



Deposited via The University of Sheffield.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/207947/>

Version: Published Version

Article:

Uttley, L., Quintana, D.S., Montgomery, P. et al. (2023) The problems with systematic reviews: a living systematic review. *Journal of Clinical Epidemiology*, 156. pp. 30-41. ISSN: 0895-4356

<https://doi.org/10.1016/j.jclinepi.2023.01.011>

Reuse

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

REVIEW ARTICLE**The problems with systematic reviews: a living systematic review**

Lesley Uttley^{a,*}, Daniel S. Quintana^{b,c,d}, Paul Montgomery^e, Christopher Carroll^a,
Matthew J. Page^f, Louise Falzon^a, Anthea Sutton^a, David Moher^g

^a*School of Health and Related Research, University of Sheffield, Sheffield, UK*

^b*Department of Psychology, University of Oslo, Oslo, Norway*

^c*NevSom, Department of Rare Disorders, Oslo University Hospital, Oslo, Norway*

^d*Norwegian Centre for Mental Disorders Research (NORMENT) and KG Jebsen Centre for Neurodevelopmental Disorders, University of Oslo, Oslo, Norway*

^e*Department of Social Policy, Sociology and Criminology, University of Birmingham, Birmingham, UK*

^f*Methods in Evidence Synthesis Unit, School of Public Health and Preventive Medicine, Monash University, Melbourne, Australia*

^g*Ottawa Hospital Research Institute, University of Ottawa, Ottawa, Canada*

Accepted 29 January 2023; Published online 14 February 2023

Abstract

Objectives: Systematic reviews and meta-analyses are proliferating as they are an important building block to inform evidence-based guidelines and decision-making. Enforcement of best practice in clinical trials is firmly on the research agenda of good clinical practice, but there is less clarity as to how evidence syntheses that combine these studies can be influenced by bad practice. Our aim was to conduct a living systematic review of articles that highlight flaws in published systematic reviews to formally document and understand these problems.

Study Design and Setting: We conducted a comprehensive assessment of all literature examining problems, which relate to published systematic reviews.

Results: The first iteration of our living systematic review (<https://systematicreviewlution.com/>) has found 485 articles documenting 67 discrete problems relating to the conduct and reporting of systematic reviews which can potentially jeopardize their reliability or validity.

Conclusion: Many hundreds of articles highlight that there are many flaws in the conduct, methods, and reporting of published systematic reviews, despite the existence and frequent application of guidelines. Considering the pivotal role that systematic reviews have in medical decision-making due to having apparently transparent, objective, and replicable processes, a failure to appreciate and regulate problems with these highly cited research designs is a threat to credible science. © 2023 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Keywords: Systematic review; Meta-analysis; Research integrity; Bias; Reproducibility; Evidence syntheses; Influence

1. Background

Systematic reviews are regarded as the gold standard in evidence synthesis. As a key tool in evidence-based medicine, social and education policy, and practice guidelines,

they influence important decisions such as patient care and resource allocation, among others. Consequently poorly conducted and/or reported systematic reviews can have wide-ranging ramifications on societal outcomes. The publication of systematic reviews is proliferating [1] and in some fields outpaces the publication of primary studies [2].

Notwithstanding the laudable purpose of systematic reviews to provide an objective and evidence-based synthesis, increasing research published across many different specialities highlights that systematic review products are commonly not as reliable as their name implies [3]. Systematic reviews are receiving scrutiny and criticism, even from their proponents, for contributing to research waste [4,5], varying massively in their reporting and/or methodological quality [6], and misleading or serving conflicted interests [1]. Although systematic reviews are upheld as

Conflict of interest: None declared.

Funding: The work underpinning the comment is funded by United Kingdom Research and Innovation (UKRI) Medical Research Council (MRC) career development awarded to Lesley Uttley (MR/T009861/1) to investigate the human influences in systematic reviews.

Licensing statement: For the purpose of open access, the author has applied a Creative Commons Attribution (CC BY) license to any Author Accepted Manuscript version arising from this submission.

* Corresponding author. School of Health and Related Research, University of Sheffield, 30 Regent Street, Sheffield S1 4DA, UK. Tel.: +44 0 114 2220782; fax: +44 0 114 2220749.

E-mail address: l.uttley@sheffield.ac.uk (L. Uttley).

<https://doi.org/10.1016/j.jclinepi.2023.01.011>

0895-4356/© 2023 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

What is new?**Key findings**

- A living systematic review finds 67 discrete problems with published systematic reviews in its first iteration.
- Qualitative analysis of the problems described in the 485 included articles relate to key domains of systematic review conduct for being comprehensive, rigorous, transparent and objective.
- This is a methodological systematic review but the ramifications impact patients as systematic reviews are relied upon to guide important healthcare decisions.

What this adds to what was known?

- This research is problem focussed and does not aim to criticise individual researchers or teams.
- This living review includes the infrastructure to incorporate an emerging evidence base.
- Elaboration of the problems, where they are covered by guidelines and the corresponding articles that support them are available via a dedicated website (<https://systematicreviewlution.com/>).

What is the implication and what should change now?

- Many problems of reporting and reporting quality are covered by existing systematic review guidelines but problems of the objectivity and the systematic review team require further scrutiny.
- As a living review we are open to feedback and engagement from the research community and beyond.
- This research aims to be a learning resource to researchers and users of systematic reviews to continually improve the reliability and validity of future systematic reviews.

methodological rigour and reporting of conduct, including the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement [9–14]. However despite the presence of these guidelines and their ostensive enforcement by many scientific publishers, criticisms of systematic review conduct pervade. There are reports by many that a vast quantity of poor-quality reviews are being published in spite of clear guidance being in place for well over a decade [15,16].

Despite a plethora of research from different research teams, disciplines, jurisdictions, and journals raising issues with systematic reviews, the messages about how to improve them are slow to permeate. In the meantime, sub-optimal systematic reviews continue to pass the scrutiny of peer review and editorial oversight and are published [17]. In an era of perverse academic incentives, publication of redundant, overlapping, unreliable, or poor quality systematic reviews are an increasingly unwelcome contributor to the research waste [18].

Scoping searches of literature on this topic indicates hundreds of articles illustrating problems with systematic reviews, such as flaws in statistical analyses or spin (exaggeration) in systematic review conclusions [17,19], but the messages from these individual studies are not being joined up to emphasize the sheer scale of the problem.

To help quantify and understand the vast amount of emerging literature indicating an array of different problems with the way systematic reviews are conducted and reported we have conducted a living systematic review [20] of problems with systematic reviews (<https://systematicreviewlution.com/>). The purpose of this article is to highlight the range and breadth of published literature documenting and examining the problems with systematic reviews. A living model of research enables collation and dissemination of relevant research in an emerging area to researchers globally and across disciplines.

The aim of this research was to comprehensively accumulate articles that have highlighted limitations with published systematic reviews and to thematically characterize the different flaws, limitations, or errors noted in the systematic review process.

2. Methods

An a priori protocol for this study was registered with PROSPERO (CRD42020181371) and posted on Open Science Framework (OSF) (<https://osf.io/2hmv9/>). The protocol was adherent to PRISMA-P [21] and revisions to the protocol were updated to both sources. Methods were informed by the Cochrane Handbook's guidelines for living systematic reviews [22], where appropriate, as a methodological, non-Cochrane systematic review (additional details in Appendix 1).

impartial, dispassionate scientific processes they can themselves introduce bias into the evidence base just as bias can infiltrate primary research. Influential subjective decisions are made by researchers at all stages of a review including question setting, searching, study selection, analysis, interpretation of findings, formulation of recommendations for future research, and in the publication of the review regardless of the findings [7,8]. In light of the importance of rigour and methodological quality required from systematic reviews, several initiatives exist to enable review authors or those appraising systematic reviews to assess

2.1. Eligibility criteria

Included studies of interest were empirical studies published in peer-reviewed academic journals including cross-sectional analyses of systematic reviews; meta-epidemiological studies, which analyze differences that certain characteristics make to systematic reviews, overviews of systematic review quality, survey/questionnaires, or qualitative studies exploring author opinion or practices in systematic reviews.

Nonempirical articles published in peer-reviewed academic journals, such as nonsystematic literature reviews, discussion pieces, and letters to the editor, which highlighted problems with published systematic reviews, were considered a separate category of evidence to empirical studies.

Studies that were ineligible from the review were clinical studies or overview systematic reviews which do not assess research conduct such as methodological or reporting quality of the included systematic reviews; protocols, books, conference proceedings, and PhD theses; good practice guidelines for systematic reviews; guideline validation studies; studies of rapid reviews; methodological studies of articles that may encompass systematic reviews but where inclusion criteria were not specific to systematic reviews and studies which did not distinguish between systematic reviews [23] and meta-analyses.

2.2. Literature searches

A comprehensive literature search strategy was designed in MEDLINE using key known papers by an experienced senior information specialist. This strategy was peer-reviewed by a second experienced senior information specialist (A.S.) using the PRESS checklist [24]. Given that citation-searching of existing known relevant articles was effective for identifying further relevant articles, a simplified version of the MEDLINE search strategy was conducted on the other databases searched. The simplified search combined thesaurus terms and title searches for types of reviews (literature reviews, systematic reviews, or meta-analyses, Cochrane reviews) with generic terms such as “problem”, “error” or “bias” to identify the main issues with systematic review conduct or reporting. The first search iteration of electronic bibliographic databases was completed in May 2022 and updated in November 2022. In both iterations databases were searched from the year 2000 to present, in order to focus on more recent systematic review conduct than historical, in the following databases.

- Ovid MEDLINE(R) and Epub Ahead of Print, In-Process and Other Non-Indexed Citations and Daily <1946 to May 25, 2022>
- Campbell Systematic Reviews journal at: <https://onlinelibrary.wiley.com/journal/18911803>
- Ovid Embase 1974–2022 May 25

- Science Citation Index Expanded (SCI-EXPANDED)—1900-present via Web of Science
- Social Sciences Citation Index—1900-present via Web of Science
- Library and Information Science Abstracts via Proquest 1969-present
- Cochrane Database of Systematic Reviews Issue 5 of 12, May 2022
- Cochrane Central Register of Controlled Trials Issue 5 of 12, May 2022.

All search strategies are provided in [Appendix 2](#). Literature searches for this living review are manually reviewed every 6 months by the project’s information specialist and will be re-implemented every 12 months until December 2024.

2.3. Supplementary search techniques

Bibliographies of all relevant studies were reviewed to identify further relevant citations. Citation searches (forward and backward) of key relevant articles were conducted to identify further relevant references and to assess how relevant studies are indexed. Due to the novelty of this topic, medical subject headings (MeSH) for “systematic reviews” were used to filter relevant articles. Literature searches to inform future iterations of this living review will be refined to incorporate new articles periodically to ensure that emerging relevant findings are included into the review.

2.4. Outcomes

The primary outcome of this review is to determine the range of “problems” (limitations, errors, or flaws) in systematic review conduct or reportings identified from eligible articles. To retrieve these data, articles were systematically coded using thematic analysis to delimit the overarching categories of commonality/differences between systematic review problems highlighted by the authors of the included articles. Systematic review problems of interest relate to their methodological and reporting quality, author conduct and decisions, and characteristics of the systematic review or team that were noted by authors examining in the published systematic reviews.

2.5. Study selection

The combined citations from literature searches were downloaded into reference management software (Endnote version 9) for independent duplicate screening. Titles of all citations were assessed in duplicate and independently (L.U., L.F.). Disagreements were resolved by discussion and consensus with reasons for final decisions of exclusion at full text recorded against citations in the Endnote database using keywords.

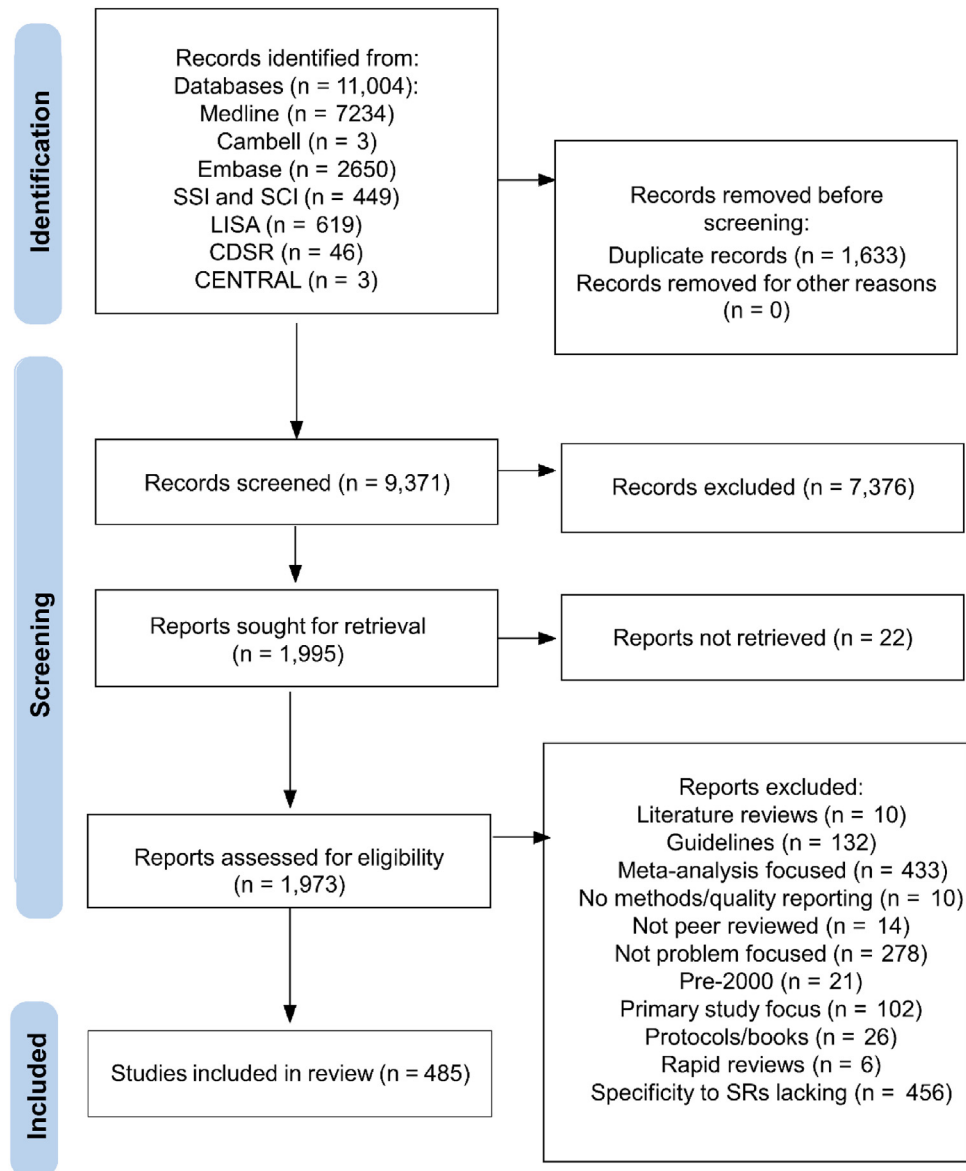


Fig. 1. PRISMA flow diagram of study identification and selection.

2.6. Data extraction

Relevant data from each article were extracted by one reviewer (L.U.) into an online web form that stores bibliographic information about the included references in a customized online relational database. The research software to develop this living systematic review was built de novo to correspond with the dedicated project website. Abstracted data were entered into a database, rather than a spreadsheet or document, which allowed included articles to be coded against multiple problems. A second member of the review team performed independent verification of all data extraction (L.F.). Data extraction items included the following: first author affiliation country, article type (empirical/editorial), full list of authors, article aim, number of citations assessed for eligibility, number of

systematic reviews finally included, a summary of findings as reported in abstracts and discussion sections of the main report, measurement against existing guidelines (e.g., PRISMA, A MeaSurement Tool to Assess systematic Reviews [AMSTAR] and so on.), whether the article examined whether the problem(s) impact the treatment effects of the systematic reviews and whether the article finds that the problem(s) led to changes in the interpretation of the included systematic review(s) results.

2.7. Quality assessment

Methodological studies of existing meta-research do not yet have validated quality appraisal tools. In the absence of a formal dedicated tool to assess the reporting and methodological quality of included studies, the key criteria deemed

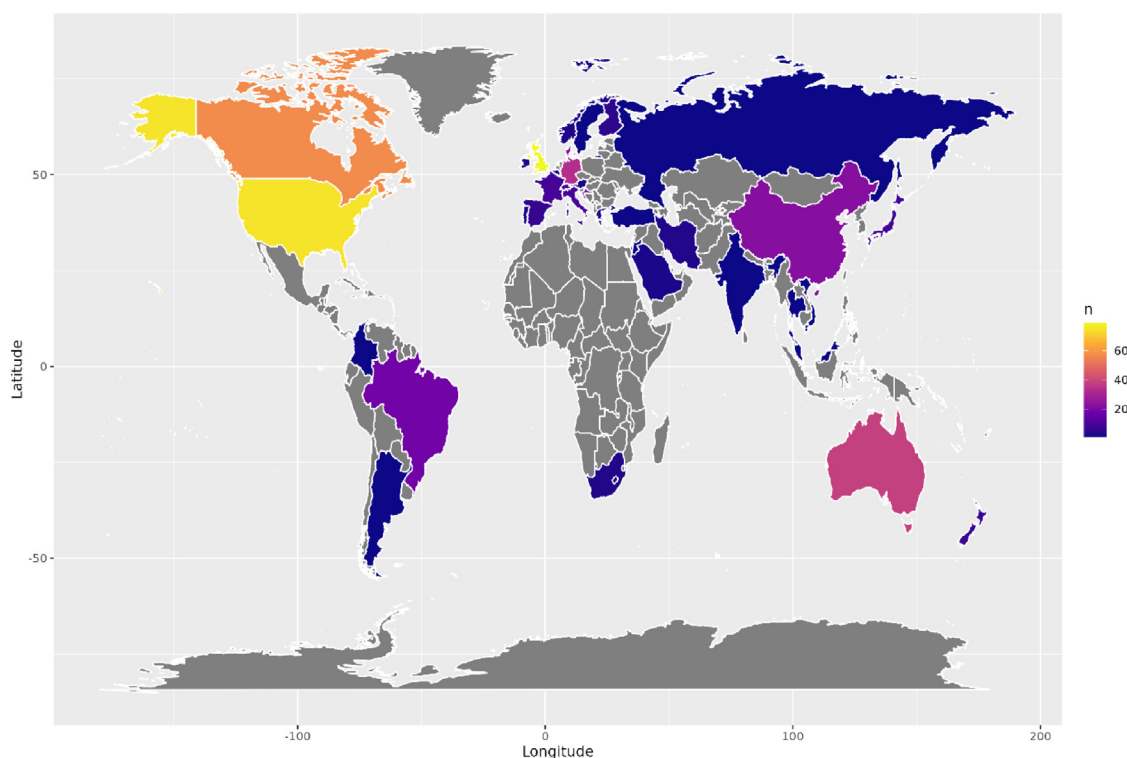


Fig. 2. Choropleth map of included article frequency by country.

relevant to study quality were extracted. Relevant quality assessment criteria extracted in this review relate to availability of a study protocol; review team expertise as assessed by authors' individual roles and institutional discipline; systematic review sampling strategy (comprehensive, random, consecutive, purposeful, and preselected); sample size power (of systematic reviews); statistical adjustments for confounding and whether the methods of the study are described in enough detail to allow replication (based on provision of searches strategies and references lists of included studies). Evaluation of the quality of nonempirical articles was also performed where quality criteria were applicable, for example in editorial articles that attempted to provide empirical data, often published as "research letters". Quality assessment was performed firstly by one reviewer (L.U.). A second member of the review team (L.F.) performed independent verification of a proportion (20%) of all quality assessment, selected at random.

2.8. Data synthesis

Included articles were reviewed in full and categorized to describe the systematic review problem/s highlighted, using thematic analysis. One reviewer generated themes/categories from the primary findings of the eligible articles inductively based on the principles of systematic review conduct to generate an initial thematic framework. A

deductive approach was then used to code all the included articles against the themes identified into the framework for conduct of systematic reviews according to reference standard guidelines [15]. Core elements representing hallmark systematic review conduct [9,10,12,13,15,25] were used as a guide to form the framework, but the thematic synthesis also proposes new themes that are not covered by current systematic review guidelines.

The qualitative summaries of the expressed themes are considered the focus of this review, rather than the quantitative distribution of the articles within themes. Nonetheless, the number of articles corresponding to each problem are recorded and data relating to the citation and source are provided within the living review.

3. Results

The yield from bibliographic database and supplementary searches were 9,371 after duplicate citations were removed. Full text assessment was performed on 1,973 articles. Four hundred and eighty-five relevant articles were included in the first iteration of this living systematic review. Bibliography checking of all included studies yielded no further eligible citations from the search period that were not retrieved by the database searches. The results of literature searching and the study selection process is shown in Figure 1.

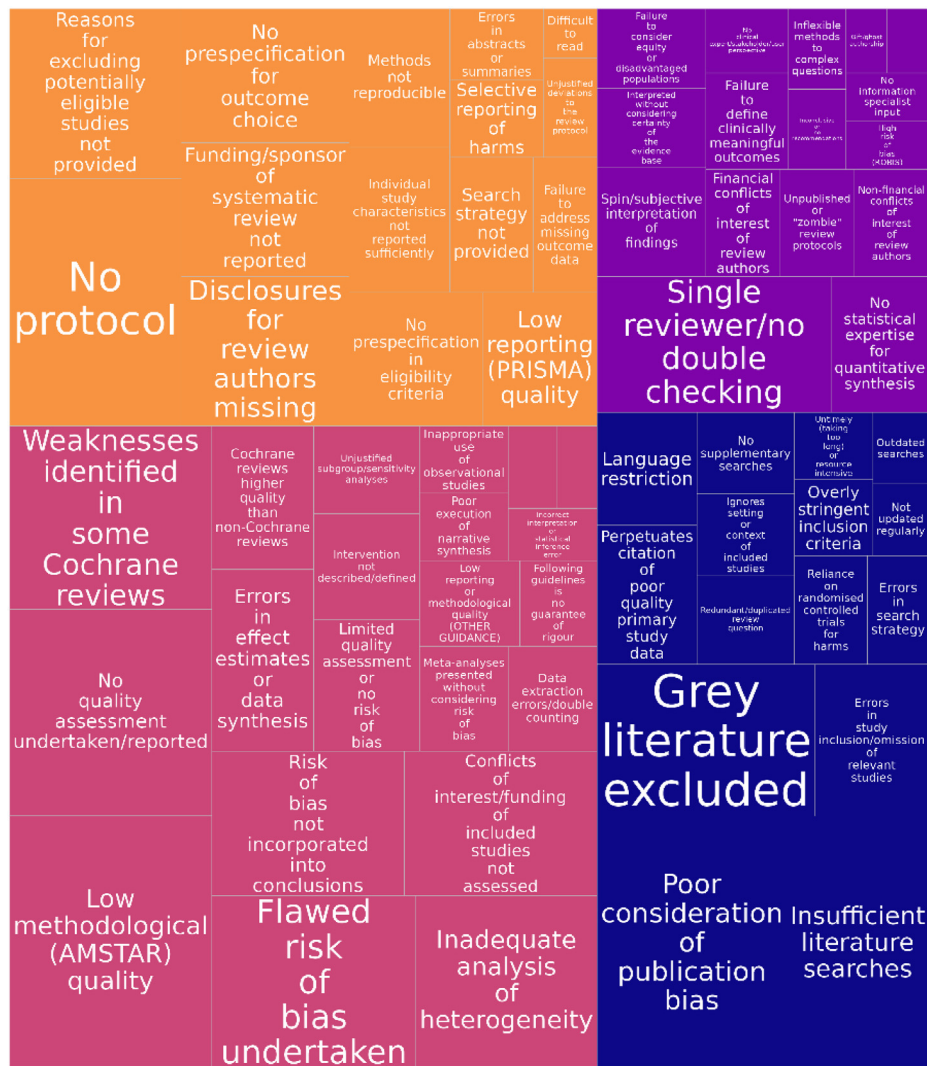


Fig. 3. Treemap of the range and frequency of problems with systematic reviews identified by domain.

The 485 included studies in the first iteration of this living systematic review were published between the years 2000 and 2021, across 314 academic journals. 410 included articles were empirical research studies, 68 were editorial articles (discussion pieces or letters to the editor) and seven were nonsystematic literature reviews. Included articles originated from 48 different countries. The geographical spread and frequency of published included articles in this review are depicted in Figure 2.

Assessment of methodological study quality of the included articles found that 55/485 (11%) had a publicly available protocol; 121/485 (25%) were performed by a multidisciplinary team; 385/485 (79%) used a fair sampling strategy; 317/485 (65%) used duplicate study assessment (two researchers) in study selection or data extraction; 15/485 (3%) performed a power calculation to obtain the sample size of systematic reviews; 49/485 (10%) performed a statistical adjustment for confounding; and 307/485 (63%)

were reported in a manner to allow the results to be replicated. Data extraction found that 67/485 (14%) of included articles attempted reanalysis of the results from the included sample of systematic reviews. Furthermore 58/485 (12%) of the included articles found that such reanalysis to lead to a qualitative change in the conclusions of the systematic reviews.

Following thematic analysis of the 485 included articles, the number of problems, limitations or flaws with published systematic reviews that were deemed to be thematically discrete is 67. Figure 3 is a treemap of the 67 problems, with the size of each cell representing the relative number of included articles citing that problem.

Problems were categorized to correspond to core elements of systematic review conduct to build a theoretical framework. Four overarching domains were derived representing hallmark characteristics of systematic review conduct.

Table 1. Comprehensive: Systematic review problems relating to the evidence included and number of corresponding articles

Problems under the “comprehensive” domain of systematic review conduct	No. of articles
Overly stringent inclusion criteria affecting external validity	9
Errors or omissions in search strategy	10
Grey literature excluded	50
Insufficient literature searches	54
Lack of supplementary searches beyond databases	12
Language restriction	17
Outdated searches	6
Errors in study inclusion or omission of relevant studies	26
Ignores setting or context of included studies which limits review applicability	12
Perpetuates citation of poor quality primary study data	21
Reliance on randomized controlled trials for harms/safety data	12
Poor consideration of publication bias	72
Not updated regularly	7
Redundant/overlapping/duplicated review question; leads to research waste	13
Untimely (taking too long) or resource intensive	8

- (i) comprehensive (i.e., including all relevant evidence);
- (ii) rigorous (i.e., using appropriate methods);
- (iii) transparent (i.e., enabling reproducibility) and
- (iv) objective (i.e., handling the review process fairly)

The problems under each domain are described herein. Full information on the articles included and the main findings of each article are available at (www.systematicreviewlution.com).

3.1. Domain 1: comprehensive

A systematic review should aim to be inclusive of all relevant studies. Problems occur when relevant studies are missed or ignored, and this can limit the systematic review's validity. Problems regarding the included systematic reviews' comprehensiveness are documented in Table 1.

3.2. Domain 2: rigorous

A systematic review uses appropriate methods to ensure it is methodologically sound. Errors in the approach to conducting the review or lack of expertise can jeopardize the internal validity of the systematic review. Problems regarding the included systematic reviews' rigour are documented in Table 2.

Table 2. Rigorous: Systematic review problems relating to the methodological conduct and number of corresponding articles

Problems under the “rigorous” domain of systematic review conduct	No. of articles
Intervention not described/defined	18
Data extraction errors and double counting	14
Inclusion of observational/nonrandomized studies	10
Conflicts of interest or funding of included studies not assessed	42
Flawed risk of bias undertaken	53
Limited quality assessment or no risk of bias	21
Meta-analyses and forest plots presented without considering risk of bias/quality	14
No quality assessment undertaken or reported	63
Errors in effect estimate calculations or data synthesis	28
Inadequate analysis of heterogeneity	47
Incorrect interpretation or statistical inference error from meta-analysis	7
Poor execution of narrative synthesis	9
Risk of bias not incorporated into conclusions of review	42
Small number of trials in meta-analyses	5
Unplanned or unjustified subgroup or sensitivity analyses	14
Cochrane reviews more rigorous/higher quality than non-Cochrane reviews	22
Following guidelines is no guarantee of a rigorous systematic review	10
Lack of guidance or consistency in systematic overview/umbrella/review of systematic reviews	6
Low methodological (AMSTAR) quality	77
Low reporting or methodological quality (OTHER GUIDANCE)	14
Weaknesses identified in some Cochrane reviews	56

3.3. Domain 3: transparent

A systematic review should report its methods with enough detail to enable replication. If the reporting quality of a review is poor or the full methods are inaccessible, then the systematic review is not reproducible. Problems regarding the included systematic reviews' transparency are documented in Table 3.

3.4. Domain 4: objective

A systematic review should be conducted fairly, by teams with appropriate expertise and not by people with vested interests. If the review is done in a perfunctory way by people who do not care about the validity of the conclusion, then the systematic review may not be reliable. Problems regarding the included systematic reviews' objectivity are documented in Table 4.

The number of articles relating to each problem, by systematic review domain are depicted in the below treemap.

Table 3. Transparent: systematic review problems relating to the reproducibility and number of corresponding articles

Problems under the “transparent” domain of systematic review conduct	No. of articles
Lack of prespecification in eligibility criteria	27
Methods not described to enable replication	21
Multiplicity of outcomes and lack of prespecification for outcome reporting	34
No registered or published protocol	64
Undocumented or unjustified deviations to the review protocol	8
Conflict of interest statement or disclosures for review authors missing	38
Funding or sponsor of systematic review not reported	34
Search strategy not provided	17
Individual study characteristics not reported sufficiently	22
Reasons for excluding potentially eligible studies not provided	44
Selective reporting of harms/safety/adverse events/side effects	11
Failure to address missing outcome data in analyses	13
Errors in systematic review abstracts or plain language summaries	10
Unwieldy/difficult to read	4
Low reporting (PRISMA) quality	23

The average number of systematic reviews included within each article is 298 and the range is between 1 systematic review to 18, 959 systematic reviews. These articles include detailed analyses of several flaws in the conduct of one influential systematic review, to rigorous analyses of thousands of Cochrane reviews. Figure 4 shows the range and ratio of the number of citations that were screened for eligibility by the 485 included articles to the number of systematic reviews ultimately included by each article. A higher ratio (lighter colour) indicates studies which assessed many citations before selecting the final sample of systematic reviews. A smaller ratio (darkest colour) indicates articles which included a sample of systematic reviews without screening them for particular characteristics.

Due to the heterogeneity in the way that samples of systematic reviews across the included articles were obtained (comprehensive bibliographic searches; random, consecutive, or purposive samples) the prevalence of problems across systematic reviews cannot currently be reliably estimated. However keywords were assigned to every included article with regards to both topic, type, and subject area by two reviewers (L.U. and L.F.) provide a tentative understanding of prevalence of themes such as scientific discipline or method across the included articles (Fig. 5). Larger word size in the figure represents higher frequency of associated keywords across the included articles to

Table 4. Objective: Systematic review problems relating to the fairness of conduct and number of corresponding articles

Problems under the “objective” domain of systematic review conduct	No. of articles
Failure to consider equity, different socioeconomic groups, or disadvantaged populations	13
Failure to define clinically meaningful outcomes	12
Inflexible methods to complex questions	7
Lack of clinical expert/stakeholder/user perspective	8
Review question not justified/important	4
Literature searches not validated by information specialist	6
Single reviewer/lack of double checking	48
Lack of statistical expertise in handling of quantitative data	20
Spin or subjective interpretation of findings	18
Inconclusive or lack of recommendations	7
Interpreted without considering certainty or overall quality of the evidence base	13
Unpublished or “zombie” reviews (the file-drawer effect)	12
Financial conflicts of interest of review authors	12
Guest/gift/ghost authorship	5
High risk of bias (ROBIS)	6
Nonfinancial conflicts of interest of review authors	12

represent how frequently problematic systematic reviews are being documented across the literature.

4. Discussion

This research has so far found that between the years 2000 and November 2022, at least 485 articles have been published which document problems with published systematic reviews. The nature of the included articles ranges between editorials highlighting concerns over the conduct or reporting of one systematic review to prespecified rigorous analyses of issues with hundreds, and sometimes thousands, of systematic reviews. This living review accumulates and organizes the many limitations of published systematic reviews. Presently, we recommend that researchers who use, conduct, or appraise systematic reviews fully digest the range and nature of the problems discovered to reflect on areas of improvement for their own current and future work. The next step for this research includes undertaking a prioritization exercise to better understand which problems are potentially fatal flaws for systematic reviews and to propose solutions. We propose using the reviews website (<https://systematicreviewlution.com/>) to learn about the problems, whether and where they are covered by existing guidelines and to explore the examples of studies that have highlighted these problems. In this respect we hope that the review will become a learning resource for those who work with systematic reviews to

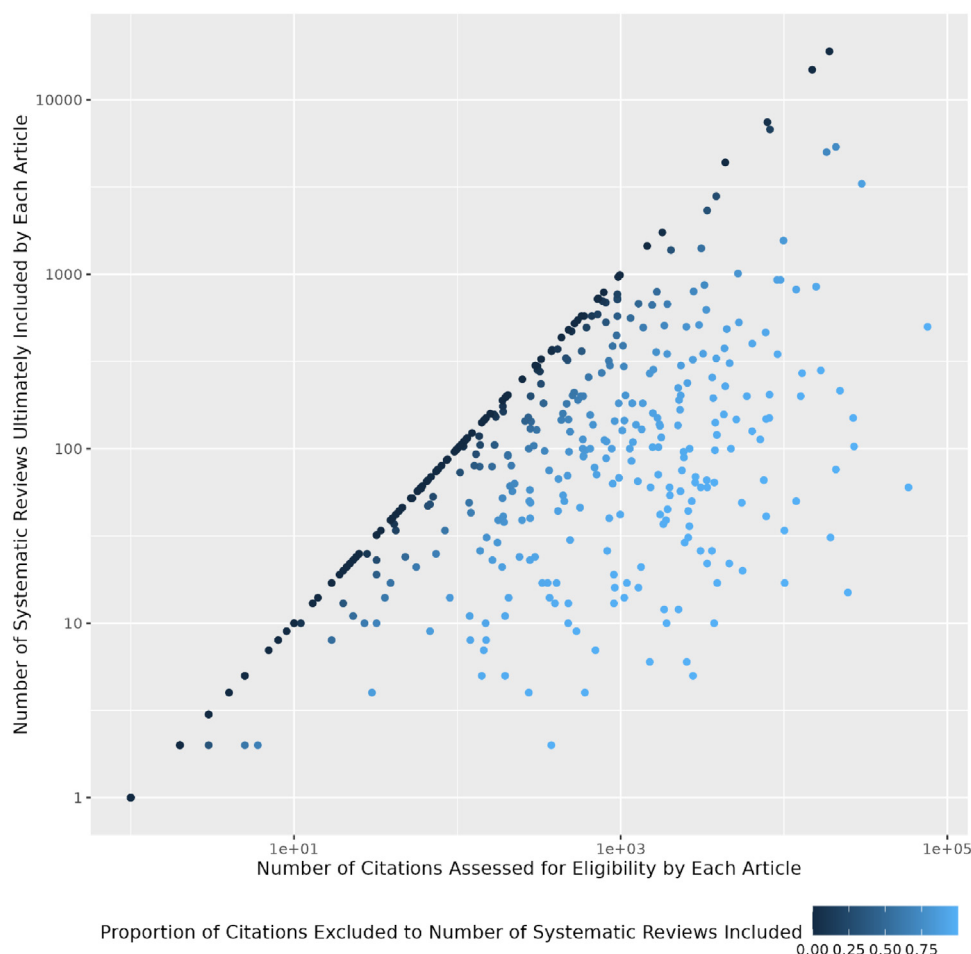


Fig. 4. Scatterplot of number of eligible/included systematic reviews in each included article.

understand the importance of adhering to the best practice methods and where guidelines do not cover all the issues.

4.1. Solutions to some of the problems identified, by domain, include

4.1.1. Comprehensive

Involvement of an experienced information specialist when designing and implementing a literature search is the key to ensuring a systematic review is inclusive of all relevant evidence. Employ reflexivity in the search process and revisit searches to implement changes when further relevant articles are identified through supplementary techniques that were not picked up by the initial search strategy. Consider in advance restrictions to the inclusion criteria that may potentially limit the applicability of the review conclusions with regards to setting or underrepresented groups.

4.1.2. Rigorous

Adhere to the best practice methodological guidelines (e.g., MECIR) wherever possible to aim for an internally valid systematic review. Have appropriate statistical

guidance in reviews where quantitative data are combined. Frontload effort in protocol development to prespecify definitions, outcomes and units of measurement, anticipating multiplicity and confounding. Plan for how heterogeneity will be handled, rather than ignoring or excluding heterogeneous data at the synthesis stage to avoid downgrading of evidence and thus better uptake of the findings in due course by policymakers and practitioners.

4.1.3. Transparent

Adhere to reporting guidelines (e.g., PRISMA), or guideline extensions for nontraditional systematic review designs to ensure reproducible outputs. Write a priori protocols and make them publicly accessible either through the International Prospective Register of Systematic Reviews (PROSPERO) registration or through other platforms such as protocols.io or OSF. Update protocol records with any deviations to intended methods and publish review outputs regardless of direction of findings. Following publication, make data publicly available using existing data sharing platforms (e.g., OSF, Figshare and GitHub) to facilitate replicability.

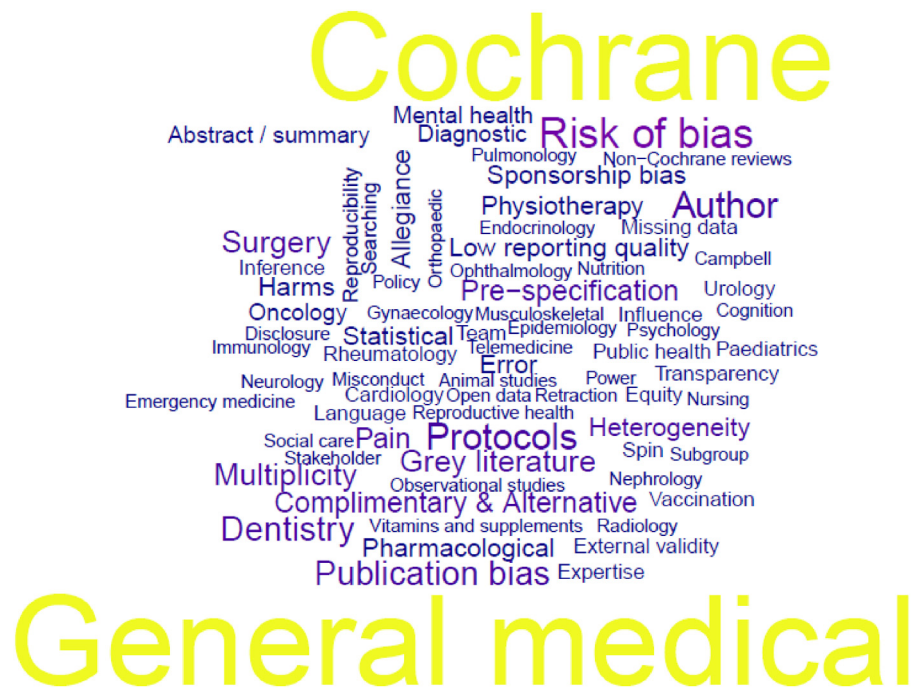


Fig. 5. Wordcloud of frequently Ascribed keywords for included articles indicating problems with published systematic reviews.

4.1.4. Objective

A suitably skilled team that includes input from both methodological and content experts is essential to producing a reliable systematic review. Confirm that research questions are important and necessary, and that outcomes are clinically meaningful to patients through public involvement to improve the external validity of reviews. Double-check data and consult statistical experts to make sure inferences made from results are accurate. Interpret results of studies in terms of quality, risk of bias, and certainty to ensure findings are couched with appropriate caution. In addition to full disclosure of all potential vested interests of review authors, management of potential researcher allegiance or sponsorship bias is crucial to ensuring that conflicted individuals are not making critical decisions that could influence systematic review conclusions.

As this is a living review we are open to feedback from the research community on the nature and structure of the analysis. Future literature searches for this review will be broadened to include contacting of systematic review-specific editorial teams and searching Cochrane colloquia. The overall aim is to highlight which specific problems pose the greatest risk to the reliability and validity of systematic review conclusions.

As many included articles in this review focus on highlighting adherence to guidelines, such as PRISMA, AMSTAR, and MECIR, there is a degree of ascertainment bias where the problems highlighted by included literature are largely guided by what these key guidelines represent as

being important. In this sense some lesser covered areas which do not feature heavily on established versions of the guidelines, such as “failure to define clinically meaningful outcomes” feature less in this review. For this reason, it should be stressed that quantity of research noted under one problem is not necessarily synonymous with importance for systematic review conduct. As established systematic review guidelines are updated, such as the PRISMA reporting guideline and as extensions to guideline are added, it is likely that the limitations of systematic reviews will expand accordingly. As the body of included literature was substantial, some important distinctions were necessary to help ensure the internal validity of the included sample. Included studies had to examine articles which claimed to be systematic reviews and were published as such. In this context studies of meta-analyses that were not conducted in the context of a Cochrane review or a systematic review were excluded. The findings of this review are applicable to meta-analyses, but it is our assertion that meta-analyses as a statistical technique should only be published if they claim to have been conducted in the context of a systematic review.

Systematic reviews are still the best available tool for examining and synthesizing evidence from multiple sources. But apparent adherence to guidelines may not guarantee research quality. This work can empower researchers to better identify the limitations of systematic reviews during peer review, critical appraisal, and to learn how to strive for best conduct in their own future systematic reviews. This living review offers an innovative alternative

to publication at one time point to highlight problems and expecting the messages of caution to be heeded by researchers, editors, and decision makers. In an era of unprecedented public interest as to how evidence is collected, examined, and presented, the messages from this research are relevant for a wide audience, including patients. We all have a stake in the reliability of systematic reviews, and we must move to improve how they are conducted and published.

4.2. Limitations

This review included only articles which pertain to systematic reviews and therefore articles examining meta-analyses which were not conducted in the context of a systematic review were excluded. Additionally, if an article was not clearly identified by title or abstract as highlighting problems with systematic reviews, it may not be detected during the screening process. The definition of “problematic” systematic reviews may be regarded as subjective in the included papers. This review focusses on the problems of systematic reviews as outputs, rather than the efficiency of the systematic review process, which may be addressed by rapid review products. For reporting or methodological guideline adherence we have imposed a relatively arbitrary threshold of <85% compliance to denote less than ideal conduct. The literature searches were limited to studies from year 2000 onward which was an arbitrary cutoff point before which there were fewer consolidated standards for systematic review conduct and reporting. Grey literature in the form of books and PhD theses were excluded from this review for feasibility purposes due to the copious amount of published data. A potential research recommendation could include focussing on unpublished data which may examine the effects of publication bias on this review and provide access to even more comprehensive evaluations of systematic review problems.

5. Conclusions

This living research, conducted independently to any systematic review body or guideline, will continue to evaluate how medical evidence is scrutinized by joining together research which highlights that published systematic reviews may not be as reliable and valid as their name suggests. The objective of this research is to continually improve the conduct and quality of systematic reviews.

Acknowledgments

The authors wish to thank Sarah Dawson, the University of Bristol for her help in designing the initial literature search strategy and Mark Dunning, the University of Sheffield, for help generating figures using data from the review.

Figures were produced in *R* ggplot2 (version 3.3.6), wordcloud (version 2.6) and treemapify (version 2.5.5).

Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jclinepi.2023.01.011>.

References

- [1] Ioannidis J. The mass production of redundant, misleading, and conflicted systematic reviews and meta-analyses. *Milbank Q* 2016; 94(3):485–514.
- [2] Niforatos JD, Chaitoff A, Weaver M, Feinstein MM, Johansen ME. Pediatric literature shift: growth of meta-analyses was 23 times greater than growth of randomized trials. *J Clin Epidemiol* 2020; 121:112–4.
- [3] Moore RA, Fisher E, Eccleston C. Systematic reviews do not (yet) represent the ‘gold standard’ of evidence: a position paper. *Eur J Pain* 2022;26:557–66.
- [4] Siontis KC, Ioannidis JPA. Replication, duplication, and waste in a quarter million systematic reviews and meta-analyses. *Circ Cardiovasc Qual Outcomes* 2018;11(12):e005212.
- [5] Chalmers I, Glasziou P. Systematic reviews and research waste. *Lancet* 2016;387:122–3.
- [6] Page MJ, Altman DG, Shamseer L, McKenzie JE, Ahmadzai N, Wolfe D, et al. Reproducible research practices are underused in systematic reviews of biomedical interventions. *J Clin Epidemiol* 2018; 94:8–18.
- [7] Uttley L, Montgomery P. The influence of the team in conducting a systematic review. *Syst Rev* 2017;6(1):149.
- [8] Kirkham JJ, Altman DG, Williamson PR. Bias due to changes in specified outcomes during the systematic review process. *PLoS One* 2010;5:e9810.
- [9] Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Ann Intern Med* 2009;151:264–9.
- [10] Whiting P, Savović J, Higgins JP, Caldwell DM, Reeves BC, Shea B, et al. ROBIS: a new tool to assess risk of bias in systematic reviews was developed. *J Clin Epidemiol* 2016;69:225–34.
- [11] Higgins JPT, Lane PW, Anagnostelis B, Anzuere-Cabrera J, Baker NF, Cappelleri JC, et al. A tool to assess the quality of a meta-analysis. *Res Synth Methods* 2013;4(4):351–66.
- [12] Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ* 2017;358:j4008.
- [13] Chandler J, Churchill R, Higgins J, Lasserson T, Tovey D. Methodological standards for the conduct of new Cochrane Intervention Reviews. London, UK: Cochrane Collaboration; 2013.
- [14] Churchill C, Higgins L, Tovey. Methodological expectations of Campbell collaboration intervention reviews (MEC2IR): DRAFT version 0.1, June 23, 2014 (updated September 2014). Adaptations on MECIR version 2.2 conduct standards. The Campbell collaboration. 2014. 2017. Available at <https://community.cochrane.org/sites/default/files/uploads/MECIR%20PRINTED%20BOOKLET%20FINAL%20v1.01.pdf>. Accessed February 3, 2023.
- [15] Page MJ, Moher D, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. *BMJ* 2021;372:n160.
- [16] Page MJ, Moher D. Mass production of systematic reviews and meta-analyses: an exercise in mega-silliness? *Milbank Q* 2016;94(3): 515–9.

- [17] Page MJ, Altman DG, McKenzie JE, Shamseer L, Ahmadzai N, Wolfe D, et al. Flaws in the application and interpretation of statistical analyses in systematic reviews of therapeutic interventions were common: a cross-sectional analysis. *J Clin Epidemiol* 2018; 95:7–18.
- [18] Roberts I, Ker K. How systematic reviews cause research waste. *Lancet* 2015;386:1536.
- [19] Imberger G, Thorlund K, Gluud C, Wetterslev J. False-positive findings in Cochrane meta-analyses with and without application of trial sequential analysis: an empirical review. *BMJ Open* 2016;6(8): e011890.
- [20] Elliott JH, Synnot A, Turner T, Simmonds M, Akl EA, McDonald S, et al. Living systematic review: 1. Introduction—the why, what, when, and how. *J Clin Epidemiol* 2017;91:23–30.
- [21] Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst Rev* 2015; 4(1):1.
- [22] Higgins JP, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al. *Cochrane handbook for systematic reviews of interventions*. Hoboken, NJ: John Wiley & Sons; 2019.
- [23] McKinley C. Library research guides: systematic review resources: what is a systematic review: temple university libraries. 2019. Available at <https://guides.temple.edu/c.php?g=78618&p=4178713>. Accessed February 3, 2023.
- [24] McGowan J, Sampson M, Salzwedel DM, Cogo E, Foerster V, Lefebvre C. PRESS peer review of electronic search strategies: 2015 guideline statement. *J Clin Epidemiol* 2016;75:40–6.
- [25] Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. *BMC Med Res Methodol* 2007;7:10.