

PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews

Matthew J Page, senior research fellow^{1,*}, David Moher, director and professor^{2,†}, Patrick M Bossuyt, professor³, Isabelle Boutron, professor⁴, Tammy C Hoffmann, professor⁵, Cynthia D Mulrow, professor⁶, Larissa Shamseer, doctoral student⁷, Jennifer M Tetzlaff, research product specialist⁸, Elie A Akl, professor⁹, Sue E Brennan, senior research fellow¹, Roger Chou, professor¹⁰, Julie Glanville, associate director¹¹, Jeremy M Grimshaw, professor¹², Asbjørn Hróbjartsson, professor¹³, Manoj M Lalu, associate scientist and assistant professor¹⁴, Tianjing Li, associate professor¹⁵, Elizabeth W Loder, professor¹⁶, Evan Mayo-Wilson, associate professor¹⁷, Steve McDonald, senior research fellow¹, Luke A McGuinness, research associate¹⁸, Lesley A Stewart, professor and director¹⁹, James Thomas, professor²⁰, Andrea C Tricco, scientist and associate professor²¹, Vivian A Welch, associate professor²², Penny Whiting, associate professor¹⁸, Joanne E McKenzie, associate professor^{1,†}

1. School of Public Health and Preventive Medicine, Monash University, Melbourne, Australia
2. Centre for Journalology, Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, Canada; School of Epidemiology and Public Health, Faculty of Medicine, University of Ottawa, Ottawa, Canada
3. Department of Clinical Epidemiology, Biostatistics and Bioinformatics, Amsterdam University Medical Centres, University of Amsterdam, Amsterdam, Netherlands
4. Université de Paris, Centre of Epidemiology and Statistics (CRESS), Inserm, F 75004 Paris, France
5. Institute for Evidence-Based Healthcare, Faculty of Health Sciences and Medicine, Bond University, Gold Coast, Australia
6. University of Texas Health Science Center at San Antonio, San Antonio, Texas, United States; *Annals of Internal Medicine*
7. School of Epidemiology and Public Health, Faculty of Medicine, University of Ottawa, Ottawa, Canada
8. Evidence Partners, Ottawa, Canada
9. Clinical Research Institute, American University of Beirut, Beirut, Lebanon; Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, Ontario, Canada
10. Department of Medical Informatics and Clinical Epidemiology, Oregon Health & Science University, Portland, Oregon, United States
11. York Health Economics Consortium (YHEC Ltd), University of York, York, UK

12. Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, Canada; School of Epidemiology and Public Health, University of Ottawa, Ottawa, Canada; Department of Medicine, University of Ottawa, Ottawa, Canada
13. Centre for Evidence-Based Medicine Odense, Odense University Hospital, Odense, Denmark; Department of Clinical Research, University of Southern Denmark, Odense, Denmark; Open Patient data Explorative Network, Odense University Hospital, Odense, Denmark
14. Department of Anesthesiology and Pain Medicine, The Ottawa Hospital, Ottawa, Canada; Clinical Epidemiology Program, Blueprint Translational Research Group, Ottawa Hospital Research Institute, Ottawa, Canada; Regenerative Medicine Program, Ottawa Hospital Research Institute, Ottawa, Canada
15. Department of Ophthalmology, School of Medicine, University of Colorado Denver, Denver, Colorado, United States; Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, United States
16. Division of Headache, Department of Neurology, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, United States; Head of Research, The BMJ, London, UK
17. Department of Epidemiology and Biostatistics, Indiana University School of Public Health-Bloomington, Bloomington, Indiana, United States
18. Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK
19. Centre for Reviews and Dissemination, University of York, York, UK
20. EPPI-Centre, UCL Social Research Institute, University College London, London, UK
21. Li Ka Shing Knowledge Institute of St. Michael's Hospital, Unity Health Toronto, Toronto, Canada; Epidemiology Division of the Dalla Lana School of Public Health and the Institute of Health Management, Policy, and Evaluation, University of Toronto, Toronto, Canada; Queen's Collaboration for Health Care Quality Joanna Briggs Institute Centre of Excellence, Queen's University, Kingston, Canada
22. Methods Centre, Bruyère Research Institute, Ottawa, Ontario, Canada; School of Epidemiology and Public Health, Faculty of Medicine, University of Ottawa, Ottawa, Canada

***Correspondence to:** Dr. Matthew Page, School of Public Health and Preventive Medicine, Monash University, 553 St Kilda Road, Melbourne, Victoria, 3004, Australia. Telephone: +61 9903 0248. Email address: matthew.page@monash.edu

†DM and JEM are joint senior authors.

Dedication: We dedicate this paper to the late Douglas G Altman and Alessandro Liberati, whose contributions were fundamental to the development and implementation of the original PRISMA statement.

ABSTRACT

The methods and results of systematic reviews should be reported in sufficient detail to allow users to assess the trustworthiness and applicability of the review findings. The Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) statement was developed to facilitate transparent and complete reporting of systematic reviews and has been updated (to PRISMA 2020) to reflect recent advances in systematic review methodology and terminology. Here, we present the explanation and elaboration paper for PRISMA 2020, where we explain why reporting of each item is recommended, present bullet points that detail the reporting recommendations, and present exemplars from published reviews. We hope that changes to the content and structure of PRISMA 2020 will facilitate uptake of the guideline and lead to more transparent, complete and accurate reporting of systematic reviews.

INTRODUCTION

Systematic reviews are essential for health care providers, policy makers and other decision makers, who without them, would be confronted by an overwhelming volume of research on which to base their decisions. To allow decision makers to assess the trustworthiness and applicability of review findings, reports of systematic reviews should be transparent and complete. Furthermore, such reporting should allow others to replicate or update reviews. The Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) statement published in 2009 (hereafter referred to as PRISMA 2009) (1, 2) was designed to help authors prepare transparent accounts of their reviews, and its recommendations have been widely endorsed and adopted (3). We have updated the PRISMA 2009 statement (to PRISMA 2020) to ensure currency and relevance and to reflect advances in systematic review methodology and terminology.

Scope of this guideline

The PRISMA 2020 statement has been designed primarily for systematic reviews of studies that evaluate the effects of health interventions, irrespective of the design of the included studies. However, the checklist items are applicable to reports of systematic reviews evaluating other non-health-related interventions (e.g. social or educational interventions), and many items are applicable to systematic reviews with objectives other than evaluating interventions (e.g. evaluating aetiology, prevalence or prognosis). PRISMA 2020 is intended for use in systematic reviews that include synthesis (e.g. pairwise meta-analysis, or other statistical synthesis methods), or do not include synthesis (e.g. because only one eligible study is identified). The PRISMA 2020 items are relevant for mixed-methods systematic reviews (which include quantitative and qualitative studies), but reporting guidelines addressing the presentation and synthesis of qualitative data should also be consulted (4, 5). PRISMA 2020 can be used for original systematic reviews, updated systematic reviews, or continually updated (“living”) systematic reviews. Extensions to the PRISMA 2009 statement have been developed to guide reporting of network meta-analyses (6), meta-analyses of individual participant data (7), systematic reviews of harms (8), systematic reviews of diagnostic test accuracy studies (9) and scoping reviews (10); for these types of reviews we recommend authors report their review in accordance with the recommendations in PRISMA 2020 along with the guidance specific to the extension.

PRISMA 2020 explanation and elaboration

PRISMA 2020 is published as a suite of three papers: a statement paper, consisting of the 27-item checklist, an expanded checklist that details reporting recommendations for each item, the PRISMA 2020 abstract checklist, and the revised flow diagram (11); a development paper, which outlines the

steps taken to update the PRISMA 2009 statement, and provides rationale for modifications to the original items (12); and this paper, the updated explanation and elaboration for PRISMA 2020. In this paper, for each item, we explain why reporting of the item is recommended, and present bullet points that detail the reporting recommendations. This structure is new to PRISMA 2020 and has been adopted to facilitate implementation of the guidance (13, 14). Authors familiar with PRISMA 2020 may opt to use the standalone statement paper (11); however, for those who are new to, or not familiar with, PRISMA 2020, we encourage use of this explanation and elaboration document. Box 1 includes a glossary of terms used throughout the PRISMA 2020 explanation and elaboration paper.

We use standardised language in the explanation and elaboration to indicate whether the reporting recommendations for each item (which we refer to as ‘elements’ throughout) are *essential* or *additional*. Essential elements should be reported in all systematic reviews (except for those preceded by “If...”, which should only be reported where applicable). These have been selected as essential because we consider their reporting important for users to assess the trustworthiness and applicability of a review’s findings, or their reporting would aid in reproducing the findings. Additional elements are those which are not essential but provide supplementary information that may enhance the completeness and usability of systematic review reports. The essential and additional elements are framed in terms of reporting the ‘presence’ of a method/result (e.g. reporting if individuals were contacted to identify studies), rather than reporting on their absence. In some instances, however, reporting the absence of a method may be helpful (e.g. “we did not contact individuals to identify studies”). We leave these decisions to the judgement of authors.

We sought examples of good reporting for each checklist item from published systematic reviews and present one for each item below; more examples are available in Table S1 in the Supplement. We have edited the examples by removing all citations within them (to avoid potential confusion with the citation for each example) and we spelled out abbreviations to aid comprehension. We encourage readers to submit evidence that informs any of the recommendations in PRISMA 2020 and any exemplars that could be added to our bank of examples of good reporting (via the PRISMA statement website: <http://www.prisma-statement.org/>).

TITLE

Item 1. Identify the report as a systematic review.

Explanation: Inclusion of “systematic review” in the title facilitates identification by potential users (e.g. patients, health care providers, policy makers) and appropriate indexing in databases. Terms such

as “review”, “literature review”, “evidence synthesis” or “knowledge synthesis” are not recommended because they do not distinguish systematic and non-systematic approaches. We also discourage using the terms “systematic review” and “meta-analysis” interchangeably because a systematic review refers to the entire set of processes used to identify, select and synthesise evidence; whereas, meta-analysis refers only to the statistical synthesis. Furthermore, a meta-analysis can be done outside the context of a systematic review (e.g. when researchers meta-analyse results from a limited set of studies that they have conducted).

Essential elements

- Identify the report as a systematic review in the title.
- Report an informative title that provides key information about the main objective or question the review addresses (for reviews of interventions, this usually includes the population and intervention(s) the review addresses).

Additional elements

- Consider providing additional information in the title, such as the method of analysis used (e.g. “a systematic review with meta-analysis”), the designs of included studies (e.g. “a systematic review of randomized trials”), or an indication that the review is an update of an existing review, or a continually updated (“living”) systematic review.

Example

“Comparison of the therapeutic effects of rivaroxaban versus warfarin in antiphospholipid syndrome: a systematic review” (15)

ABSTRACT

Item 2. See the PRISMA 2020 for Abstracts checklist (see Box 2).

Explanation: An abstract providing key information about the review question, methods, results and implications of the findings should help readers decide whether to access the full report (16). For some readers, the abstract may be all that they have access to. Therefore, it is critical that results are presented for all main outcomes for the primary review question(s), regardless of the statistical significance, magnitude, or direction of effect. Terms presented in the abstract will be used to index the systematic review in bibliographic databases. Therefore, reporting key words that accurately describe the review question (e.g. population, interventions, outcomes) is recommended.

Essential elements

- Report an abstract addressing each item in the PRISMA 2020 for Abstracts checklist (see Box 2).

Example

“Title: Psychological interventions for common mental disorders in women experiencing intimate partner violence in low-income and middle-income countries: a systematic review and meta-analysis. Background: Evidence on the effectiveness of psychological interventions for women with common mental disorders (CMDs) who also experience intimate partner violence is scarce. We aimed to test our hypothesis that exposure to intimate partner violence would reduce intervention effectiveness for CMDs in low-income and middle-income countries (LMICs). Methods: For this systematic review and meta-analysis, we searched MEDLINE, Embase, PsycINFO, Web of Knowledge, Scopus, CINAHL, LILACS, ScieELO, Cochrane, PubMed databases, trials registries, 3ie, Google Scholar, and forward and backward citations for studies published between database inception and Aug 16, 2019. All randomised controlled trials (RCTs) of psychological interventions for CMDs in LMICs which measured intimate partner violence were included, without language or date restrictions. We approached study authors to obtain unpublished aggregate subgroup data for women who did and did not report intimate partner violence. We did separate random-effects meta-analyses for anxiety, depression, post-traumatic stress disorder (PTSD), and psychological distress outcomes. Evidence from randomised controlled trials was synthesised as differences between standardised mean differences (SMDs) for change in symptoms, comparing women who did and who did not report intimate partner violence via random-effects meta-analyses. The quality of the evidence was assessed with the Cochrane risk of bias tool. This study is registered on PROSPERO, number CRD42017078611. Findings: Of 8122 records identified, 21 were eligible and data were available for 15 RCTs, all of which had a low to moderate risk of overall bias. Anxiety (five interventions, 728 participants) showed a greater response to intervention among women reporting intimate partner violence than among those who did not (difference in standardised mean differences [dSMD] 0.31, 95% CI 0.04 to 0.57, I²=49.4%). No differences in response to intervention were seen in women reporting intimate partner violence for PTSD (eight interventions, n=1436; dSMD 0.14, 95% CI -0.06 to 0.33, I²=42.6%), depression (12 interventions, n=2940; 0.10, -0.04 to 0.25, I²=49.3%), and psychological distress (four interventions, n=1591; 0.07, -0.05 to 0.18, I²=0.0%, p=0.681). Interpretation: Psychological interventions treat anxiety effectively in women with current or recent intimate partner violence exposure in LMICs when delivered by appropriately trained and supervised health-care staff, even when not tailored for this population or targeting intimate partner violence directly. Future research should investigate whether adapting evidence-based psychological interventions for CMDs to address intimate partner violence

enhances their acceptability, feasibility, and effectiveness in LMICs. Funding: UK National Institute for Health Research ASSET and King's IoPPN Clinician Investigator Scholarship.” (17)

RATIONALE

Item 3. Describe the rationale for the review in the context of existing knowledge.

Explanation: Describing the rationale should help readers understand why the review was conducted and what the review might add to existing knowledge.

Essential elements

- Describe the current state of knowledge and its uncertainties.
- Articulate why it is important to do the review.
- If other systematic reviews addressing the same (or a largely similar) question are available, explain why the current review was considered necessary (e.g. previous reviews are out of date or have discordant results; new review methods are available to address the review question; existing reviews are methodologically flawed; or the current review was commissioned to inform a guideline or policy for a particular organisation). If the review is an update or replication of a particular systematic review, indicate this and cite the previous review.
- If the review examines the effects of interventions, also briefly describe how the intervention(s) examined might work.

Additional elements

- If there is complexity in the intervention or context of its delivery (or both) (e.g. multi-component interventions, interventions targeting the population and individual level, equity considerations (18)), consider presenting a logic model (sometimes referred to as a conceptual framework or theory of change) to visually display the hypothesised relationship between intervention components and outcomes (19, 20).

Example

“To contain widespread infection and to reduce morbidity and mortality among health-care workers and others in contact with potentially infected people, jurisdictions have issued conflicting advice about physical or social distancing. Use of face masks with or without eye protection to achieve additional protection is debated in the mainstream media and by public health authorities, in particular the use of face masks for the general population; moreover, optimum use of face masks in health-care settings, which have been used for decades for infection prevention, is facing challenges amid personal

protective equipment (PPE) shortages. Any recommendations about social or physical distancing, and the use of face masks, should be based on the best available evidence. Evidence has been reviewed for other respiratory viral infections, mainly seasonal influenza, but no comprehensive review is available of information on SARS-CoV-2 or related betacoronaviruses that have caused epidemics, such as severe acute respiratory syndrome (SARS) or Middle East respiratory syndrome (MERS). We, therefore, systematically reviewed the effect of physical distance, face masks, and eye protection on transmission of SARS-CoV-2, SARS-CoV, and MERS-CoV.” (21)

OBJECTIVES

Item 4. Provide an explicit statement of the objective(s) or question(s) the review addresses.

Explanation: An explicit and concise statement of the review objective(s) or question(s) will help readers understand the scope of the review and assess whether the methods used in the review (e.g. eligibility criteria, search methods, data items, and the comparisons used in the synthesis) adequately address the objective(s). Such statements may be written in the form of objectives (“the objectives of the review were to examine the effects of...”) or as questions (“what are the effects of...?”) (19).

Essential elements

- Provide an explicit statement of all objective(s) or question(s) the review addresses, expressed in terms of a relevant question formulation framework (see Table S1 of Booth et al. (22) for various frameworks).
- If the purpose is to evaluate the effects of interventions, use the Population, Intervention, Comparator, Outcome (PICO) framework or one of its variants, to state the comparisons that will be made.

Example

“Objectives: To evaluate the benefits and harms of down-titration (dose reduction, discontinuation, or disease activity-guided dose tapering) of anti-tumour necrosis factor-blocking agents (adalimumab, certolizumab pegol, etanercept, golimumab, infliximab) on disease activity, functioning, costs, safety, and radiographic damage compared with usual care in people with rheumatoid arthritis and low disease activity.” (23)

ELIGIBILITY CRITERIA

Item 5. Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses.

Explanation: Specifying the criteria used to decide what evidence was eligible or ineligible in sufficient detail should enable readers to understand the scope of the review and verify inclusion decisions (24). The PICO framework is commonly used to structure the reporting of eligibility criteria for reviews of interventions (25). In addition to specifying the review PICO, the intervention, outcome and population groups that were used in the syntheses, need to be identified and defined (26). For example, in a review examining the effects of psychological interventions for smoking cessation in pregnancy, the authors specified intervention groups (counselling, health education, feedback, incentive-based interventions, social support, and exercise) and the defining components of each group (27).

Essential elements

- Specify all study characteristics used to decide whether a study was eligible for inclusion in the review, that is, components described in the PICO framework or one of its variants, and other characteristics, such as eligible study design(s) and setting(s), and minimum duration of follow-up.
- Specify eligibility criteria with regard to report characteristics, such as year of dissemination, language, and report status (e.g. whether reports, such as unpublished manuscripts and conference abstracts, were eligible for inclusion).
- Clearly indicate if studies were ineligible because the outcomes of interest were not measured, or ineligible because the results for the outcome of interest were not reported. Reporting that studies were excluded if they had “no relevant outcome data” is ambiguous and should be avoided (28).
- Specify any groups used in the synthesis (e.g. intervention, outcome and population groups) and link these to the comparisons specified in the objectives (item #4).

Additional elements

- Consider providing rationales for any notable restrictions to study eligibility. For example, authors might explain that the review was restricted to studies published from 2000 onward, because that was the year the device was first available.

Example

“Population: We included randomized controlled trials of adult (age ≥ 18 years) patients undergoing non-cardiac surgery, excluding organ transplantation surgery (as findings in patients who need immunosuppression may not be generalisable to others).”

Intervention: We considered all perioperative care interventions identified by the search if they were protocolised (therapies were systematically provided to patients according to pre-defined algorithm or plan) and were started and completed during the perioperative pathway (that is, during preoperative preparation for surgery, intraoperative care, or inpatient postoperative recovery). Examples of interventions that we did or did not deem perioperative in nature included long term preoperative drug treatment (not included, as not started and completed during the perioperative pathway) and perioperative physiotherapy interventions (included, as both started and completed during the perioperative pathway). We excluded studies in which the intervention was directly related to surgical technique.

Outcomes: To be included, a trial had to use a defined clinical outcome relating to postoperative pulmonary complications, such as “pneumonia” diagnosed according to the Centers for Disease Control and Prevention’s definition. Randomized controlled trials reporting solely physiological (for example, lung volumes and flow measurements) or biochemical (for example, lung inflammatory markers) outcomes are valuable but neither patient centric nor necessarily clinically relevant, and we therefore excluded them. We applied no language restrictions. Our primary outcome measure was the incidence of postoperative pulmonary complications, with postoperative pulmonary complications being defined as the composite of any of respiratory infection, respiratory failure, pleural effusion, atelectasis, or pneumothorax...Where a composite postoperative pulmonary complication was not reported, we contacted corresponding authors via email to request additional information, including primary data.” (29)

INFORMATION SOURCES

Item 6. Specify all databases, registers, websites, organisations, reference lists and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted.

Explanation: Authors should provide a detailed description of the information sources, such as bibliographic databases, registers and reference lists that were searched or consulted, including the dates when each source was last searched, to allow readers to assess the completeness and currency of the systematic review, and facilitate updating (30). Authors should fully report the ‘what, when and how’ of the sources searched; the ‘what’ and ‘when’ are covered in item #6 and the ‘how’ is covered in item #7. Further guidance and examples about searching can be found in PRISMA-Search, an extension to the PRISMA statement for reporting literature searches in systematic reviews (31).

Essential elements

- Specify all databases, registers, websites, organisations, reference lists and other sources searched or consulted to identify studies.
- Specify the date when each source was last searched or consulted.
- If bibliographic databases were searched, specify for each database its name (e.g. MEDLINE, CINAHL), the interface or platform through which the database was searched (e.g. Ovid, EBSCOhost), and the dates of coverage (where this information is provided).
- If study registers (e.g. ClinicalTrials.gov), regulatory databases (e.g. Drugs@FDA) and other online repositories (e.g. SIDER Side Effect Resource) were searched, specify the name of each source and any date restrictions that were applied.
- If websites, search engines or other online sources were browsed or searched, specify the name of each source.
- If organisations or manufacturers were contacted to identify studies, specify the name of each source.
- If individuals were contacted to identify studies, specify the types of individuals contacted (e.g. authors of studies included in the review or researchers with expertise in the area).
- If reference lists were examined, specify the types of references examined (e.g. references of studies included in the systematic review, or references of systematic reviews on the same or a similar topic).
- If cited or citing reference searches (also called backward and forward citation searching) were conducted, specify the bibliographic details of the reports to which citation searching was applied, the citation index or platform used (e.g. Web of Science), and the date the citation searching was done.
- If journals or conference proceedings were consulted, specify the names of each source, the dates covered and how they were searched (e.g. handsearching or browsing online).

Example

“On 21 December 2017, MAJ searched 16 health, social care, education and legal databases, the names and date coverage of which are given in the Table 1...We also carried out a ‘snowball’ search to identify additional studies by searching the reference lists of publications eligible for full-text review and using Google Scholar to identify and screen studies citing them...On 26 April 2018, we conducted a search of Google Scholar and additional supplementary searches for publications on websites of 10 relevant organisations (including government departments, charities, think-tanks and research institutes). Full details of these supplementary searches can be found in the Additional file. Finally, we updated the database search on 7 May 2019, and the snowball and additional searches on 10 May

2019 as detailed in the Additional file. We used the same search method, except that we narrowed the searches to 2017 onwards.” (32)

SEARCH STRATEGY

Item 7. Present the full search strategies for all databases, registers and websites, including any filters and limits used.

Explanation: Reporting the full details of all search strategies (e.g. the full line by line search strategy as run in each database) should enhance the transparency of the systematic review, improve replicability, and enable a review to be more easily updated (30, 33). As well as reporting the search strategies, a description of the search strategy development process can help readers judge how far the strategy is likely to have identified all studies relevant to the review’s inclusion criteria. The description of the search strategy development process might include details of the approaches used to identify keywords, synonyms or subject indexing terms used in the search strategies, or any processes used to validate or peer review the search strategies. Empirical evidence suggests that peer review of search strategies is associated with improvements to search strategies, leading to retrieval of additional relevant records (34). Further guidance and examples of reporting search strategies can be found in PRISMA-Search (31).

Essential elements

- Provide the full line by line search strategy as run in each database with a sophisticated interface (such as Ovid), or the sequence of terms that were used to search simpler interfaces, such as search engines or websites. Make all search strategies publicly accessible, by uploading them as a supplementary file (if the systematic review is published in an open-access journal), or by depositing the search strategies in a permanent, open-access repository (e.g. Open Science Framework).
- Describe any limits applied to the search strategy (e.g. date or language) and justify these by linking back to the review’s eligibility criteria.
- If published approaches, such as search filters designed to retrieve specific types of records (e.g. filter for randomized trials) (35) or search strategies from other systematic reviews, were used, cite them. If published approaches were adapted, for example if existing search filters were amended, note the changes made.
- If natural language processing or text frequency analysis tools were used to identify keywords, synonyms or subject indexing terms to use in the search strategy (36, 37), specify the tool(s) used.

- If the search strategy was validated, for example by evaluating whether it could identify a set of clearly eligible studies, report the validation process used and specify which studies were included in the validation set (30).
- If the search strategy was peer reviewed, report the peer review process used and specify any tool used such as the Peer Review of Electronic Search Strategies (PRESS) checklist (38).
- If languages other than English were used to carry out searches, specify the languages used.
- If the search strategy structure adopted was not based on a PICO-style approach, describe the final conceptual structure and any explorations that were undertaken to achieve it (e.g. use of a multi-faceted approach that uses a series of searches, with different combinations of concepts, to capture a complex research question, or use of a variety of different search approaches to compensate for when a specific concept is difficult to define) (30).

Example

Note: the following is an abridged version of an example presented in full in Table S1.

“MEDLINE(R) In-Process & Other Non-Indexed Citations and Ovid MEDLINE were searched via OvidSP. The database coverage was 1946 to present and the databases were searched on 29 August 2013.

1. *Urinary Bladder, Overactive/*
2. *((overactiv\$ or over-activ\$ or hyperactiv\$ or hyper-activ\$ or unstable or instability or incontinen\$) adj3 bladder\$).ti,ab.*
3. *(OAB or OABS or IOAB or IOABS).ti,ab.*
4. *(urge syndrome\$ or urge frequenc\$).ti,ab.*
5. *((overactiv\$ or over-activ\$ or hyperactiv\$ or hyper-activ\$ or unstable or instability) adj3 detrusor\$).ti,ab.*
6. *Urination Disorders/*
7. *exp Urinary Incontinence/*
8. *Urinary Bladder Diseases/*
9. *(urge\$ adj3 incontinen\$).ti,ab.*
10. *(urin\$ adj3 (incontinen\$ or leak\$ or urgen\$ or frequen\$)).ti,ab.*
11. *(urin\$ adj3 (disorder\$ or dysfunct\$)).ti,ab.*
12. *(detrusor\$ adj3 (hyperreflexia\$ or hyper-reflexia\$ or hypertoni\$ or hyper-toni\$)).ti,ab.*
13. *(void\$ adj3 (disorder\$ or dysfunct\$)).ti,ab.*
14. *(micturition\$ adj3 (disorder\$ or dysfunct\$)).ti,ab.*
15. *exp Enuresis/*
16. *Nocturia/*

17. *(nocturia or nycturia or enuresis).ti,ab.*
18. *or/1-17*
19. *(mirabegron or betmiga\$ or myrbetriq\$ or betanis\$ or YM-178 or YM178 or 223673-61-8 or "223673618" or MVR3JL3B2V).ti,ab,rn.*
20. *exp Electric Stimulation Therapy/*
21. *Electric Stimulation/*
22. *((sacral or S3) adj3 (stimulat\$ or modulats\$)).ti,ab.*
23. *(neuromodulat\$ or neuro-modulat\$ or neural modulats\$ or electromodulat\$ or electro-modulats\$ or neurostimulat\$ or neuro-stimulat\$ or neural stimulat\$ or electrostimulat\$ or electro-stimulat\$).ti,ab.*
24. *(InterStim or SNS).ti,ab.*
25. *((electric\$ or nerve\$1) adj3 (stimulat\$ or modulats\$)).ti,ab.*
26. *(electric\$ therap\$ or electrotherap\$ or electro-therap\$).ti,ab.*
27. *TENS.ti,ab.*
28. *exp Electrodes/*
29. *electrode\$1.ti,ab.*
30. *((implant\$ or insert\$) adj3 pulse generator\$).ti,ab.*
31. *((implant\$ or insert\$) adj3 (neuroprosthesis\$ or neuro-prosthesis\$ or neural prosthesis\$)).ti,ab.*
32. *PTNS.ti,ab.*
33. *(SANS or Stoller Afferent or urosurg\$).ti,ab.*
34. *(evaluat\$ adj3 peripheral nerve\$).ti,ab.*
35. *exp Botulinum Toxins/*
36. *(botulinum\$ or botox\$ or onabotulinumtoxin\$ or 1309378-01-5 or "1309378015").ti,ab,rn.*
37. *or/19-36*
38. *18 and 37*
39. *randomized controlled trial.pt.*
40. *controlled clinical trial.pt.*
41. *random\$.ti,ab.*
42. *placebo.ti,ab.*
43. *drug therapy.fs.*
44. *trial.ti,ab.*
45. *groups.ab.*
46. *or/39-45*
47. *38 and 46*

48. *animals/ not humans/*

49. *47 not 48*

50. *limit 49 to english language*

Search strategy development process: Five known relevant studies were used to identify records within databases. Candidate search terms were identified by looking at words in the titles, abstracts and subject indexing of those records. A draft search strategy was developed using those terms and additional search terms were identified from the results of that strategy. Search terms were also identified and checked using the PubMed PubReMiner word frequency analysis tool. The MEDLINE strategy makes use of the Cochrane RCT filter reported in the Cochrane Handbook v5.2. As per the eligibility criteria the strategy was limited to English language studies. The search strategy was validated by testing whether it could identify the five known relevant studies and also three further studies included in two systematic reviews identified as part of the strategy development process. All eight studies were identified by the search strategies in MEDLINE and Embase. The strategy was developed by an information specialist and the final strategies were peer reviewed by an experienced information specialist within our team. Peer review involved proofreading the syntax and spelling and overall structure, but did not make use of the PRESS checklist.” (39)

SELECTION PROCESS

Item 8. Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they worked independently, and if applicable, details of automation tools used in the process.

Explanation: Study selection is typically a multi-stage process in which potentially eligible studies are first identified from screening titles and abstracts, then assessed through full-text review and, where necessary, contact with study investigators. Increasingly, a mix of screening approaches might be applied (e.g. automation to eliminate records before screening or prioritise records during screening). In addition to automation, authors increasingly have access to screening decisions that are made by people independent of the author team (e.g. crowdsourcing) (see Box 3). Authors should describe in detail the process for deciding how records retrieved by the search were considered for inclusion in the review, to enable readers to assess the potential for errors in selection (40-43).

Essential elements for systematic reviews regardless of the selection processes used

- Report how many reviewers screened each record (title/abstract) and each report retrieved, whether multiple reviewers worked independently at each stage of screening or not (e.g. records

screened by one reviewer and exclusions verified by another), and any processes used to resolve disagreements between screeners (e.g. referral to a third reviewer or by consensus).

- Report any processes used to obtain or confirm relevant information from study investigators.
- If articles required translation into another language to determine their eligibility, report how these articles were translated, for example by asking a native speaker or by using software programs.

Essential elements for systematic reviews using automation tools in the selection process

- Report how automation tools were integrated within the overall study selection process; for example, whether records were excluded based solely on a machine assessment, or whether machine assessments were used to double-check human decisions.
- If an externally derived machine learning classifier was applied (e.g. Cochrane RCT Classifier), either to eliminate records or to replace a single screener, include a reference or URL to the version used. If the classifier was used to eliminate records *before screening*, report the number eliminated in the PRISMA flow diagram as ‘Records marked as ineligible by automation tools’.
- If an internally derived machine learning classifier was used to assist with the screening process, identify the software/classifier and version, describe how it was used (e.g. to remove records or replace a single screener) and trained (if relevant), and what internal or external validation was done to understand the risk of missed studies or incorrect classifications. For example, authors might state that the classifier was trained on the set of records generated for the review in question (as may be the case when updating reviews) and specify which thresholds were applied to remove records.
- If machine learning algorithms were used to prioritise screening (whereby unscreened records are continually re-ordered based on screening decisions), state the software used and provide details of any screening rules applied (e.g. screening stopped altogether leaving some records to be excluded based on automated assessment alone, or screening switched from double to single screening, once a pre-specified number or proportion of consecutive records was eliminated).

Essential elements for systematic reviews using crowdsourcing or previous ‘known’ assessments in the selection process

- If crowdsourcing was used to screen records, provide details of the platform used and specify how it was integrated within the overall study selection process.
- If datasets of already-screened records were used to eliminate records retrieved by the search from further consideration, briefly describe the derivation of these datasets. For example, if prior

work has already determined that a given record does not meet the eligibility criteria, it can be removed without manual checking. This is the case for Cochrane's Screen4Me service in which an increasingly large dataset of records that are known not to represent randomized trials can be used to eliminate any matching records from further consideration.

Example

“Three researchers (AP, HB-R, FG) independently reviewed titles and abstracts of the first 100 records and discussed inconsistencies until consensus was obtained. Then, in pairs, the researchers independently screened titles and abstracts of all articles retrieved. In case of disagreement, consensus on which articles to screen full-text was reached by discussion. If necessary, the third researcher was consulted to make the final decision. Next, two researchers (AP, HB-R) independently screened full-text articles for inclusion. Again, in case of disagreement, consensus was reached on inclusion or exclusion by discussion and if necessary, the third researcher (FG) was consulted.” (44)

For examples of systematic reviews using automation tools, crowdsourcing or previous 'known' assessments in the selection process, see Table S1 in the Supplement.

DATA COLLECTION PROCESS

Item 9. Specify the methods used to collect data from reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and if applicable, details of automation tools used in the process.

Explanation: Authors should report the methods used to collect data from reports of included studies, to enable readers to assess the potential for errors in the data presented (45-47).

Essential elements

- Report how many reviewers collected data from each report, whether multiple reviewers worked independently or not (e.g. data collected by one reviewer and checked by another) (48), and any processes used to resolve disagreements between data collectors.
- Report any processes used to obtain or confirm relevant data from study investigators (e.g. how they were contacted, what data were sought, and success in obtaining the necessary information).
- If any automation tools were used to collect data, report how the tool was used (e.g. machine learning models to extract sentences from articles relevant to the PICO characteristics) (49, 50),

how the tool was trained, and what internal or external validation was done to understand the risk of incorrect extractions.

- If articles required translation into another language to enable data collection, report how these articles were translated, for example by asking a native speaker or by using software programs (51).
- If any software was used to extract data from figures (52), specify the software used.
- If any decision rules were used to select data from multiple reports corresponding to a study, and any steps were taken to resolve inconsistencies across reports, report the rules and steps used (53).

Example

“We designed a data extraction form based on that used by Lumley 2009, which two review authors (RC and TC) used to extract data from eligible studies. Extracted data were compared, with any discrepancies being resolved through discussion. RC entered data into Review Manager 5 software (Review Manager 2014), double checking this for accuracy. When information regarding any of the above was unclear, we contacted authors of the reports to provide further details.” (54)

DATA ITEMS

Item 10a. List and define all outcomes for which data were sought. Specify whether all results that were compatible with each outcome domain in each study were sought (e.g. for all measures, time points, analyses), and if not, the methods used to decide which results to collect.

Explanation: Defining outcomes in systematic reviews generally involves specifying outcome domains (e.g. pain, quality of life, adverse events such as nausea) and the time frame of measurement (e.g. less than 6 months) (26). Included studies may report multiple results that are eligible for inclusion within the review outcome definition (55, 56). For example, a study may report results for two measures of pain (e.g. the McGill Pain Questionnaire and the Brief Pain Inventory), at two time points (e.g. 4 weeks, 8 weeks), all of which are compatible with a review outcome defined as ‘pain <6 months’. Multiple results compatible with an outcome domain in a study might also arise when study investigators report results based on multiple analysis populations (e.g. all participants randomized, all participants receiving a specific amount of treatment), methods for handling missing data (e.g. multiple imputation, last-observation-carried-forward) or methods for handling confounding (e.g. adjustment for different covariates) (56-58).

Reviewers might seek all results that were compatible with each outcome definition from each study; or, use a process to select a subset of the results (53, 58). Examples of processes to select results include selecting the outcome definition that: (a) was most common across studies; (b) the review authors considered “best” according to a prespecified hierarchy; or (c) the study investigators considered most important (e.g. the study’s primary outcome). It is important to specify the methods that were used to select the results when multiple were available, so that users are able to judge the appropriateness of those methods and whether there is potential for bias in the selection of results. Reviewers may make changes to the inclusion or definition of the outcome domains, or to the importance given to them in the review (e.g. an outcome listed as ‘important’ in the protocol, is considered ‘critical’ in the review). Providing a rationale for the change allows readers to assess the legitimacy of the change and whether it has potential to introduce bias in the review process (59).

Essential elements

- List and define the outcome domains and time frame of measurement for which data were sought.
- Specify whether all results that were compatible with each outcome domain in each study were sought, and if not, what process was used to select results within eligible domains.
- If any changes were made to the inclusion or definition of the outcome domains, or to the importance given to them in the review, specify the changes, along with a rationale.
- If any changes were made to the processes used to select results within eligible outcome domains, specify the changes, along with a rationale.

Additional elements

- Consider specifying which outcome domains were considered the most important for interpreting the review’s conclusions (e.g. ‘critical’ versus ‘important’ outcomes) and provide rationale for the labelling (e.g. “a recent core outcome set identified the outcomes labelled ‘critical’ as being the most important to patients”).

Example

Note: the following is an abridged version of an example presented in full in Table S1.

“Eligible outcomes were broadly categorised as follows:

- *Cognitive function*
 - *Global cognitive function*
 - *Domain-specific cognitive function (especially domains that reflect specific alcohol-related neuropathologies, such as psychomotor speed and working memory)*

- *Clinical diagnoses of cognitive impairment*
 - *Mild cognitive impairment (also referred to as mild neurocognitive disorders)*

Any measure of cognitive function was eligible for inclusion. The tests or diagnostic criteria used in each study should have had evidence of validity and reliability for the assessment of mild cognitive impairment, but studies were not excluded on this basis...Results could be reported as an overall test score that provides a composite measure across multiple areas of cognitive ability (i.e. global cognitive function), sub-scales that provide a measure of domain-specific cognitive function or cognitive abilities (e.g. processing speed, memory), or both...Studies with a minimum follow-up of 6 months were eligible, a time frame chosen to ensure that studies were designed to examine more persistent effects of alcohol consumption...No restrictions were placed on the number of points at which the outcome was measured, but the length of follow-up and number of measurement points (including a baseline measure of cognition) was considered when interpreting study findings and in deciding which outcomes were similar enough to combine for synthesis.

We anticipated that individual studies would report data for multiple cognitive outcomes. Specifically, a single study may report results:

- *For multiple constructs related to cognitive function, for example, global cognitive function and cognitive ability on specific domains (e.g. memory, attention, problem-solving, language);*
- *Using multiple methods or tools to measure the same or similar outcome, for example reporting measures of global cognitive function using both the Mini-Mental State Examination and the Montreal Cognitive Assessment;*
- *At multiple time points, for example, at 1, 5, and 10 years.*

Where multiple cognition outcomes were reported, we selected one outcome for inclusion in analyses and for reporting the main outcomes (e.g. for GRADEing), choosing the result that provided the most complete information for analysis. Where multiple results remained, we listed all available outcomes (without results) and asked our content expert to independently rank these based on relevance to the review question, and the validity and reliability of the measures used. Measures of global cognitive function were prioritised, followed by measures of memory, then executive function. In the circumstance where results from multiple multivariable models were presented, we extracted associations from the most fully adjusted model, except in the case where an analysis adjusted for a possible intermediary along the causal pathway (i.e. post-baseline measures of prognostic factors (e.g. smoking, drug use, hypertension)).” (60)

Item 10b. List and define all other variables for which data were sought (e.g. participant and intervention characteristics, funding sources). Describe any assumptions made about any missing or unclear information.

Explanation: Authors should report the data and information collected from the studies so that readers can understand the type of the information sought and to inform data collection in other similar reviews. Variables of interest might include characteristics of the study (e.g. countries, settings, number of centres, funding sources, registration status), characteristics of the study design (e.g. randomized or non-randomized), characteristics of participants (e.g. age, sex, socioeconomic status), number of participants enrolled and included in analyses, the results (e.g. summary statistics, estimates of effect and measures of precision, factors adjusted for in analyses), and competing interests of study authors. For reviews of interventions, authors may also collect data on characteristics of the interventions (e.g. what interventions and comparators were delivered, how they were delivered, by whom, where, and for how long).

Essential elements

- List and define all other variables for which data were sought. It may be sufficient to report a brief summary of information collected if the data collection and dictionary forms are made available (e.g. as additional files or deposited in a publicly available repository).
- Describe any assumptions made about any missing or unclear information from the studies. For example, in a study that includes “children and adolescents”, for which the investigators did not specify the age range, authors might assume that the oldest participants would be 18 years, based on what was observed in similar studies included in the review, and should report that assumption.
- If a tool was used to inform which data items to collect (e.g. the Tool for Addressing Conflicts of Interest in Trials (TACIT) (61, 62), or a tool for recording intervention details (63-65)), cite the tool used.

Example

“We collected data on:

- *the report: author, year, and source of publication;*
- *the study: sample characteristics, social demography, and definition and criteria used for depression;*
- *the participants: stroke sequence (first ever vs recurrent), social situation, time elapsed since stroke onset, history of psychiatric illness, current neurological status, current treatment for depression, and history of coronary artery disease;*

- *the research design and features: sampling mechanism, treatment assignment mechanism, adherence, non-response, and length of follow up;*
- *the intervention: type, duration, dose, timing, and mode of delivery.” (66)*

STUDY RISK OF BIAS ASSESSMENT

Item 11. Specify the methods used to assess risk of bias in the included studies, including details of the tool(s) used, how many reviewers assessed each study and whether they worked independently, and if applicable, details of automation tools used in the process.

Explanation: Users of reviews need to know the risk of bias in the included studies to appropriately interpret the evidence. Numerous tools have been developed to assess study limitations for various designs (67). However, many tools have been criticised because of their content (which may extend beyond assessing study limitations that have the potential to bias findings) and the way in which the items are combined (e.g. scales where items are combined to yield a numerical score) (see Box 4) (62). Reporting details of the selected tool enables readers to assess whether the tool focuses solely on items that have the potential to bias findings, or not. Reporting details of how studies were assessed (e.g. by one or two authors) allows readers to assess the potential for errors in the assessments (46). Reporting how risk of bias assessments were incorporated into the analysis is addressed in Items #13e and #13f.

Essential elements

- Specify the tool(s) (and version) used to assess risk of bias in the included studies.
- Specify the methodological domains/components/items of the risk of bias tool(s) used.
- Report whether an overall risk of bias judgement that summarised across domains/components/items was made, and if so, what rules were used to reach an overall judgement.
- If any adaptations to an existing tool to assess risk of bias in studies were made (e.g. omitting or modifying items), specify the adaptations.
- If a new risk of bias tool was developed for use in the review, describe the content of the tool and make it publicly accessible.
- Report how many reviewers assessed risk of bias in each study, whether multiple reviewers worked independently or not (e.g. assessments performed by one reviewer and checked by another), and any processes used to resolve disagreements between assessors.
- Report any processes used to obtain or confirm relevant information from study investigators.

- If an automation tool was used to assess risk of bias in studies, report how the automation tool was used (e.g. machine learning models to extract sentences from articles relevant to risk of bias (68)), how the tool was trained, and details on the tool's performance and internal validation.

Example

"We assessed risk of bias in the included studies using the revised Cochrane 'Risk of bias' tool for randomised trials (RoB 2.0) (Higgins 2016a), employing the additional guidance for cluster-randomised and cross-over trials (Eldridge 2016; Higgins 2016b). RoB 2.0 addresses five specific domains: (1) bias arising from the randomisation process; (2) bias due to deviations from intended interventions; (3) bias due to missing outcome data; (4) bias in measurement of the outcome; and (5) bias in selection of the reported result. Two review authors independently applied the tool to each included study, and recorded supporting information and justifications for judgements of risk of bias for each domain (low; high; some concerns). Any discrepancies in judgements of risk of bias or justifications for judgements were resolved by discussion to reach consensus between the two review authors, with a third review author acting as an arbiter if necessary. Following guidance given for RoB 2.0 (Section 1.3.4) (Higgins 2016a), we derived an overall summary 'Risk of bias' judgement (low; some concerns; high) for each specific outcome, whereby the overall RoB for each study was determined by the highest RoB level in any of the domains that were assessed." (69)

EFFECT MEASURES

Item 12. Specify for each outcome the effect measure(s) (e.g. risk ratio, mean difference) used in the synthesis or presentation of results.

Explanation: To interpret a synthesized or study result, users need to know what effect measure was used. Effect measures refer to statistical constructs that compare outcome data between two groups. For instance, a risk ratio is an example of an effect measure that might be used for dichotomous outcomes (70). The chosen effect measure has implications for interpretation of the findings and might impact the meta-analysis results (e.g. heterogeneity (71)). Authors might use one effect measure to synthesise results and then re-express the synthesised results using another effect measure. For example, for meta-analyses of standardized mean differences, authors might re-express the combined results in units of a well-known measurement scale, and for meta-analyses of risk ratios or odds ratios, authors might re-express results in absolute terms (e.g. risk difference) (72).

Essential elements

- Specify for each outcome (or type of outcome [e.g. binary, continuous]), the effect measure(s) (e.g. risk ratio, mean difference) used in the synthesis or presentation of results.
- If synthesized results were re-expressed to a different effect measure, report the methods used to re-express results (e.g. meta-analysing risk ratios and computing an absolute risk reduction based on an assumed comparator risk).

Additional elements

- Consider providing justification for the choice of effect measure. For example, a standardized mean difference may have been chosen because multiple instruments or scales were used across studies to measure the same outcome domain (e.g. different instruments to assess depression).

Example

“We planned to analyse dichotomous outcomes by calculating the risk ratio (RR) of a successful outcome (i.e. improvement in relevant variables) for each trial...Because the included resilience-training studies used different measurement scales to assess resilience and related constructs, we used standardised mean difference (SMD) effect sizes (Cohen's d) and their 95% confidence intervals (CIs) for continuous data in pair-wise meta-analyses.” (73)

SYNTHESIS METHODS

Item 13a. Describe the processes used to decide which studies were eligible for each synthesis.

Explanation: Prior to undertaking any statistical synthesis (item #13d), decisions must be made about which studies are eligible for each planned synthesis (item #5). These decisions will likely involve subjective judgements, yet the processes used and information to support the decisions are often absent from reviews. Reporting the processes (whether formal or informal), and any supporting information, is recommended for transparency of the decisions made in grouping studies for synthesis. Structured approaches may involve the tabulation and coding of the main characteristics of the populations, interventions and outcomes (74). For example, in a review examining the effects of psychological interventions for smoking cessation in pregnancy, the main intervention component of each study was coded as one of the following based on pre-specified criteria: counselling, health education, feedback, incentive-based interventions, social support, and exercise (27). This coding provided the basis for determining which studies were eligible for each planned synthesis (e.g. incentive-based interventions versus usual care). Similar coding processes can be applied to populations and outcomes.

Essential elements

- Describe the processes used to decide which studies were eligible for each synthesis.

Example

“Given the complexity of the interventions being investigated, we attempted to categorize the included interventions along four dimensions: (1) was housing provided to the participants as part of the intervention; (2) to what degree was the tenants’ residence in the provided housing dependent on, for example, sobriety, treatment attendance, etc.; (3) if housing was provided, was it segregated from the larger community, or scattered around the city; and (4) if case management services were provided as part of the intervention, to what degree of intensity. We created categories of interventions based on the above dimensions:

1. *Case management only*
2. *Abstinence-contingent housing*
3. *Non-abstinence-contingent housing*
4. *Housing vouchers*
5. *Residential treatment with case management*

Some of the interventions had multiple components (e.g. abstinence-contingent housing with case management). These interventions were categorized according to the main component (the component that the primary authors emphasized). They were also placed in separate analyses. We then organized the studies according to which comparison intervention was used (any of the above interventions, or usual services).” (75)

Item 13b. Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics, or data conversions.

Explanation: Authors may need to prepare the data collected from studies so that it is suitable for presentation or to be included in a synthesis. This could involve algebraic manipulation to convert reported statistics to required statistics (e.g. converting standard errors to standard deviations) (70), transforming effect estimates (e.g. converting standardized mean differences to odds ratios (76)), or imputing missing summary data (e.g. missing standard deviations for continuous outcomes, intra-cluster correlations in cluster randomized trials) (77-79). Reporting the methods required to prepare the data will allow readers to judge the appropriateness of the methods used, the assumptions made, and aid in attempts to replicate the synthesis.

Essential elements

- Report any methods required to prepare the data collected from studies for presentation or synthesis, such as handling of missing summary statistics, or data conversions.

Example

“We used cluster-adjusted estimates from cluster randomised controlled trials (c-RCTs) where available. If the studies had not adjusted for clustering, we attempted to adjust their standard errors using the methods described in the Cochrane Handbook for Systematic Reviews of Interventions (Higgins 2019), using an estimate of the intra-cluster correlation coefficient (ICC) derived from the trial. If the trial did not report the cluster-adjusted estimated or the ICC, we imputed an ICC from a similar study included in the review, adjusting if the nature or size of the clusters was different (e.g. households compared to classrooms). We assessed any imputed ICCs using sensitivity analysis.” (80)

Item 13c. Describe any methods used to tabulate or visually display results of individual studies and syntheses.

Explanation: Presentation of study results using tabulation and visual display is important for transparency (particularly so for reviews, or outcomes within reviews, where a meta-analysis has not been undertaken) and facilitates the identification of patterns in the data. Tables may be used to present results from individual studies or from a synthesis (e.g. Summary of Findings table (81, 82); see item #22). The purpose of tabulating data varies, but commonly includes the complete and transparent reporting of the results, or comparing the results across study characteristics (83). Different purposes will likely lead to different table structures. Reporting the chosen structure(s), along with details of the data presented (e.g. effect estimates), can aid users in understanding the basis and rationale for the structure (e.g. “Table have been structured by outcome domain, within which studies are ordered from low to high risk of bias to increase the prominence of the most trustworthy evidence.”).

The principal graphical method for meta-analysis is the forest plot, which displays the effect estimates and confidence intervals of each study, and often the summary estimate (84, 85). Other graphs that aim to display information about the magnitude or direction of effects might be considered when a forest plot cannot be used due to incompletely reported effect estimates (e.g. no measure of precision reported) (83, 86). Careful choice and design of graphs is required so that they effectively and accurately represent the data (84).

Essential elements

- Report chosen tabular structure(s) used to display results of individual studies and syntheses, along with details of the data presented.
- Report chosen graphical methods used to visually display results of individual studies and syntheses.

Additional elements

- If studies are ordered or grouped within tables or graphs, consider reporting the basis for the chosen ordering/grouping.
- If non-standard graphs were used, consider reporting the rationale for selecting the chosen graph.

Example

“Meta-analyses could not be undertaken due to the heterogeneity of interventions, settings, study designs and outcome measures. Albatross plots were created to provide a graphical overview of the data for interventions with more than five data points for an outcome. Albatross plots are a scatter plot of p -values against the total number of individuals in each study. Small p -values from negative associations appear at the left of the plot, small p -values from positive associations at the right, and studies with null results towards the middle. The plot allows p -values to be interpreted in the context of the study sample size; effect contours show a standardised effect size (expressed as relative risk—RR) for a given p -value and study size, providing an indication of the overall magnitude of any association. We estimated an overall magnitude of association from these contours, but this should be interpreted cautiously.” (87)

Item 13d. Describe any methods used to synthesize results and provide a rationale for the choice(s). If meta-analysis was performed, describe the model(s), method(s) to identify the presence and extent of statistical heterogeneity, and software package(s) used.

Explanation: Various statistical methods are available to synthesize results, the most common of which is meta-analysis of effect estimates (see Box 5). Meta-analysis is used to synthesize effect estimates across studies, yielding a summary estimate. Different meta-analysis models are available, with the random-effects and fixed-effect models being in widespread use. Model choice can importantly affect the summary estimate and its confidence interval; hence, rationale for the selected model should be provided (see Box 5). For random-effects models, many methods are available, and their performance has been shown to differ depending on the characteristics of the meta-analysis (e.g. the number and size of the included studies (88, 89)).

When study data are not amenable to meta-analysis of effect estimates, alternative statistical synthesis methods (e.g. calculating the median effect across studies, combining P values) or structured summaries might be used (83, 90). Additional guidance for reporting alternative statistical synthesis methods is available (see Synthesis Without Meta-analysis (SWiM) reporting guideline (91)).

Regardless of the chosen synthesis method(s), authors should provide sufficient detail such that readers are able to assess the appropriateness of the selected methods and could reproduce the reported results (with access to the data).

Essential elements

- If statistical synthesis methods were used, reference the software, packages and version numbers used to implement synthesis methods (e.g. *metan* in Stata (92), *metafor* (version 2.1-0) in R (93)).
- If it was not possible to conduct a meta-analysis, describe and justify the synthesis methods (e.g. combining P values was used because no or minimal information beyond P values and direction of effect was reported in the studies) or summary approach used.
- If meta-analysis was done, specify:
 - the meta-analysis model (fixed-effect, fixed-effects or random-effects) and provide rationale for the selected model.
 - the method used (e.g. Mantel-Haenszel, inverse-variance) (94).
 - any methods used to identify or quantify statistical heterogeneity (e.g. visual inspection of results, a formal statistical test for heterogeneity (94), heterogeneity variance (τ^2), inconsistency (e.g. I^2 (95)), and prediction intervals (96)).
- If a random-effects meta-analysis model was used, specify:
 - the between-study (heterogeneity) variance estimator used (e.g. DerSimonian and Laird, restricted maximum likelihood (REML)).
 - the method used to calculate the confidence interval for the summary effect (e.g. Wald-type confidence interval, Hartung-Knapp-Sidik-Jonkman (97)).
- If a Bayesian approach to meta-analysis was used, describe the prior distributions about quantities of interest (e.g. intervention effect being analysed, amount of heterogeneity in results across studies) (94).
- If multiple effect estimates from a study were included in a meta-analysis (as may arise, for example, when a study reports multiple outcomes eligible for inclusion in a particular meta-analysis), describe the method(s) used to model or account for the statistical dependency (e.g. multivariate meta-analysis, multilevel models or robust variance estimation) (26, 58).

- If a planned synthesis was not considered possible or appropriate, report this and the reason for that decision.

Additional elements

- If a random-effects meta-analysis model was used, consider specifying other details about the methods used, such as the method for calculating confidence limits for the heterogeneity variance.

Example 1: meta-analysis

“As the effects of functional appliance treatment were deemed to be highly variable according to patient age, sex, individual maturation of the maxillofacial structures, and appliance characteristics, a random-effects model was chosen to calculate the average distribution of treatment effects that can be expected. A restricted maximum likelihood random-effects variance estimator was used instead of the older DerSimonian-Laird one, following recent guidance. Random-effects 95% prediction intervals were to be calculated for meta-analyses with at least three studies to aid in their interpretation by quantifying expected treatment effects in a future clinical setting. The extent and impact of between-study heterogeneity were assessed by inspecting the forest plots and by calculating the tau-squared and the I-squared statistics, respectively. The 95% CIs (uncertainty intervals) around tau-squared and the I-squared were calculated to judge our confidence about these metrics. We arbitrarily adopted the I-squared thresholds of >75% to be considered as signs of considerable heterogeneity, but we also judged the evidence for this heterogeneity (through the uncertainty intervals) and the localization on the forest plot...All analyses were run in Stata SE 14.0 (StataCorp, College Station, TX) by one author.”
(98)

Example 2: calculating the median effect across studies

“We based our primary analyses upon consideration of dichotomous process adherence measures (for example, the proportion of patients managed according to evidence-based recommendations). In order to provide a quantitative assessment of the effects associated with reminders without resorting to numerous assumptions or conveying a misleading degree of confidence in the results, we used the median improvement in dichotomous process adherence measures across studies...With each study represented by a single median outcome, we calculated the median effect size and interquartile range across all included studies for that comparison.” (99)

Item 13e. Describe any methods used to explore possible causes of heterogeneity among study results.

Explanation: If authors used methods to explore possible causes of variation of results across studies (i.e. statistical heterogeneity), such as subgroup analysis or meta-regression (see Box 5), they should provide sufficient details so that readers are able to assess the appropriateness of the selected methods and could reproduce the reported results (with access to the data). Such methods might be used to explore whether, for example, participant or intervention characteristics, or risk of bias of the included studies, explain variation in results.

Essential elements

- If methods were used to explore possible causes of statistical heterogeneity, specify the method used (e.g. subgroup analysis, meta-regression).
- If subgroup analysis or meta-regression was performed, specify for each:
 - which factors were explored, levels of those factors, and which direction of effect modification was expected and why (where possible).
 - whether analyses were conducted using study-level variables (i.e. where each study is included in one subgroup only), within-study contrasts (i.e. where data on subsets of participants within a study are available, allowing the study to be included in more than one subgroup), or some combination of the above (100).
 - how subgroup effects were compared (e.g. statistical test for interaction for subgroup analyses (94)).
- If other methods were used to explore heterogeneity because data were not amenable to meta-analysis of effect estimates, describe the methods used (e.g. structuring tables to examine variation in results across studies based on subpopulation, key intervention components or contextual factors), along with the factors and levels (83, 91).
- If any analyses used to explore heterogeneity were not pre-specified, identify them as such.

Example

“Given a sufficient number of trials, we used unadjusted and adjusted mixed-effects meta-regression analyses to assess whether variation among studies in smoking cessation effect size was moderated by tailoring of the intervention for disadvantaged groups. The resulting regression coefficient indicates how the outcome variable (log risk ratio (RR) for smoking cessation) changes when interventions take a socioeconomic-position-tailored versus non-socioeconomic-tailored approach. A statistically significant ($p < 0.05$) coefficient indicates that there is a linear association between the effect estimate for smoking cessation and the explanatory variable. More moderators (study-level variables) can be included in the model, which might account for part of the heterogeneity in the true effects. We pre-

planned an adjusted model to include important study covariates related to the intensity and delivery of the intervention (number of sessions delivered (above median vs below median), whether interventions involved a trained smoking cessation specialist (yes vs no), and use of pharmacotherapy in the intervention group (yes vs no). These covariates were included a priori as potential confounders given that programmes tailored to socioeconomic position might include more intervention sessions or components or be delivered by different professionals with varying experience. The regression coefficient estimates how the intervention effect in the socioeconomic-position-tailored subgroup differs from the reference group of non-socioeconomic-position-tailored interventions.” (101)

Item 13f. Describe any sensitivity analyses conducted to assess robustness of the synthesized results.

Explanation: If authors performed sensitivity analyses to assess robustness of the synthesized results to decisions made during the review process (see Box 5), they should provide sufficient details so that readers are able to assess the appropriateness of the analyses and could reproduce the reported results (with access to the data). Ideally, sensitivity analyses should be pre-specified in the protocol, but unexpected issues may emerge during the review process that necessitate their use.

Essential elements

- If sensitivity analyses were performed, provide details of each analysis (e.g. removal of studies at high risk of bias, use of an alternative meta-analysis model).
- If any sensitivity analyses were not pre-specified, identify them as such.

Example

“We conducted sensitivity meta-analyses restricted to trials with recent publication (2000 or later); overall low risk of bias (low risk of bias in all seven criteria); and enrolment of generally healthy women (rather than those with a specific clinical diagnosis). To incorporate trials with zero events in both intervention and control arms (which are automatically dropped from analyses of pooled relative risks), we also did sensitivity analyses for dichotomous outcomes in which we added a continuity correction of 0.5 to zero cells.” (102)

REPORTING BIAS ASSESSMENT

Item 14. Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases).

Explanation: The validity of a synthesis may be threatened when the available results differ systematically from the missing results. This is known as ‘bias due to missing results’ and arises from

‘reporting biases’ such as selective non-publication and selective non-reporting of results (see Box 4) (103). Direct methods for assessing the risk of bias due to missing results include comparing outcomes and analyses pre-specified in study registers, protocols, and statistical analysis plans with results that were available in study reports. Statistical and graphical methods exist to assess whether the observed data suggest potential for missing results (e.g. contour enhanced funnel plots, Egger’s test) and how robust the synthesis is to different assumptions about the nature of potentially missing results (e.g. selection models) (104-107). Tools (e.g. checklists, scales, or domain-based tools) that prompt users to consider some or all of these approaches are available (103, 104). Therefore, reporting any methods (tools, graphical, statistical or other) used to assess risk of bias due to missing results is recommended so that readers are able to assess how appropriate the methods were. The process by which assessments were conducted should also be reported to enable readers to assess the potential for errors and facilitate replicability.

Essential elements

- Specify the methods (tool, graphical, statistical or other) used to assess the risk of bias due to missing results in a synthesis (arising from reporting biases).
- If risk of bias due to missing results was assessed using an existing tool, specify the methodological components/domains/items of the tool, and the process used to reach a judgement of overall risk of bias.
- If any adaptations to an existing tool to assess risk of bias due to missing results were made (e.g. omitting or modifying items), specify the adaptations.
- If a new tool to assess risk of bias due to missing results was developed for use in the review, describe the content of the tool and make it publicly accessible.
- Report how many reviewers assessed risk of bias due to missing results in a synthesis, whether multiple reviewers worked independently, and any processes used to resolve disagreements between assessors.
- Report any processes used to obtain or confirm relevant information from study investigators.
- If an automation tool was used to assess risk of bias due to missing results, report how the tool was used, how the tool was trained, and details on the tool’s performance and internal validation.

Example

“To assess small-study effects, we planned to generate funnel plots for meta-analyses including at least 10 trials of varying size. If asymmetry in the funnel plot was detected, we planned to review the characteristics of the trials to assess whether the asymmetry was likely due to publication bias or other

factors such as methodological or clinical heterogeneity of the trials. To assess outcome reporting bias, we compared the outcomes specified in trial protocols with the outcomes reported in the corresponding trial publications; if trial protocols were unavailable, we compared the outcomes reported in the methods and results sections of the trial publications.” (108)

CERTAINTY ASSESSMENT

Item 15. Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome.

Explanation: Authors typically use some criteria to decide how certain (or confident) they are in the body of evidence for each important outcome. Common factors considered include precision of the effect estimate (or sample size), consistency of findings across studies, study design limitations and missing results (risk of bias), and how directly the studies address the question. Tools and frameworks can be used to provide a systematic, explicit approach to assessing these factors and provide a common approach and terminology for communicating certainty (109-112). For example, using the GRADE approach, authors will first apply criteria to assess each GRADE domain (imprecision, inconsistency, risk of bias and so forth) and then make an overall judgement of whether the evidence supporting a result is of high, moderate, low or very low certainty. Reporting the factors considered and the criteria used to assess each factor enables readers to determine which factors fed into reviewers’ assessment of certainty. Reporting the process by which assessments were conducted enables readers to assess the potential for errors and facilitates replication.

Essential elements

- Specify the tool or system (and version) used to assess certainty in the body of evidence.
- Report the factors considered (e.g. precision of the effect estimate, consistency of findings across studies) and the criteria used to assess each factor when assessing certainty in the body of evidence.
- Describe the decision rules used to arrive at an overall judgement of the level of certainty (e.g. high, moderate, low, very low), together with the intended interpretation (or definition) of each level of certainty (109).
- If applicable, report any review-specific considerations for assessing certainty, such as thresholds used to assess imprecision and the rationale for these thresholds (113).
- If any adaptations to an existing tool or system to assess certainty were made, specify the adaptations in sufficient detail that the approach is replicable.

- Report how many reviewers assessed the certainty of evidence, whether multiple reviewers worked independently, and any processes used to resolve disagreements between assessors.
- Report any processes used to obtain or confirm relevant information from investigators.
- If an automation tool was used to support the assessment of certainty, report how the automation tool was used, how the tool was trained, and details on the tool's performance and internal validation.
- Describe methods for reporting the results of assessments of certainty, such as the use of Summary of Findings tables (see item #22).
- If standard phrases that incorporate the certainty of evidence were used (e.g. "hip protectors *probably* reduce the risk of hip fracture slightly") (114), report the intended interpretation of each phrase and the reference for the source guidance.

Where a published system is adhered to, it may be sufficient to briefly describe the factors considered and the decision rules for reaching an overall judgement and reference the source guidance for full details of assessment criteria.

Example

"Two people (AM, JS) independently assessed the certainty of the evidence. We used the five GRADE considerations (study limitations, consistency of effect, imprecision, indirectness, and publication bias) to assess the certainty of the body of evidence as it related to the studies that contributed data to the meta-analyses for the prespecified outcomes. We assessed the certainty of evidence as high, moderate, low, or very low. We considered the following criteria for upgrading the certainty of evidence, if appropriate: large effect, dose-response gradient, and plausible confounding effect. We used the methods and recommendations described in sections 8.5 and 8.7, and chapters 11 and 12, of the Cochrane Handbook for Systematic Reviews of Interventions. We used GRADEpro GDT software to prepare the 'Summary of findings' tables (GRADEpro GDT 2015). We justified all decisions to down- or up-grade the certainty of studies using footnotes, and we provided comments to aid the reader's understanding of the results where necessary." (115)

STUDY SELECTION

Item 16a. Describe the results of the search and selection process, from the number of records identified in the search to the number of studies included in the review, ideally using a flow diagram (see Figure 1).

Explanation: Review authors should report, ideally with a flow diagram (see Figure 1), the results of the search and selection process so that readers can understand the flow of retrieved records through

to inclusion in the review. Such information is useful for future systematic review teams seeking to estimate resource requirements and for information specialists in evaluating their searches (116, 117). The flow diagram in Figure 1 provides a template of the flow of records through the review separated by source, although other layouts may be preferable depending on the information sources consulted (53).

Essential elements

- Report, ideally using a flow diagram, the number of: records identified; records excluded before screening (e.g. because they were duplicates, or deemed ineligible by machine classifiers); records screened; records excluded after screening titles or titles and abstracts; reports retrieved for detailed evaluation; potentially eligible reports that were not retrievable; retrieved reports that did not meet inclusion criteria and the primary reasons for exclusion (e.g. ineligible study design, ineligible population); and the number of studies and reports included in the review. If applicable, authors should also report the number of ongoing studies and associated reports identified.
- If the review is an update of a previous review, report results of the search and selection process for the current review and specify the number of studies included in the previous review. An additional box could be added to the flow diagram indicating the number of studies included in the previous review (see Figure 1) (118).
- If applicable, indicate in the PRISMA flow diagram how many records were excluded by a human and how many by automation tools.

Example

“We found 1,333 records in databases searching. After duplicates removal, we screened 1,092 records, from which we reviewed 34 full-text documents, and finally included six papers [each cited]. Later, we searched documents that cited any of the initially included studies as well as the references of the initially included studies. However, no extra articles that fulfilled inclusion criteria were found in these searches (a flow diagram is available at <https://doi.org/10.1371/journal.pone.0233220>).”
(119)

Item 16b. Cite studies that met many but not all inclusion criteria (‘near-misses’) and explain why they were excluded.

Explanation: Identifying the excluded records allows readers to make an assessment of the validity and applicability of the systematic review (30, 120). At a minimum, a list of excluded studies that readers might reasonably expect to see included (i.e. ‘near misses’) with citation and a reason for

exclusion, should be reported. This would include studies meeting most inclusion criteria (e.g. those with appropriate intervention and population but an ineligible control or study design). It is also useful to list studies that were potentially relevant but for which the full text or data essential to inform eligibility were not accessible. This information can be reported in the text or as list/table in the report or in an online supplement. Potentially contentious exclusions should be clearly stated in the report.

Essential elements

- Cite studies that met many but not all inclusion criteria ('near-misses') and explain why they were excluded.

Example

"We excluded seven studies from our review (Bosiers 2015; ConSeQuent; DEBATE-ISR; EXCITE ISR; NCT00481780; NCT02832024; RELINE), and we listed reasons for exclusion in the Characteristics of excluded studies tables. We excluded studies because they compared stenting in Bosiers 2015 and RELINE, laser atherectomy in EXCITE ISR, or cutting balloon angioplasty in NCT00481780 versus uncoated balloon angioplasty for in-stent restenosis. The ConSeQuent trial compared DEB versus uncoated balloon angioplasty for native vessel restenosis rather than in-stent restenosis. The DEBATE-ISR study compared a prospective cohort of patients receiving DEB therapy for in-stent restenosis against a historical cohort of diabetic patients. Finally, the NCT02832024 study compared stent deployment versus atherectomy versus uncoated balloon angioplasty alone for in-stent restenosis." (121)

STUDY CHARACTERISTICS

Item 17. Cite each included study and present its characteristics.

Explanation: Reporting the details of the included studies allows readers to understand the characteristics of studies that have addressed the review question(s) and is therefore important for understanding the applicability of the review. Characteristics of interest might include study design features, characteristics of participants, how outcomes were ascertained (e.g. smoking cessation self-reported or biochemically validated, or specific harms systematically assessed or reported by participants as they emerged), funding source and competing interests of study authors. Presenting the key characteristics of each study in a table or figure can facilitate comparison of characteristics across the studies (74). Citing each study enables retrieval of relevant reports if desired.

For systematic reviews of interventions, presenting an additional table that summarises the intervention details for each study (e.g. using the template based on the Template for Intervention Description and Replication (TIDieR) (63) has several benefits. An intervention summary table helps readers compare the characteristics of the interventions and consider those that may be feasible for implementation in their setting; highlights missing or unavailable details; shows which studies did not specify certain characteristics as part of the intervention; and highlights characteristics that have not been investigated in existing studies (63, 65).

Essential elements

- Cite each included study.
- Present the key characteristics of each study in a table or figure (considering a format that will facilitate comparison of characteristics across the studies).

Additional elements

- If the review examines the effects of interventions, consider presenting an additional table that summarises the intervention details for each study.

Example

In a review examining the association between aspirin use and fracture risk, the authors included a table presenting for each included study the citation, study design, country, sample size, setting, mean age, percentage of females, number of years follow-up, exposure details and outcomes assessed (Table 2) (122).

RISK OF BIAS IN STUDIES

Item 18. Present assessments of risk of bias for each included study.

Explanation: For readers to understand the internal validity of a systematic review's results, they need to know the risk of bias in results of each included study. Reporting only summary data (e.g. "two of eight studies successfully blinded participants") is inadequate because it fails to inform readers which studies had each particular methodological shortcoming. A more informative approach is to present tables or figures indicating for each study the risk of bias in each domain/component/item assessed (e.g. blinding of outcome assessors, missing outcome data), so that users can understand what factors led to the overall study-level risk of bias judgement (62, 123).

Essential elements

- Present tables or figures indicating for each study the risk of bias in each domain/component/item assessed and overall study-level risk of bias.
- Present justification for each risk of bias judgement, for example in the form of relevant quotations from reports of included studies.

Additional elements

- If assessments of risk of bias were done for specific outcomes or results in each study, consider displaying risk of bias judgements on a forest plot, next to the study results, so that the limitations of studies contributing to a particular meta-analysis are evident (see Sterne et al. (124) for an example forest plot).

Example

“We used the RoB 2.0 tool to assess risk of bias for each of the included studies. A summary of these assessments is provided in Table 3. In terms of overall risk of bias, there were concerns about risk of bias for the majority of studies (20/24), with two of these assessed as at high risk of bias (Musher-Eizenman 2010; Wansink 2013a). A text summary is provided below for each of the six individual components of the 'Risk of bias' assessment. Justifications for assessments are available at the following (<http://dx.doi.org/10.6084/m9.figshare.9159824>)” (69).

RESULTS OF INDIVIDUAL STUDIES

Item 19. For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) an effect estimate and its precision (e.g. confidence/credible interval), ideally using structured tables or plots.

Explanation: Presenting data from individual studies facilitates understanding of each study's contribution to the findings and reuse of the data by others seeking to perform additional analyses or perform an update of the review. There are different ways of presenting results of individual studies (e.g. table, forest plot) (83, 90). Visual display of results supports interpretation by readers, while tabulation of the results makes it easier for others to reuse the data.

Displaying summary statistics by group is helpful, because it allows an assessment of the severity of the problem in the studies (e.g. level of depression symptoms), which is not available from between-group results (i.e. effect estimates) (125). However, there are some scenarios where presentation of simple summary statistics for each group may be misleading. For example, in the case of cluster-randomized designs, the observed number of events and sample size in each group does not reflect

the effective sample size; that is, the sample size adjusted for correlation amongst observations. However, providing the estimated proportion of events (or another summary statistic) per group, will be helpful (126). The effect estimates from models that appropriately adjust for clustering (and other design features) should be reported and included in the meta-analysis in such instances.

Essential elements

- For all outcomes, irrespective of whether statistical synthesis was undertaken, present for each study summary statistics for each group (where appropriate). For dichotomous outcomes, report the number of participants with and without the events for each group; or the number with the event and the total for each group (e.g. 12/45). For continuous outcomes, report the mean, standard deviation and sample size of each group.
- For all outcomes, irrespective of whether statistical synthesis was undertaken, present for each study an effect estimate and its precision (e.g. standard error or 95% confidence/credible interval). For example, for time-to-event outcomes, present a hazard ratio and its confidence interval.
- If study-level data is presented visually or reported in the text (or both), also present a tabular display of the results.
- If results were obtained from multiple sources (e.g. journal article, study register entry, clinical study report, correspondence with authors), report the source of the data. This need not be overly burdensome. For example, a statement indicating that unless otherwise specified, all data came from the primary reference for each included study, would suffice. Alternatively, this could be achieved by, for example, presenting the origin of each data point in footnotes, in a column of the data table, or as a hyperlink to relevant text highlighted in reports (e.g. using SRDR Data Abstraction Assistant (127)).
- If applicable, indicate which results were not reported directly and had to be computed or estimated from other information (see item #13b).

Example

For an example of individual study results presented for a dichotomous outcome, see Figure 2. For an example of individual study results presented for a continuous outcome, see Figure 3. (128)

RESULTS OF SYNTHESSES

Item 20a. For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies.

Explanation: Many systematic review reports include narrative summaries of the characteristics and risk of bias across *all* included studies (25). However, such general summaries are not useful when the studies contributing to each synthesis vary, and particularly when there are many studies. For example, one meta-analysis might include three studies of participants aged 30 years on average, whereas another meta-analysis might include 10 studies of participants aged 60 years on average; in this case, knowing the mean age per synthesis is more meaningful than the overall mean age across all 13 studies. Providing a brief summary of the characteristics and risk of bias among studies contributing to each synthesis (meta-analysis or other) should help readers understand the applicability and risk of bias in the synthesized result. Furthermore, a summary at the level of the synthesis is more usable since it obviates the need for readers to refer to multiple sections of the review in order to interpret results (74).

Essential elements

- Provide a brief summary of the characteristics and risk of bias among studies contributing to each synthesis (meta-analysis or other). The summary should focus only on study characteristics that help in interpreting the results (especially those that suggest the evidence addresses only a restricted part of the review question, or indirectly addresses the question). If the same set of studies contribute to more than one synthesis, or if the same risk of bias issues are relevant across studies for different syntheses, such a summary need be provided once only.
- Indicate which studies were included in each synthesis (e.g. by listing each study in a forest plot or table or citing studies in the text).

Example

“Nine randomized controlled trials (RCTs) directly compared delirium incidence between haloperidol and placebo groups [9 studies cited]. These RCTs enrolled 3,408 patients in both surgical and medical intensive care and non-intensive care unit settings and used a variety of validated delirium detection instruments. Five of the trials were low risk of bias [5 studies cited], three had unclear risk of bias [3 studies cited], and one had high risk of bias owing to lack of blinding and allocation concealment [1 study cited]. Intravenous haloperidol was administered in all except two trials; in those two exceptions, oral doses were given [two studies cited]. These nine trials were pooled, as they each identified new onset of delirium (incidence) within the week after exposure to prophylactic haloperidol or placebo.”
(129)

Item 20b. Present results of all statistical syntheses conducted. If meta-analysis was done, present for each the summary estimate and its precision (e.g. confidence/credible interval) and measures of statistical heterogeneity. If comparing groups, describe the direction of the effect.

Explanation: Users of reviews rely on the reporting of all statistical syntheses conducted so that they have complete and unbiased evidence on which to base their decisions. Studies examining selective reporting of results in systematic reviews have found that 11% to 22% of reviews did not present results for at least one pre-specified outcome of the review (130-133).

Essential elements

- Report results of all statistical syntheses described in the protocol and all syntheses conducted that were not pre-specified.
- If meta-analysis was conducted, report for each:
 - the summary estimate and its precision (e.g. standard error or 95% confidence/credible interval).
 - measures of statistical heterogeneity (e.g. τ^2 , I^2 , prediction interval).
- If other statistical synthesis methods were used (e.g. summarising effect estimates, combining P values), report the synthesized result and a measure of precision (or equivalent information, for example, the number of studies and total sample size).
- If the statistical synthesis method does not yield an estimate of effect (e.g. as is the case when P values are combined), report the relevant statistics (e.g. P value from the statistical test), along with an interpretation of the result that is consistent with the question addressed by the synthesis method (e.g. “There was strong evidence of benefit of the intervention *in at least one study* (P < 0.001, 10 studies)” when P values have been combined) (83).
- If comparing groups, describe the direction of effect (e.g. fewer events in the intervention group, or higher pain in the comparator group).
- If synthesising mean differences, specify for each synthesis, where applicable, the unit of measurement (e.g. kilograms or pounds for weight), the upper and lower limits of the measurement scale (e.g. anchors range from 0 to 10), direction of benefit (e.g. higher scores denote higher severity of pain), and the minimal clinically important difference, if known.

Example

“Twelve studies, including a total of 159,086 patients, reported on the rate of major bleeding complications. Aspirin use was associated with a 46% relative risk increase of major bleeding

complications (risk ratio 1.46; 95% CI, 1.30-1.64; p < 0.00001; I² = 31%; absolute risk increase 0.077%; number needed to treat to harm 1295)" (134)

Item 20c. Present results of all investigations of possible causes of heterogeneity among study results.

Explanation: Presenting results from all investigations of possible causes of heterogeneity among study results is important for users of reviews and for future research. For users, understanding the factors that may, and equally, may not, explain variability in the effect estimates, may inform decision making. Similarly, presenting all results is important for designing future studies. For example, the results may help to generate hypotheses about potential modifying factors that can be tested in future studies, or help identify 'active' intervention ingredients that might be combined and tested in a future randomized trial. Selective reporting of the results leads to an incomplete representation of the evidence that risks misdirecting decision making and future research.

Essential elements

- If investigations of possible causes of heterogeneity were conducted:
 - present results regardless of the statistical significance, magnitude, or direction of effect modification.
 - identify the studies contributing to each subgroup.
 - report results with due consideration to the observational nature of the analysis and risk of confounding due to other factors (135, 136).
- If subgroup analysis was conducted, report for each analysis the exact P value for a test for interaction, as well as, within each subgroup, the summary estimates, their precision (e.g. standard error or 95% confidence/credible interval) and measures of heterogeneity. Results from subgroup analyses might usefully be presented graphically (see Fisher et al. (100)).
- If meta-regression was conducted, report for each analysis the exact P value for the regression coefficient and its precision.
- If informal methods (i.e. those that do not involve a formal statistical test) were used to investigate heterogeneity – which may arise particularly when the data are not amenable to meta-analysis – describe the results observed. For example, present a table that groups study results by dosage or overall risk of bias, and comment on any patterns observed (91).

Additional elements

- If subgroup analysis was conducted, consider presenting the estimate for the difference between subgroups and its precision.
- If meta-regression was conducted, consider presenting a meta-regression scatterplot with the study effect estimates plotted against the potential effect modifier (135).

Example

“Among the 4 trials that recruited critically ill patients who were and were not receiving invasive mechanical ventilation at randomization, the association between corticosteroids and lower mortality was less marked in patients receiving invasive mechanical ventilation (ratio of odds ratios (ORs), 4.34 [95% CI, 1.46-12.91]; $P = .008$ based on within-trial estimates combined across trials); however, only 401 patients (120 deaths) contributed to this comparison...All trials contributed data according to age group and sex. For the association between corticosteroids and mortality, the OR was 0.69 (95% CI, 0.51-0.93) among 880 patients older than 60 years, the OR was 0.67 (95% CI, 0.48-0.94) among 821 patients aged 60 years or younger (ratio of ORs, 1.02 [95% CI, 0.63-1.65], $P = .94$), the OR was 0.66 (95% CI, 0.51-0.84) among 1215 men, and the OR was 0.66 (95% CI, 0.43-0.99) among 488 women (ratio of ORs, 1.07 [95% CI, 0.58-1.98], $P = .84$).” (137)

Item 20d. Present results of all sensitivity analyses conducted to assess the robustness of the synthesized results.

Explanation: Presenting results of sensitivity analyses conducted allows readers to assess how robust the synthesized results were to decisions made during the review process. Reporting results of all sensitivity analyses is important; presentation of a subset, based on the nature of the results, risks introducing bias due to selective reporting. Forest plots are a useful way to present results of sensitivity analyses; however, generally, these may be best placed in an appendix, with the main forest plots presented in the main report, to not reduce readability. An exception may be when sensitivity analyses reveal the results are not robust to decisions made during the review process.

Essential elements

- If any sensitivity analyses were conducted:
 - report the results for each sensitivity analysis.
 - comment on how robust the main analysis was given the results of all corresponding sensitivity analyses.

Additional elements

- If any sensitivity analyses were conducted, consider:
 - presenting results in tables that indicate: (i) the summary effect estimate, a measure of precision (and potentially other relevant statistics, for example, I^2 statistic) and contributing studies for the original meta-analysis; (ii) the same information for the sensitivity analysis; and (iii) details of the original and sensitivity analysis assumptions.
 - presenting results of sensitivity analyses visually using forest plots.

Example

“Sensitivity analyses that removed studies with potential bias showed consistent results with the primary meta-analyses (risk ratio 1.00 for undetectable HIV-1 RNA, 1.00 for virological failure, 0.98 for severe adverse effects, and 1.02 for AIDS defining events; supplement 3E, 3F, 3H, and 3I, respectively). Such sensitivity analyses were not performed for other outcomes because none of the studies reporting them was at a high risk of bias. Sensitivity analysis that pooled the outcome data reported at 48 weeks, which also showed consistent results, was performed for undetectable HIV-1 RNA and increase in CD4 T cell count only (supplement 3J and 3K) and not for other outcomes owing to lack of relevant data. When the standard deviations for increase in CD4 T cell count were replaced by those estimated by different methods, the results of figure 3 either remained similar (that is, quadruple and triple arms not statistically different) or favoured triple therapies (supplement 2).” (128)

RISK OF REPORTING BIASES IN SYNTHESSES

Item 21. Present assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis assessed.

Explanation: Presenting assessments of the risk of bias due to missing results in syntheses allows readers to assess potential threats to the trustworthiness of a systematic review’s results. Providing the evidence used to support judgements of risk of bias allows readers to determine the validity of the assessments.

Essential elements

- Present assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis assessed.
- If a tool was used to assess risk of bias due to missing results in a synthesis, present responses to questions in the tool, judgements about risk of bias and any information used to support such judgements, to help readers understand why particular judgements were made.

- If a funnel plot was generated to evaluate small-study effects (one cause of which is reporting biases), present the plot and specify the effect estimate and measure of precision used in the plot (presented typically on the horizontal axis and vertical axis, respectively (138)). If a contour-enhanced funnel plot was generated, specify the ‘milestones’ of statistical significance that the plotted contour lines represent ($P = 0.01, 0.05, 0.1$, etc.) (139).
- If a test for funnel plot asymmetry was used, report the exact P value observed for the test, and potentially other relevant statistics, for example the standardised normal deviate, from which the P value is derived (138).
- If any sensitivity analyses seeking to explore the potential impact of missing results on the synthesis were conducted, present results of each analysis (see item #20d), compare them with results of the primary analysis, and report results with due consideration of the limitations of the statistical method (106).

Additional elements

- If studies were assessed for selective non-reporting of results by comparing outcomes and analyses pre-specified in study registers, protocols, and statistical analysis plans with results that were available in study reports, consider presenting a matrix (with rows as studies and columns as syntheses) to present the availability of study results (107).
- If an assessment of selective non-reporting of results reveals that some studies are missing from the synthesis, consider displaying the studies with missing results underneath a forest plot or including a table with the available study results (e.g. see Page et al. (103) for an example forest plot).

Example

“Clinical global impression of change was assessed in Doody 2008, NCT00912288, CONCERT and CONNECTION using the CIBIC-Plus. However, we were only able to extract results from Doody 2008 [because no results for CIBIC-Plus were reported in the other three studies]...The authors reported small but significant improvements on the CIBIC-Plus for 183 patients (89 on latrepirdine and 94 on placebo) favouring latrepirdine following the 26-week primary endpoint (MD -0.60, 95% CI -0.89 to -0.31, $P < 0.001$). Similar results were found at the additional 52-week follow-up (MD -0.70, 95% CI -1.01 to -0.39, $P < 0.001$). However, we considered this to be low quality evidence due to imprecision and reporting bias. Thus, we could not draw conclusions about the efficacy of latrepirdine in terms of changes in clinical impression.” (140)

CERTAINTY OF EVIDENCE

Item 22. Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed.

Explanation: An important feature of systems for assessing certainty, such as GRADE, is explicit reporting of both the level of certainty (or confidence) in the evidence and the basis for judgments (81, 82, 111). Evidence summary tables, such as GRADE Summary of Findings tables, are an effective and efficient way to report assessments of the certainty of evidence (81, 111, 141, 142).

Essential elements

- Report the overall level of certainty in the body of evidence (e.g. high, moderate, low, or very low) for each important outcome.
- Provide an explanation of reasons for downgrading (or upgrading) the evidence (e.g. in footnotes to an evidence summary table). Explanations for each judgement should be concise, informative, relevant to the target audience, easy to understand, and accurate (i.e. addressing criteria specified in the methods guidance) (143).
- Communicate certainty in the evidence wherever results are reported (i.e. abstract, evidence summary tables, results, conclusions). Use a format appropriate for the section of the review. For example, in text, certainty might be reported explicitly in a sentence (e.g. ‘*Moderate-certainty evidence* (downgraded for bias) indicates that...’) or in brackets alongside an effect estimate (e.g. [RR 1.17, 95% CI 0.81 to 1.68; 4 studies, 1781 participants; *moderate certainty evidence*]). When interpreting results in Summary of Findings tables or conclusions, certainty might be communicated implicitly using standard phrases (e.g. “hip protectors *probably* reduce the risk of hip fracture slightly”) (114).

Additional elements

- Consider including evidence summary tables, such as GRADE Summary of Findings tables.

Example

“Compared with non-operative treatment, low-certainty evidence indicates surgery (repair with subacromial decompression) may have little or no effect on function at 12 months. The evidence was downgraded two steps, once for bias and once for imprecision – the 95% CIs overlap minimal important difference in favour of surgery at this time point.” A summary of findings table presents the same information as the text above, with footnotes explaining judgements. (108)

DISCUSSION

Item 23a. Provide a general interpretation of the results in the context of other evidence.

Explanation: Discussing how the results of the review relate to other relevant evidence should help readers interpret the findings. For example, authors might compare the current results to results of other similar systematic reviews (e.g. that addressed the same question using different methods or that addressed slightly different questions) and explore possible reasons for discordant results. Similarly, authors might summarise additional information relevant to decision makers that was not explored in the review, such as findings of studies evaluating the cost-effectiveness of the intervention or surveys gauging the values and preferences of patients.

Essential elements

- Provide a general interpretation of the results in the context of other evidence.

Example

“Although we need to exercise caution in interpreting these findings because of the small number of studies, these findings nonetheless appear to be largely in line with the recent systematic review on what works to improve education outcomes in low- and middle-income countries of Snilstveit et al. (2012). They found that structured pedagogical interventions may be among the effective approaches to improve learning outcomes in low- and middle-income countries. This is consistent with our findings that teacher training is only effective in improving early grade literacy outcomes when it is combined with teacher coaching. The finding is also consistent with our result that technology in education programs may have at best no effects unless they are combined with a focus on pedagogical practices. In line with our study, Snilstveit et al. (2012) also do not find evidence for statistically significant effects of the one-laptop-per-child program. These results are consistent with the results of a meta-analysis showing that technology in education programs are not effective when not accompanied by parent or student training (McEwan, 2015). However, neither Snilstveit et al. (2012) nor McEwan (2015) find evidence for negative effects of the one-laptop-per-child program on early grade literacy outcomes.”

(144)

Item 23b. Discuss any limitations of the evidence included in the review.

Explanation: Discussing the completeness, applicability and uncertainties in the evidence included in the review should help readers interpret the findings appropriately. For example, authors might acknowledge that they identified few eligible studies or studies with a small number of participants, leading to imprecise estimates; have concerns about risk of bias in studies or missing results; or

identified studies that only partially, or indirectly, address the review question, leading to concerns about their relevance and applicability to particular patients, settings or other target audiences. The assessments of certainty (or confidence) in the body of evidence (item #22) can support the discussion of such limitations.

Essential elements

- Discuss any limitations of the evidence included in the review.

Example

“Study populations were young, and few studies measured longitudinal exposure. The included studies were often limited by selection bias, recall bias, small sample of marijuana-only smokers, reporting of outcomes on marijuana users and tobacco users combined, and inadequate follow-up for the development of cancer...Most studies poorly assessed exposure, and some studies did not report details on exposure, preventing meta-analysis for several outcomes.” (145)

Item 23c. Discuss any limitations of the review processes used.

Explanation: Discussing limitations, avoidable or unavoidable, in the review process should help readers understand the trustworthiness of the review findings. For example, authors might acknowledge the decision to restrict eligibility to studies in English only, search only a small number of databases, have only one reviewer screen records or collect data, or not contact study authors to clarify unclear information. They might also acknowledge that they were unable to access all potentially eligible study reports, or carry out some of the planned analyses because of insufficient data (146, 147). While some limitations may impact the validity of the review findings, others may not.

Essential elements

- Discuss any limitations of the review processes used, and comment on the potential impact of each limitation.

Example

“Because of time constraints...we dually screened only 30% of the titles and abstracts; for the rest, we used single screening. A recent study showed that single abstract screening misses up to 13% of relevant studies (Gartlehner 2020). In addition, single review authors rated risk of bias, conducted data extraction and rated certainty of evidence. A second review author checked the plausibility of decisions and the correctness of data. Because these steps were not conducted dually and independently, we

introduced some risk of error...Nevertheless, we are confident that none of these methodological limitations would change the overall conclusions of this review. Furthermore, we limited publications to English and Chinese languages. Because COVID-19 has become a rapidly evolving pandemic, we might have missed recent publications in languages of countries that have become heavily affected in the meantime (e.g. Italian or Spanish).” (148)

Item 23d. Discuss implications of the results for practice, policy, and future research.

Explanation: There are many potential end-users of a systematic review (e.g. patients, health care providers, researchers, insurers and policy makers), each of whom will want to know what actions they should take given the review findings. Patients and health care providers may be primarily interested in the balance of benefits and harms, while policy makers and administrators may value data on organizational impact and resource utilization. For reviews of interventions, authors might clarify trade-offs between benefits and harms and how the values attached to the most important outcomes of the review might lead different people to make different decisions. In addition, rather than making recommendations for practice or policy that apply universally, authors might discuss factors that are important in translating the evidence to different settings and factors that may modify the magnitude of effects.

Explicit recommendations for future research, as opposed to general statements such as “More research on this question is needed”, can better direct the questions future studies should address and the methods that should be used. For example, authors might consider describing the type of understudied participants who should be enrolled in future studies, the specific interventions that could be compared, suggested outcome measures to use, and ideal study design features to employ.

Essential elements

- Discuss implications of the results for practice and policy.
- Make explicit recommendations for future research.

Example

“Implications for practice and policy: Findings from this review indicate that bystander programs have significant beneficial effects on bystander intervention behaviour. This provides important evidence of the effectiveness of mandated programs on college campuses. Additionally, the fact that our (preliminary) moderator analyses found program effects on bystander intervention to be similar for adolescents and college students suggests early implementation of bystander programs (i.e. in

secondary schools with adolescents) may be warranted. Importantly, although we found that bystander programs had a significant beneficial effect on bystander intervention behaviour, we found no evidence that these programs had an effect on participants' sexual assault perpetration. Bystander programs may therefore be appropriate for targeting bystander behaviour, but may not be appropriate for targeting the behaviour of potential perpetrators. Additionally, effects of bystander programs on bystander intervention behaviour diminished by 6-month post-intervention. Thus, programs effects may be prolonged by the implementation of booster sessions conducted prior to 6 months post-intervention.

Implications for research: Findings from this review suggest there is a fairly strong body of research assessing the effects of bystander programs on attitudes and behaviours. However, there are a couple of important questions worth further exploration...Our understanding of the causal mechanisms of program effects on bystander behaviour would benefit from further analysis (e.g., path analysis mapping relationships between specific knowledge/attitude effects and bystander intervention)...Our understanding of the differential effects of gendered versus gender neutral programs would benefit from the design and implementation of high-quality primary studies that make direct comparisons between these two types of programs (e.g., RCTs comparing the effects of two active treatment arms that differ in their gendered approach)...Our understanding of bystander programs' generalizability to non-US contexts would be greatly enhanced by high quality research conducted across the world.” (149)

REGISTRATION AND PROTOCOL

Item 24a. Provide registration information for the review, including register name and registration number, or state that the review was not registered.

Explanation: Stating where the systematic review was registered (e.g. PROSPERO, Open Science Framework) and the registration number or DOI for the register entry (see Box 6) facilitates identification of the systematic review in the register. This allows readers to compare what was pre-specified with what was eventually reported in the review and decide if any deviations may have introduced bias. Reporting registration information also facilitates linking of publications related to the same systematic review (e.g. when a review is presented at a conference and published in a journal) (150).

Essential elements

- Provide registration information for the review, including register name and registration number, or state that the review was not registered.

Example

“...this systematic review has been registered in the international prospective register of systematic reviews (PROSPERO) under the registration number: CRD42019128569” (151)

Item 24b. Indicate where the review protocol can be accessed, or state that a protocol was not prepared.

Explanation: The review protocol may contain information about the methods that is not provided in the final review report (see Box 6). Providing a citation, DOI or link to the review protocol allows readers to locate the protocol more easily. Comparison of the methods pre-specified in the review protocol with what was eventually done allows readers to assess whether any deviations may have introduced bias (152). If the review protocol was not published or deposited in a public repository, we recommend providing the contact details of the author responsible for sharing the protocol. If authors did not prepare a review protocol, this should be stated, to prevent users spending time trying to locate the document.

Essential elements

- Indicate where the review protocol can be accessed (e.g. by providing a citation, DOI or link), or state that a protocol was not prepared.

Example

“...this systematic review and meta-analysis protocol has been published elsewhere [citation for the protocol provided].” (153)

Item 24c. Describe and explain any amendments to information provided at registration or in the protocol.

Explanation: Careful consideration of a review’s methodological and analytical approach early on is likely to lessen unnecessary changes after protocol development (154). However, it is difficult to anticipate all scenarios that will arise, necessitating some clarifications, modifications and changes to the protocol (e.g. data available may not be amenable to the planned meta-analysis) (152, 155). For reasons of transparency, authors should report details of any amendments. Amendments could be recorded in various places, including the full text of review, a supplementary file, or as amendments to the published protocol or registration record.

Essential elements

- Report details of any amendments to information provided at registration or in the protocol, noting: (a) the amendment itself; (b) the reason for the amendment; and (c) the stage of the review process at which the amendment was implemented.

Example

“Differences from protocol: We modified the lower limit for age in our eligibility criteria from 12 years of age to 10 years of age because the age of adolescence was reduced. We used the WHO measures for severe anaemia, defined by haemoglobin levels < 80 g/L instead of < 70 g/L as stated in the protocol. We decided to add adverse events to our list of primary outcomes (instead of secondary) and we changed reinfection rate to a secondary outcome.” (156)

SUPPORT

Item 25. Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the review.

Explanation: As with any research report, authors should be transparent about the sources of support received to conduct the review. For example, funders may provide salary to researchers to undertake the review, the services of an information specialist to conduct searches, or access to commercial databases that would otherwise not have been available. Authors may have also obtained support from a translation service to translate articles or in-kind use of software to manage or analyse the study data. In some reviews, the funder or sponsor (i.e. individual or organisation assuming responsibility for the initiation and management of the review) may have contributed to defining the review question, determining eligibility of studies, collecting data, analysing data, interpreting results, or approving the final review report. There is potential for bias in the review findings arising from such involvement, particularly when the funder or sponsor have an interest in obtaining a particular result (157).

Essential elements

- Describe sources of financial or non-financial support for the review, specifying relevant grant ID numbers for each funder. If no specific financial or non-financial support was received, this should be stated.
- Describe the role of the funders or sponsors (or both) in the review. If funders or sponsors had no role in the review, this should be declared, for example by stating, “The funders had no role in the

design of the review, data collection and analysis, decision to publish, or preparation of the manuscript.”

Example

“Funding/Support: This research was funded under contract HHS290201500009i, Task Order 7, from the Agency for Healthcare Research and Quality (AHRQ), US Department of Health and Human Services, under a contract to support the US Preventive Services Task Force (USPSTF). Role of the Funder/Sponsor: Investigators worked with USPSTF members and AHRQ staff to develop the scope, analytic framework, and key questions for this review. AHRQ had no role in study selection, quality assessment, or synthesis. AHRQ staff provided project oversight, reviewed the report to ensure that the analysis met methodological standards, and distributed the draft for peer review. Otherwise, AHRQ had no role in the conduct of the study; collection, management, analysis, and interpretation of the data; and preparation, review, or approval of the manuscript findings. The opinions expressed in this document are those of the authors and do not reflect the official position of AHRQ or the US Department of Health and Human Services.” (158)

COMPETING INTERESTS

Item 26. Declare any competing interests of review authors.

Explanation: Authors of a systematic review may have relationships with organisations or entities with an interest in the review findings (e.g. an author may serve as a consultant for a company manufacturing the drug or device under review) (159). Such relationships or activities are examples of a competing interest (or conflict of interest), which can negatively affect the integrity and credibility of systematic reviews. For example, evidence suggests that systematic reviews with financial competing interests more often have conclusions favourable to the experimental intervention than systematic reviews without financial competing interests (160). Information about authors’ relationships or activities that readers could consider pertinent or to have influenced the review should be disclosed using the format requested by the publishing entity (e.g. using the International Committee of Medical Journal Editors (ICMJE) disclosure form) (161). Authors should report how competing interests were managed for particular review processes. For example, if a review author was an author of an included study, they may have been prevented from assessing the risk of bias in the study results.

Essential elements

- Disclose any of the authors' relationships or activities that readers could consider pertinent or to have influenced the review.
- If any authors had competing interests, report how they were managed for particular review processes.

Example

"Declarations of interest: R Buchbinder was a principal investigator of Buchbinder 2009. D Kallmes was a principal investigator of Kallmes 2009 and Evans 2015. D Kallmes participated in IDE trial for Benvenue Medical spinal augmentation device. He is a stockholder, Marblehead Medical, LLC, Development of spine augmentation devices. He holds a spinal fusion patent license, unrelated to spinal augmentation/vertebroplasty. R Buchbinder and D Kallmes did not perform risk of bias assessments for their own or any other placebo-controlled trials included in the review." (162)

AVAILABILITY OF DATA, CODE AND OTHER MATERIALS

Item 27. Report which of the following are publicly available and where they can be found: template data collection forms; data extracted from included studies; data used for all analyses; analytic code; any other materials used in the review.

Explanation: Sharing of data, analytic code and other materials enables others to reuse the data, check the data for errors, attempt to reproduce the findings, and understand more about the analysis than may be provided by descriptions of methods (163, 164). Support for sharing of data, analytic code and other materials is growing, including from patients (165) and journal editors, including BMJ and PLOS Medicine (166).

Sharing of data, analytic code and other materials relevant to a systematic review includes making various items publicly available, such as the template data collection forms; all data extracted from included studies; a file indicating necessary data conversions; the clean dataset(s) used for all analyses in a format ready for reuse (e.g. CSV file); metadata (e.g. complete descriptions of variable names, README files describing each file shared); analytic code used in software with a command-line interface, or complete descriptions of the steps used in point-and-click software to run all analyses. Other materials might include more detailed information about the intervention delivered in the primary studies that are otherwise not available, such as a video of the specific cognitive behavioural therapy supplied by the study investigators to reviewers (63). Similarly, other material might include a list of all citations screened and any decisions about eligibility.

Because sharing of data, analytic code and other materials is not yet universal in health and medical research (166), even interested authors may not know how to make their materials publicly available. Data, analytic code and other materials can be uploaded to one of several publicly accessible repositories (e.g. Open Science Framework, Dryad, figshare). The Systematic Review Data Repository (<https://srdp.ahrq.gov/>) is another example of a platform for sharing materials specific to the systematic review community (167). All of these open repositories should be given consideration, particularly if the completed review is to be considered for publication in a paywalled journal. The Findable, Accessible, Interoperable, Reusable (FAIR) data principles are also a useful resource for authors to consult (168), as they provide guidance on the best way to share information.

There are some situations where authors might not be able to share review materials, such as when the review team are custodians rather than owners of individual participant data, or when there are legal or licencing restrictions. For example, records exported directly from bibliographic databases (e.g. Ovid MEDLINE) typically include copyrighted material; authors should read the licensing terms of the databases they search to see what they can share and to consider the copyright legislation of their countries.

Essential elements

- Report which of the following are publicly available: template data collection forms; data extracted from included studies; data used for all analyses; analytic code; any other materials used in the review.
- If any of the above materials are publicly available, report where they can be found (e.g. provide a link to files deposited in a public repository).
- If data, analytic code, or other materials will be made available upon request, provide the contact details of the author responsible for sharing the materials and describe the circumstances under which such materials will be shared.

Example

“All meta-analytic data and all codebooks and analysis scripts (for Mplus and R) are publicly available at the study’s associated page on the Open Science Framework (<https://osf.io/r8a24/>)...The precise sources (table, section, or paragraph) for each estimate are described in notes in the master data spreadsheet, available on the Open Science Framework page for this study (<https://osf.io/r8a24/>)”
(169)

CONCLUSION TO PRISMA 2020 EXPLANATION AND ELABORATION

This explanation and elaboration paper has been designed to assist authors seeking comprehensive guidance on what to include in systematic review reports. We hope that use of this resource will lead to more transparent, complete and accurate reporting of systematic reviews, thus facilitating evidence-based decision-making.

DECLARATIONS

Acknowledgements

We thank the following contributors who completed the survey to inform discussions at the development meeting: Xavier Armoiry, Edoardo Aromataris, Ana Patricia Ayala, Ethan M Balk, Virginia Barbour, Elaine Beller, Jesse A Berlin, Lisa Bero, Zhao-Xiang Bian, Jean Joel Bigna, Ferrán Catalá-López, Anna Chaimani, Mike Clarke, Tammy Clifford, Ioana A Cristea, Miranda Cumpston, Sofia Dias, Corinna Dressler, Ivan D Florez, Joel J Gagnier, Chantelle Garritty, Long Ge, Davina Gherzi, Sean Grant, Gordon Guyatt, Neal R Haddaway, Julian PT Higgins, Sally Hopewell, Brian Hutton, Jamie J Kirkham, Jos Kleijnen, Julia Koricheva, Joey SW Kwong, Toby J Lasserson, Julia H Littell, Yoon K Loke, Malcolm R Macleod, Chris G Maher, Ana Marušić, Dimitris Mavridis, Jessie McGowan, Matthew DF McInnes, Philippa Middleton, Karel G Moons, Zachary Munn, Jane Noyes, Barbara Nußbaumer-Streit, Donald L Patrick, Tatiana Pereira-Cenci, Ba' Pham, Bob Phillips, Dawid Pieper, Michelle Pollock, Daniel S Quintana, Drummond Rennie, Melissa L Rethlefsen, Hannah R Rothstein, Maroeska M Rovers, Rebecca Ryan, Georgia Salanti, Ian J Saldanha, Margaret Sampson, Nancy Santesso, Rafael Sarkis-Onofre, Jelena Savović, Christopher H Schmid, Kenneth F Schulz, Guido Schwarzer, Beverley J Shea, Paul G Shekelle, Farhad Shokraneh, Mark Simmonds, Nicole Skoetz, Sharon E Straus, Anneliese Synnot, Emily E Tanner-Smith, Brett D Thombs, Hilary Thomson, Alexander Tsertsvadze, Peter Tugwell, Tari Turner, Lesley Uttley, Jeffrey C Valentine, Matt Vassar, Areti Angeliki Veroniki, Meera Viswanathan, Cole Wayant, Paul Whaley, and Kehu Yang. We thank the following contributors who provided feedback on a preliminary version of the PRISMA 2020 checklist: Jo Abbott, Fionn Büttner, Patricia Correia-Santos, Victoria Freeman, Emily A Hennessy, Rakibul Islam, Amalia (Emily) Karahalios, Kasper Krommes, Andreas Lundh, Dafne Port Nascimento, Davina Robson, Catherine Schenck-Yglesias, Mary M Scott, Sarah Tanveer and Pavel Zhelnov.

Funding

MJP is supported by an Australian Research Council Discovery Early Career Researcher Award (DE200101618) and was previously supported by an Australian National Health and Medical Research Council (NHMRC) Early Career Fellowship (1088535) during the conduct of this research. JEM is supported by an Australian NHMRC Career Development Fellowship (1143429). TCH is supported by an Australian NHMRC Senior Research Fellowship (1154607). JMT is supported by Evidence Partners Inc. JMG is supported by a Tier 1 Canada Research Chair in Health Knowledge Transfer and Uptake. MML is supported by The Ottawa Hospital Anesthesia Alternate Funds Association and a Faculty of Medicine Junior Research Chair. TL is supported by funding from the National Eye Institute (UG1EY020522), National Institutes of Health, United States. LAM is supported by a National Institute for Health Research Doctoral Research Fellowship (DRF-2018-11-ST2-048). ACT is supported by a Tier 2 Canada Research Chair in Knowledge Synthesis. DM is supported in part by a University Research Chair, University of Ottawa. The funders had no role in the study design, data collection and analysis, or preparation of the manuscript.

Competing interests

All authors have completed the ICMJE uniform disclosure form at <http://www.icmje.org/conflicts-of-interest/> and declare: EL is head of research for the *BMJ*, MJP is an editorial board member for *PLOS Medicine*, ACT is an associate editor and MJP, TL, EMW, and DM are editorial board members for the *Journal of Clinical Epidemiology*, and DM and LAS were editors in chief, LS, JMT and ACT are associate editors, and JG is an editorial board member for *Systematic Reviews*; none of these authors were involved in the peer review process or decision to publish. TCH has received personal fees from Elsevier outside the submitted work. EMW has received personal fees from the *American Journal for Public Health*, for which he is the editor for systematic reviews. VW is editor in chief of the Campbell Collaboration which produces systematic reviews and co-convenor of the Campbell and Cochrane equity methods group. DM is chair of the EQUATOR Network, IB is adjunct director of the French EQUATOR Centre and TCH is co-director of the Australasian EQUATOR Centre, which advocate for the use of reporting guidelines to improve the quality of reporting in research articles. JMT received salary from Evidence Partners Inc., creators of DistillerSR software for systematic reviews; Evidence Partners Inc. was not involved in the design or outcomes of the statement and the views expressed solely represent those of the author.

Author contributions

All authors declare to meet the ICMJE conditions for authorship. MJP, JEM, PMB, IB, TCH, CDM, LS and DM conceived this paper and designed the literature review and survey conducted to inform the

guideline content. MJP conducted the literature review, administered the survey and analysed the data for both. MJP prepared all materials for the development meeting. MJP and JEM presented proposals at the development meeting. All authors except for TCH, JMT, EAA, SEB and LAM attended the development meeting. MJP and JEM took and consolidated notes from the development meeting. MJP and JEM led the drafting and editing of the article. JEM, PMB, IB, TCH, LS, JMT, EAA, SEB, RC, JG, AH, TL, EMW, SM, LAM, LAS, JT, ACT, PW and DM drafted particular sections of the article. All authors were involved in revising the article critically for important intellectual content. All authors approved the final version of the article. MJP is the guarantor of this work.

References

1. Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Medicine / Public Library of Science*. 2009;6(7):e1000097.
2. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *PLoS Medicine / Public Library of Science*. 2009;6(7):e1000100.
3. Page MJ, Moher D. Evaluations of the uptake and impact of the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) Statement and extensions: a scoping review. *Systematic reviews*. 2017;6(1):263.
4. Tong A, Flemming K, McInnes E, Oliver S, Craig J. Enhancing transparency in reporting the synthesis of qualitative research: ENTREQ. *BMC medical research methodology*. 2012;12:181.
5. France EF, Cunningham M, Ring N, Uny I, Duncan EAS, Jepson RG, et al. Improving reporting of meta-ethnography: the eMERGe reporting guidance. *BMC medical research methodology*. 2019;19(1):25.
6. Hutton B, Salanti G, Caldwell DM, Chaimani A, Schmid CH, Cameron C, et al. The PRISMA extension statement for reporting of systematic reviews incorporating network meta-analyses of health care interventions: checklist and explanations. *Ann Intern Med*. 2015;162(11):777-84.
7. Stewart LA, Clarke M, Rovers M, Riley RD, Simmonds M, Stewart G, et al. Preferred Reporting Items for Systematic Review and Meta-Analyses of individual participant data: the PRISMA-IPD Statement. *JAMA*. 2015;313(16):1657-65.
8. Zorzela L, Loke YK, Ioannidis JP, Golder S, Santaguida P, Altman DG, et al. PRISMA harms checklist: improving harms reporting in systematic reviews. *BMJ*. 2016;352:i157.
9. McInnes MF, Moher D, Thombbs BD, McGrath TA, Bossuyt PM, and the PRISMA-DTA Group. Preferred reporting items for a systematic review and meta-analysis of diagnostic test accuracy studies: The PRISMA-DTA statement. *JAMA*. 2018;319(4):388-96.
10. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-SCR): Checklist and explanation. *Ann Intern Med*. 2018;169(7):467-73.
11. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. Submitted. 2020.

12. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. Updating guidance for reporting systematic reviews: development of the PRISMA 2020 statement. Submitted. 2020.
13. Barnes C, Boutron I, Giraudeau B, Porcher R, Altman DG, Ravaud P. Impact of an online writing aid tool for writing a randomized trial report: the COBWEB (Consort-based WEB tool) randomized controlled trial. *BMC medicine*. 2015;13:221.
14. Chauvin A, Ravaud P, Moher D, Schriger D, Hopewell S, Shanahan D, et al. Accuracy in detecting inadequate research reporting by early career peer reviewers using an online CONSORT-based peer-review tool (COBPeer) versus the usual peer-review process: a cross-sectional diagnostic study. *BMC medicine*. 2019;17(1):205.
15. Khodashahi M, Rezaieyazdi Z, Sahebari M. Comparison of the Therapeutic Effects of Rivaroxaban Versus Warfarin in Antiphospholipid Syndrome: A Systematic Review. *Archives of rheumatology*. 2020;35(1):107-16.
16. Beller EM, Glasziou PP, Altman DG, Hopewell S, Bastian H, Chalmers I, et al. PRISMA for Abstracts: Reporting Systematic Reviews in Journal and Conference Abstracts. *PLoS medicine*. 2013;10:e1001419.
17. Keynejad RC, Hanlon C, Howard LM. Psychological interventions for common mental disorders in women experiencing intimate partner violence in low-income and middle-income countries: a systematic review and meta-analysis. *The Lancet Psychiatry*. 2020;7(2):173-90.
18. Welch V, Petticrew M, Petkovic J, Moher D, Waters E, White H, et al. Extending the PRISMA statement to equity-focused systematic reviews (PRISMA-E 2012): explanation and elaboration. *J Clin Epidemiol*. 2016;70:68-89.
19. Thomas J, Kneale D, McKenzie JE, Brennan SE, Bhaumik S. Chapter 2: Determining the scope of the review and the questions it will address. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al., editors. *Cochrane Handbook for Systematic Reviews of Interventions*. Version 6.0 (updated July 2019). Available from <https://training.cochrane.org/handbook>: Cochrane; 2019.
20. Rehfuess EA, Booth A, Brereton L, Burns J, Gerhardus A, Mozygemba K, et al. Towards a taxonomy of logic models in systematic reviews and health technology assessments: A priori, staged, and iterative approaches. *Research synthesis methods*. 2018;9(1):13-24.
21. Chu DK, Akl EA, Duda S, Solo K, Yaacoub S, Schünemann HJ. Physical distancing, face masks, and eye protection to prevent person-to-person transmission of SARS-CoV-2 and COVID-19: a systematic review and meta-analysis. *Lancet*. 2020;395(10242):1973-87.

22. Booth A, Noyes J, Flemming K, Moore G, Tuncalp O, Shakibazadeh E. Formulating questions to explore complex interventions within qualitative evidence synthesis. *BMJ global health*. 2019;4(Suppl 1):e001107.
23. Verhoef LM, van den Bemt BJB, van der Maas A, Vriezekolk JE, Hulscher ME, van den Hoogen FHJ, et al. Down - titration and discontinuation strategies of tumour necrosis factor - blocking agents for rheumatoid arthritis in patients with low disease activity. *Cochrane Database of Systematic Reviews*. 2019(5):CD010455.
24. IOM (Institute of Medicine). *Finding What Works in Health Care: Standards for Systematic Reviews*. Washington, DC: The National Academies Press; 2011.
25. Page MJ, Shamseer L, Altman DG, Tetzlaff J, Sampson M, Tricco AC, et al. Epidemiology and Reporting Characteristics of Systematic Reviews of Biomedical Research: A Cross-Sectional Study. *PLoS medicine*. 2016;13(5):e1002028.
26. McKenzie JE, Brennan SE, Ryan RE, Thomson HJ, Johnston RV, Thomas J. Chapter 3: Defining the criteria for including studies and how they will be grouped for the synthesis. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al., editors. *Cochrane Handbook for Systematic Reviews of Interventions*. Version 6.0 (updated July 2019). Available from <https://training.cochrane.org/handbook>: Cochrane; 2019.
27. Chamberlain C, O'Mara - Eves A, Porter J, Coleman T, Perlen SM, Thomas J, et al. Psychosocial interventions for supporting women to stop smoking in pregnancy. *Cochrane Database of Systematic Reviews*. 2017(2):CD001055.
28. Dwan KM, Williamson PR, Kirkham JJ. Do systematic reviews still exclude studies with "no relevant outcome data"? *BMJ*. 2017;358:j3919.
29. Odor PM, Bampoe S, Gilhooly D, Creagh-Brown B, Moonesinghe SR. Perioperative interventions for prevention of postoperative pulmonary complications: systematic review and meta-analysis. *BMJ*. 2020;368:m540.
30. Lefebvre C, Glanville J, Briscoe S, Littlewood A, Marshall C, Metzendorf M-I, et al. Chapter 4: Searching for and selecting studies. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al., editors. *Cochrane Handbook for Systematic Reviews of Interventions*. Version 6.0 (updated July 2019). Available from <https://training.cochrane.org/handbook>: Cochrane; 2019.
31. Rethlefsen ML, Kirtley S, Waffenschmidt S, Ayala AP, Moher D, Page MJ, et al. PRISMA-S: an extension to the PRISMA statement for reporting literature searches in systematic reviews <https://doi.org/10.31219/osf.io/sfc38>. OSF Preprints 2020.

32. Jay MA, Mc Grath-Lone L. Educational outcomes of children in contact with social care in England: a systematic review. *Systematic reviews*. 2019;8(1):155.
33. Faggion CM, Jr., Huivin R, Aranda L, Pandis N, Alarcon M. The search and selection for primary studies in systematic reviews published in dental journals indexed in MEDLINE was not fully reproducible. *J Clin Epidemiol*. 2018;98:53-61.
34. Spry C, Mierzwinski-Urban M. The impact of the peer review of literature search strategies in support of rapid review reports. *Research synthesis methods*. 2018;9(4):521-6.
35. ISSG Search Filter Resource [Internet]. Glanville J, Lefebvre C, Wright K, editors. York (UK): The InterTASC Information Specialists' Sub-Group; 2008 [updated 2020 February 4; cited 2020 April 27]. Available from: <https://sites.google.com/a/york.ac.uk/issg-search-filters-resource/home>.
36. Stansfield C, O'Mara-Eves A, Thomas J. Text mining for search term development in systematic reviewing: A discussion of some methods and challenges. *Research synthesis methods*. 2017;8(3):355-65.
37. Glanville J. Text mining for information specialists. In: Levay P, Craven J, editors. *Systematic Searching: Practical ideas for improving results*. London, UK: Facet Publishing; 2019.
38. 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.
39. Freemantle N, Ginsberg DA, McCool R, Fleetwood K, Arber M, Khalaf K, et al. Comparative assessment of onabotulinumtoxinA and mirabegron for overactive bladder: an indirect treatment comparison. *BMJ Open*. 2016;6(2):e009122.
40. Wang Z, Nayfeh T, Tetzlaff J, O'Blenis P, Murad MH. Error rates of human reviewers during abstract screening in systematic reviews. *PLoS one*. 2020;15(1):e0227742.
41. Waffenschmidt S, Knellingen M, Sieben W, Bühn S, Pieper D. Single screening versus conventional double screening for study selection in systematic reviews: a methodological systematic review. *BMC medical research methodology*. 2019;19(1):132.
42. Shemilt I, Khan N, Park S, Thomas J. Use of cost-effectiveness analysis to compare the efficiency of study identification methods in systematic reviews. *Systematic reviews*. 2016;5(1):140.
43. Gartlehner G, Affengruber L, Titscher V, Noel-Storr A, Dooley G, Ballarini N, et al. Single-reviewer abstract screening missed 13 percent of relevant studies: a crowd-based, randomized controlled trial. *J Clin Epidemiol*. 2020;121:20-8.
44. Bomhof-Roordink H, Gartner FR, Stiggelbout AM, Pieterse AH. Key components of shared decision making models: a systematic review. *BMJ Open*. 2019;9(12):e031763.

45. Li T, Saldanha IJ, Jap J, Smith BT, Canner J, Hutfless SM, et al. A randomized trial provided new evidence on the accuracy and efficiency of traditional vs. electronically annotated abstraction approaches in systematic reviews. *J Clin Epidemiol*. 2019;115:77-89.
46. Robson RC, Pham B, Hwee J, Thomas SM, Rios P, Page MJ, et al. Few studies exist examining methods for selecting studies, abstracting data, and appraising quality in a systematic review. *J Clin Epidemiol*. 2019;106:121-35.
47. E JY, Saldanha IJ, Canner J, Schmid CH, Le JT, Li T. Adjudication rather than experience of data abstraction matters more in reducing errors in abstracting data in systematic reviews. *Research synthesis methods*. 2020.
48. Li T, Higgins JPT, Deeks JJ. Chapter 5: Collecting data. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al., editors. *Cochrane Handbook for Systematic Reviews of Interventions*. Version 6.0 (updated July 2019). Available from <https://training.cochrane.org/handbook>: Cochrane; 2019.
49. Marshall IJ, Wallace BC. Toward systematic review automation: a practical guide to using machine learning tools in research synthesis. *Systematic reviews*. 2019;8(1):163.
50. Jonnalagadda SR, Goyal P, Huffman MD. Automating data extraction in systematic reviews: a systematic review. *Systematic reviews*. 2015;4:78.
51. Jackson JL, Kuriyama A, Anton A, Choi A, Fournier J-P, Geier A-K, et al. The Accuracy of Google Translate for Abstracting Data From Non-English-Language Trials for Systematic Reviews. *Ann Intern Med*. 2019.
52. Jelacic Kadic A, Vucic K, Dosenovic S, Sapunar D, Puljak L. Extracting data from figures with software was faster, with higher interrater reliability than manual extraction. *J Clin Epidemiol*. 2016;74:119-23.
53. Mayo-Wilson E, Li T, Fusco N, Dickersin K. Practical guidance for using multiple data sources in systematic reviews and meta-analyses (with examples from the MUDS study). *Research synthesis methods*. 2018;9(1):2-12.
54. Claire R, Chamberlain C, Davey MA, Cooper SE, Berlin I, Leonardi - Bee J, et al. Pharmacological interventions for promoting smoking cessation during pregnancy. *Cochrane Database of Systematic Reviews*. 2020(3):CD010078.
55. Page MJ, Forbes A, Chau M, Green SE, McKenzie JE. Investigation of bias in meta-analyses due to selective inclusion of trial effect estimates: empirical study. *BMJ Open*. 2016;6(4):e011863.

56. Mayo-Wilson E, Fusco N, Li T, Hong H, Canner JK, Dickersin K. Multiple outcomes and analyses in clinical trials create challenges for interpretation and research synthesis. *J Clin Epidemiol.* 2017;86:39-50.
57. Mayo-Wilson E, Li T, Fusco N, Bertizzolo L, Canner JK, Cowley T, et al. Cherry-picking by trialists and meta-analysts can drive conclusions about intervention efficacy. *J Clin Epidemiol.* 2017;91:95-110.
58. Lopez-Lopez JA, Page MJ, Lipsey MW, Higgins JPT. Dealing with effect size multiplicity in systematic reviews and meta-analyses. *Research synthesis methods.* 2018;9(3):336-51.
59. Page MJ, McKenzie JE, Kirkham J, Dwan K, Kramer S, Green S, et al. Bias due to selective inclusion and reporting of outcomes and analyses in systematic reviews of randomised trials of healthcare interventions. *Cochrane Database of Systematic Reviews.* 2014(10):MR000035.
60. Brennan SE, McDonald S, Page MJ, Reid J, Ward S, Forbes AB, et al. Long-term effects of alcohol consumption on cognitive function: a systematic review and dose-response analysis of evidence published between 2007 and 2018. *Systematic reviews.* 2020;9(1):33.
61. Lundh A, Boutron I, Stewart L, Hróbjartsson A. What to do with a clinical trial with conflicts of interest. *BMJ evidence-based medicine.* 2019.
62. Boutron I, Page MJ, Higgins JPT, Altman DG, Lundh A, Hróbjartsson A. Chapter 7: Considering bias and conflicts of interest among the included studies. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al., editors. *Cochrane Handbook for Systematic Reviews of Interventions.* Version 6.0 (updated July 2019). Available from <https://training.cochrane.org/handbook>: Cochrane; 2019.
63. Hoffmann TC, Oxman AD, Ioannidis JP, Moher D, Lasserson TJ, Tovey DI, et al. Enhancing the usability of systematic reviews by improving the consideration and description of interventions. *BMJ.* 2017;358:j2998.
64. Lewin S, Hendry M, Chandler J, Oxman AD, Michie S, Shepperd S, et al. Assessing the complexity of interventions within systematic reviews: development, content and use of a new tool (iCAT_SR). *BMC medical research methodology.* 2017;17(1):76.
65. Montgomery P, Underhill K, Gardner F, Operario D, Mayo-Wilson E. The Oxford Implementation Index: a new tool for incorporating implementation data into systematic reviews and meta-analyses. *J Clin Epidemiol.* 2013;66(8):874-82.
66. Allida S, Cox KL, Hsieh CF, Lang H, House A, Hackett ML. Pharmacological, psychological, and non - invasive brain stimulation interventions for treating depression after stroke. *Cochrane Database of Systematic Reviews.* 2020(1):CD003437.

67. Bai A, Shukla VK, Bak G, Wells G. Quality Assessment Tools Project Report. Ottawa: Canadian Agency for Drugs and Technologies in Health; 2012.
68. Marshall IJ, Kuiper J, Wallace BC. RobotReviewer: evaluation of a system for automatically assessing bias in clinical trials. *J Am Med Inform Assoc.* 2016;23(1):193-201.
69. Hollands GJ, Carter P, Anwer S, King SE, Jebb SA, Ogilvie D, et al. Altering the availability or proximity of food, alcohol, and tobacco products to change their selection and consumption. *Cochrane Database of Systematic Reviews.* 2019(9):CD012573.
70. Higgins JPT, Li T, Deeks JJ. Chapter 6: Choosing effect measures and computing estimates of effect. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al., editors. *Cochrane Handbook for Systematic Reviews of Interventions.* Version 6.0 (updated July 2019). Available from <https://training.cochrane.org/handbook>: Cochrane; 2019.
71. Deeks JJ. Issues in the selection of a summary statistic for meta-analysis of clinical trials with binary outcomes. *Stat Med.* 2002;21(11):1575-600.
72. Schünemann HJ, Vist GE, Higgins JPT, Santesso N, Deeks JJ, Glasziou P, et al. Chapter 15: Interpreting results and drawing conclusions. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al., editors. *Cochrane Handbook for Systematic Reviews of Interventions.* Version 6.0 (updated July 2019). Available from <https://training.cochrane.org/handbook>: Cochrane; 2019.
73. Kunzler AM, Helmreich I, König J, Chmitorz A, Wessa M, Binder H, et al. Psychological interventions to foster resilience in healthcare students. *Cochrane Database of Systematic Reviews.* 2020(7):CD013684.
74. McKenzie JE, Brennan SE, Ryan RE, Thomson HJ, Johnston RV. Chapter 9: Summarizing study characteristics and preparing for synthesis. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al., editors. *Cochrane Handbook for Systematic Reviews of Interventions.* Version 6.0 (updated July 2019). Available from <https://training.cochrane.org/handbook>: Cochrane; 2019.
75. Munthe-Kaas HM, Berg RC, Blaasvær N. Effectiveness of interventions to reduce homelessness: a systematic review and meta-analysis. *Campbell Systematic Reviews.* 2018;14(1):1-281.
76. da Costa BR, Rutjes AW, Johnston BC, Reichenbach S, Nuesch E, Tonia T, et al. Methods to convert continuous outcomes into odds ratios of treatment response and numbers needed to treat: meta-epidemiological study. *Int J Epidemiol.* 2012;41(5):1445-59.

77. Weir CJ, Butcher I, Assi V, Lewis SC, Murray GD, Langhorne P, et al. Dealing with missing standard deviation and mean values in meta-analysis of continuous outcomes: a systematic review. *BMC medical research methodology*. 2018;18(1):25.
78. Mavridis D, White IR. Dealing with missing outcome data in meta-analysis. *Research synthesis methods*. 2020;11(1):2-13.
79. Kahale LA, Diab B, Brignardello-Petersen R, Agarwal A, Mustafa RA, Kwong J, et al. Systematic reviews do not adequately report, or address missing outcome data in their analyses: a methodological survey. *J Clin Epidemiol*. 2018.
80. Das JK, Salam RA, Mahmood SB, Moin A, Kumar R, Mukhtar K, et al. Food fortification with multiple micronutrients: impact on health outcomes in general population. *Cochrane Database of Systematic Reviews*. 2019(12):CD011400.
81. Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, et al. GRADE guidelines: 1. Introduction—GRADE evidence profiles and summary of findings tables. *J Clin Epidemiol*. 2011;64(4):383-94.
82. Schünemann HJ, Higgins JPT, Vist GE, Glasziou P, Akl EA, Skoetz N, et al. Chapter 14: Completing ‘Summary of findings’ tables and grading the certainty of the evidence. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al., editors. *Cochrane Handbook for Systematic Reviews of Interventions*. Version 6.0 (updated July 2019). Available from <https://training.cochrane.org/handbook>: Cochrane; 2019.
83. McKenzie JE, Brennan SE. Chapter 12: Synthesizing and presenting findings using other methods. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al., editors. *Cochrane Handbook for Systematic Reviews of Interventions*. Version 6.0 (updated July 2019). Available from <https://training.cochrane.org/handbook>: Cochrane; 2019.
84. Anzures-Cabrera J, Higgins JP. Graphical displays for meta-analysis: An overview with suggestions for practice. *Research synthesis methods*. 2010;1(1):66-80.
85. Lewis S, Clarke M. Forest plots: trying to see the wood and the trees. *BMJ*. 2001;322(7300):1479-80.
86. Kossmeier M, Tran US, Voracek M. Charting the landscape of graphical displays for meta-analysis and systematic reviews: a comprehensive review, taxonomy, and feature analysis. *BMC medical research methodology*. 2020;20(1):26.
87. Hunter BM, Harrison S, Portela A, Bick D. The effects of cash transfers and vouchers on the use and quality of maternity care services: A systematic review. *PloS one*. 2017;12(3):e0173068-e.

88. Veroniki AA, Jackson D, Viechtbauer W, Bender R, Bowden J, Knapp G, et al. Methods to estimate the between-study variance and its uncertainty in meta-analysis. *Research synthesis methods*. 2016;7(1):55-79.
89. Langan D, Higgins JPT, Jackson D, Bowden J, Veroniki AA, Kontopantelis E, et al. A comparison of heterogeneity variance estimators in simulated random-effects meta-analyses. *Research synthesis methods*. 2019;10(1):83-98.
90. Higgins JPT, Lopez-Lopez JA, Becker BJ, Davies SR, Dawson S, Grimshaw JM, et al. Synthesising quantitative evidence in systematic reviews of complex health interventions. *BMJ global health*. 2019;4(Suppl 1):e000858.
91. Campbell M, McKenzie JE, Sowden A, Katikireddi SV, Brennan SE, Ellis S, et al. Synthesis without meta-analysis (SWiM) in systematic reviews: reporting guideline. *BMJ*. 2020;368:l6890.
92. Harris R, Bradburn M, Deeks J, Harbord R, Altman D, Sterne J. metan: fixed- and random-effects meta-analysis. *Stata Journal*. 2008;8(1):3-28.
93. Viechtbauer W. Conducting Meta-Analyses in R with the metafor Package. *Journal of Statistical Software*. 2010;36(3):1-48.
94. Deeks JJ, Higgins JPT, Altman DG. Chapter 10: Analysing data and undertaking meta-analyses. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al., editors. *Cochrane Handbook for Systematic Reviews of Interventions*. Version 6.0 (updated July 2019). Available from <https://training.cochrane.org/handbook>: Cochrane; 2019.
95. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ*. 2003;327(7414):557-60.
96. Riley RD, Higgins JP, Deeks JJ. Interpretation of random effects meta-analyses. *BMJ*. 2011;342:d549.
97. Veroniki AA, Jackson D, Bender R, Kuss O, Langan D, Higgins JPT, et al. Methods to calculate uncertainty in the estimated overall effect size from a random-effects meta-analysis. *Research synthesis methods*. 2019;10(1):23-43.
98. Kyburz KS, Eliades T, Papageorgiou SN. What effect does functional appliance treatment have on the temporomandibular joint? A systematic review with meta-analysis. *Progress in Orthodontics*. 2019;20(1):32.
99. Pantoja T, Grimshaw JM, Colomer N, Castañon C, Leniz Martelli J. Manually - generated reminders delivered on paper: effects on professional practice and patient outcomes. *Cochrane Database of Systematic Reviews*. 2019(12):CD001174.

100. Fisher DJ, Carpenter JR, Morris TP, Freeman SC, Tierney JF. Meta-analytical methods to identify who benefits most from treatments: daft, deluded, or deft approach? *BMJ*. 2017;356:j573.
101. Kock L, Brown J, Hiscock R, Tattan-Birch H, Smith C, Shahab L. Individual-level behavioural smoking cessation interventions tailored for disadvantaged socioeconomic position: a systematic review and meta-regression. *The Lancet Public Health*. 2019;4(12):e628-e44.
102. Roth DE, Leung M, Mesfin E, Qamar H, Watterworth J, Papp E. Vitamin D supplementation during pregnancy: state of the evidence from a systematic review of randomised trials. *BMJ*. 2017;359:j5237.
103. Page MJ, Higgins JPT, Sterne JAC. Chapter 13: Assessing risk of bias due to missing results in a synthesis. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al., editors. *Cochrane Handbook for Systematic Reviews of Interventions*. Version 6.0 (updated July 2019). Available from <https://training.cochrane.org/handbook>: Cochrane; 2019.
104. Page MJ, McKenzie JE, Higgins JPT. Tools for assessing risk of reporting biases in studies and syntheses of studies: a systematic review. *BMJ Open*. 2018;8(3):e019703.
105. Mueller KF, Meerpohl JJ, Briel M, Antes G, von Elm E, Lang B, et al. Methods for detecting, quantifying and adjusting for dissemination bias in meta-analysis are described. *J Clin Epidemiol*. 2016;80:25-33.
106. Vevea JL, Coburn K, Sutton A. Publication bias. In: Cooper H, Hedges LV, Valentine JC, editors. *The Handbook of Research Synthesis and Meta-Analysis*: Russell Sage Foundation; 2019. p. 383-430.
107. Kirkham JJ, Altman DG, Chan AW, Gamble C, Dwan KM, Williamson PR. Outcome reporting bias in trials: a methodological approach for assessment and adjustment in systematic reviews. *BMJ*. 2018;362:k3802.
108. Karjalainen TV, Jain NB, Heikkinen J, Johnston RV, Page CM, Buchbinder R. Surgery for rotator cuff tears. *Cochrane Database of Systematic Reviews*. 2019(12):CD013502.
109. Balshem H, Helfand M, Schünemann HJ, Oxman AD, Kunz R, Brozek J, et al. GRADE guidelines: 3. Rating the quality of evidence. *J Clin Epidemiol*. 2011;64(4):401-6.
110. Schunemann HJ, Brozek J, Guyatt G, Oxman AD, editors. *Handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach*. Accessed 5 July 2016. Hamilton, Canada: McMaster University; 2013.
111. Berkman ND, Lohr KN, Ansari MT, Balk EM, Kane R, McDonagh M, et al. Grading the strength of a body of evidence when assessing health care interventions: an EPC update. *J Clin Epidemiol*. 2015;68(11):1312-24.

112. Nikolakopoulou A, Higgins JPT, Papakonstantinou T, Chaimani A, Del Giovane C, Egger M, et al. CINeMA: An approach for assessing confidence in the results of a network meta-analysis. *PLoS medicine*. 2020;17(4):e1003082.
113. Hultcrantz M, Rind D, Akl EA, Treweek S, Mustafa RA, Iorio A, et al. The GRADE Working Group clarifies the construct of certainty of evidence. *J Clin Epidemiol*. 2017;87:4-13.
114. Santesso N, Glenton C, Dahm P, Garner P, Akl EA, Alper B, et al. GRADE guidelines 26: informative statements to communicate the findings of systematic reviews of interventions. *J Clin Epidemiol*. 2020;119:126-35.
115. Mudano AS, Tugwell P, Wells GA, Singh JA. Tai Chi for rheumatoid arthritis. *Cochrane Database of Systematic Reviews*. 2019(9):CD004849.
116. Sampson M, Tetzlaff J, Urquhart C. Precision of healthcare systematic review searches in a cross-sectional sample. *Research synthesis methods*. 2011;2(2):119-25.
117. Haddaway NR, Westgate MJ. Predicting the time needed for environmental systematic reviews and systematic maps. *Conservation biology : the journal of the Society for Conservation Biology*. 2019;33(2):434-43.
118. Stovold E, Beecher D, Foxlee R, Noel-Storr A. Study flow diagrams in Cochrane systematic review updates: an adapted PRISMA flow diagram. *Systematic reviews*. 2014;3:54.
119. Chambergo-Michilot D, Tellez WA, Becerra-Chauca N, Zafra-Tanaka JH, Taype-Rondan A. Text message reminders for improving sun protection habits: A systematic review. *PloS one*. 2020;15(5):e0233220.
120. McDonagh M, Peterson K, Raina P, Chang S, Shekelle P. AHRQ Methods for Effective Health Care: Avoiding Bias in Selecting Studies. *Methods Guide for Effectiveness and Comparative Effectiveness Reviews*. Rockville (MD): Agency for Healthcare Research and Quality (US); 2008.
121. Kayssi A, Al - Jundi W, Papia G, Kucey DS, Forbes T, Rajan DK, et al. Drug - eluting balloon angioplasty versus uncoated balloon angioplasty for the treatment of in - stent restenosis of the femoropopliteal arteries. *Cochrane Database of Systematic Reviews*. 2019(1):CD012510.
122. Barker AL, Soh SE, Sanders KM, Pasco J, Khosla S, Ebeling PR, et al. Aspirin and fracture risk: a systematic review and exploratory meta-analysis of observational studies. *BMJ Open*. 2020;10(2):e026876.
123. McGuinness LA, Higgins JPT. Risk-of-bias VISualization (robvis): An R package and Shiny web app for visualizing risk-of-bias assessments. *Research synthesis methods*. 2020.
124. Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ*. 2019;366:l4898.

125. Altman DG, Cates C. The need for individual trial results in reports of systematic reviews. *BMJ*. 2001;323:776.
126. Hemming K, Taljaard M, McKenzie JE, Hooper R, Copas A, Thompson JA, et al. Reporting of stepped wedge cluster randomised trials: extension of the CONSORT 2010 statement with explanation and elaboration. *BMJ*. 2018;363:k1614.
127. Jap J, Saldanha IJ, Smith BT, Lau J, Schmid CH, Li T. Features and functioning of Data Abstraction Assistant, a software application for data abstraction during systematic reviews. *Research synthesis methods*. 2019;10(1):2-14.
128. Feng Q, Zhou A, Zou H, Ingle S, May MT, Cai W, et al. Quadruple versus triple combination antiretroviral therapies for treatment naive people with HIV: systematic review and meta-analysis of randomised controlled trials. *BMJ*. 2019;366:l4179.
129. Neufeld KJ, Needham DM, Oh ES, Wilson LM, Nikooie R, Zhang A, et al. AHRQ Comparative Effectiveness Reviews. Antipsychotics for the Prevention and Treatment of Delirium. Rockville (MD): Agency for Healthcare Research and Quality (US); 2019.
130. Silagy CA, Middleton P, Hopewell S. Publishing protocols of systematic reviews: comparing what was done to what was planned. *JAMA*. 2002;287(21):2831-4.
131. Parmelli E, Liberati A, D'Amico R. Reporting of outcomes in systematic reviews: comparison of protocols and published systematic reviews (abstract). XV Cochrane Colloquium; 2007 Oct 23-27; São Paulo, Brazil. 2007:118-9.
132. Dwan K, Kirkham JJ, Williamson PR, Gamble C. Selective reporting of outcomes in randomised controlled trials in systematic reviews of cystic fibrosis. *BMJ Open*. 2013;3(6).
133. Shah K, Egan G, Huan LN, Kirkham J, Reid E, Tejani AM. Outcome reporting bias in Cochrane systematic reviews: a cross-sectional analysis. *BMJ Open*. 2020;10(3):e032497.
134. Gelbenegger G, Postula M, Pecen L, Halvorsen S, Lesiak M, Schoergenhofer C, et al. Aspirin for primary prevention of cardiovascular disease: a meta-analysis with a particular focus on subgroups. *BMC medicine*. 2019;17(1):198.
135. Thompson SG, Higgins JP. How should meta-regression analyses be undertaken and interpreted? *Stat Med*. 2002;21(11):1559-73.
136. Schandelmaier S, Briel M, Varadhan R, Schmid CH, Devasenapathy N, Hayward RA, et al. Development of the Instrument to assess the Credibility of Effect Modification Analyses (ICEMAN) in randomized controlled trials and meta-analyses. *CMAJ*. 2020;192(32):E901-e6.
137. Sterne JAC, Murthy S, Diaz JV, Slutsky AS, Villar J, Angus DC, et al. Association Between Administration of Systemic Corticosteroids and Mortality Among Critically Ill Patients With COVID-19: A Meta-analysis. *JAMA*. 2020.

138. Sterne JA, Sutton AJ, Ioannidis JP, Terrin N, Jones DR, Lau J, et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ*. 2011;343:d4002.
139. Peters JL, Sutton AJ, Jones DR, Abrams KR, Rushton L. Contour-enhanced meta-analysis funnel plots help distinguish publication bias from other causes of asymmetry. *J Clin Epidemiol*. 2008;61(10):991-6.
140. Chau S, Herrmann N, Ruthirakuhan MT, Chen JJ, Lanctot KL. Latrepirdine for Alzheimer's disease. *The Cochrane database of systematic reviews*. 2015(4):CD009524.
141. Guyatt GH, Oxman AD, Santesso N, Helfand M, Vist G, Kunz R, et al. GRADE guidelines: 12. Preparing Summary of Findings tables—binary outcomes. *J Clin Epidemiol*. 2013;66(2):158-72.
142. Guyatt GH, Thorlund K, Oxman AD, Walter SD, Patrick D, Furukawa TA, et al. GRADE guidelines: 13. Preparing Summary of Findings tables and evidence profiles—continuous outcomes. *J Clin Epidemiol*. 2013;66(2):173-83.
143. Santesso N, Carrasco-Labra A, Langendam M, Brignardello-Petersen R, Mustafa RA, Heus P, et al. Improving GRADE Evidence Tables part 3: Guidance for useful GRADE certainty in the evidence judgments through explanatory footnotes. *J Clin Epidemiol*. 2016;74:28-39.
144. Stone R, de Hoop T, Coombes A, Nakamura P. What works to improve early grade literacy in Latin America and the Caribbean? A systematic review and meta-analysis. *Campbell Systematic Reviews*. 2020;16(1):e1067.
145. Ghasemiesfe M, Barrow B, Leonard S, Keyhani S, Korenstein D. Association Between Marijuana Use and Risk of Cancer: A Systematic Review and Meta-analysis. *JAMA network open*. 2019;2(11):e1916318.
146. Whiting P, Savovic 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.
147. 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.
148. Nussbaumer-Streit B, Mayr V, Dobrescu AI, Chapman A, Persad E, Klerings I, et al. Quarantine alone or in combination with other public health measures to control COVID-19: a rapid review. *The Cochrane database of systematic reviews*. 2020;4(4):CD013574.
149. Kettrey HH, Marx RA, Tanner-Smith EE. Effects of bystander programs on the prevention of sexual assault among adolescents and college students: A systematic review. *Campbell Systematic Reviews*. 2019;15(1-2):e1013.

150. Sender D, Clark J, Hoffmann TC. Analysis of articles directly related to randomized trials finds poor protocol availability and inconsistent linking of articles. *J Clin Epidemiol.* 2020;124:69-74.
151. Jud L, Fotouhi J, Andronic O, Aichmair A, Osgood G, Navab N, et al. Applicability of augmented reality in orthopedic surgery - A systematic review. *BMC Musculoskelet Disord.* 2020;21(1):103.
152. Koensgen N, Rombey T, Allers K, Mathes T, Hoffmann F, Pieper D. Comparison of non-Cochrane systematic reviews and their published protocols: Differences occurred frequently but were seldom explained. *J Clin Epidemiol.* 2019;110:34-41.
153. Semahegn A, Torpey K, Manu A, Assefa N, Tesfaye G, Ankomah A. Psychotropic medication non-adherence and its associated factors among patients with major psychiatric disorders: a systematic review and meta-analysis. *Systematic reviews.* 2020;9(1):17.
154. Shamseer L, Moher D, Clarke M, Gherzi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. *BMJ.* 2015;349:g7647.
155. Pieper D, Allers K, Mathes T, Hoffmann F, Klerings I, Rombey T, et al. Comparison of protocols and registry entries to published reports for systematic reviews. *Cochrane Database of Systematic Reviews.* 2020(2).
156. Tanjong Ghogomu E, Suresh S, Rayco-Solon P, Hossain A, McGowan J, Peña-Rosas JP, et al. Deworming in non-pregnant adolescent girls and adult women: a systematic review and meta-analysis. *Systematic reviews.* 2018;7(1):239.
157. Lundh A, Lexchin J, Mintzes B, Schroll JB, Bero L. Industry sponsorship and research outcome. *Cochrane Database of Systematic Reviews.* 2017(2):MR000033.
158. Chou R, Dana T, Fu R, Zakher B, Wagner J, Ramirez S, et al. Screening for Hepatitis C Virus Infection in Adolescents and Adults: Updated Evidence Report and Systematic Review for the US Preventive Services Task Force. *JAMA.* 2020.
159. Hakoum MB, Anouti S, Al-Gibbawi M, Abou-Jaoude EA, Hasbani DJ, Lopes LC, et al. Reporting of financial and non-financial conflicts of interest by authors of systematic reviews: a methodological survey. *BMJ Open.* 2016;6(8):e011997.
160. Hansen C, Lundh A, Rasmussen K, Hrobjartsson A. Financial conflicts of interest in systematic reviews: associations with results, conclusions, and methodological quality. *The Cochrane database of systematic reviews.* 2019(8):MR000047.

161. Taichman DB, Backus J, Baethge C, Bauchner H, Flanagin A, Florenzano F, et al. A Disclosure Form for Work Submitted to Medical Journals: A Proposal From the International Committee of Medical Journal Editors. *Ann Intern Med.* 2020.
162. Buchbinder R, Johnston RV, Rischin KJ, Homik J, Jones CA, Golmohammadi K, et al. Percutaneous vertebroplasty for osteoporotic vertebral compression fracture. *Cochrane Database of Systematic Reviews.* 2018(11):CD006349.
163. Haddaway NR. Open Synthesis: on the need for evidence synthesis to embrace Open Science. *Environmental Evidence.* 2018;7(1):26.
164. Goldacre B, Morton CE, DeVito NJ. Why researchers should share their analytic code. *BMJ.* 2019;367:l6365.
165. Mello MM, Lieou V, Goodman SN. Clinical Trial Participants' Views of the Risks and Benefits of Data Sharing. *N Engl J Med.* 2018;378(23):2202-11.
166. Naudet F, Sakarovitch C, Janiaud P, Cristea I, Fanelli D, Moher D, et al. Data sharing and reanalysis of randomized controlled trials in leading biomedical journals with a full data sharing policy: survey of studies published in *The BMJ* and *PLOS Medicine*. *BMJ.* 2018;360:k400.
167. Saldanha IJ, Smith BT, Ntzani E, Jap J, Balk EM, Lau J. The Systematic Review Data Repository (SRDR): descriptive characteristics of publicly available data and opportunities for research. *Systematic reviews.* 2019;8(1):334.
168. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data.* 2016;3:160018.
169. Ritchie SJ, Tucker-Drob EM. How Much Does Education Improve Intelligence? A Meta-Analysis. *Psychol Sci.* 2018;29(8):1358-69.
170. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al., editors. *Cochrane Handbook for Systematic Reviews of Interventions: Version 6.0 (updated July 2019)*. Available from <https://training.cochrane.org/handbook>. Cochrane; 2019.
171. O'Mara-Eves A, Thomas J, McNaught J, Miwa M, Ananiadou S. Using text mining for study identification in systematic reviews: a systematic review of current approaches. *Systematic reviews.* 2015;4:5.
172. Marshall IJ, Noel-Storr A, Kuiper J, Thomas J, Wallace BC. Machine learning for identifying Randomized Controlled Trials: An evaluation and practitioner's guide. *Research synthesis methods.* 2018.

173. Noel-Storr A. Working with a new kind of team: harnessing the wisdom of the crowd in trial identification. *EFSA journal European Food Safety Authority*. 2019;17(Suppl 1):e170715.
174. Nama N, Sampson M, Barrowman N, Sandarage R, Menon K, Macartney G, et al. Crowdsourcing the Citation Screening Process for Systematic Reviews: Validation Study. *Journal of medical Internet research*. 2019;21(4):e12953.
175. Buttner F, Winters M, Delahunt E, Elbers R, Lura CB, Khan KM, et al. Identifying the 'incredible'! Part 1: assessing the risk of bias in outcomes included in systematic reviews. *Br J Sports Med*. 2020;54(13):798-800.
176. Moher D, Jadad AR, Nichol G, Penman M, Tugwell P, Walsh S. Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists. *Control Clin Trials*. 1995;16(1):62-73.
177. Armijo Olivo S, Macedo LG, Gadotti IC, Fuentes J, Stanton T, Magee DJ. Scales to assess the quality of randomized controlled trials: a systematic review. *Phys Ther*. 2008;88(2):156-75.
178. Page MJ, Higgins JP, Clayton G, Sterne JA, Hrobjartsson A, Savovic J. Empirical Evidence of Study Design Biases in Randomized Trials: Systematic Review of Meta-Epidemiological Studies. *PloS one*. 2016;11(7):e0159267.
179. Chan AW, Song F, Vickers A, Jefferson T, Dickersin K, Gotzsche PC, et al. Increasing value and reducing waste: addressing inaccessible research. *Lancet*. 2014;383(9913):257-66.
180. Dwan K, Gamble C, Williamson PR, Kirkham JJ. Systematic review of the empirical evidence of study publication bias and outcome reporting bias - an updated review. *PloS one*. 2013;8(7):e66844.
181. Whiting P, Wolff R, Mallett S, Simera I, Savovic J. A proposed framework for developing quality assessment tools. *Systematic reviews*. 2017;6(1):204.
182. Sterne JA, Hernan MA, Reeves BC, Savovic J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ*. 2016;355:i4919.
183. McKenzie JE, Beller EM, Forbes AB. Introduction to systematic reviews and meta-analysis. *Respirology*. 2016;21(4):626-37.
184. Rice K, Higgins J, Lumley T. A re - evaluation of fixed effect(s) meta - analysis. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*. 2018;181(1):205-27.
185. 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.

186. Mavridis D, Salanti G. A practical introduction to multivariate meta-analysis. *Stat Methods Med Res.* 2013;22(2):133-58.
187. Konstantopoulos S. Fixed effects and variance components estimation in three-level meta-analysis. *Research synthesis methods.* 2011;2(1):61-76.
188. Hedges LV, Tipton E, Johnson MC. Robust variance estimation in meta-regression with dependent effect size estimates. *Research synthesis methods.* 2010;1(1):39-65.
189. Stewart L, Moher D, Shekelle P. Why prospective registration of systematic reviews makes sense. *Systematic reviews.* 2012;1:7.
190. Page MJ, Shamseer L, Tricco AC. Registration of systematic reviews in PROSPERO: 30,000 records and counting. *Systematic reviews.* 2018;7(1):32.
191. Booth A, Clarke M, Dooley G, Gherzi D, Moher D, Petticrew M, et al. The nuts and bolts of PROSPERO: an international prospective register of systematic reviews. *Systematic reviews.* 2012;1:2.
192. Moher D, Shamseer L, Clarke M, Gherzi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic reviews.* 2015;4:1.
193. Boers M. Graphics and statistics for cardiology: designing effective tables for presentation and publication. *Heart.* 2018;104(3):192-200.

Box 1. Glossary

Systematic review: A review that uses explicit, systematic methods to collate and synthesize findings of studies that address a clearly formulated question (170).

Statistical synthesis: The combination of quantitative results of two or more studies. This encompasses meta-analysis of effect estimates (described below) and other methods, such as combining P values, calculating the range and distribution of observed effects, and vote counting based on the direction of effect (see McKenzie and Brennan (83) for a description of each method).

Meta-analysis of effect estimates: A statistical technique used to synthesize results when study effect estimates and their variances are available, yielding a quantitative summary of results (83).

Outcome: An event or measurement collected for participants in a study (e.g. quality of life, mortality).

Result: The combination of a point estimate (such as a mean difference, risk ratio or proportion) and a measure of its precision (such as a confidence/credible interval) for a particular outcome.

Report: A document (paper or electronic) supplying information about a particular study. It could be a journal article, preprint, conference abstract, study register entry, clinical study report, dissertation, unpublished manuscript, government report or any other document providing relevant information.

Record: The title or abstract (or both) of a report indexed in a database or website (e.g. a title or abstract for an article indexed in MEDLINE). Records that refer to the same report (e.g., the same journal article) are “duplicates”; however, records that refer to reports that are merely similar (e.g. a similar abstract submitted to two different conferences) should be considered unique.

Study: An investigation, such as a clinical trial, that includes a defined group of participants and one or more interventions and outcomes. A “study” might have multiple reports. For example, reports could include the protocol, statistical analysis plan, baseline characteristics, results for the primary outcome, results for harms, results for secondary outcomes, and results for additional mediator and moderator analyses.

Box 2. Items in the PRISMA 2020 for Abstracts checklist

The PRISMA 2020 for Abstracts checklist retains the same items as those included in the PRISMA for Abstracts statement published in 2013 (16), but has been revised to make the wording consistent with the PRISMA 2020 statement and includes a new item recommending authors specify the methods used to present and synthesize results (item #6). The checklist includes the following 12 items:

1. Identify the report as a systematic review.
2. Provide an explicit statement of the main objective(s) or question(s) the review addresses.
3. Specify the inclusion and exclusion criteria for the review.
4. Specify the information sources (e.g. databases, registers) used to identify studies and the date when each was last searched.
5. Specify the methods used to assess risk of bias in the included studies.
6. Specify the methods used to present and synthesize results.
7. Give the total number of included studies and participants and summarise relevant characteristics of studies.
8. Present results for main outcomes, preferably indicating the number of included studies and participants for each. If meta-analysis was done, report the summary estimate and confidence/credible interval. If comparing groups, indicate the direction of the effect (i.e. which group is favoured).
9. Provide a brief summary of the limitations of the evidence included in the review (e.g. study risk of bias, inconsistency and imprecision).
10. Provide a general interpretation of the results and important implications.
11. Specify the primary source of funding for the review.
12. Provide the register name and registration number.

Box 3. Study selection methods

Several approaches to selecting studies exist. In this Box, we comment on the advantages and disadvantages of each.

Assessment of each record by one reviewer: Single screening is an efficient use of time and resources, but there is a higher risk of missing relevant studies (40, 41, 43).

Assessment of records by more than one reviewer: Double screening can vary from duplicate checking of all records (by two or more reviewers independently) to a second reviewer checking a sample only (e.g. random sample of screened records, or all excluded records). This approach may be more reliable than single screening but at the expense of increased reviewer time, given the time needed to resolve discrepancies (40, 41, 43).

Priority screening to focus early screening effort on most relevant records: Instead of screening records in year, title, author or random order, machine learning is used to identify relevant studies earlier in the screening process than would otherwise be the case. Priority screening is an iterative process in which the machine continually reassesses unscreened records for relevance. This approach can increase review efficiency by enabling the review team to start on subsequent steps of the review while less relevant records are still being screened. Both single and multiple reviewer assessments can be combined with priority screening (42, 171).

Priority screening with the automatic elimination of less relevant records: Once the most relevant records have been identified using priority screening, teams may choose to stop screening based on the assumption that the remaining records are unlikely to be relevant. However, there is a risk of erroneously excluding relevant studies because of uncertainty about when it is safe to stop screening; the balance between efficiency gains and risk tolerance will be review-specific (42, 171).

Machine learning classifiers: Machine learning classifiers are statistical models that use training data to rank records according to their relevance. They can be calibrated to achieve a given level of recall, thus enabling reviewers to implement screening rules, such as eliminating records or replacing double with single screening. Because the performance of classifiers is highly dependent on the data used to build them, classifiers should only be used to classify records for which they are designed (171, 172).

Previous 'known' assessments: Screening decisions for records that have already been manually checked can be reused to exclude the same records from being reassessed, provided the eligibility criteria are the same. For example, groups that maintain registers of controlled trials to facilitate systematic reviews can avoid continually rescreening the same records by matching and then including/excluding those records from further consideration.

Crowdsourcing: Crowdsourcing involves recruiting (usually via the internet) a large group of individuals to contribute to a task or project, such as screening records. If crowdsourcing is integrated with other study selection approaches, the specific platforms used should have well-established and documented agreement algorithms, and data on crowd accuracy and reliability (173, 174).

Box 4. Assessment of risk of bias in studies and bias due to missing results*Terminology*

The terms 'quality assessment' and 'critical appraisal' are often used to describe the process of evaluating the methodological conduct or reporting of studies (67). In PRISMA 2020, we distinguish 'quality' from 'risk of bias' and have focused the relevant items and elaborations on the latter. Risk of bias refers to the potential for study findings to systematically deviate from the truth due to methodological flaws in the design, conduct or analysis (62). Quality is not well defined, but has been shown to encompass constructs beyond those that may bias the findings, including, for example, imprecision, reporting completeness, ethics and applicability (175-177). In systematic reviews, focus should be given to the design, conduct and analysis features that may lead to important bias in the findings.

Different types of risk of bias

In PRISMA 2020, two aspects of risk of bias are considered. The first aspect is risk of bias in the results of the individual studies included in a systematic review. Empirical evidence and theoretical considerations suggest that several features of study design are associated with larger intervention effect estimates in studies; these features include inadequate generation and concealment of a random sequence to assign participants to groups, substantial loss to follow-up of participants, and unblinded outcome assessment (178).

The second aspect is risk of bias in the result of a synthesis (e.g. meta-analysis) due to missing studies or results within studies. Missing studies/results may introduce bias when the decision to publish a study/result is influenced by the observed P value, magnitude or direction of the effect (103). For example, studies with statistically non-significant results may not have been submitted for publication (publication bias), or particular results that were statistically non-significant may have been omitted from study reports (selective non-reporting bias) (179, 180).

Tools for assessing risk of bias

Many tools have been developed to assess the risk of bias in studies (67, 176, 177) or bias due to missing results (104). Existing tools typically take the form of composite scales and domain-based tools (176, 181). Composite scales include multiple items which each have a numeric score attached, from which an overall summary score might be calculated. Domain-based tools require users to judge risk of bias within specific domains, and to record the information on which each judgement was based (62, 124, 182). Specifying the components/domains in the tool used in the review can help readers determine whether the tool focuses on risk of bias only or addresses other 'quality' constructs. Presenting assessments for each component/domain in the tool is preferable to reporting a single 'quality score', because it enables users to understand the specific components/domains that are at risk of bias in each study.

Incorporating assessments of risk of bias in studies into the analysis

The risk of bias in included studies should be considered in the presentation and interpretation of results of individual studies and syntheses. Different analytic strategies may be used to examine whether the risks of bias of the studies may influence the study results: (i) restricting the primary analysis to studies judged to be at low risk of bias (sensitivity analysis); (ii) stratifying studies according to risk of bias using subgroup analysis or meta-regression; or (iii) adjusting the result from each study in an attempt to remove the bias. Further details about each approach are available elsewhere (62).

Box 5. Meta-analysis and its extensions

Meta-analysis is a statistical technique used to synthesize results when study effect estimates and their variances are available, yielding a quantitative summary of results (94). The method facilitates interpretation that would otherwise be difficult to achieve if, for example, a narrative summary of each result was presented, particularly as the number of studies increases. Furthermore, meta-analysis increases the chance of detecting a clinically important effect as statistically significant, if it exists, and increases the precision of the estimated effect (183).

Meta-analysis models and methods

The summary estimate is a weighted average of the study effect estimates, where the study weights are determined primarily by the meta-analysis model. The two most common meta-analysis models are the fixed-effect and random-effects models (94). The assumption underlying the *fixed-effect* model is that there is one true (common) intervention effect, and that the observed differences in results across studies reflect random variation only. This model is sometimes referred to as the 'common-effects' or 'equal-effects' model (94). A fixed-effect model can also be interpreted under a different assumption, that the true intervention effects are different and unrelated. This model is referred to as the *fixed-effects* model (184). The *random-effects* model assumes that there is not one true intervention effect but rather, a *distribution* of true intervention effects, and that the observed differences in results across studies reflect real differences in the effects of an intervention (183). The *random-effects* and *fixed-effects* model are similar in that they assume the true intervention effects are different, but they differ in that the *random-effects* model assumes the effects are related through a distribution, while the *fixed-effects* model does not make this assumption.

Many considerations may influence an author's choice of meta-analysis model. For example, their choice may be based on the clinical and methodological diversity of the included studies and the expectation that the underlying intervention effects will differ (potentially leading to selection of a random-effects model), or, concern about small-study effects, the tendency for smaller studies to show different effects to larger ones (138) (potentially leading to fitting of both a random-effects and fixed-effect model). Sometimes authors select a model based on the heterogeneity statistics observed (e.g. switch from a fixed-effect to a random-effects model if the I^2 statistic was >50%) (185). However, this practice is strongly discouraged.

There are different methods available to assign weights in fixed-effect or random-effects meta-analyses (e.g. Mantel-Haenszel, inverse-variance) (94). For random-effects meta-analyses, there are also different ways to estimate the between-study variance (e.g. DerSimonian and Laird, restricted maximum likelihood (REML)) and calculate the confidence interval for the summary effect (e.g. Wald-type confidence interval, Hartung-Knapp-Sidik-Jonkman (97)). Readers are referred to Deeks et al. (94) for further information on how to select a particular meta-analysis model and method.

Subgroup analyses, meta-regression and sensitivity analyses

Extensions to meta-analysis, including subgroup analysis and meta-regression, are available to explore causes of variation of results across studies (i.e. statistical heterogeneity) (94). Subgroup analyses involve splitting studies or participant data into subgroups and comparing the effects of the subgroups. Meta-regression is an extension of subgroup analysis, which allows for the effect of continuous and categorical variables to be investigated (135). Authors might use either type of analysis to explore, for example, whether the intervention effect estimate varied with different participant characteristics (e.g. mild versus severe disease) or intervention characteristics (e.g. high versus low dose of a drug).

Sensitivity analyses are undertaken to examine the robustness of findings to decisions made during the review process. This involves repeating an analysis, but using different decisions to those originally made, and informally comparing the findings (94). For example, sensitivity analyses might

have been done to examine the impact on the meta-analysis of including results from conference abstracts that have never been published in full, including studies where a majority (but not all) participants were in a particular age range, including studies at high risk of bias, or using a fixed-effect versus random-effects meta-analysis model.

Sensitivity analyses differ from subgroup analyses. Sensitivity analyses consist of making informal comparisons between different ways of estimating the same effect, whereas subgroup analyses consist of formally undertaking a statistical comparison across the subgroups (94).

Extensions to meta-analysis that model or account for dependency

In most meta-analyses, effect estimates from independent studies are combined. Standard meta-analysis methods are appropriate for this situation, since an underlying assumption is that the effect estimates are independent. However, standard meta-analysis methods are not appropriate when the effect estimates are correlated. Correlated effect estimates arise when multiple effect estimates from a single study are calculated using some, or all, of the same participants, and are included in the same meta-analysis. For example, where multiple effect estimates from a multi-arm trial are included in the same meta-analysis, or effect estimates for multiple outcomes from the same study are included. For this situation, a range of methods are available that appropriately model or account for the dependency of the effect estimates. These methods include multivariate meta-analysis (186), multilevel models (187) or robust variance estimation (188). See Lopez-Lopez for further discussion (58).

Box 6. Systematic review registration and protocols

Registration aims to reduce bias, increase transparency, facilitate scrutiny and improve trustworthiness of systematic reviews (189, 190). Registration also aims to reduce unintended duplication; researchers planning a new review should search register listings to identify similar completed or ongoing reviews before deciding whether their review is needed, noting that planned duplication may be justified (189).

A registration entry captures key elements of the review protocol and is submitted to a host register ideally before starting the review. The register maintains a permanent public record of this information along with any subsequent amendments (date-stamped), and issues a unique number used to link the registration entry to completed review publications (191). Publicly recording details of inclusion and exclusion criteria, planned outcomes and syntheses enables peer reviewers, journal editors and readers to compare the completed review with what was planned, identify any deviations and decide whether these may have introduced bias.

PROSPERO (www.crd.york.ac.uk/prospero/) currently registers systematic reviews with direct health outcomes. It also accepts systematic reviews of animal studies that have direct implications for human health, and methodology reviews which have direct bearing on human health or systematic review conduct. Reviews not meeting the criteria for inclusion in PROSPERO could be registered elsewhere, for example in the Open Science Framework (OSF) repository. Both PROSPERO and OSF allow for registration without cost.

A review protocol is distinct from a register entry for a review. A review protocol outlines in detail the pre-planned objectives and methods intended to be used to conduct the review, helping to anticipate/avoid potential problems before embarking on a review and providing a methodical approach to prevent arbitrary decision-making during the review process (154). Systematic reviewers are encouraged to report their protocols in accordance with the PRISMA guidance for protocols (PRISMA-P) (192). PRISMA-P consists of a checklist (192) accompanied by a detailed guidance document providing researchers with a step-by-step approach for documenting a systematic review protocol (154).

A review protocol should be a public document in order to facilitate future purposeful replications or updates of the review and to help future users evaluate whether selective reporting and potential bias were present in the review process (154). Review protocols can be made public through one of several routes. One option is to upload a PDF of the protocol to the corresponding PROSPERO registration record so they are linked in perpetuity. Another option is to make a protocol a document with its own unique identifier (i.e. a DOI) so it can be cited across various documents including the PROSPERO registration record and in the full text of the completed review. To achieve this, reviewers may opt to publish a protocol in an open access journal (e.g. *Systematic Reviews*, *BMJ Open*, or a journal using the Registered Reports publishing framework (<https://cos.io/rr/>)), where it will benefit from external feedback prior to publication; or, deposit a protocol in an open access repository (e.g. Open Science Framework Registries, Zenodo).

Table 1. Example of Item 6 (Information sources). The table displays for each database consulted its name (e.g. MEDLINE), the interface or platform through which the database was searched (e.g. Ovid), and the dates of coverage. Reproduced from Jay et al. (32).

Database	Coverage
Ovid	
Medline and Epub Ahead of Print, In-Process & Other Non-Index Citations, Daily and Versions	1946 to present
Embase and Embase Classic	1947 to present
PsycInfo	1806 to present
Social Policy & Practice	1890s to present
Scopus	1788 to present
EBSCOhost	
British Education Index	1929 to present
Education Abstracts	1983 to present
	1995 to present (books)
The Education Resources Information Center	1966 to present
Index to Legal Periodicals and Books	1979 to present
ProQuest Central	
The Education Database	1988 to present
Social Science Database	1942 to present
The Applied Social Sciences Index & Abstracts	1987 to present
The International Bibliography of the Social Sciences	1951 to present
The Sociology Database	1985 to present
Sociological Abstracts	1952 to present
Westlaw UK	1986 to present

Table 2. Example of Item 17 (Study characteristics). The table displays for each included study the citation, study design, country, sample size, setting, mean age, percentage of females, number of years follow-up, exposure details and outcomes assessed. Reproduced from Barker et al. (122).

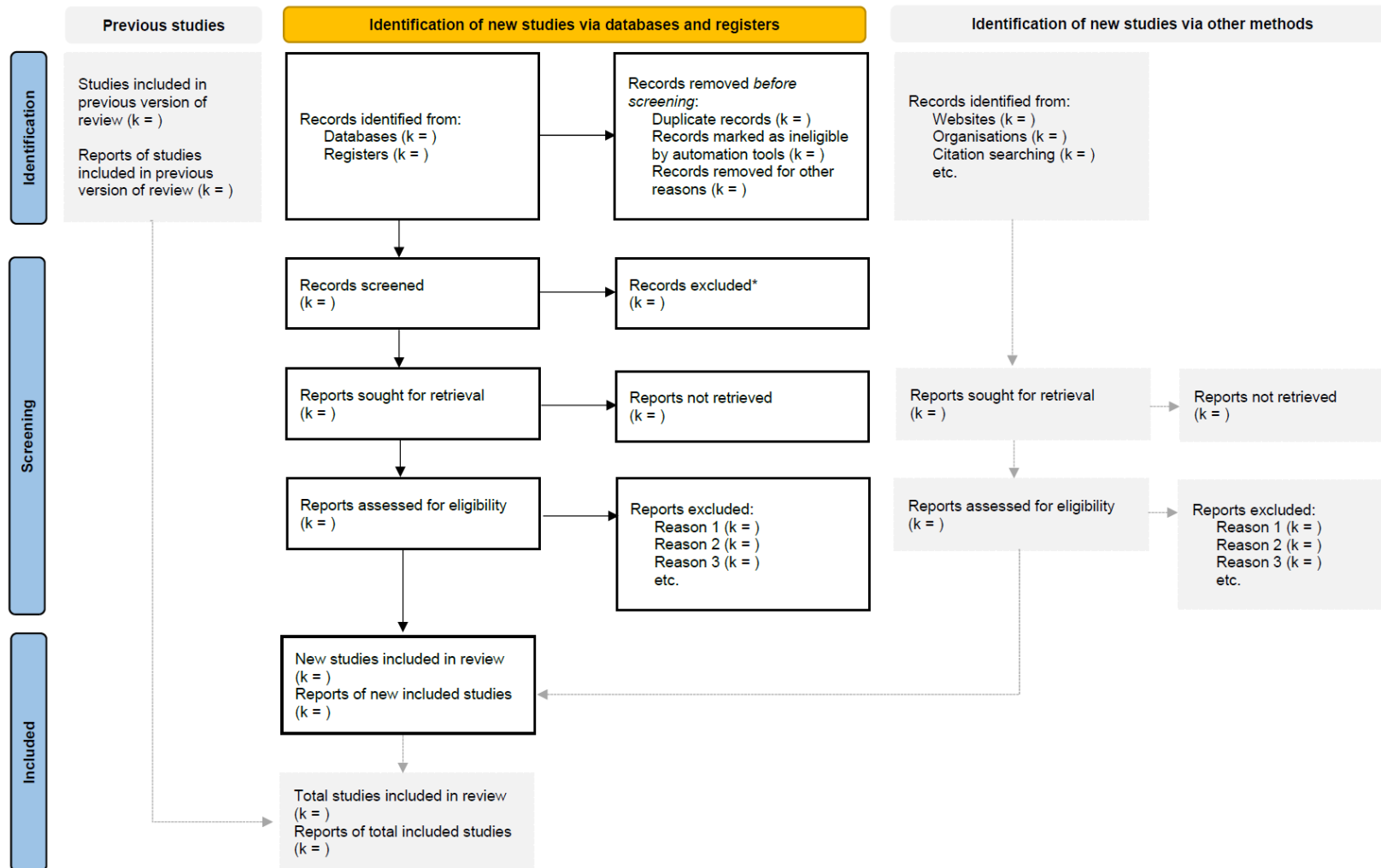
<i>Study ID</i> Author (year)	Study design	<i>Population</i>		Source of participants	Age, mean	Female, %	Follow-up (years)	<i>Exposure to aspirin</i>		<i>Outcomes</i>	
		Country	Sample size					Identification	Dose	Fracture	BMD
Bauer (1996)	Cohort	USA	7786	Community	73.1 74.1	100	1.6	Self-report	1–4 times/week 5–7 times/week NR	✓	✓
Bleicher (2011)	Cross-sectional	Australia	1705	Community	77.0	0	–	Medication verified in clinic	NR	–	✓
Bonten (2017)	Cross-sectional	Netherlands	854	Community	59.0	34	–	Medication verified in clinic	30–125 mg/day	✓	✓
Carbone (2003)	Cross-sectional	USA	2853	Community	73.6	50	–	Medication verified in clinic	328 mg/day	✓	✓
Chuang (2016)	Case-control	Taiwan	555	Community	74.0	61	5	Prescription history	106 mg	✓	–
Dobnig (2007)	Cohort	Austria	1664	Nursing homes	–	100	2	Not reported	Not reported	✓	–
Hill (2008)	Cross-sectional	Trinidad and Tobago	340	Community	63.9	100	–	Medication verified in clinic	≥3 times/week	–	✓
Hill (2008)	Cross-sectional	Trinidad and Tobago	2501	Community	56.3	0	–	Self-report	NR	–	✓
Lane (1997)	Cross-sectional	USA	499	Community	73.6	100	–	Self-report	5–7 days/week	–	✓
Vestergaard (2006, 2012)	Case-control	Denmark	498 617	Community	43.4	52	1	Prescription history	≤150 mg/day	✓	–
Vestergaard (2012)	Cohort	Denmark	2016	Community	50.8	100	10	Self-report	325 mg/day	✓	✓

Table 3. Example of Item 18 (Risk of bias in studies). The table displays for each included study the risk-of-bias judgement for each of six domains of bias, and for the overall risk of bias in two results (selection of a product, consumption of a product); the following is an abridged version of the table presented in the review. Reproduced from Hollands et al. (69).

Study	Bias arising from the randomisation process	Bias arising from the timing of identification and recruitment of individual participants in relation to timing of randomisation (CRCT only)	Bias due to deviations from intended interventions	Bias due to missing outcome data	Bias in measurement of the outcome	Bias in selection of the reported result	Overall risk of bias (selection of a product)	Overall risk of bias (consumption of a product)
Fiske 2004	Some concerns	Low risk	Low risk	Low risk	Low risk	Low risk	Some concerns	Not applicable
Foster 2014	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk	Not applicable
Kocken 2012	Some concerns	Low risk	Low risk	Low risk	Low risk	Low risk	Some concerns	Not applicable
Pechey 2019	Some concerns	Not applicable	Low risk	Low risk	Low risk	Low risk	Some concerns	Not applicable
Roe 2013	Some concerns	Not applicable	Low risk	Low risk	Low risk	Low risk	Some concerns	Some concerns
Stubbs 2001	Some concerns	Not applicable	Low risk	Low risk	Low risk	Low risk	Not applicable	Some concerns

CRCT: cluster-randomised controlled trials. Justifications for assessments are available at the following (<http://dx.doi.org/10.6084/m9.figshare.9159824>).

Figure 1. PRISMA 2020 flow diagram template for systematic reviews



*If automation tools were used, indicate how many records were excluded by a human and how many were excluded by automation tools. The new design is adapted from flow diagrams proposed by Boers (193) and Mayo-Wilson et al. (53) and Stovold et al. (118). The boxes in grey should only be completed if applicable; otherwise they should be removed from the flow diagram. Note that a “report” could be a journal article, preprint, conference abstract, study register entry, clinical study report, dissertation, unpublished manuscript, government report or any other document providing relevant information.

Figure 2. Example of Item 19 (Results of individual studies). The figure displays for each study included in the meta-analysis the summary statistics (number of events and sample size) for the quadruple and triple combination antiretroviral therapies (cART) groups, and the risk ratio (RR) and its 95% confidence interval for the dichotomous outcome, undetectable HIV-1 RNA. Reproduced from Feng et al. (128).

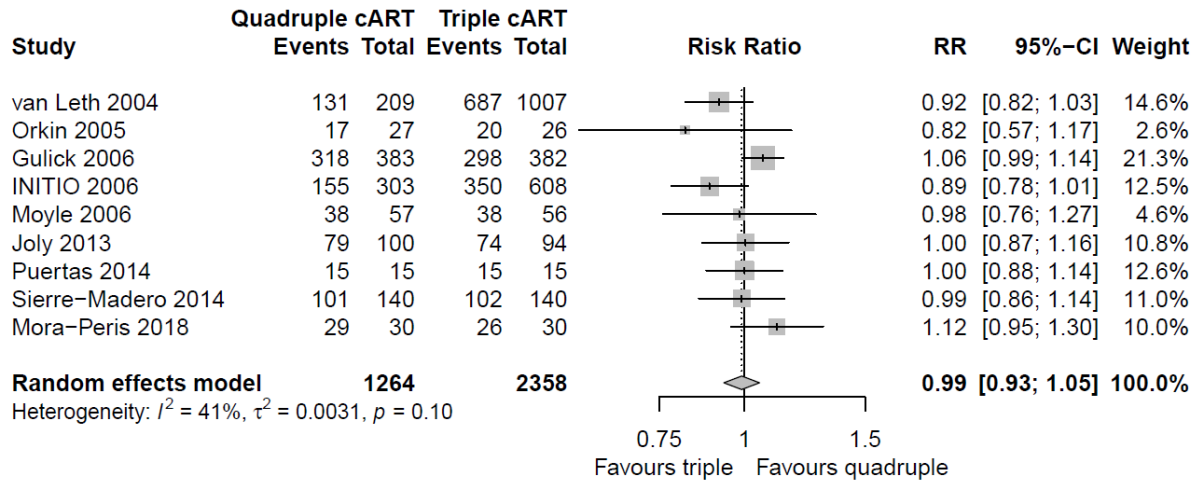


Figure 3. Example of Item 19 (Results of individual studies). The figure displays for each study included in the meta-analysis the summary statistics (mean, standard deviation (SD) and sample size) for the quadruple and triple combination antiretroviral therapies (cART) groups, and the mean difference (MD) and its 95% confidence interval for the continuous outcome, CD4 T cell count (cells/ μ L). Reproduced from Feng et al. (128).

