, Al-Rifai RH, Chou BR, Gobba F, Ivanov ID, Leppink N, Loney T, Pega F, Peters CE, Prüss-Üstün AM, Silva Paulo M, Ujita Y, Wittlich M, and Modenese A. 2019. "WHO/ILO work-related burden of disease and injury: Protocol for systematic reviews of occupational exposure to solar ultraviolet radiation and of the effect of occupational exposure to solar ultraviolet radiation on cataract." Environ Int 125:542–553. doi: 10.1016/j.envint.2018.10.001. [PubMed: 30737039]

Tsai P, and Hatfield T. 2011. "Global Benefits From the Phaseout of Leaded Fuel." Journal of Environmental Health 74 (5):8–15.

U.S. Environmental Protection Agency. 2011. The Benefits and Costs of the Clean Air Act From 1990 to 2020. Washington, DC.

Vandenberg LN, Ågerstrand M, Beronius A, Beausoleil C, Å Bergman, Bero LA, Bornehag, Boyer CS, Cooper GS, Cotgreave I, Gee D, Grandjean P, Guyton KZ, Hass U, Heindel JJ, Jobling S, Kidd KA, Kortenkamp A, Macleod MR, Martin OV, Norinder U, Scheringer M, Thayer KA, Toppari J, Whaley P, Woodruff TJ, and Rudén C. 2016. "A proposed framework for the systematic review and integrated assessment (SYRINA) of endocrine disrupting chemicals." Environ Health 15 (1):74. doi: 10.1186/s12940-016-0156-6. [PubMed: 27412149]

Vesterinen HM, Johnson PI, Atchley DS, Sutton P, Lam J, Zlatnik MG, Sen S, and Woodruff TJ. 2015. "Fetal growth and maternal glomerular filtration rate: a systematic review." J Matern Fetal Neonatal Med 28 (18):2176–81. doi: 10.3109/14767058.2014.980809. [PubMed: 25382561]

Vrijheid M, Casas M, Gascon M, Valvi D, and Nieuwenhuijsen M. 2016. "Environmental pollutants and child health-A review of recent concerns." Int J Hyg Environ Health 219 (4–5):331–42. doi: 10.1016/j.ijheh.2016.05.001. [PubMed: 27216159]

Whaley P 2013. Systematic review and the future of evidence in chemicals policy. The Policy from Science Project.

Whaley P, Halsall C, Agerstrand M, Aiassa E, Benford D, Bilotta G, Coggon D, Collins C, Dempsey C, Duarte-Davidson R, FitzGerald R, Galay-Burgos M, Gee D, Hoffmann S, Lam J, Lasserson T, Levy L, Lipworth S, Ross SM, Martin O, Meads C, Meyer-Baron M, Miller J, Pease C, Rooney A, Sapiets A, Stewart G, and Taylor D. 2016. "Implementing systematic review techniques in chemical risk assessment: Challenges, opportunities and recommendations." Environ Int 92–93:556–64. doi: 10.1016/j.envint.2015.11.002.

Whaley P, Letcher RJ, Covaci A, and Alcock R. 2016. "Raising the standard of systematic reviews published in Environment International." Environ Int 97:274–276. doi: 10.1016/j.envint.2016.08.007. [PubMed: 27567414]

Woodruff TJ, and Sutton P. 2014. “The Navigation Guide systematic review methodology: a rigorous and transparent method for translating environmental health science into better health outcomes.” Environ Health Perspect 122 (10):1007–14. doi: 10.1289/ehp.1307175. [PubMed: 24968373]

Woodruff TJ, Sutton P, and Navigation Guide Work Group. 2011. “An evidence-based medicine methodology to bridge the gap between clinical and environmental health sciences.” Health Aff (Millwood) 30 (5):931–7. doi: 10.1377/hlthaff.2010.1219. [PubMed: 21555477]

Yao Y, Liang W, Zhu L, Duan Y, Jin Y, and He L. 2015. “Relationship between the concentration of formaldehyde in the air and asthma in children: a meta-analysis.” Int J Clin Exp Med 8 (6):8358–62. [PubMed: 26309488]

Highlights

1. Environmental health science is transitioning to systematic review (SR) methods
2. Using the Literature Review Appraisal Toolkit (LRAT) we evaluated SRs and non SRs
3. SRs received a higher % of satisfactory ratings in all domains compared to non SRs
4. SRs that used empirical SR methods achieved satisfactory ratings in all domains
5. Ongoing development and implementation of empirical SR methods are required

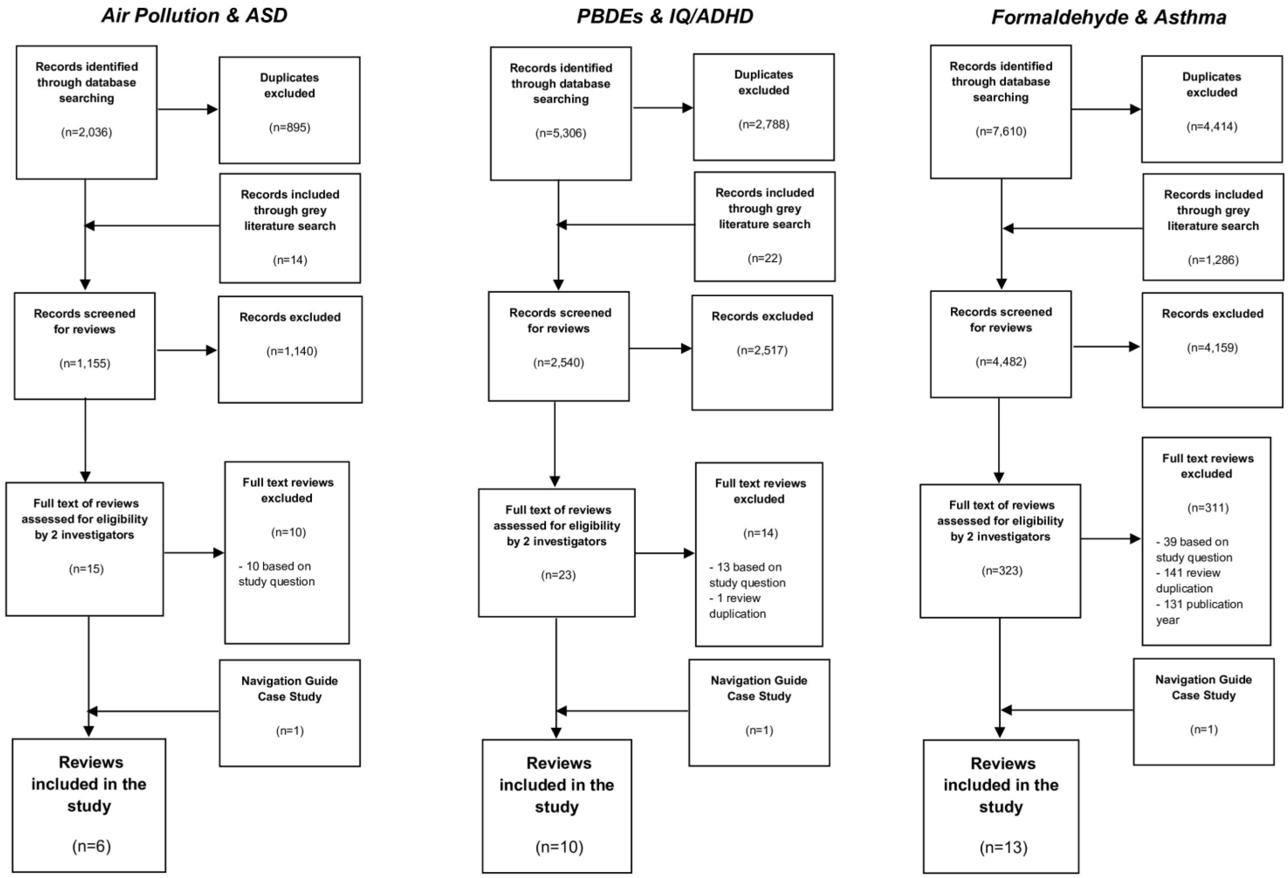

Figure 1.
Study Flow Diagrams for all case study topics.

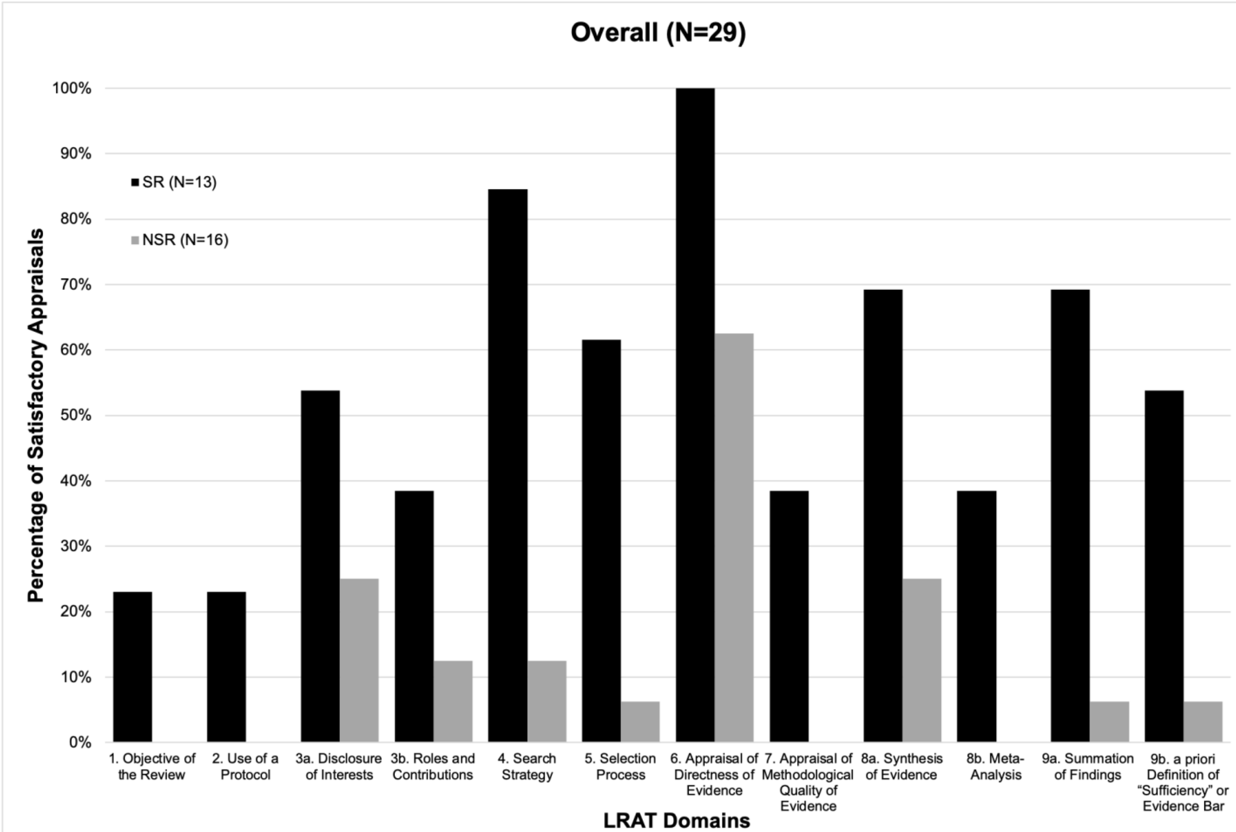

Figure 2.

Adherent Reviews for all Case Study Topics. Represented is the percentage of systematic and non-systematic reviews receiving a "satisfactory" rating for each LRAT domain overall.

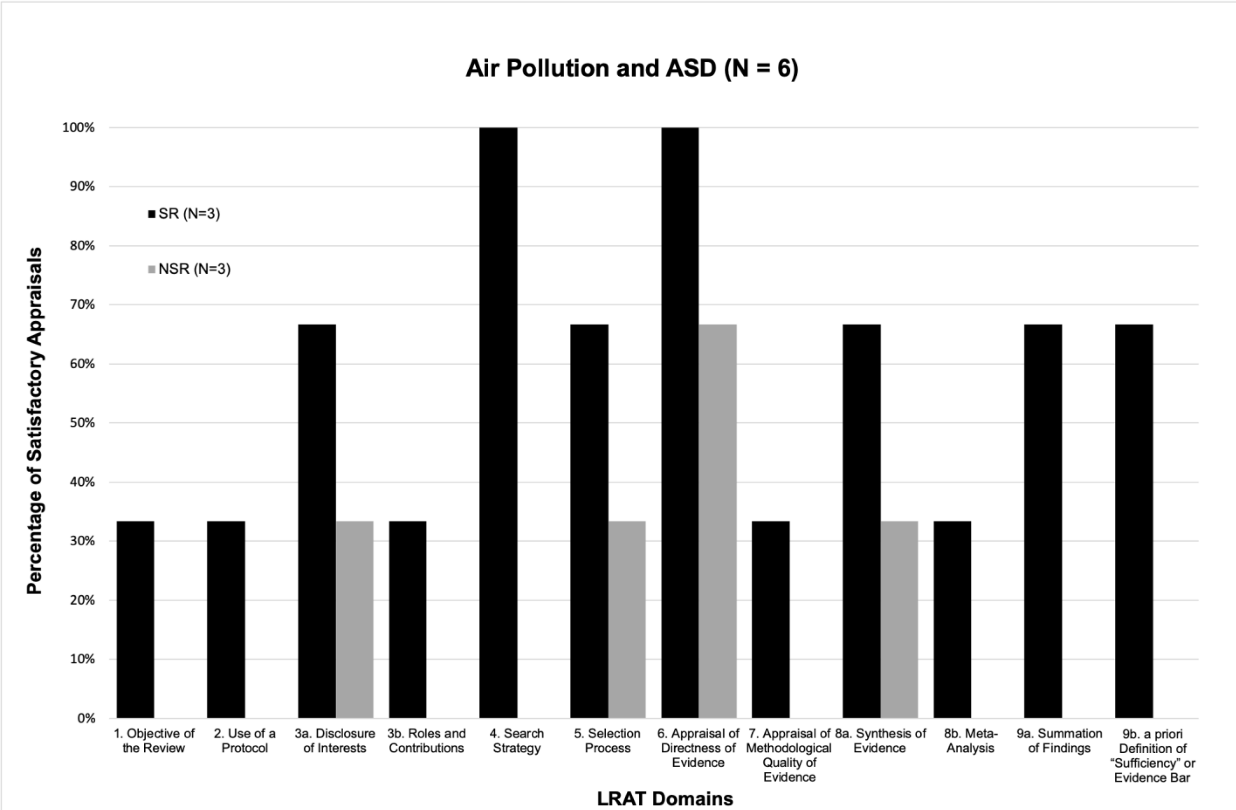

Figure 3.

Adherent Reviews for the Air Pollution and ASD Case Study Topic. Represented is the percentage of systematic and non-systematic reviews receiving a “satisfactory” rating for each LRAT domain that examined the relationship between air pollution and ASD.

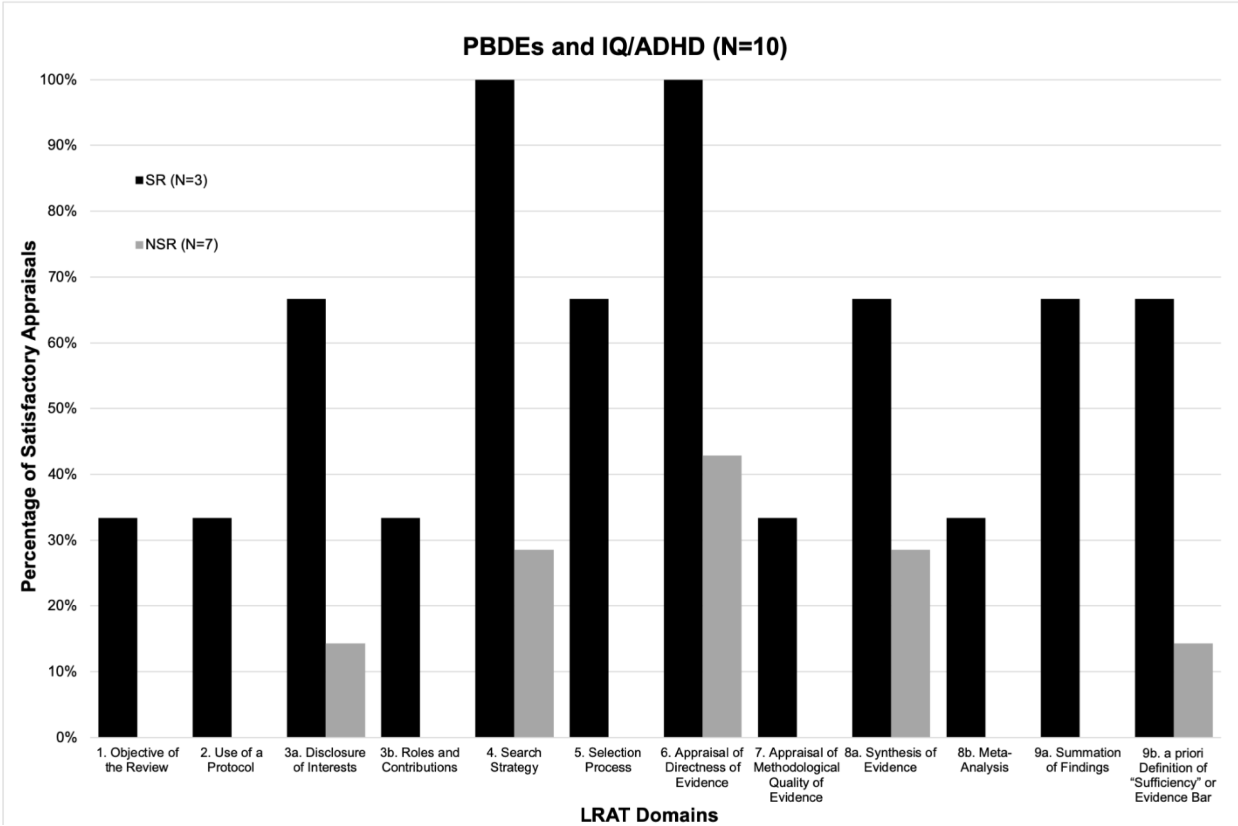

Figure 4.

Adherent Reviews for the PBDEs and IQ and/or ADHD Case Study Topic. Represented is the percentage of systematic and non-systematic reviews receiving a "satisfactory" rating for each LRAT domain that examined the relationship between PBDEs and IQ/ADHD

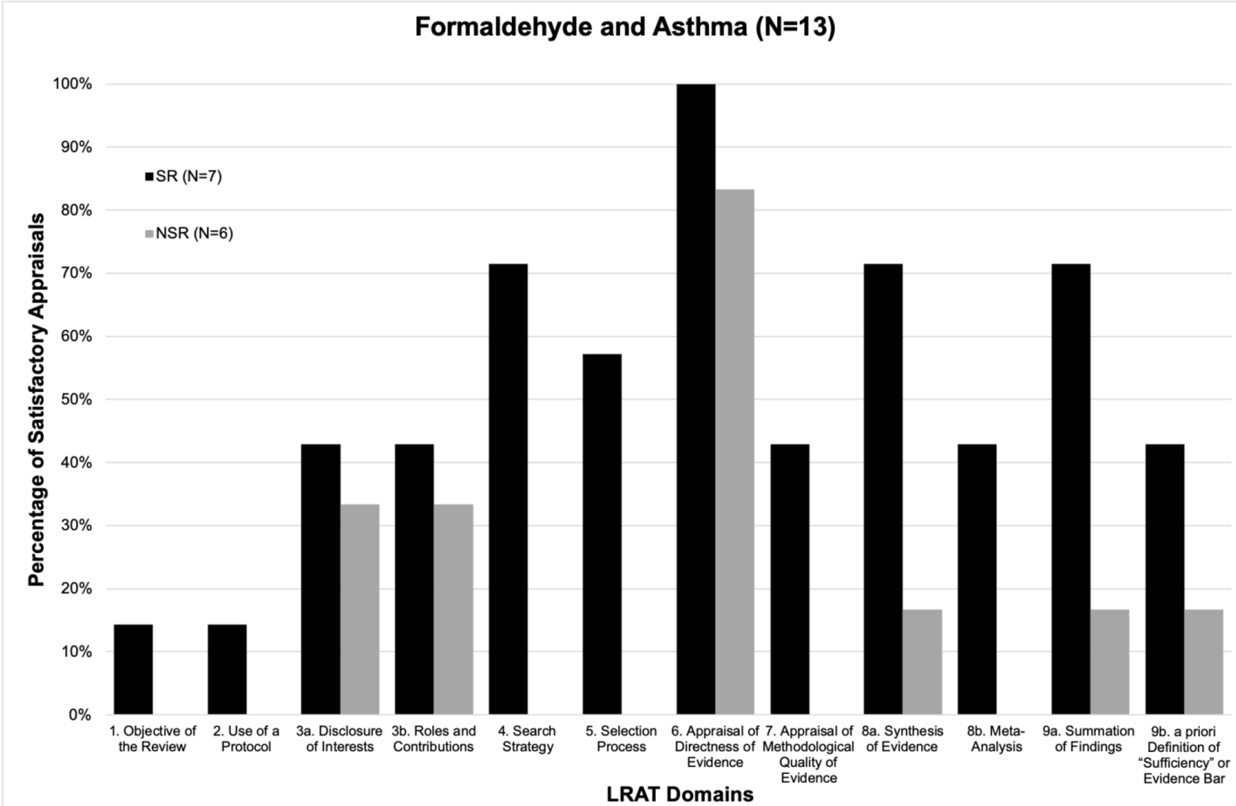

Figure 5.

Adherent Reviews for the Formaldehyde and Asthma Case-Study Topic. Represented is the percentage of systematic and non-systematic reviews receiving a "satisfactory" rating for each LRAT domain that examined the relationship between formaldehyde and Asthma

Table 1.

Description of the domains under which the utility, validity and transparency of a hypothesis-driven literature review can be appraised using the Literature Review Appraisal Toolkit and modifications used in this paper[a]

Domain[b]	Description of the Domain	Clarifications/Revisions
1. The Objective of the review	Does it answer a clear question of sufficient relevance to the controversy on which it is trying to shed light?	Rated as "unsatisfactory" if there was no "PECO" (Population, exposure, comparator, outcome) or comparable statement.
2. The Use of a Protocol	Does the review follow a pre-conceived plan for finding and analyzing evidence relevant to its objective?	Rated as "unsatisfactory" if no protocol was mentioned or if PRISMA or MOOSE checklists were cited as the "protocol."
3a. Disclosure of Interests 3b. Roles and Contributions	Are the interests of the authors of the review, and records of how each author contributed to the review process, sufficient to allow their input into the review to be placed in their academic and societal context?	We modified this domain to capture two important, but separate questions: a) was there an adequate author conflict of interest statement? *Domain name:* 'Disclosures of interests'; and b) were the roles and contributions of each author reported separately (Yes/No)? *Domain name:* 'Roles and contributions'
4. Search Strategy	Did the review locate all the evidence which might have been relevant to the review's objective?	Rated as "satisfactory" if full terms and multiple databases were cited even if the paper lacked a chart or other indication of the numbers of papers retrieved and excluded.
5. Selection Process	Did the review employ a screening process which included for analysis all the studies of actual relevance to the review's objective?	Rated as "unclear" if it was not stated that 2 or more reviewers screened the articles retrieved in the search.
6. Appraisal of Directness of evidence (external validity)	Did the review present and consistently apply a valid scheme for giving greater weight to findings of studies of more direct relevance to the review objective?	Rated as "satisfactory" if only human data were considered.
7. Appraisal of Methodological Quality of Evidence	Did the review present and consistently apply a valid scheme for giving greater weight to findings of studies which were of more robust methodological quality?	Rated as "unclear" if no mention was made of a risk of bias assessment of included papers and "unsatisfactory" if an inadequate risk of bias assessment was explicitly mentioned.
8a. Synthesis of Evidence 8b. Meta-analysis	Did the authors combine, according to a valid methodology, the results, directness and methodological quality of evidence into a statement of what is and is not known regarding the answer to the objective of the review?	We modified this domain to capture two important, but separate questions: a) did the authors combine, according to a valid methodology, the results, directness and methodological quality of evidence into a statement of what is and is not known regarding the answer to the objective of the review? *Domain name:* 'Synthesis of evidence'; and b) did the review authors conduct a meta-analysis to summarize the results before synthesizing the evidence (Yes/No)? *Domain name:* 'Meta-analysis'
9a. Summation of Findings 9b. 'a priori definition of 'sufficiency' or Evidence bar'	Do the concluding and summary sections of the review present a succinct summary of the findings of the review which accurately reflect its material content?	We modified this domain to capture two important, but separate questions: a) did the concluding and summary sections of the review present a succinct summary of the findings of the review which accurately reflect its material content? *Domain name:* 'Summation of findings'; and b) did the review authors include an *a priori* definition of "sufficiency" or other explicit evidence bar for its conclusions Yes/No)? *Domain name:* 'a priori definition of 'sufficiency' or Evidence bar'

a) We used a modified version of the tool to assess the evidence review methods—all modifications are captured in the Clarifications/Revisions column

b) For each LRAT domain there were three possible appraisal ratings: (1) Satisfactory: conducted according to a clear, valid, and consistent procedure; (2) Unclear: insufficient documentation to allow evaluation; or (3) Unsatisfactory: positive evidence of invalid or inconsistent procedure. For the three domains (three, eight and nine) where we adapted the tool, there were two possible ratings for the 'b' question, 'Yes' or 'No.'