BMJ Open

Identifying and addressing conflicting results across multiple discordant systematic reviews on the same topic: protocol for a replication study of the Jadad algorithm

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2021-054223
Article Type:	Protocol
Date Submitted by the Author:	10-Jun-2021
Complete List of Authors:	Lunny, Carole; University of Toronto Dalla Lana School of Public Health, Epidemiology Division and Institute for Health Policy, Management, and Evaluation Thirugnanasampanthar, Sai Surabi; University of Toronto Dalla Lana School of Public Health, Epidemiology Division; St Michael's Hospital Li Ka Shing Knowledge Institute, Knowledge Translation Program Kanji, Salmaan; The Ottawa Hospital, Department of Pharmacy; University of Ottawa, Faculty of Medicine Ferri, Nicola; University of Bologna, Department of Biomedical and Neuromotor Sciences Thabet, Pierre; Hôpital Montfort Pieper, Dawid; Witten/Herdecke University, Cologne, Germany Whitelaw, Sera; McMaster University, Department of Health Research Methods, Evidence, and Impact Tasnim, Sara; The University of British Columbia Faculty of Medicine, Anesthesiology, Pharmacology & Therapeutics Nelson, Harrison; Queen's University Faculty of Health Sciences Reid, Emma; Nova Scotia Health Authority Zhang, Jia He (Janet); The University of British Columbia Kalkat, Banveer; The University of British Columbia, Faculty of Pharmaceutical Sciences Chi, Yuan; Beijing Yealth Technology Co., Ltd, Yealth Network Thompson, Jacqueline; University of Birmingham Institute of Applied Health Research Abdoulrezzak, Reema; The University of British Columbia, Faculty of Pharmaceutical Sciences Zheng, Di Wen (Wendy); The University of British Columbia, Faculty of Pharmaceutical Sciences Pangka, Lindy; The University of British Columbia, Faculty of Pharmaceutical Sciences Wang, Dian (Xin Ran); The University of British Columbia, Faculty of Pharmaceutical Sciences Safavi , Parisa; The University of British Columbia, Faculty of Pharmaceutical Sciences Sooch, Anmol; The University of British Columbia, Faculty of Pharmaceutical Sciences Kang, Kevin; The University of British Columbia, Faculty of Pharmaceutical Sciences Tricco, Andrea; Li Ka Shing Knowledge Institute of St Michael's Hospital,

	Knowledge Translation Program; University of Toronto Dalla Lana School of Public Health, Epidemiology Division and Institute for Health Policy, Management, and Evaluation
Keywords:	PUBLIC HEALTH, STATISTICS & RESEARCH METHODS, EPIDEMIOLOGY

Title: Identifying and addressing conflicting results across multiple discordant systematic reviews on the same topic: protocol for a replication study of the Jadad algorithm

Authors:

- Lunny C, PhD[1]*, carole.lunny@ubc.ca, https://orcid.org/0000-0002-7825-6765, St Michael's Hospital, Unity Health Toronto and the Cochrane Hypertension Review Group, University of British Columbia, Vancouver, BC, Canada, V6T 1Z3
- Sai Surabi Thirugnanasampanthar, BHSc Saisurabi.Thirugnanasampanthar@unityhealth.to, Knowledge Translation Program, Li Ka Shing Knowledge Institute, St. Michael's Hospital, Unity Health and the Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada
- Kanji S, Pharm.D., skanji@toh.ca, The Ottawa Hospital and Ottawa Hospital Research Institute
- Ferri N, PT Nicola.ferri11@unibo.it, Department of Biomedical and Neuromotor Sciences (DIBINEM), Alma Mater Studiorum, University of Bologna, 40138, Bologna, Italy
- Thabet P, B. Pharm, pierrethabet@montfort.on.ca, Hôpital Montfort, Ottawa, Canada
- Pieper D, PhD, Dawid.Pieper@uni-wh.de, Institute for Research in Operative Medicine, Witten/Herdecke University, Germany
- Whitelaw S, MSc whitelsm@mcmaster.ca, Department of Health Research Methods, Evidence and Impact, McMaster University, Hamilton, ON, Canada
- Tasnim S, MSc, stasnim@mail.ubc.ca, Cochrane Hypertension Review Group, Therapeutics Initiative, Department of Anesthesiology, Pharmacology & Therapeutics, Faculty of Medicine, University of British Columbia, 2176 Health Science Mall, Vancouver, BC, Canada, V6T 1Z3
- Harrison Nelson, harrison.nelson@queensu.ca, https://orcid.org/0000-0001-8306-5448, Faculty of Health Sciences, Queen's University, Kingston, ON, Canada
- Emma Reid, Pharm.D, emmak.reid@nshealth.ca, Nova Scotia Health, Halifax, Nova Scotia, Canada
- Jia He (Janet) Zhang, BSc, jiahe.janet@gmail.com, University of British Columbia, Vancouver, BC, Canada
- Banveer Kalkat, banveerk@student.ubc.ca, Faculty of Pharmaceutical Sciences, University of British Columbia, Vancouver, BC, Canada
- Yuan Chi. yuan.chi@outlook.com, Beijing Yealth Technology Co., Ltd, Yealth Network, Beijing, China

- Jacqueline Y Thompson, MPH j.y.thompson@bham.ac.uk, https://orcid.org/0000-0002-9775-361X, Institute of Applied Health Research, University of Birmingham, Edgbaston, Birmingham B15 2TT
- Reema Abdoulrezzak, BSc reemaabd@student.ubc.ca, Faculty of Pharmaceutical Sciences, University of British Columbia, Vancouver, BC, Canada
- Di Wen (Wendy) Zheng, BSc wendyzhe@student.ubc.ca, Faculty of Pharmaceutical Sciences, University of British Columbia, Vancouver, BC, Canada
- Lindy R.S. Pangka, BSc lpangka@shaw.ca, Faculty of Pharmaceutical Sciences, University of British Columbia, Vancouver, BC, Canada
- Dian (Xin Ran) Wang, BSc dianw@student.ubc.ca, Faculty of Pharmaceutical Sciences, University of British Columbia, Vancouver, BC, Canada
- Parisa Safavi, BSc psafavi@student.ubc.ca, Faculty of Pharmaceutical Sciences, University of British Columbia, Vancouver, BC, Canada
- Anmol Sooch, BSc asooch19@student.ubc.ca, Faculty of Pharmaceutical Sciences, University of British Columbia, Vancouver, BC, Canada
- Kevin T. Kang, BSc ktw9412@student.ubc.ca, Faculty of Pharmaceutical Sciences, University of British Columbia, Vancouver, BC, Canada
- Tricco AC, PhD, 1.Knowledge Translation Program, Li Ka Shing Knowledge Institute, St. Michael's Hospital, Unity Health Toronto, 209 Victoria St, M5B 1T8, Toronto, Ontario, Canada.; 2. Epidemiology Division and Institute for Health Policy, Management, and Evaluation, Dalla Lana School of Public Health, University of Toronto, 155 College St Room 500, M5T 3M7, Toronto Ontario, Canada; 3.Queen's Collaboration for Health Care Quality Joanna Briggs Institute Centre of Excellence, School of Nursing, Queen's University, 99 University Ave, K7L 3N6, Kingston, Ontario, Canada

*Corresponding author

Lunny C, PhD[1]*, carole.lunny@ubc.ca, https://orcid.org/0000-0002-7825-6765, St Michael's Hospital, Unity Health Toronto and the Cochrane Hypertension Review Group, University of British Columbia, Vancouver, BC, Canada, V6T 1Z3

ABSTRACT

Introduction: An increasing growth of systematic reviews (SRs) presents notable challenges for decision-makers seeking to answer clinical questions. Overviews of systematic reviews aim to address these challenges by summarising results of SRs and making sense of potentially discrepant SR results and conclusions. In 1997, an algorithm was created by Jadad to assess discordance in results across SRs on the same topic. Since this tool pre-dates the advent of overviews, it has been inconsistently applied in this context. Our study aims to (a) replicate assessments done in a sample of overviews using the Jadad algorithm to determine if the same SR would have been chosen, (b) evaluate the Jadad algorithm in terms of utility, efficiency, and comprehensiveness, and (c) describe how overviews address discordance in results across multiple SRs.

Methods and Analysis: We will use a database of 1218 overviews (2000-2020) created from a bibliometric study as the basis of our search for overviews assessing discordance. This bibliometric study searched MEDLINE (Ovid), Epistemonikos, and Cochrane Database for overviews. We will include any overviews using Jadad (1997) or another method to assess discordance. The first 30 overviews screened at the full-text stage by two independent reviewers will be included. We will replicate Jadad assessments in overviews. We will compare our outcomes qualitatively and evaluate the differences between our Jadad assessment of discordance and the overviews' assessment.

Ethics and Dissemination: No ethics approval was required as no human subjects were involved. In addition to publishing in an open-access journal, we will disseminate evidence summaries through formal and informal conferences, academic websites, and across social media platforms. This is the first study to comprehensively evaluate and replicate Jadad algorithm assessments of discordance in SRs.

Abstract: 281 words

Strengths and limitations of this study

- This is the first proposed empirical study to use a systematic approach to evaluate authors' assessment of discordance across SRs in overviews and replicate Jadad algorithm assessments from a sample of overviews
- When assessing discordant SRs using the Jadad algorithm, there will be subjectivity in the judgements potentially introducing variability in the results. For example, some steps in the Jadad algorithm were vague in description, making it difficult to interpret.
- Therefore, a strength of the proposed study will be that we used an iterative process among authors to develop decision rules for the interpretation and application of each step in the Jadad algorithm. Upon completion and dissemination of this study, other overview authors will have more detailed guidance on how to apply the Jadad algorithm when addressing discordance in overviews.
- We have transparently described our interpretation of each step in the Jadad algorithm to aid in replication.

1.0 BACKGROUND

Information overload is an increasing problem for health practitioners, researchers, and decision-makers. Global research output is growing rapidly, and the number of published systematic reviews (SRs) being produced yearly is also expanding [1]. Between January and October 2020, 807 SRs on COVID-19 alone were published in PubMed [2], and the rate of growth from 1995 to 2017 in SRs was found to be 4676% [3]. Challenges in dealing with growth in SRs include identification of high quality, comprehensive, and recent reports on the topic of interest.

'Overviews of systematic reviews' (henceforth called overviews) evolved in response to these challenges [4-7]. Overviews summarise the results of SRs, and help make sense of potentially conflicting, discrepant, and overlapping results and conclusions of SRs on the same topic [8-12].

A major advantage of overviews is to identify and explain the discordance between conflicting SRs on the same topic [13-15]. We define discordance as when SRs with similar research questions and inclusion criteria report different results or conclusions. A common method for dealing with multiple SRs with discordant results is to specify methodological criteria to select only one SR (e.g., select the highest quality and most comprehensive SR) [12]. However, many other methods have been proposed, including statistical approaches to address discordance in results across SRs [11, 12].

As early as 1997, Jadad et al. [16] developed a decision tree (i.e. an algorithm) to assess discordance in results across a sample of SRs on the same topic to aid healthcare providers in making clinical decisions. The Jadad decision tree guides users through a methodological assessment of SR components to identify potential causes of discordance and ultimately choose the best SR among multiple on the same topic. Jadad et al.'s methods appear to be the only algorithm available to both explain and deal with discordance across SRs.

Despite the availability of this tool since 1997, it has not been universally adopted and has been inconsistently applied when used [17-19]. The aim of this study is to comprehensively replicate and evaluate the Jadad algorithm for assessing discordance across SRs.

2.0 OBJECTIVES

Our study objectives are:

1) Describe how overviews address discordance in results across multiple SRs (study 1)
2) Replicate Jadad assessments from published overviews to identify sources of discordance and to determine if the same review(s) was chosen as the "best available evidence" (study 2)
3) Evaluate the Jadad algorithm in terms of utility, efficiency, and comprehensiveness (study 2)

3.0 METHODS

3.1. Study design

This is a methods study in the knowledge synthesis field. We followed systematic review guidance for the study selection and data extraction stages [20]. Our protocol will be registered and located on the Open Science Framework at https://osf.io/nbcta/.

3.2 Search and selection of overviews

3.2.1 Database of 1218 overviews

We will use a database of 1218 overviews published between January 1, 2000 and December 30, 2020 created from a bibliometric study [21] as the basis of our search for overviews using Jadad or another method to assess discordance. This bibliometric study searched the MEDLINE (Ovid), Epistemonikos, and the Cochrane Database of Systematic Reviews for overviews. The database Epistemonikos contains both published and unpublished reports. Overviews included in the bibliometric study contain these characteristics: (a) synthesised the results of SRs, (b) systematically searched for evidence in a minimum of two databases, and (c) conducted a search using a combination of text words and MeSH terms. The included overviews also had to have a methods section in the main body of the paper and focused on health interventions. To identify overviews assessing discordance using the database of 1218 overviews, we will use the EndNote search function and Boolean logic to search for the following words: overlap*[title/abstract] or discrepan*[title/abstract] or discord*[title/abstract] or concord*[title/abstract] or conflict*[title/abstract] or Jadad [abstract].

3.2.2 Medline (Ovid) search strategy January to April 2021

We will update this search for overviews published in the first quarter of 2021 with an Ovid MEDLINE search using the following search string: ("systematic reviews".tiab or "meta-analyses".tiab) AND (overlap.tiab or discrepant.tiab or discordant.tiab or difference.tiab or conflicting.tiab or Jadad.ab). Our search was conducted on April 18, 2021.

3.3 Screening overviews

3.3.1 Process for screening overviews

Citations identified by our searches will be assigned a random number and screened sequentially. The first 30 overviews screened at full-text and meeting our eligibility criteria will be included.

All authors will pilot the screening form on 20 overviews to ensure high levels of agreement and common definitions of eligibility criteria.

Two authors will independently screen overviews as full-text publications. Discrepancies will be resolved by consensus, and arbitration by a third reviewer when necessary.

3.3.2 Stage 1 screening criteria

We first include all overviews aiming to assess discordant results across SRs on the same topic, which may or may not assess discordant interpretations and conclusions using any approach (e.g. [22-25]).

If a study meets stage 1 criteria, it will be included in Study 1.

3.3.3 Stage 2 screening criteria

From this sample, we will then screen overviews based on the following inclusion criteria:

- Must have included a minimum of 2 SRs with a meta-analysis of randomised controlled trials (RCTs), but may have included other study types beyond RCTs; and
- Used the Jadad algorithm.

If a study meets stage 2 screening, it will move onto stage 3 screening.

3.3.4 Stage 3 screening

After stage 2 screening is complete, we will screen based on the authorship team. When the same 'core' authors (first, last, and/or corresponding) conduct two or more of the identified overviews, we will only include one of the multiple overviews (i.e. the most recent study will be selected). Our rationale is that author groups use the same methods to assess discordance (e.g. Mascarenas [18] and Chalmers [15]).

We will include overviews in any language and publication status, published anytime. We will use Google translate to interpret non-English studies for screening and assessment. The resulting publications will form the set for study 2.

In the case where the authors of this study are also authors of one or several of the included overviews, those authors will not conduct screening, extraction, Jadad assessment, or analysis of the overview in question.

Full-texts of all SRs included in an overview will be obtained.

3.4 Extraction of the primary intervention and outcome

3.4.1 Identification of the primary outcome from the overview

As a first step in assessing discordance, we will identify the primary outcome from each overview. The primary outcome will be extracted when it is explicitly defined in the title, abstract, objectives, introduction or methods sections [26, 27]. If the primary outcome is not found in any of these sections, we will extract it as the outcome that is reported in a power calculation; or the first outcome mentioned in the manuscript [26, 27] or the most serious outcome. If the primary outcome cannot be identified by any of these approaches, we will consider that the article did not specify primary outcomes and the study will be excluded, and replaced with the next overview in our database.

3.4.2 Identification of the primary intervention from the overview

If multiple interventions are present, we will choose the intervention comparison with the "experimental" intervention vs. placebo or standard of care for the primary outcome, or the first intervention highlighted in the title or abstract of the overview [27].

3.4.3 Identification of included SRs with meta-analysis of RCTs addressing the primary intervention/outcome of the overview

Once the primary outcome is identified, we will next identify how many SRs with meta-analysis of RCTs were included in the overview that address the primary outcome and intervention. It is this sample of SRs with meta-analysis of RCTs that will be the focus of our Jadad assessments.

3.4.4 Process to identify primary intervention and outcome

Two authors will extract the primary intervention and outcome, and disagreements will be discussed until consensus is reached.

3.5 Blinding of results in the included SRs with meta-analysis of RCTs

Observer bias, sometimes called “detection bias” or “ascertainment bias,” occurs when outcome assessments are systematically influenced by the assessors’ conscious or unconscious predispositions [28]. Blinded outcome assessors are used in trials to avoid such bias. One empirical study found evidence of high risk of substantial bias when authors failed to blind outcome assessors in trials [28], whereas another did not [29]. In our study, it is important that reviewers are blinded to the overview authors' result of the Jadad assessment, as unblinding might predispose them to unconsciously choose the same review as the overview authors.

We will blind the following components containing study results of the Jadad assessment and conclusions: abstract, highlights, results of the Jadad assessment, and discussion/conclusions section. Blinding will be achieved via deletion using the paywalled Adobe Acrobat Pro or the freeware PDFCandy (https://pdfcandy.com). One author will blind the overview results and will not be involved in the Jadad assessment pertaining to those results. Assessors will be instructed not to search for and read the included overviews prior to, and during, assessment.

3.6 Piloting Jadad assessment prior to full assessment

A pilot practice exercise will be conducted by all assessors prior to the Jadad assessments, to ensure consistent assessments across reviewers. Two Jadad assessments will be piloted by each reviewer and compared to a second to identify discrepancies that are to be resolved through discussion. Any necessary revisions to the assessment (sections 3.7 and 3.8) will be noted.

3.7 Jadad assessments of discordance across SRs

While the Jadad paper provides an algorithm intended to identify and address discordance between SRs in an overview, there is limited guidance within the manuscript regarding the application/operationalization of the algorithm. Absence of this detailed guidance leaves room for subjective (mis)interpretation and ultimately confusion when it comes time to use the algorithm. To address this, we engaged in an iterative process of interpretation and implementation of the algorithm step by step. This process involved virtual meetings whereby consensus was sought for decision rules at each step of the algorithm to ensure consistency in both interpretation and application. Feedback was solicited and decision rules were accordingly adjusted until consensus was achieved. This tool underwent pilot testing as described in 3.6 where feedback was further solicited and adjustments were made.

Two researchers will independently assess each set of SRs in the included overview using the Jadad algorithm, starting with Step C (**Figure 1**). Information and data from the overview will be used, and when data is not reported, we will consult the full text of the included SRs. The Jadad decision tree assesses and compares sources of inconsistency between SRs with meta-analyses, including differences in clinical questions, inclusion and exclusion criteria, extracted data, methodological quality assessments, data combining, and statistical analysis methods.

Step A is to examine the multiple reviews matching the overview question using a PICO framework. If the research questions are not identical, then step B indicates choosing the review closest to the decision makers' research question and no further assessment is necessary. If multiple reviews are found with the same PICO, then step C should be investigated. As we are using overviews examining discordance as our sample, we will start at Step C in the Jadad decision tree.

Here we detail our interpretation of the Jadad algorithm for each step in assessing the discordance in a group of SRs with similar PICO elements. If an overview or the included review does not report a method, we will indicate it as "not reported," and it will not be chosen for that step.

Step D and G follow from Step C. Steps E, F, H. and I are completed depending on the decisions at Steps D and G, respectively.

"Meeting" a step means a review met the criteria in the sub-step or step that is highest in the hierarchy. For example, a review that meets E3 criteria fulfills criteria A and B, which is the highest in our hierarchy.

3.7.2 Step C - Do the reviews include the same trials for the primary intervention and outcome?

We will determine if the RCTs are similar across reviews by either finding this information in the overview, or extracting all RCTs from the included reviews using an excel matrix to list the reviews at the top, and trials in the left rows. The RCTs will be mapped to the reviews in order of publication date (earliest trials at the top). Using this matrix, we will determine if the reviews include the same or different trials.

3.7.3 Step D - Are the reviews of the same quality?

If the reviews contain the same trials, then the assessor moves to Step D – assess whether the reviews are the same methodological quality. We will either (a) extract the risk of bias/quality assessments from the overview if the overview used AMSTAR [30], AMSTAR 2 [31], or ROBIS [32]. If the overview authors used any other tool or method to assess the risk of bias/quality of the reviews or did not assess the risk of bias/quality assessment at all, we will conduct our own assessment using the ROBIS tool [32].

3.7.4 Step E – Assess and compare data extraction, clinical heterogeneity, and data synthesis

If the reviews are the same risk of bias/quality, then the next step is Step E, to assess and compare data extraction, clinical heterogeneity, and data synthesis across the reviews.

Step E1 - Assess and compare the data extraction methods across reviews

For this step, Jadad states, "If reviews differ [in outcomes reported], the decision-maker should identify the review that takes into account the outcome measures most relevant to the problem that he or she is solving." We interpret this step as selecting the review that (A) matches the overview's primary outcome.

Jadad then writes that reviews that conduct independent extractions by two reviewers are of the highest quality. We therefore decided that reviews that (B) used an independent data extraction process using two review authors should be chosen. If a ROBIS assessment is done, then the latter point will be mapped to ROBIS 3.1. "Were efforts made to minimise error in data collection?"

Decision rules:

#1. Reviews that meet criteria A and B are highest in our hierarchy

#2. Reviews that meet criteria A are second highest in our hierarchy

#3. Reviews that meet criteria B are third highest in our hierarchy

Step E2 – Assess and compare clinical heterogeneity of the included RCTs across reviews

Clinical heterogeneity is assessed at the review level by examining the research question pertaining to the primary outcome and the eligibility criteria PICO elements of each included RCT to see if they are sufficiently similar. If the PICO across RCTs are similar, then clinical heterogeneity is minimal, and reviews can progress with pooling study results in a meta-analysis. If a ROBIS assessment is done, this question is mapped to ROBIS 4.3 "Was the synthesis appropriate given

the nature and similarity in the research questions, study designs, and outcomes across included SRs?"

If a review states that (A) they assessed for clinical (e.g., PICO) heterogeneity across RCTs (in the methods or results sections), then this will be the review that is chosen at this step. Example of a review reporting a clinical heterogeneity assessment: "If we found 3 or more systematic reviews with similar study populations, treatment interventions, and outcome assessments, we conducted quantitative analyses (Gaynes 2014)". If authors reported and described clinical heterogeneity in the manuscript, then rule (B) authors that judged the clinical heterogeneity assessment to be minimal or low with rationale, will be chosen at this step.

Decision rule:
#1. Reviews that meet criteria A and B are highest in the hierarchy
#2. Reviews that meet criteria A are second highest in our hierarchy

Step E3 – Assess and compare data analysis methods across reviews

Jadad et al. are purposefully vague when describing how to judge whether a meta-analysis was appropriately conducted. For this step, we interpret it as reviews reported conducting an: (A) appropriate weighted technique to combine study results (i.e. used a fixed or random random-effects model) and (B) whether authors conducted an investigation of statistical heterogeneity (i.e. by reporting I^2, tau^2, or chi^2) (**Figure 2**).

Decision rules for if the presence or absence of heterogeneity is present in the meta-analysis:
#1. Reviews that meet criteria A and B are highest in our hierarchy
#2. Reviews that meet criteria A only are second highest in our hierarchy
#3. Reviews that meet criteria B only are third highest in our hierarchy (this decision can be ignored if heterogeneity is not observed)

Decision rules for Step E.

#1. Reviews that meet Step E1, E2, and E3 are highest in our hierarchy
#2. Reviews that meet Step E1 and E2 second highest in our hierarchy
#3. Reviews that meet Step E1 third highest in our hierarchy
#4. Reviews that meet Step E2 and E3 fourth highest in our hierarchy

#5. Reviews that meet Step E2 fifth highest in our hierarchy

#6. Reviews that meet Step E3 sixth highest in our hierarchy

Note: Reporting only Steps E1, E2 or E3 is not considered a systematic approach to evidence synthesis.

3.7.5 Step F - Select the review with the lowest risk of bias, or the highest quality

From the risk of bias/quality assessment conducted through Step D, we will choose the review with the lowest risk of bias judgement, or highest quality assessment rating. ROBIS contains a last phase where reviewers are asked to summarise concerns identified in each domain and describe whether the conclusions were supported by the evidence. Based on these last decisions, a final review rating will be made based on high, low or unclear risk of bias. For our Jadad assessment, we will choose a binary rating of either high risk or low risk of bias. Any reviews assessed as 'Unclear' risk of bias will be deemed as high risk. When using the assessments of risk of bias/quality of reviews from the included overviews, we will choose the rating of the authors. If uncertainty exists, we will re-assess the included reviews using ROBIS.

3.7.6 Step G - Do the reviews have the same eligibility criteria?

If the reviews do not include the same trials, then decision-makers are directed to turn to Step G – assess whether the reviews have the same eligibility criteria (**Figure 3**). The overview may contain text in a methods section, or a characteristics of included reviews table where the PICO eligibility criteria can be extracted and assessed. If this is not the case, then the PICO eligibility criteria will be extracted from the included reviews by two authors independently and then compared to resolve any discrepancies.

2.8.7 Step H - Assess and compare the search strategies and the application of the eligibility criteria across reviews

If the reviews contain the same eligibility criteria, then Step H is to assess and compare the search strategies and the application of the eligibility criteria across reviews (**Figure 4**).

Step H1 - Assess and compare the search strategies across reviews

In this step, Jadad et al.'s recommendations are vague, although they make reference to comprehensive search strategies as being less biased. We interpret this step as authors explicitly describing their search strategy such that it can be replicated. To meet this interpretation, our

criteria are that reviews: (A) search 2 or more databases, (B) search the grey literature; and (C) include a full search algorithm (may be attached as an appendix or included in the manuscript).

Decision rules:

#1. Reviews that meet criteria A, B and C are highest in our hierarchy

#2. Reviews that meet criteria A and B are second highest in our hierarchy

#3. Reviews that meet criteria A and C are third highest in our hierarchy

#4. Reviews that meet criteria B and C are fourth highest in our hierarchy (unlikely scenario)

#5. Reviews that meet criteria A only are fifth highest in our hierarchy

SCENARIOS for Step H1

• 3 reviews are identified for our Jadad assessment

Criteria to choose a systematic review at Step H1: (A) 2 or more databases – (B) searched grey literature --(C) full search in appendix

Scenario 1

Review 1: A and B, but not C (decision rule #2)

Review 2: A and B but not C (decision rule #2)

Review 3: A and C, but not B (decision rule #3)

Conclusion: No review meets ALL of our criteria; which do we choose? Based on our decision rules, we choose BOTH Review 1 and 2

Scenario 2

Review 1: A, but neither B nor C (decision rule #5)

Review 2: A and B, but not C (decision rule #2)

Review 3: NeitherA, B, nor C (does not report the search methods)

Conclusion: No review meets ALL of our criteria; which do we choose? Based on our decision rules, we choose Review 2

Step H2 - Assess and compare the application of the eligibility criteria across reviews

In this sub-step, Jadad indicates that we should choose the review with the most explicit and reproducible inclusion criteria, which is ambiguous. Jadad states, “Reviews with the same

selection criteria may include different trials because of differences in the application of the criteria, which are due to random or systematic error. Decision-makers should regard as more rigorous those reviews with explicit, reproducible inclusion criteria. Such criteria are likely to reduce bias in the selection of studies" [16]. We did not know if this meant clearly reproducible PICO eligibility criteria, as this would be a repeat to Step G, whether the eligibility criteria were applied consistently by reviews (i.e. compare eligibility criteria to included RCTs' PICO to see if they indeed met the eligibility criteria), or if this meant (A) independently screening of title, abstracts, and full text against the eligibility criteria by two reviewers. We selected the latter criteria when choosing from the included reviews in an overview.

Decision rules:

#1. Reviews that meet criteria A is highest in our hierarchy

Decision rules for Step H:

#1. Reviews that meet Step H1 and H2 highest in our hierarchy

#2. Reviews that meet Step H1 second highest in our hierarchy

#3. Reviews that meet Step H2 third highest in our hierarchy

3.7.8 Step I – Assess and compare the publication status, quality, language restrictions of the included RCTs, and analysis of data on individual patients

If the reviews do not have the same eligibility criteria, then the next step, Step I, is to assess and compare the publication status, quality, language restrictions of the included RCTs, and analysis of data on individual patients across the reviews (**Figure 5**). This step maps to ROBIS item 1.5, namely, "Were any restrictions in eligibility criteria based on appropriate information sources (e.g. publication status or format, language, availability of data) [32]?"

Step I1 – Assess and compare the publication status of the included RCTs across reviews

In the absence of clear guidance, we interpret this step as "choose the review that searches for and includes both published and unpublished data (grey literature)." Published studies are defined as any study or data published in a peer-reviewed medical journal. Unpublished data is defined as any information that is difficult to locate and obtained from non-peer-reviewed sources such as websites (e.g. World Health Organisation website, CADTH), clinical trial registries (e.g. clinicaltrials.gov), thesis and dissertation databases, and other unpublished data registries (e.g.

LILIACS). Our interpretation is that reviews are chosen at this step that search for: (A) studies published in peer-reviewed medical journals, and (B) reports/documents/content that are not published in medical journals.

Decision rules:
#1. Reviews that meet criteria A and B are highest in our hierarchy
#2. Reviews that meet criteria A are second highest in our hierarchy
#3. Reviews that meet criteria B are third highest in our hierarchy
Note: Reporting only A or B is not considered a systematic search.

Step I2 – Assess and compare the methods used to assess the quality of the included RCTs across reviews

In this step, the Jadad paper recommends assessing the appropriateness of the *methods used* to assess the quality of the included RCTs across reviews. This item maps to ROBIS item 3.4, "Was the risk of bias/quality of RCTs formally assessed using appropriate criteria?" Here we interpret this item as to whether the review authors used the Cochrane risk of bias tool (version 1 or 2). All other RCT quality assessment tools are inappropriate because they are out of date and omit important biases (e.g. Agency for Healthcare Research and Quality (AHRQ) 2012 [33] omits allocation concealment). However, the Cochrane risk of bias tool was only published in October 2008. Therefore, we applied a decision rule: for reviews dated 2012 (giving one year for awareness of the tool to reach researchers) and later, the Cochrane risk of bias tool is considered the gold standard. For reviews dated 2009 or earlier, we considered the Jadad scale [34] and Schulz [35] to be the most common scales used between 1995 and 2011. Other tools will be considered on a case-by-case basis.

As a decision hierarchy, to meet the minimum criteria for this step, a review will have (A) assessed the risk of bias of RCTs using any tool or approaches, and (B) used the Cochrane risk of bias tool version 1 or 2 (if dated 2009 or later). If several reviews are included that meet these two criteria, the review that (C) integrates the risk of bias assessments into the results or discussion section (i.e. discusses the risk of bias in terms of high and low risk of bias studies, reports a subgroup or sensitivity analysis) will be chosen.

Decision rules:

#1. Reviews that meet criteria A, B and C are highest in our hierarchy

#2. Reviews that meet criteria B and C are second highest in our hierarchy

#3. Reviews that meet criteria A and B are third highest in our hierarchy

#4. Reviews that meet criteria A and C are fourth highest in our hierarchy (unlikely scenario)

#5. Reviews that meet criteria A only are fifth highest in our hierarchy

SCENARIOS for Step I2

• 3 reviews are identified for our Jadad assessment

Scenario 1

Review 1: A and B but not C (decision rule #3)

Review 2: A and B but not C (decision rule #3)

Review 3: A and C, but not B (decision rule #4)

Conclusion: No review meets ALL of our criteria; which do we choose? Based on our decision rules, we choose BOTH Review 1 and 2

Scenario 2

Review 1: A, but neither B nor C (decision rule #5)

Review 2: A and B, but not C (decision rule #3)

Review 3: NeitherA, B, nor C (does not report the search methods)

Conclusion: No review meets ALL of our criteria; which do we choose? Based on our decision rules, we choose Review 2

Step 13 - Assess and compare any language restrictions across reviews

In this step, Jadad indicates that reviews with (A) no language restrictions in eligibility criteria should be prioritised and chosen over those that only include English language RCTs. This step maps to ROBIS item 1.5, namely, "Were any restrictions in eligibility criteria based on sources of info appropriate (e.g. publication status or format, language, availability of data)?"

Decision rule:

#1. Reviews that meet criteria A are highest in our hierarchy

Step I4 – Choose the analysis of data on individual patients

If (A) an individual patient data meta-analysis was identified in the overview, Jadad et al. recommend this review be chosen over reviews with pairwise meta-analysis.

Decision rule:

#1. Reviews that meet criteria A are highest in our hierarchy

Decision rules for Step I:

#1. If there is an IDP meta-analysis (Step I4), then this review is the highest in our hierarchy

#2. Reviews that meet Step I1, I2, and I3 are second highest in our hierarchy

#3. Reviews that meet Step I1 and I2 third highest in our hierarchy

#4. Reviews that meet Step I2 and I3 fourth highest in our hierarchy

#5. Reviews that meet Step I1 and I3 fifth highest in our hierarchy

#6. Reviews that meet Step I1 is sixth highest in our hierarchy

#7. Reviews that meet Step I2 is seventh highest in our hierarchy

#8. Reviews that meet Step I3 is eighth highest in our hierarchy

Note: Reporting only Steps I1, I2 or I3 is not considered a systematic approach to evidence synthesis.

3.8 Study outcomes

3.8.1 Evaluation of whether the overview authors:

- Used the Jadad decision tree to assess discordance
- Examine and record reasons for discordance (i.e. authors did not use Jadad)
- Use other approaches to deal with discordance (specify)
- Present discordance in tables and figures.

3.8.2 Results from our discordance assessment and overview authors assessments:

- Utility: Is the Jadad decision tree easy to use? (see section 2.9.3)
- Efficiency: How much time does it take to do one Jadad assessment?
- Comprehensiveness: is the Jadad algorithm comprehensive? Is it missing methods that might explain discordance (e.g. publication recency)?
- Jadad cohort: Frequency of discordant or concordant Jadad assessments between (a) overview authors assessment, and (b) our assessment (i.e. choosing the same SR).

- Non-Jadad cohort: Frequency of discordant or concordant discordance assessments between (a) overview authors assessment, and (b) our assessment (i.e. choosing the same SR).
- Comparison of overview authors stated sources of discordance and our identified sources of discordance.

3.8.3 "Ease of use" outcome measure

Each Jadad assessment will be assessed for "ease of use" by each assessor. Each Jadad assessment will be rated and colored (green, yellow, red) based on how easy or difficult the assessment was judged to be for the user. The rating is based on the following rubric:

- The step can be accomplished easily by the reviewer, due to low cognitive load or because it's a recognised method or approach.
- The step requires a notable degree of cognitive load by the reviewer but can generally be accomplished with some effort.
- The step is difficult for the reviewer, due to significant cognitive load or confusion; some reviewers would likely fail or abandon the task at this point.

The lower the score, the easier the step is to complete.

3.9 Data extraction

3.9.1 Overview level extraction

The outcomes from section 3.8 will be extracted, along with the following information from the overview:

1. Study characteristics (lead author's name, publication year).
3. Research question (objectives, health condition of treatment, PICO eligibility criteria: participant, intervention/comparison, and primary outcome).
4. Methods (how the overview authors assessed discordance among the reviews [Jadad or other approach], how they operationalised Jadad, steps where they identified discordance, number of included SRs, type of SR (e.g. individual patient data meta-analysis, SR with narrative summary, SR with meta-analysis of RCTs), type of analysis (narrative summary or meta-analysis), risk of bias/quality assessment (e.g. AMSTAR, AMSTAR 2, or ROBIS tool), risk of bias judgments, and whether risk of bias/quality assessment was integrated into the synthesis.

5. Results (results of discordance assessment, effect size and confidence intervals (CIs), number of total participants in treatment and control groups, number of total events in treatment and control groups), direction of study results (favourable or not favourable).
6. Conclusions (difference between results and conclusions defined as if they disagreed in direction (results, not favourable; conclusions, favourable), authors' result interpretation (quote from abstract and discussion section about the primary outcome result and conclusion).

3.9.2 Direction of effect, the magnitude of effect, and statistical significance

We defined direction of effect, the magnitude of effect, and statistical significance according to Jadad's definitions [16], namely:

- Direction of effect: One review favours the intervention, and another favours the comparison
- Magnitude of effect: To interpret the effect of an intervention in terms of whether it is sufficiently large to be termed clinically significant, we must describe the magnitude of the effect of the intervention. Effect size is a quantitative measure of the magnitude of the experimental effect. The larger the effect size, the stronger the relationship between the two variables. When two reviews differ in the magnitude of effect, one might have an effect estimated at a 30% reduction in mortality, where a second has a 5% reduction in mortality.
- Statistical significance indicates the probability associated with the null hypothesis, but it does not determine whether the result is important, meaningful, substantive, large enough to care about, or clinically significant. The p-value incorporates information about both the sample size and effect size, and thus interpretation, comparison of statistical significance across reviews with different sample sizes, and effect estimates is tricky. For example, if one study yielded a p-value of 0.05 and another yielded a p-value of 0.01, then in the absence of any additional information, a knowledge user might assume that the effect size was stronger in the latter case. In fact, though, if the study with a $p = 0.05$ used a sample of 10 per group and the second ($p = 0.01$) used 50 per group, then the effect size would have been substantially larger in the study with the modest p-value (a 40-point effect as compared with a 25-point effect).

3.9.3 SR level extraction

The outcomes from section 3.8 will be extracted, along with the following information from the included SRs:

1. Study characteristics (lead author's name, publication year).

2. Research question (objectives, health condition of treatment, PICO eligibility criteria: participant, intervention/comparison, and primary outcome; language restrictions and restrictions on publication status in eligibility criteria; citation of previous SRs/meta-analyses in background or discussion).
3. Search methodology (the name and number of databases searched, grey literature search details, the search period, language restrictions, restrictions on publication status, included full search in an appendix).
4. Methods (number and first author/year of included RCTs, effect metric (OR, RR, MD) and CIs, whether review authors assessed the clinical (PICO) heterogeneity across RCTs (in the methods or results sections), analysis method (appropriate weighted technique to combine study results [i.e. used a fixed or random random-effects model], investigation of statistical heterogeneity [i.e. by reporting I2, tau2 or chi2], and if heterogeneity is present, then the authors investigated the causes of any heterogeneity [i.e. by reporting subgroup, sensitivity, or meta-regression analyses]), risk of bias/quality assessment (e.g. Cochrane risk of bias tool v1 or v2), risk of bias/quality judgment for each RCT, and whether the RCT quality/risk of bias assessment was integrated into the synthesis; 2 reviewers independently screened studies, extracted data and assessed risk of bias with process for resolving discrepancies found when comparing.
5. Results (effect size and CI, number of total participants in treatment and control groups, number of total events in treatment and control groups), direction of study results (favourable or not favourable).
6. Conclusions (difference between results and conclusions defined as if they disagreed in direction (results, not favourable; conclusions, favourable), authors' result interpretation (quote from abstract and discussion section about the primary outcome result and conclusion).

Two authors will extract studies independently at full-text, and in the case of discrepant decisions, will discuss until consensus is reached.

3.10 Data analysis

We will assess and compare our outcomes (a) narratively for qualitative data, (b) using frequencies and percentages for categorical data, and (c) using median and interquartile range (IQR) for continuous data." Our analysis will be organised by our study outcomes in tables and in figures. We will discuss differences in the assessment of discordance across overviews using Jadad and overviews not using Jadad.

4.0 PATIENT AND PUBLIC INVOLVEMENT

Patients or the public were not involved in the design of our research protocol.

5.0 ETHICS, DISSEMINATION, STRENGTHS AND LIMITATIONS

5.1 Ethics

No ethics approval was required as no human subjects were involved.

5.2 Dissemination

We will disseminate evidence summaries through academic and social media platforms and websites (e.g. Twitter, Researchgate). We will publish in an open-access journal, and present at various formal and informal venues such as the Therapeutics Initiative Methods Speaker Series, and academic conferences such as Guidelines International Network conference, Cochrane Colloquium, and Public Health.

5.3 Strengths and limitations

We aim to use a systematic approach to evaluate authors' assessment of discordance across SRs in overviews and replicate Jadad algorithm assessments from a sample of overviews. We suspect that one reason for the inconsistent use of the Jadad algorithm in the existing literature may be due to the limited guidance available in the original Jadad manuscript on implementing the algorithm. We believe the greatest strength of the proposed study will be that we used an iterative process among authors to develop decision rules for the interpretation and application of each step in the Jadad algorithm. Upon completion and dissemination of this study, other overview authors will have more detailed guidance on how to apply the Jadad algorithm when addressing discordance in overviews. Furthermore, our study adopted a systematic and transparent approach to address the objectives outlined in our protocol using systematic review guidance [20]. A comprehensive search strategy, including a search of the grey literature, was employed with no restrictions to overviews' language and publication status to yield relevant studies and minimise publication bias. To minimise error, screening, extractions, and assessments will be completed by two independent reviewers, and subsequently compared. Any discrepancies will be resolved upon consensus, and if necessary, with the involvement of a third reviewer. To mitigate observer bias, reviewers were blinded to the overviews' Jadad assessments.

Despite several strengths, this study also has limitations that are to be noted. Due to the feasibility of operationalising this study, only the first 30 eligible overviews will be included. In

addition, by only focusing on overviews specific to health interventions, we may potentially have overlooked a number of relevant overviews that may have assessed discordance and/or employed the Jadad algorithm. When utilising the tools for assessment (i.e. risk of bias tools or Jadad decision tree), there will be subjectivity in the judgements potentially introducing variability in the results. To overcome this limitation, pilot screening and pilot assessments will be completed by reviewers and assessed to ensure consistency in understanding of the screening criteria, definitions of extracted items, and steps in the Jadad assessment.

Some steps in the Jadad algorithm were vague in description (e.g. step I2), making it difficult to interpret. For example, step I2 assesses the methods used to assess the quality of included RCTs across reviews. Some authors could interpret this as assessing if review authors used the Cochrane risk of bias to assess the quality of RCTs across review. However, the Cochrane risk of bias tool was only published in 2008. To minimise the incorrect interpretation of this step, we solicited feedback at the protocol stage from all authors in its interpretation prior to piloting. Thus, after discussion with authors, we applied a decision rule that for reviews dates 2009 and earlier, we considered the Jadad scale and allocation concealment to be the most common scale between 1995 and 2009, in addition to other tools being considered on a case-by-case basis. This was done for all steps. Additionally, during piloting we will amend our interpretation and instructions on how to operationalise the Jadad algorithm to ensure consistent application.

Figure legend

Figure 1. Jadad (1997) decision tree. Legend: Step A is to examine the multiple reviews matching the overviews' question using a PICO framework. If the research questions are not identical, then step B indicates choosing the review closest to the decision makers' research question and no further assessment is necessary. If multiple reviews are found with the same PICO, then step C should be investigated.

Figure 2. Our Approach to Operationalizing Step E of the Jadad Algorithm

Figure 3. Our Approach to Operationalising Step G of the Jadad Algorithm

Figure 4. Our Approach to Operationalising Step H of the Jadad Algorithm

Figure 5. Our Approach to Operationalising Step I of the Jadad Algorithm

Declarations

No declarations.

Consent for publication

Not applicable

Availability of data and materials

Not applicable

Competing interests

The authors declare that they have no competing interests.

Funding

No grants or funding were successful for this important study. Dr. Andrea Tricco currently holds a Tier 2 Canada Research Chair in Knowledge Synthesis and an Ontario Ministry of Research and Innovation Early Researcher Award. Dr. Carole Lunny is funded by a CIHR project grant for her postdoctoral research.

Authors' contributions

CL conceived of the study; all authors contributed to the design of the study; All authors edited the manuscript; and all authors read and approved the final manuscript.

References

1. Bastian H, Glasziou P, Chalmers IJPm: **Seventy-five trials and eleven systematic reviews a day: how will we ever keep up?** 2010, **7**(9):e1000326.
2. Taito S, Kataoka Y, Ariie T, Oide S, Tsujimoto Y: **Assessment of the Publication Trends of COVID-19 Systematic Reviews and Randomized Controlled Trials**. *Annals of Clinical Epidemiology* 2021, **3**(2):56-58.
3. Niforatos JD, Weaver M, Johansen ME: **Assessment of Publication Trends of Systematic Reviews and Randomised Clinical Trials, 1995 to 2017**. *JAMA Internal Medicine* 2019, **179**(11):1593-1594.
4. Becker LA, Oxman, A.D.: **Chapter 22: Overviews of reviews**. In: *Cochrane Handbook for Systematic Reviews of Interventions.* edn. Edited by Higgins JPT GSE. Hoboken, NJ: John Wiley & Sons; 2008: 607-631.
5. Pieper D, Antoine SL, Mathes T, Neugebauer EA, Eikermann M: **Systematic review finds overlapping reviews were not mentioned in every other overview**. *J Clin Epidemiol* 2014, **67**(4):368-375.
6. Pieper D, Mathes T, Eikermann M: **Impact of choice of quality appraisal tool for systematic reviews in overviews**. *J Evid Based Med* 2014, **7**(2):72-78.
7. Smith V, Devane D, Begley CM, Clarke M: **Methodology in conducting a systematic review of systematic reviews of healthcare interventions**. *BMC Med Res Methodol* 2011, **11**(1):15.
8. Caird J, Sutcliffe K, Kwan I, Dickson K, Thomas J: **Mediating policy-relevant evidence at speed: are systematic reviews of systematic reviews a useful approach?** *Evidence & Policy: A Journal of Research, Debate and Practice* 2015, **11**(1):81-97.
9. Hunt H, Pollock A, Campbell P, Estcourt L, Brunton GJSr: **An introduction to overviews of reviews: planning a relevant research question and objective for an overview**. 2018, **7**(1):39.
10. Ioannidis JPA: **The Mass Production of Redundant, Misleading, and Conflicted Systematic Reviews and Meta-analyses**. *The Milbank quarterly* 2016, **94**(3):485-514.
11. Lunny C, Brennan SE, McDonald S, McKenzie JE: **Toward a comprehensive evidence map of overview of systematic review methods: paper 1-purpose, eligibility, search and data extraction**. *Syst Rev* 2017, **6**(1):231.
12. Lunny C, Brennan SE, McDonald S, McKenzie JE: **Toward a comprehensive evidence map of overview of systematic review methods: paper 2-risk of bias assessment; synthesis, presentation and summary of the findings; and assessment of the certainty of the evidence**. *Syst Rev* 2018, **7**(1):159.
13. Bolland MJ, Grey A: **A case study of discordant overlapping meta-analyses: vitamin d supplements and fracture**. *PloS one* 2014, **9**(12):e115934.
14. Campbell J, Bellamy N, Gee T: **Differences between systematic reviews/meta-analyses of hyaluronic acid/hyaluronan/hylan in osteoarthritis of the knee**. *Osteoarthritis and cartilage / OARS, Osteoarthritis Research Society* 2007, **15**(12):1424-1436.
15. Chalmers PN, Mascarenhas R, Leroux T, Sayegh ET, Verma NN, Cole BJ, Romeo AA: **Do arthroscopic and open stabilisation techniques restore equivalent stability to the shoulder in the setting of anterior glenohumeral instability? a systematic review of overlapping meta-analyses**. *Arthroscopy : the journal of arthroscopic & related surgery : official publication of the Arthroscopy Association of North America and the International Arthroscopy Association* 2015, **31**(2):355-363.
16. Jadad AR, Cook DJ, Browman GP: **A guide to interpreting discordant systematic reviews**. *CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne* 1997, **156**(10):1411-1416.

17. Li Q, Wang C, Huo Y, Jia Z, Wang X: **Minimally invasive versus open surgery for acute Achilles tendon rupture: a systematic review of overlapping meta-analyses**. *Journal of orthopaedic surgery and research* 2016, **11**(1):65.
18. Mascarenhas R, Chalmers PN, Sayegh ET, Bhandari M, Verma NN, Cole BJ, Romeo AA: **Is double-row rotator cuff repair clinically superior to single-row rotator cuff repair: a systematic review of overlapping meta-analyses**. *Arthroscopy : the journal of arthroscopic & related surgery : official publication of the Arthroscopy Association of North America and the International Arthroscopy Association* 2014, **30**(9):1156-1165.
19. Zhao JG, Wang J, Long L: **Surgical Versus Conservative Treatments for Displaced Midshaft Clavicular Fractures: A Systematic Review of Overlapping Meta-Analyses**. *Medicine* 2015, **94**(26):e1057.
20. Higgins JP: **Cochrane handbook for systematic reviews of interventions version 5.0. 1. The Cochrane Collaboration**. *http://www cochrane-handbook org* 2008.
21. Lunny C, Neelakant T, Chen A, Shinger G, Stevens A, Tasnim S, Sadeghipouya S, Adams S, Zheng YW, Lin L: **Bibliometric study of 'overviews of systematic reviews': evaluation of prevalence, citation and impact factor**. 2021.
22. Fabiano GA, Schatz NK, Aloe AM, Chacko A, Chronis-Tuscano A: **A systematic review of meta-analyses of psychosocial treatment for attention-deficit/hyperactivity disorder**. *Clin Child Fam Psychol Rev* 2015, **18**(1):77-97.
23. Finnane A, Janda M, Hayes SC: **Review of the evidence of lymphedema treatment effect**. *American journal of physical medicine & rehabilitation* 2015, **94**(6):483-498.
24. Lau R, Stevenson F, Ong BN, Dziedzic K, Treweek S, Eldridge S, Everitt H, Kennedy A, Qureshi N, Rogers A: **Achieving change in primary care—effectiveness of strategies for improving implementation of complex interventions: systematic review of reviews**. *BMJ open* 2015, **5**(12):e009993.
25. López NJ, Uribe S, Martinez B: **Effect of periodontal treatment on preterm birth rate: a systematic review of meta-analyses**. *Periodontology 2000* 2015, **67**(1):87-130.
26. Moher D, Dulberg CS, Wells GA: **Statistical power, sample size, and their reporting in randomised controlled trials**. *Jama* 1994, **272**(2):122-124.
27. Tricco AC, Tetzlaff J, Pham B, Brehaut J, Moher D: **Non-Cochrane vs. Cochrane reviews were twice as likely to have positive conclusion statements: cross-sectional study**. *J Clin Epidemiol* 2009, **62**(4):380-386.e381.
28. Hróbjartsson A, Thomsen ASS, Emanuelsson F, Tendal B, Hilden J, Boutron I, Ravaud P, Brorson S: **Observer bias in randomised clinical trials with measurement scale outcomes: a systematic review of trials with both blinded and nonblinded assessors**. *CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne* 2013, **185**(4):E201-E211.
29. Morissette K, Tricco AC, Horsley T, Chen MH, Moher D: **Blinded versus unblinded assessments of risk of bias in studies included in a systematic review**. *Cochrane Database of Systematic Reviews* 2011(9).
30. Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, Porter AC, Tugwell P, Moher D, Bouter LM: **Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews**. *BMC Med Res Methodol* 2007, **7**(1):1-7.
31. Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J, Moher D, Tugwell P, Welch V, Kristjansson E: **AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both**. *bmj* 2017, **358**.
32. Whiting P SJ, Churchill R.: **Introduction to ROBIS, a new tool to assess the risk of bias in a systematic review**. In: *23rd Cochrane Colloquium: 2015; Vienna, Austria*: John Wiley & Sons; 2015.
33. Viswanathan M, Patnode CD, Berkman ND, Bass EB, Chang S, Hartling L, Murad MH, Treadwell JR, Kane RL: **Assessing the risk of bias in systematic reviews of health care**

interventions. *Methods guide for effectiveness and comparative effectiveness reviews [Internet]* 2017.

34. Jadad AR, Moher D, Klassen TP: **Guides for reading and interpreting systematic reviews: II. How did the authors find the studies and assess their quality?** *Archives of pediatrics & adolescent medicine* 1998, **152**(8):812-817.

35. Schulz KF, Chalmers I, Hayes RJ, Altman DG: **Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials**. *Jama* 1995, **273**(5):408-412.

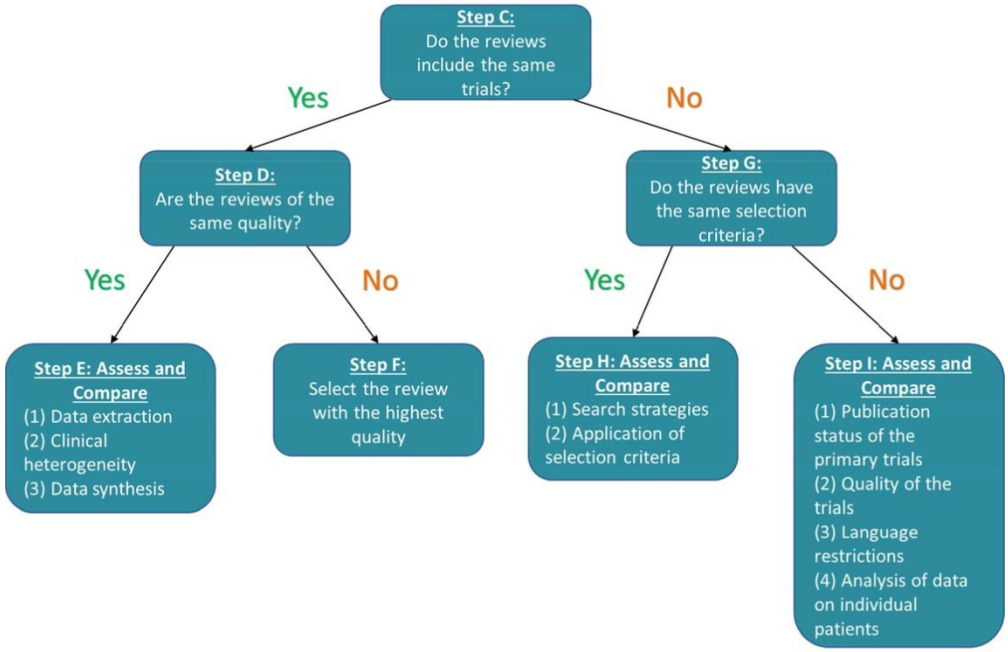

Figure 1. Jadad (1997) decision tree.

Legend: Step A is to examine the multiple reviews matching the overviews' question using a PICO framework. If the research questions are not identical, then step B indicates choosing the review closest to the decision makers' research question and no further assessment is necessary. If multiple reviews are found with the same PICO, then step C should be investigated.

194x132mm (300 x 300 DPI)

Figure 2. Our Approach to Operationalizing Step E of the Jadad Algorithm.

164x208mm (300 x 300 DPI)

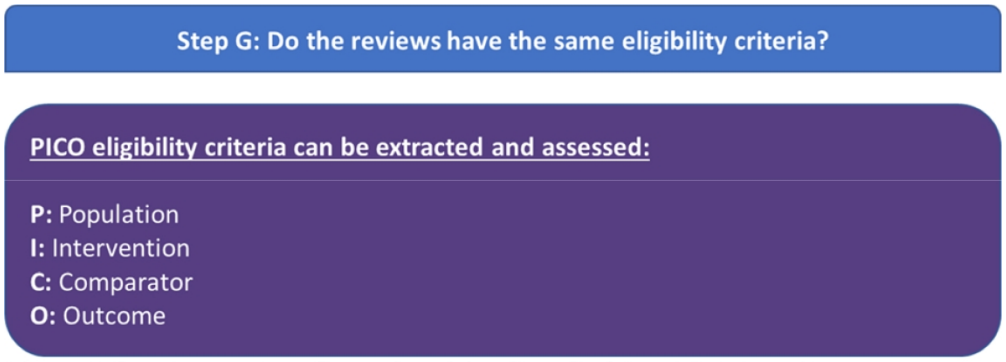

Figure 3. Our Approach to Operationalising Step G of the Jadad Algorithm

160x61mm (300 x 300 DPI)

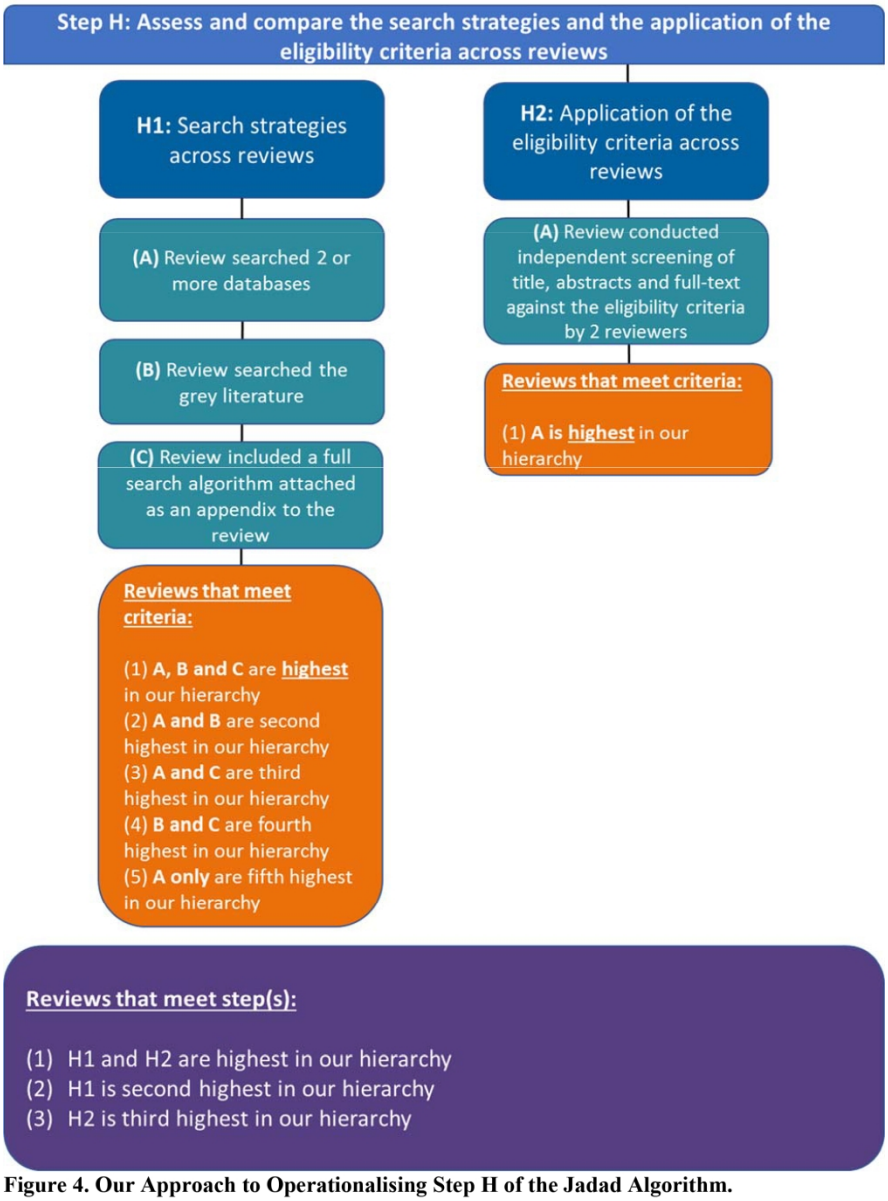

Figure 4. Our Approach to Operationalising Step H of the Jadad Algorithm.

166x226mm (300 x 300 DPI)

Figure 5. Our Approach to Operationalising Step I of the Jadad Algorithm.

175x220mm (300 x 300 DPI)

BMJ Open

Identifying and addressing conflicting results across multiple discordant systematic reviews on the same question: Protocol for a replication study of the Jadad algorithm

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2021-054223.R1
Article Type:	Protocol
Date Submitted by the Author:	30-Nov-2021

Complete List of Authors:	Lunny, Carole; University of Toronto Dalla Lana School of Public Health, Epidemiology Division and Institute for Health Policy, Management, and Evaluation Thirugnanasampanthar, Sai Surabi; University of Toronto Dalla Lana School of Public Health, Epidemiology Division; St Michael's Hospital Li Ka Shing Knowledge Institute, Knowledge Translation Program Kanji, Salmaan; The Ottawa Hospital, Department of Pharmacy; University of Ottawa, Faculty of Medicine Ferri, Nicola; University of Bologna, Department of Biomedical and Neuromotor Sciences Thabet, Pierre; Hôpital Montfort Pieper, Dawid; Witten/Herdecke University, Cologne, Germany Tasnim, Sara; The University of British Columbia Faculty of Medicine, Anesthesiology, Pharmacology & Therapeutics Nelson, Harrison; Queen's University Faculty of Health Sciences Reid, Emma; Nova Scotia Health Authority Zhang, Jia He (Janet); The University of British Columbia Kalkat, Banveer; The University of British Columbia, Faculty of Pharmaceutical Sciences Chi, Yuan; Beijing Yealth Technology Co., Ltd, Yealth Network Thompson, Jacqueline; University of Birmingham Institute of Applied Health Research Abdoulrezzak, Reema; The University of British Columbia, Faculty of Pharmaceutical Sciences Zheng, Di Wen (Wendy); The University of British Columbia, Faculty of Pharmaceutical Sciences Pangka, Lindy; The University of British Columbia, Faculty of Pharmaceutical Sciences Wang, Dian (Xin Ran); The University of British Columbia, Faculty of Pharmaceutical Sciences Safavi , Parisa; The University of British Columbia, Faculty of Pharmaceutical Sciences Sooch, Anmol; The University of British Columbia, Faculty of Pharmaceutical Sciences Kang, Kevin; The University of British Columbia, Faculty of Pharmaceutical Sciences Whitelaw, Sera; McGill University, Faculty of Medicine and Health Sciences

	Tricco, Andrea; Li Ka Shing Knowledge Institute of St Michael's Hospital, Knowledge Translation Program; University of Toronto Dalla Lana School of Public Health, Epidemiology Division and Institute for Health Policy, Management, and Evaluation
Primary Subject Heading:	Research methods
Secondary Subject Heading:	Research methods
Keywords:	PUBLIC HEALTH, STATISTICS & RESEARCH METHODS, EPIDEMIOLOGY

Title: Identifying and addressing conflicting results across multiple discordant systematic reviews on the same question: Protocol for a replication study of the Jadad algorithm

Authors:

- Lunny C, PhD[1]*, carole.lunny@ubc.ca, https://orcid.org/0000-0002-7825-6765, St Michael's Hospital, Unity Health Toronto and the Cochrane Hypertension Review Group, University of British Columbia, Vancouver, BC, Canada, V6T 1Z3
- Sai Surabi Thirugnanasampanthar, BHSc Saisurabi.Thirugnanasampanthar@unityhealth.to, Knowledge Translation Program, Li Ka Shing Knowledge Institute, St. Michael's Hospital, Unity Health and the Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada
- Kanji S, Pharm.D., skanji@toh.ca, The Ottawa Hospital and Ottawa Hospital Research Institute
- Ferri N, PT Nicola.ferri11@unibo.it, Department of Biomedical and Neuromotor Sciences (DIBINEM), Alma Mater Studiorum, University of Bologna, 40138, Bologna, Italy
- Thabet P, B. Pharm, pierrethabet@montfort.on.ca, Hôpital Montfort, Ottawa, Canada
- Pieper D, PhD, Dawid.Pieper@ mhb-fontane.de , Institute for Research in Operative Medicine, Witten/Herdecke University, Germany
- Tasnim S, MSc, stasnim@mail.ubc.ca, Cochrane Hypertension Review Group, Therapeutics Initiative, Department of Anesthesiology, Pharmacology & Therapeutics, Faculty of Medicine, University of British Columbia, 2176 Health Science Mall, Vancouver, BC, Canada, V6T 1Z3
- Harrison Nelson, harrison.nelson@queensu.ca, https://orcid.org/0000-0001-8306-5448, Faculty of Health Sciences, Queen's University, Kingston, ON, Canada
- Emma Reid, Pharm.D, emmak.reid@nshealth.ca, Nova Scotia Health, Halifax, Nova Scotia, Canada
- Jia He (Janet) Zhang, BSc, jiahe.janet@gmail.com, University of British Columbia, Vancouver, BC, Canada
- Banveer Kalkat, banveerk@student.ubc.ca, Faculty of Pharmaceutical Sciences, University of British Columbia, Vancouver, BC, Canada
- Yuan Chi. yuan.chi@outlook.com, Beijing Yealth Technology Co., Ltd, Yealth Network, Beijing, China
- Jacqueline Y Thompson, MPH j.y.thompson@bham.ac.uk, https://orcid.org/0000-0002-9775-361X, Institute of Applied Health Research, University of Birmingham, Edgbaston, Birmingham B15 2TT

- Reema Abdoulrezzak, BSc reemaabd@student.ubc.ca, Faculty of Pharmaceutical Sciences, University of British Columbia, Vancouver, BC, Canada
- Di Wen (Wendy) Zheng, BSc wendyzhe@student.ubc.ca, Faculty of Pharmaceutical Sciences, University of British Columbia, Vancouver, BC, Canada
- Lindy R.S. Pangka, BSc lpangka@shaw.ca, Faculty of Pharmaceutical Sciences, University of British Columbia, Vancouver, BC, Canada
- Dian (Xin Ran) Wang, BSc dianw@student.ubc.ca, Faculty of Pharmaceutical Sciences, University of British Columbia, Vancouver, BC, Canada
- Parisa Safavi, BSc psafavi@student.ubc.ca, Faculty of Pharmaceutical Sciences, University of British Columbia, Vancouver, BC, Canada
- Anmol Sooch, BSc asooch19@student.ubc.ca, Faculty of Pharmaceutical Sciences, University of British Columbia, Vancouver, BC, Canada
- Kevin T. Kang, BSc ktw9412@student.ubc.ca, Faculty of Pharmaceutical Sciences, University of British Columbia, Vancouver, BC, Canada
- Whitelaw S, MSc sera.whitelaw@mail.mcgill.ca, Faculty of Medicine and Health Sciences, McGill University, Montreal, QC, Canada
- Tricco AC, PhD, 1.Knowledge Translation Program, Li Ka Shing Knowledge Institute, St. Michael's Hospital, Unity Health Toronto, 209 Victoria St, M5B 1T8, Toronto, Ontario, Canada.; 2. Epidemiology Division and Institute for Health Policy, Management, and Evaluation, Dalla Lana School of Public Health, University of Toronto, 155 College St Room 500, M5T 3M7, Toronto Ontario, Canada; 3.Queen's Collaboration for Health Care Quality Joanna Briggs Institute Centre of Excellence, School of Nursing, Queen's University, 99 University Ave, K7L 3N6, Kingston, Ontario, Canada

*Corresponding author

Lunny C, PhD[1]*, carole.lunny@ubc.ca, https://orcid.org/0000-0002-7825-6765, St Michael's Hospital, Unity Health Toronto and the Cochrane Hypertension Review Group, University of British Columbia, Vancouver, BC, Canada, V6T 1Z3

ABSTRACT

Introduction: An increasing growth of systematic reviews (SRs) presents notable challenges for decision-makers seeking to answer clinical questions. In 1997, an algorithm was created by Jadad to assess discordance in results across SRs on the same question. Our study aims to (a) replicate assessments done in a sample of studies using the Jadad algorithm to determine if the same SR would have been chosen, (b) evaluate the Jadad algorithm in terms of utility, efficiency, and comprehensiveness, and (c) describe how authors address discordance in results across multiple SRs.

Methods and Analysis: We will use a database of 1218 overviews (2000-2020) created from a bibliometric study as the basis of our search for studies assessing discordance (called Discordant Reviews). This bibliometric study searched MEDLINE (Ovid), Epistemonikos, and Cochrane Database of Systematic Reviews for overviews. We will include any study using Jadad (1997) or another method to assess discordance. The first 30 studies screened at the full-text stage by two independent reviewers will be included. We will replicate the authors' Jadad assessments. We will compare our outcomes qualitatively and evaluate the differences between our Jadad assessment of discordance and the authors' assessment.

Ethics and Dissemination: No ethics approval was required as no human subjects were involved. In addition to publishing in an open-access journal, we will disseminate evidence summaries through formal and informal conferences, academic websites, and across social media platforms. This is the first study to comprehensively evaluate and replicate Jadad algorithm assessments of discordance across multiple SRs.

Abstract: 281 words

Strengths and limitations of this study

- This is the first proposed empirical study to use a systematic approach to evaluate authors' assessment of discordance across SRs and replicate Jadad algorithm assessments
- We believe the greatest strength of the proposed study will be that we used an iterative process among authors to develop decision rules for the interpretation and application of each step in the Jadad algorithm.
- To mitigate observer bias, reviewers were blinded to the Discordant Review authors' Jadad assessments.
- In our search update, we only searched MEDLINE (Ovid) which would have limited the number of potentially relevant studies found.

1.0 BACKGROUND

Information overload is an increasing problem for health practitioners, researchers, and decision-makers. Global research output is growing rapidly, and the number of published systematic reviews (SRs) being produced yearly is also expanding [1]. Between January and October 2020, 807 SRs on COVID-19 alone were published in PubMed [2], and the rate of growth from 1995 to 2017 in SRs was found to be 4676% [3]. Challenges in dealing with growth in SRs include identification of high quality, comprehensive, and recent reports on the topic of interest.

'Overviews of systematic reviews' (henceforth called overviews) evolved in response to these challenges [4-7]. Overviews summarise the results of SRs, and help make sense of potentially conflicting, discrepant, and overlapping results and conclusions of SRs on the same question [8-15].. Overviews may also include systematic reviews with concordant results; hence we have named studies that identify and explain the discordance between conflicting SRs on the same question as "Discordant Reviews". These Discordant Reviews are often called systematic reviews of overlapping meta-analyses (e.g. [13, 16]), conflicting results of meta-analyses [17, 18], and discordant meta-analyses [19-21].

We define discordance as when SRs with identical or /very similar clinical, public heath or policy questions report different results for the same outcome. A common method for dealing with multiple SRs with discordant results is to specify methodological criteria to select only one SR (e.g., select the highest quality and most comprehensive SR) [12]. However, many other methods have been proposed, including statistical approaches to address discordance in results across SRs [11, 12].

As early as 1997, Jadad et al. [22] developed a decision tree (i.e. an algorithm) to assess discordance in results across a sample of SRs on the same question to aid healthcare providers in making clinical decisions. The Jadad algorithm guides users through a methodological assessment of SR components to identify potential causes of discordance and ultimately choose the best SR across multiple. Jadad et al.'s . [22] appears to be the only formal algorithm available to both explain and deal with discordance across SRs. Despite the availability of this tool since 1997, it has not been universally adopted and has been inconsistently applied when used [23-25]. The aim of this study is to comprehensively replicate and evaluate the Jadad algorithm for assessing discordance across SRs.

2.0 OBJECTIVES

Our study objectives are:

1) Describe how Discordant Reviews address discordance in results across multiple SRs using content analysis (study 1)

2) Replicate Jadad assessments from published Discordant Reviews to identify sources of discordance and to determine if the same SR(s) was chosen as the "best available evidence" (study 2)
3) Evaluate the Jadad algorithm in terms of utility, efficiency, and comprehensiveness (study 2)

3.0 METHODS

3.1. Study design

This is a methods study in the knowledge synthesis field. We followed SR guidance for the study selection and data extraction stages [26]. Our protocol will be registered and located on the Open Science Framework at https://osf.io/nbcta/.

3.2 Search and selection of Discordant Reviews

3.2.1 Database of 1218 overviews

We will use a database of 1218 overviews published between January 1, 2000 and December 30, 2020 created from a bibliometric study [27] as the basis of our search for Discordant Reviews using Jadad or another method to assess discordance. A validated search filter for overviews [28] was used in MEDLINE (Ovid), Epistemonikos, and the Cochrane Database of Systematic Reviews (CDSR) for overviews. In an empirical methods study of the retrieval sensitivity of 6 databases, the combination of MEDLINE and Epistemonikos retrieved 95.2% of all systematic reviews [29]. As a rationale, we believe this combination would retrieve an equal proportion of overviews. The database Epistemonikos contains both published and unpublished reports. We searched the CDSR through the website interface using the filter for Cochrane reviews. Overviews included in the bibliometric study contain these characteristics: (a) synthesised the results of SRs, (b) systematically searched for evidence in a minimum of two databases, and (c) conducted a search using a combination of text words and MeSH terms. The included overviews also had to have a methods section in the main body of the paper and focused on health interventions. To identify studies assessing discordance using the database of 1218 overviews, we will use the EndNote search function and Boolean logic to search for the following words: overlap*[title/abstract] or discrepan*[title/abstract] or discord*[title/abstract] or concord*[title/abstract] or conflict*[title/abstract] or Jadad [abstract].

3.2.2 Medline (Ovid) search strategy January to April 2021

We will update this search with an Ovid MEDLINE search using the following search string: ("systematic reviews".ti,ab. or "meta-analyses".ti,ab.) AND (overlap.ti,ab or discrepant.ti,ab or discordant.ti,ab. or difference.ti,ab. or conflicting.ti,ab. or Jadad.ab.). Our search was conducted on April 18, 2021.

3.3 Screening Discordant Reviews

3.3.1 Process for screening Discordant Reviews

Citations identified by our searches will be assigned a random number and screened sequentially. The first 30 Discordant Reviews screened at full-text and meeting our eligibility criteria will be included.

All authors will pilot the screening form on 20 Discordant Reviews to ensure high levels of agreement and common definitions of eligibility criteria.

Two authors will independently screen Discordant Reviews as full-text publications. Discrepancies will be resolved by consensus, and arbitration by a third reviewer when necessary.

3.3.2 Stage 1 screening criteria

We first include all reports aiming to assess discordant results across SRs on the same question. Studies assessing discordance can assess (a) discordant results, or (b) discordant interpretations of the results and conclusions. Both studies examining (a) and (b) were eligible (e.g. [30-33]).

If a study meets stage 1 criteria, it will be included in Study 1.

3.3.3 Stage 2 screening criteria

From this sample, we will then screen Discordant Reviews based on the following inclusion criteria:

- Included a minimum of two SRs with a meta-analysis of randomised controlled trials (RCTs), but may have included other study types beyond RCTs; and
- Used the Jadad algorithm.

If a study meets stage 2 screening, it will move onto stage 3 screening.

3.3.4 Stage 3 screening

After stage 2 screening is complete, we will screen based on the authorship team. When the same 'core' authors (first, last, and/or corresponding) conduct two or more of the identified Discordant Reviews, we will only include one of the multiple Discordant Reviews (i.e. the most recent study will be selected). Our rationale is that author groups use the same methods to assess discordance (e.g. Mascarenas [24] and Chalmers [15]).

We will include Discordant Reviews in any language and publication status, published anytime. We will use Google translate to interpret non-English studies for screening and assessment. The resulting publications will form the set for study 2.

In the case where the authors of this study are also authors of one or several of the included Discordant Reviews, those authors will not conduct screening, extraction, Jadad assessment, or analysis of the Discordant Review in question.

Full-texts of all SRs included in an Discordant Review will be obtained.

3.4 Extraction of the primary intervention and outcome

3.4.1 Identification of the primary outcome from the Discordant Review

As a first step in assessing discordance, we will identify the primary outcome from each Discordant Review. The primary outcome will be extracted when it is explicitly defined in the title, abstract, objectives, introduction or methods sections [34, 35]. If the primary outcome is not found in any of these sections, we will extract it as the first outcome mentioned in the manuscript [34, 35]. If the primary outcome cannot be identified by any of these approaches, we will consider that the article did not specify primary outcomes and the study will be excluded, and replaced with the next Discordant Review in our database.

3.4.2 Identification of the primary intervention from the Discordant Review

Identification of the primary outcome and intervention is a two-step process. As a first step, we will identify the primary intervention associated with the primary outcome from each Discordant Review. Then we will extract the primary outcome and intervention from the included systematic reviews when doing the Jadad assessments. If this is unclear, we will choose the first intervention highlighted in the title or abstract of the Discordant Review [35].

3.4.3 Identification of included SRs with meta-analysis of RCTs addressing the primary intervention/outcome of the Discordant Review

Once the primary outcome is identified, we will next identify how many SRs with meta-analysis of RCTs were included in the Discordant Review that address the primary outcome and intervention. It is this sample of SRs with meta-analysis of RCTs that will be the focus of our Jadad assessments.

3.4.4 Process to identify primary intervention and outcome

Two authors will extract the primary intervention and outcome, and disagreements will be discussed until consensus is reached.

3.5 Blinding of results in the included SRs

Observer bias, sometimes called "detection bias" or "ascertainment bias," occurs when outcome assessments are systematically influenced by the assessors' conscious or unconscious predispositions [36]. Blinded outcome assessors are used in trials to avoid such bias. One empirical study found evidence of high risk of substantial bias when authors failed to blind outcome assessors in trials [36], whereas another did not [37]. In our study, it is important that reviewers are blinded to the Discordant Review authors' result of the Jadad assessment, as unblinding might predispose them to unconsciously choose the same SR as the Discordant Review authors.

We will blind the following components containing study results of the Jadad assessment and conclusions: abstract, highlights, results of the Jadad assessment, and discussion/conclusions section. Blinding will be achieved via deletion using the paywalled Adobe Acrobat Pro or the freeware PDFCandy (https://pdfcandy.com). One author will blind the Discordant Review results and will not be involved in the Jadad assessment pertaining to those results. Assessors will be instructed not to search for and read the included Discordant Reviews prior to, and during, assessment.

3.6 Piloting Jadad assessment prior to full assessment

A pilot practice exercise will be conducted by all assessors prior to the Jadad assessments, to ensure consistent assessments across reviewers. Two Jadad assessments will be piloted by each reviewer and compared to a second to identify discrepancies that are to be resolved through discussion. Any necessary revisions to the assessment (sections 3.7 and 3.8) will be noted.

3.7 Jadad assessments of discordance across SRs

While the Jadad paper provides an algorithm intended to identify and address discordance between SRs in an Discordant Review, there is limited guidance within the manuscript regarding the application/operationalization of the algorithm. Absence of this detailed guidance leaves room for subjective (mis)interpretation and ultimately confusion when it comes time to use the algorithm. To address this, we engaged in an iterative process of interpretation and implementation of the algorithm step by step. This process involved virtual meetings whereby consensus was sought for decision rules at each step of the algorithm to ensure consistency in both interpretation and application. Feedback was

solicited and decision rules were accordingly adjusted until consensus was achieved. This tool underwent pilot testing as described in 3.6 where feedback was further solicited and adjustments were made.

The Jadad decision tree assesses and compares sources of inconsistency between SRs with meta-analyses, including differences in clinical, public health, or policy questions, inclusion and exclusion criteria, extracted data, methodological quality assessments, data combining, and statistical analysis methods.

Step A is to examine the multiple SRs matching the Discordant Review question using a PICO framework. If the clinical, public health, or policy questions are not identical, then step B indicates choosing the SR closest to the decision makers' research question and no further assessment is necessary. If multiple SRs are found with the same PICO, then step C should be investigated. As we are using Discordant Reviews as our sample, we will start at Step C in the Jadad decision tree.

Figure 1. Jadad (1997) decision tree.

Here we detail our interpretation of the Jadad algorithm for each step in assessing the discordance in a group of SRs with similar PICO elements. Two researchers will independently assess each set of SRs in the included Discordant Review using the Jadad algorithm, starting with Step C (Figure 1). Information and data from the Discordant Review will be used first if reported, and when data is not reported, we will consult the full text of the included SRs. If a Discordant Review or the included SRs does not report a method, we will indicate it as "not reported," and it will not be chosen for that step.

Step D and G follow from Step C. Steps E, F, H. and I are completed depending on the decisions at Steps D and G, respectively.

"Meeting" a step means a SR met the criteria in the sub-step or step that is highest in the hierarchy. For example, a SR that meets E3 criteria fulfills criteria A and B, which is the highest in our hierarchy.

3.7.2 Step C - Do the reviews include the same trials for the primary intervention and outcome?

We will determine if the RCTs are similar across SRs by either finding this information in the Discordant Review, or extracting all RCTs from the included SRs using an excel matrix to list the SRs

at the top, and trials in the left rows. The RCTs will be mapped to the SRs in order of publication date (earliest trials at the top). Using this matrix, we will determine if the SRs include the same or different trials.

3.7.3 Step D - Are the reviews of the same quality?

If the SRs contain the same trials, then the assessor moves to Step D – assess whether the SRs are the same methodological quality. We will either (a) extract the risk of bias/quality assessments from the Discordant Review if the Discordant Review used AMSTAR [38], AMSTAR 2 [39], or ROBIS [40]. AMSTAR [38] and the updated AMSTAR 2 [39] are tools to assess methodological quality (i.e. quality of conduct and reporting) and ROBIS [40] is a tool used to assess the risk of bias at the systematic-review level. Review-level biases include selective outcome reporting (e.g. only describing the statistically significant, and not describing all outcomes) and publication bias (e.g. published studies are more likely to report positive results). If the Discordant Review authors used any other tool or method to assess the risk of bias/quality of the SRs or did not assess the risk of bias/quality assessment at all, we will conduct our own assessment using the ROBIS tool [40].

3.7.4 Step E – Assess and compare data extraction, clinical heterogeneity, and data synthesis

If the SRs are the same risk of bias/quality, then the next step is Step E, to assess and compare data extraction, clinical heterogeneity, and data synthesis across the SRs.

Step E1 - Assess and compare the data extraction methods across reviews

For this step, Jadad states, "If reviews differ [in outcomes reported], the decision-maker should identify the review that takes into account the outcome measures most relevant to the problem that he or she is solving." We interpret this step as selecting the SR that (A) matches the Discordant Review's primary outcome.

Jadad then writes that SRs that conduct independent extractions by two reviewers are of the highest quality. We therefore decided that SRs that (B) used an independent data extraction process using two SR authors should be chosen. If a ROBIS assessment is done, then the latter point will be mapped to ROBIS 3.1. "Were efforts made to minimise error in data collection?"

Decision rules:

#1. Reviews that meet criteria A and B are highest in our hierarchy

#2. Reviews that meet criteria A are second highest in our hierarchy

#3. Reviews that meet criteria B are third highest in our hierarchy

Step E2 – Assess and compare clinical heterogeneity of the included RCTs across reviews

Clinical heterogeneity is assessed at the SR level by examining the clinical, public health, or policy question pertaining to the primary outcome and the eligibility criteria PICO elements of each included RCT to see if they are sufficiently similar. If the PICO across RCTs are similar, then clinical heterogeneity is minimal, and SRs can progress with pooling study results in a meta-analysis. If a ROBIS assessment is done, this question is mapped to ROBIS 4.3 "Was the synthesis appropriate given the nature and similarity in the clinical, public health, or policy questions, study designs, and outcomes across included SRs?"

If a SR states that (A) they assessed for clinical (e.g., PICO) heterogeneity across RCTs (in the methods or results sections), then this will be the SR that is chosen at this step. Example of a SR reporting a clinical heterogeneity assessment: "If we found 3 or more systematic SRs with similar study populations, treatment interventions, and outcome assessments, we conducted quantitative analyses (Gaynes 2014)". If authors reported and described clinical heterogeneity in the manuscript, then rule (B) authors that judged the clinical heterogeneity assessment to be minimal or low with rationale, will be chosen at this step.

Decision rule:

#1. Reviews that meet criteria A and B are highest in the hierarchy

#2. Reviews that meet criteria A are second highest in our hierarchy

Step E3 – Assess and compare data analysis methods across reviews

Jadad et al. are purposefully vague when describing how to judge whether a meta-analysis was appropriately conducted. T **his step was interpreted as reviews that conducted: (A) an appropriate weighted technique to combine study results (i.e. fixed or random effects model) and (B) an investigation of statistical heterogeneity** (i.e. by reporting I^2, tau^2, or chi^2) (**Figure 2**).

Decision rules for if the presence or absence of heterogeneity is present in the meta-analysis:

#1. Reviews that meet criteria A and B are highest in our hierarchy

#2. Reviews that meet criteria A only are second highest in our hierarchy

#3. Reviews that meet criteria B only are third highest in our hierarchy (this decision can be ignored if heterogeneity is not observed)

Decision rules for Step E.

#1. Reviews that meet Step E1, E2, and E3 are highest in our hierarchy

#2. Reviews that meet Step E1 and E2 second highest in our hierarchy

#3. Reviews that meet Step E1 third highest in our hierarchy

#4. Reviews that meet Step E2 and E3 fourth highest in our hierarchy

#5. Reviews that meet Step E2 fifth highest in our hierarchy

#6. Reviews that meet Step E3 sixth highest in our hierarchy

Note: Reporting only Steps E1, E2 or E3 is not considered a systematic approach to evidence synthesis.

Figure 2. Our Approach to Operationalizing Step E of the Jadad Algorithm.

3.7.5 Step F - Select the review with the lowest risk of bias, or the highest quality

From the risk of bias/quality assessment conducted through Step D, we will choose the SR with the lowest risk of bias judgement, or highest quality assessment rating. ROBIS contains a last phase where reviewers are asked to summarise concerns identified in each domain and describe whether the conclusions were supported by the evidence. Based on these last decisions, a final SR rating will be made based on high, low or unclear risk of bias. For our Jadad assessment, we will choose a binary rating of either high risk or low risk of bias. Any SRs assessed as 'Unclear' risk of bias will be deemed as high risk. When using the assessments of risk of bias/quality of SRs from the included Discordant Reviews, we will choose the rating of the authors. If uncertainty exists, we will re-assess the included SRs using ROBIS.

3.7.6 Step G - Do the reviews have the same eligibility criteria?

If the SRs do not include the same trials, then decision-makers are directed to turn to Step G – assess whether the SRs have the same eligibility criteria (**Figure 3**). The Discordant Review may contain text in a methods section, or a characteristics of included SRs table where the PICO eligibility criteria can

be extracted and assessed. If this is not the case, then the PICO eligibility criteria will be extracted from the included SRs by two authors independently and then compared to resolve any discrepancies.

Figure 3. Our Approach to Operationalising Step G of the Jadad Algorithm

2.8.7 Step H - Assess and compare the search strategies and the application of the eligibility criteria across reviews

If the SRs contain the same eligibility criteria, then Step H is to assess and compare the search strategies and the application of the eligibility criteria across SRs (**Figure 4**).

Step H1 - Assess and compare the search strategies across reviews

In this step, Jadad et al.'s recommendations are vague, although they make reference to comprehensive search strategies as being less biased. We interpret this step as authors explicitly describing their search strategy such that it can be replicated. To meet this interpretation, our criteria are that SRs: (A) search 2 or more databases, (B) search the grey literature; and (C) include a full search algorithm (may be attached as an appendix or included in the manuscript).

Decision rules:

#1. Reviews that meet criteria A, B and C are highest in our hierarchy

#2. Reviews that meet criteria A and B are second highest in our hierarchy

#3. Reviews that meet criteria A and C are third highest in our hierarchy

#4. Reviews that meet criteria B and C are fourth highest in our hierarchy (unlikely scenario)

#5. Reviews that meet criteria A only are fifth highest in our hierarchy

<u>SCENARIOS for Step H1</u>

• 3 SRs are identified for our Jadad assessment

Criteria to choose a systematic SR at Step H1: (A) 2 or more databases – (B) searched grey literature -- (C) full search in appendix

Scenario 1

Review 1: A and B, but not C (decision rule #2)

Review 2: A and B but not C (decision rule #2)

Review 3: A and C, but not B (decision rule #3)

Conclusion: No SR meets ALL of our criteria; which do we choose? Based on our decision rules, we choose BOTH Review 1 and 2

Scenario 2

Review 1: A, but neither B nor C (decision rule #5)

Review 2: A and B, but not C (decision rule #2)

Review 3: Neither A, B, nor C (does not report the search methods)

Conclusion: No SR meets ALL of our criteria; which do we choose? Based on our decision rules, we choose Review 2

Step H2 - Assess and compare the application of the eligibility criteria across reviews

In this sub-step, Jadad indicates that we should choose the SR with the most explicit and reproducible inclusion criteria, which is ambiguous. Jadad states, "SRs with the same selection criteria may include different trials because of differences in the application of the criteria, which are due to random or systematic error. Decision-makers should regard as more rigorous those SRs with explicit, reproducible inclusion criteria. Such criteria are likely to reduce bias in the selection of studies" [22]. We did not know if this meant clearly reproducible PICO eligibility criteria, as this would be a repeat to Step G, whether the eligibility criteria were applied consistently by SRs (i.e. compare eligibility criteria to included RCTs' PICO to see if they indeed met the eligibility criteria), or if this meant (A) independently screening of title, abstracts, and full text against the eligibility criteria by two reviewers. We selected the latter criteria when choosing from the included SRs in an Discordant Review.

Decision rules:

#1. Reviews that meet criteria A is highest in our hierarchy

Decision rules for Step H:

#1. Reviews that meet Step H1 and H2 highest in our hierarchy

#2. Reviews that meet Step H1 second highest in our hierarchy

#3. Reviews that meet Step H2 third highest in our hierarchy

Figure 4. Our Approach to Operationalising Step H of the Jadad Algorithm.

3.7.8 Step I – Assess and compare the publication status, quality, language restrictions of the included RCTs, and analysis of data on individual patients

If the SRs do not have the same eligibility criteria, then the next step, Step I, is to assess and compare the publication status, quality, language restrictions of the included RCTs, and analysis of data on individual patients across the SRs (**Figure 5**). This step maps to ROBIS item 1.5, namely, "Were any restrictions in eligibility criteria based on appropriate information sources (e.g. publication status or format, language, availability of data) [40]?"

Step I1 – Assess and compare the publication status of the included RCTs across reviews

In the absence of clear guidance, we interpret this step as "choose the SR that searches for and includes both published and unpublished data (grey literature)." Published studies are defined as any study or data published in a peer-reviewed medical journal. Unpublished data is defined as any information that is difficult to locate and obtained from non-peer-reviewed sources such as websites (e.g. World Health Organisation website, CADTH), clinical trial registries (e.g. clinicaltrials.gov), thesis and dissertation databases, and other unpublished data registries (e.g. LILIACS). Our interpretation is that SRs are chosen at this step that search for: (A) studies published in peer-reviewed medical journals, and (B) reports/documents/content that are not published in medical journals.

Decision rules:
#1. Reviews that meet criteria A and B are highest in our hierarchy
#2. Reviews that meet criteria A are second highest in our hierarchy
#3. Reviews that meet criteria B are third highest in our hierarchy
Note: Reporting only A or B is not considered a systematic search.

Step I2 – Assess and compare the methods used to assess the quality of the included RCTs across reviews

In this step, the Jadad paper recommends assessing the appropriateness of the *methods used* to assess the quality of the included RCTs across SRs. This item maps to ROBIS item 3.4, "Was the risk of bias/quality of RCTs formally assessed using appropriate criteria?" Here we interpret this

item as to whether the SR authors used the Cochrane risk of bias tool (version 1 or 2). All other RCT quality assessment tools are inappropriate because they are out of date and omit important biases (e.g. Agency for Healthcare Research and Quality (AHRQ) 2012 [41] omits allocation concealment). However, the Cochrane risk of bias tool was only published in October 2008. Therefore, we applied a decision rule: for SRs dated 2012 (giving one year for awareness of the tool to reach researchers) and later, the Cochrane risk of bias tool is considered the gold standard. For SRs dated 2009 or earlier, we considered the Jadad [42] and Schulz [43] scales to be the most common scales used between 1995 and 2011. Other tools will be considered on a case-by-case basis.

As a decision hierarchy, to meet the minimum criteria for this step, a SR will have (A) assessed the risk of bias of RCTs using any tool or approaches, and (B) used the Cochrane risk of bias tool version 1 or 2 (if dated 2009 or later). If several SRs are included that meet these two criteria, the SR that (C) integrates the risk of bias assessments into the results or discussion section (i.e. discusses the risk of bias in terms of high and low risk of bias studies, reports a subgroup or sensitivity analysis) will be chosen.

Decision rules:

#1. Reviews that meet criteria A, B and C are highest in our hierarchy

#2. Reviews that meet criteria B and C are second highest in our hierarchy

#3. Reviews that meet criteria A and B are third highest in our hierarchy

#4. Reviews that meet criteria A and C are fourth highest in our hierarchy (unlikely scenario)

#5. Reviews that meet criteria A only are fifth highest in our hierarchy

SCENARIOS for Step I2

• 3 SRs are identified for our Jadad assessment

Scenario 1

Review 1: A and B but not C (decision rule #3)

Review 2: A and B but not C (decision rule #3)

Review 3: A and C, but not B (decision rule #4)

Conclusion: No SR meets ALL of our criteria; which do we choose? Based on our decision rules, we choose BOTH Review 1 and 2

Scenario 2

Review 1: A, but neither B nor C (decision rule #5)

Review 2: A and B, but not C (decision rule #3)

Review 3: Neither A, B, nor C (does not report the search methods)

Conclusion: No SR meets ALL of our criteria; which do we choose? Based on our decision rules, we choose Review 2

Step 13 - Assess and compare any language restrictions across reviews

In this step, Jadad indicates that SRs with (A) no language restrictions in eligibility criteria should be prioritised and chosen over those that only include English language RCTs. This step maps to ROBIS item 1.5, namely, "Were any restrictions in eligibility criteria based on sources of info appropriate (e.g. publication status or format, language, availability of data)?"

Decision rule:

#1. Reviews that meet criteria A are highest in our hierarchy

Step I4 – Choose the analysis of data on individual patients

If (A) an individual patient data meta-analysis was identified in the Discordant Review, Jadad et al. recommend this SR be chosen over SRs with pairwise meta-analysis.

Decision rule:

#1. Reviews that meet criteria A are highest in our hierarchy

Decision rules for Step I:

#1. If there is an IDP meta-analysis (Step I4), then this SR is the highest in our hierarchy

#2. Reviews that meet Step I1, I2, and I3 are second highest in our hierarchy

#3. Reviews that meet Step I1 and I2 third highest in our hierarchy

#4. Reviews that meet Step I2 and I3 fourth highest in our hierarchy

#5. Reviews that meet Step I1 and I3 fifth highest in our hierarchy

#6. Reviews that meet Step I1 is sixth highest in our hierarchy

#7. Reviews that meet Step I2 is seventh highest in our hierarchy

#8. Reviews that meet Step I3 is eighth highest in our hierarchy

Note: Reporting only Steps I1, I2 or I3 is not considered a systematic approach to evidence synthesis.

Figure 5. Our Approach to Operationalising Step I of the Jadad Algorithm

3.8 Study outcomes

3.8.1 Evaluation of whether the Discordant Review authors:

- Used the Jadad decision tree to assess discordance
- Examine and record reasons for discordance (i.e. authors did not use Jadad)
- Use other approaches to deal with discordance (specify)
- Present discordance in tables and figures.

3.8.2 Results from our discordance assessment and Discordant Review authors' assessments:

- Utility: Is the Jadad decision tree easy to use? (see section 2.9.3)
- Efficiency: How much time does it take to do one Jadad assessment?
- Comprehensiveness: is the Jadad algorithm comprehensive? Is it missing methods that might explain discordance (e.g. publication recency)?
- Jadad cohort: Frequency of disagreement or agreement across Jadad assessments between (a) Discordant Review authors' assessment, and (b) our assessment (i.e. choosing the same SR).
- Non-Jadad cohort: Frequency of disagreement or agreement assessments between (a) Discordant Review authors' assessment, and (b) our assessment (i.e. choosing the same SR).
- Comparison of Discordant Review authors stated sources of discordance and our identified sources of discordance.

3.8.3 "Ease of use" outcome measure

Each Jadad assessment will be assessed for "ease of use" by each assessor. Each Jadad assessment will be rated and colored (green, yellow, red) based on how easy or difficult the assessment was judged to be for the user. The rating is based on the following rubric:

- The step can be accomplished easily by the reviewer, due to low cognitive load or because it's a recognised method or approach.
- The step requires a notable degree of cognitive load by the reviewer but can generally be accomplished with some effort.
- The step is difficult for the reviewer, due to significant cognitive load or confusion; some reviewers would likely fail or abandon the task at this point.

The lower the score, the easier the step is to complete.

3.9 Data extraction

3.9.1 Discordant review level extraction

The outcomes from section 3.8 will be extracted, along with the following information from the Discordant Review:

1. Study characteristics (lead author's name, publication year).
3. Clinical, public health, or policy question (objectives, health condition of treatment, PICO eligibility criteria: participant, intervention/comparison, and primary outcome).
4. Methods (how the Discordant Review authors assessed discordance among the SRs [Jadad or other approach], how they operationalised Jadad, steps where they identified discordance, number of included SRs, type of SR (e.g. individual patient data meta-analysis, SR with narrative summary, SR with meta-analysis of RCTs), type of analysis (narrative summary or meta-analysis), risk of bias/quality assessment (e.g. AMSTAR, AMSTAR 2, or ROBIS tool), risk of bias judgments, and whether risk of bias/quality assessment was integrated into the synthesis.
5. Results (results of discordance assessment, effect size and confidence intervals (CIs), number of total participants in treatment and control groups, number of total events in treatment and control groups), direction of study results (favourable or not favourable).
6. Conclusions (difference between results and conclusions defined as if they disagreed in direction (results, not favourable; conclusions, favourable), authors' result interpretation (quote from abstract and discussion section about the primary outcome result and conclusion).

3.9.3 SR level extraction

The outcomes from section 3.8 will be extracted, along with the following information from the included SRs:

1. Study characteristics (lead author's name, publication year).
2. Clinical, public health, or policy question (objectives, health condition of treatment, PICO eligibility criteria: participant, intervention/comparison, and primary outcome; language restrictions and restrictions on publication status in eligibility criteria; citation of previous SRs/meta-analyses in background or discussion).
3. Search methodology (the name and number of databases searched, grey literature search details, the search period, language restrictions, restrictions on publication status, included full search in an appendix).
4. Methods (number and first author/year of included RCTs, effect metric (OR, RR, MD) and CIs, whether SR authors assessed the clinical (PICO) heterogeneity across RCTs (in the methods or results sections), analysis method (appropriate weighted technique to combine study results [i.e. used a fixed or random random-effects model], investigation of statistical heterogeneity [i.e. by reporting I2, tau2 or chi2], and if heterogeneity is present, then the authors investigated the causes of any heterogeneity [i.e. by reporting subgroup, sensitivity, or meta-regression analyses]), risk of bias/quality assessment (e.g. Cochrane risk of bias tool v1 or v2), risk of bias/quality judgment for each RCT, and whether the RCT quality/risk of bias assessment was integrated into the synthesis; 2 reviewers independently screened studies, extracted data and assessed risk of bias with process for resolving discrepancies found when comparing.
5. Results (effect size and CI, number of total participants in treatment and control groups, number of total events in treatment and control groups), direction of study results (favourable or not favourable).
6. Conclusions (difference between results and conclusions defined as if they disagreed in direction (results, not favourable; conclusions, favourable), authors' result interpretation (quote from abstract and discussion section about the primary outcome result and conclusion).

Two authors will extract studies independently at full-text, and in the case of discrepant decisions, will discuss until consensus is reached.

3.10 Data analysis

We will assess and compare our outcomes (a) narratively for qualitative data, (b) using frequencies and percentages for categorical data, and (c) using median and interquartile range (IQR) for continuous data. Our analysis will be organised by our study outcomes in tables and in figures. We will discuss

differences in the assessment of discordance across Discordant Reviews using Jadad and Discordant Reviews not using Jadad.

4.0 PATIENT AND PUBLIC INVOLVEMENT

Patients nor the public were not involved in the design of our research protocol.

5.0 ETHICS, DISSEMINATION, STRENGTHS AND LIMITATIONS

5.1 Ethics

No ethics approval was required as no human subjects were involved.

5.2 Dissemination

We will disseminate evidence summaries through academic and social media platforms and websites (e.g. Twitter, Researchgate). We will publish in an open-access journal, and present at various formal and informal venues such as the Therapeutics Initiative Methods Speaker Series, and academic conferences such as Guidelines International Network conference, Cochrane Colloquium, and Public Health.

5.3 Strengths and limitations

We aim to use a systematic approach to evaluate authors' assessment of discordance across SRs in Discordant Reviews and replicate Jadad algorithm assessments from a sample of Discordant Reviews. We suspect that one reason for the inconsistent use of the Jadad algorithm in the existing literature may be due to the limited guidance available in the original Jadad manuscript on implementing the algorithm. We believe the greatest strength of the proposed study will be that we used an iterative process among authors to develop decision rules for the interpretation and application of each step in the Jadad algorithm. Upon completion and dissemination of this study, other Discordant Review authors will have more detailed guidance on how to apply the Jadad algorithm when addressing discordance in Discordant Reviews. Furthermore, our study adopted a systematic and transparent approach to address the objectives outlined in our protocol using SR guidance [26]. A comprehensive search strategy, including a search of the grey literature, was employed with no restrictions on language and publication status to yield relevant studies and minimise publication bias. To minimise error, screening, extractions, and assessments will be completed by two independent reviewers, and subsequently compared. Any discrepancies will be resolved upon consensus, and if necessary, with the

involvement of a third reviewer. To mitigate observer bias, reviewers were blinded to the Discordant Reviews' Jadad assessments.

Despite several strengths, this study also has limitations that are to be noted. In our search update, we only searched MEDLINE (Ovid) which would have limited the number of potentially relevant studies found. However, since the aim of our methods study is to replicate Jadad algorithm assessments, we do not think updating our search would make a difference to the robustness of the results. Our search strategy may have been more complete by adding terms such as inconsistencies or concerns, and truncated variations. Moreover, the searches were not carried out with the guidance from a trained medical/health librarian and/or peer reviewed. We did not, however, aim to retrieve all Discordant Reviews in the literature, as we could only feasibly replicate 30 such studies.

In addition, by only focusing on Discordant Reviews specific to health interventions, we may potentially have overlooked a number of relevant Discordant Reviews that may have assessed discordance and/or employed the Jadad algorithm. When utilising the tools for assessment (i.e. risk of bias tools or Jadad decision tree), there will be subjectivity in the judgements potentially introducing variability in the results. To overcome this limitation, pilot screening and pilot assessments will be completed by reviewers and assessed to ensure consistency in understanding of the screening criteria, definitions of extracted items, and steps in the Jadad assessment.

Some steps in the Jadad algorithm were vague in description (e.g. step I2), making it difficult to interpret. For example, step I2 assesses the methods used to assess the quality of included RCTs across SRs. Some authors could interpret this as assessing if SR authors used the Cochrane risk of bias to assess the quality of RCTs across SRs. However, the Cochrane risk of bias tool was only published in 2008. To minimise the incorrect interpretation of this step, we solicited feedback at the protocol stage from all authors in its interpretation prior to piloting. Thus, after discussion with authors, we applied a decision rule that for SRs dates 2009 and earlier, we considered the Jadad [42] and Schulz [43] scales to be the most common scale between 1995 and 2009, in addition to other tools being considered on a case-by-case basis. This was done for all steps. Additionally, during piloting we will amend our interpretation and instructions on how to operationalise the Jadad algorithm to ensure consistent application.

Figure legend

Figure 1. Jadad (1997) decision tree.

Figure 2. Our Approach to Operationalizing Step E of the Jadad Algorithm

Figure 3. Our Approach to Operationalising Step G of the Jadad Algorithm

Figure 4. Our Approach to Operationalising Step H of the Jadad Algorithm

Figure 5. Our Approach to Operationalising Step I of the Jadad Algorithm

Declarations

No declarations.

Consent for publication

Not applicable

Availability of data and materials

Not applicable

Competing interests

The authors declare that they have no competing interests.

Funding

No grants or funding were successful for this important study. Dr. Andrea Tricco currently holds a Tier 2 Canada Research Chair in Knowledge Synthesis and an Ontario Ministry of Research and Innovation Early Researcher Award. Dr. Carole Lunny is funded by a CIHR project grant for her postdoctoral research.

Authors' contributions

CL conceived of the study; CL, NF, ACT, SW, KK, AS, PS, DW, LP, WZ, RA, JT, YC, BK, JZ, ER, HN, ST, DP, PT, SK, SST contributed to the design of the study; CL drafted the manuscript; CL, NF,

ACT, SW, KK, AS, PS, DW, LP, WZ, RA, JT, YC, BK, JZ, ER, HN, ST, DP, PT, SK, SST edited the manuscript and read and approved the final manuscript.

References

1. Bastian H, Glasziou P, Chalmers IJPm: **Seventy-five trials and eleven systematic reviews a day: how will we ever keep up?** 2010, **7**(9):e1000326.
2. Taito S, Kataoka Y, Ariie T, Oide S, Tsujimoto Y: **Assessment of the Publication Trends of COVID-19 Systematic Reviews and Randomized Controlled Trials**. *Annals of Clinical Epidemiology* 2021, **3**(2):56-58.
3. Niforatos JD, Weaver M, Johansen ME: **Assessment of Publication Trends of Systematic Reviews and Randomized Clinical Trials, 1995 to 2017**. *JAMA Internal Medicine* 2019, **179**(11):1593-1594.
4. Becker LA, Oxman, A.D.: **Chapter 22: Overviews of reviews**. In: *Cochrane Handbook for Systematic Reviews of Interventions.* edn. Edited by Higgins JPT GSE. Hoboken, NJ: John Wiley & Sons; 2008: 607-631.
5. Pieper D, Antoine SL, Mathes T, Neugebauer EA, Eikermann M: **Systematic review finds overlapping reviews were not mentioned in every other overview**. *J Clin Epidemiol* 2014, **67**(4):368-375.
6. Pieper D, Mathes T, Eikermann M: **Impact of choice of quality appraisal tool for systematic reviews in overviews**. *J Evid Based Med* 2014, **7**(2):72-78.
7. Smith V, Devane D, Begley CM, Clarke M: **Methodology in conducting a systematic review of systematic reviews of healthcare interventions**. *BMC Med Res Methodol* 2011, **11**(1):15.
8. Caird J, Sutcliffe K, Kwan I, Dickson K, Thomas J: **Mediating policy-relevant evidence at speed: are systematic reviews of systematic reviews a useful approach?** *Evidence & Policy: A Journal of Research, Debate and Practice* 2015, **11**(1):81-97.
9. Hunt H, Pollock A, Campbell P, Estcourt L, Brunton GJSr: **An introduction to overviews of reviews: planning a relevant research question and objective for an overview**. 2018, **7**(1):39.
10. Ioannidis JPA: **The Mass Production of Redundant, Misleading, and Conflicted Systematic Reviews and Meta-analyses**. *The Milbank quarterly* 2016, **94**(3):485-514.
11. Lunny C, Brennan SE, McDonald S, McKenzie JE: **Toward a comprehensive evidence map of overview of systematic review methods: paper 1-purpose, eligibility, search and data extraction**. *Syst Rev* 2017, **6**(1):231.
12. Lunny C, Brennan SE, McDonald S, McKenzie JE: **Toward a comprehensive evidence map of overview of systematic review methods: paper 2-risk of bias assessment; synthesis, presentation and summary of the findings; and assessment of the certainty of the evidence**. *Syst Rev* 2018, **7**(1):159.
13. Bolland MJ, Grey A: **A case study of discordant overlapping meta-analyses: vitamin d supplements and fracture**. *PloS one* 2014, **9**(12):e115934.
14. Campbell J, Bellamy N, Gee T: **Differences between systematic reviews/meta-analyses of hyaluronic acid/hyaluronan/hylan in osteoarthritis of the knee**. *Osteoarthritis and cartilage / OARS, Osteoarthritis Research Society* 2007, **15**(12):1424-1436.
15. Chalmers PN, Mascarenhas R, Leroux T, Sayegh ET, Verma NN, Cole BJ, Romeo AA: **Do arthroscopic and open stabilization techniques restore equivalent stability to the shoulder in the setting of anterior glenohumeral instability? a systematic review of overlapping meta-analyses**. *Arthroscopy : the journal of arthroscopic & related surgery : official publication of the Arthroscopy Association of North America and the International Arthroscopy Association* 2015, **31**(2):355-363.
16. Riva N, Puljak L, Moja L, Ageno W, Schünemann H, Magrini N, Squizzato A: **Multiple overlapping systematic reviews facilitate the origin of disputes: the case of thrombolytic therapy for pulmonary embolism**. *J Clin Epidemiol* 2018, **97**:1-13.

17. Vavken P, Dorotka R: **A systematic review of conflicting meta-analyses in orthopaedic surgery**. *Clinical Orthopaedics and Related Research®* 2009, **467**(10):2723-2735.
18. Peters WD, Panchbhavi VK: **Primary Arthrodesis Versus Open Reduction and Internal Fixation Outcomes for Lisfranc Injuries: An Analysis of Conflicting Meta-analysis Results**. *Foot & Ankle Orthopaedics* 2020, **5**(4):2473011420S2473000383.
19. Susantitaphong P, Jaber BL: **Understanding discordant meta-analyses of convective dialytic therapies for chronic kidney failure**. *American Journal of Kidney Diseases* 2014, **63**(6):888-891.
20. Teehan GS, Liangos O, Lau J, Levey AS, Pereira BJ, Jaber BL: **Dialysis membrane and modality in Acute Renal Failure: Understanding discordant meta-analyses**. In: *Seminars in dialysis: 2003*: Wiley Online Library; 2003: 356-360.
21. Druyts E, Thorlund K, Humphreys S, Lion M, Cooper CL, Mills EJ: **Interpreting discordant indirect and multiple treatment comparison meta-analyses: an evaluation of direct acting antivirals for chronic hepatitis C infection**. *Clinical epidemiology* 2013, **5**:173-183.
22. Jadad AR, Cook DJ, Browman GP: **A guide to interpreting discordant systematic reviews**. *CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne* 1997, **156**(10):1411-1416.
23. Li Q, Wang C, Huo Y, Jia Z, Wang X: **Minimally invasive versus open surgery for acute Achilles tendon rupture: a systematic review of overlapping meta-analyses**. *Journal of orthopaedic surgery and research* 2016, **11**(1):65.
24. Mascarenhas R, Chalmers PN, Sayegh ET, Bhandari M, Verma NN, Cole BJ, Romeo AA: **Is double-row rotator cuff repair clinically superior to single-row rotator cuff repair: a systematic review of overlapping meta-analyses**. *Arthroscopy : the journal of arthroscopic & related surgery : official publication of the Arthroscopy Association of North America and the International Arthroscopy Association* 2014, **30**(9):1156-1165.
25. Zhao JG, Wang J, Long L: **Surgical Versus Conservative Treatments for Displaced Midshaft Clavicular Fractures: A Systematic Review of Overlapping Meta-Analyses**. *Medicine* 2015, **94**(26):e1057.
26. Higgins JP: **Cochrane handbook for systematic reviews of interventions version 5.0. 1. The Cochrane Collaboration**. *http://www cochrane-handbook org* 2008.
27. Lunny C, Neelakant T, Chen A, Shinger G, Stevens A, Tasnim S, Sadeghipouya S, Adams S, Zheng YW, Lin L *et al*: **Bibliometric study of 'overviews of systematic reviews' of health interventions: Evaluation of prevalence, citation and journal impact factor**. *Res Synth Methods* 2021.
28. Lunny C, McKenzie JE, McDonald S: **Retrieval of overviews of systematic reviews in MEDLINE was improved by the development of an objectively derived and validated search strategy**. *Journal of clinical epidemiology* 2016, **74**:107-118.
29. Goossen K, Hess S, Lunny C, Pieper D: **Database combinations to retrieve systematic reviews in overviews of reviews: a methodological study**. *BMC medical research methodology* 2020, **20**(1):138-138.
30. Fabiano GA, Schatz NK, Aloe AM, Chacko A, Chronis-Tuscano A: **A systematic review of meta-analyses of psychosocial treatment for attention-deficit/hyperactivity disorder**. *Clin Child Fam Psychol Rev* 2015, **18**(1):77-97.
31. Finnane A, Janda M, Hayes SC: **Review of the evidence of lymphedema treatment effect**. *American journal of physical medicine & rehabilitation* 2015, **94**(6):483-498.
32. Lau R, Stevenson F, Ong BN, Dziedzic K, Treweek S, Eldridge S, Everitt H, Kennedy A, Qureshi N, Rogers A: **Achieving change in primary care—effectiveness of strategies for improving implementation of complex interventions: systematic review of reviews**. *BMJ open* 2015, **5**(12):e009993.
33. López NJ, Uribe S, Martinez B: **Effect of periodontal treatment on preterm birth rate: a systematic review of meta-analyses**. *Periodontology 2000* 2015, **67**(1):87-130.
34. Moher D, Dulberg CS, Wells GA: **Statistical power, sample size, and their reporting in randomized controlled trials**. *Jama* 1994, **272**(2):122-124.

35. Tricco AC, Tetzlaff J, Pham B, Brehaut J, Moher D: **Non-Cochrane vs. Cochrane reviews were twice as likely to have positive conclusion statements: cross-sectional study**. *J Clin Epidemiol* 2009, **62**(4):380-386.e381.
36. Hróbjartsson A, Thomsen ASS, Emanuelsson F, Tendal B, Hilden J, Boutron I, Ravaud P, Brorson S: **Observer bias in randomized clinical trials with measurement scale outcomes: a systematic review of trials with both blinded and nonblinded assessors**. *CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne* 2013, **185**(4):E201-E211.
37. Morissette K, Tricco AC, Horsley T, Chen MH, Moher D: **Blinded versus unblinded assessments of risk of bias in studies included in a systematic review**. *Cochrane Database of Systematic Reviews* 2011(9).
38. Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, Porter AC, Tugwell P, Moher D, Bouter LM: **Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews**. *BMC Med Res Methodol* 2007, **7**(1):1-7.
39. Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J, Moher D, Tugwell P, Welch V, Kristjansson E: **AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both**. *bmj* 2017, **358**.
40. Whiting P SJ, Churchill R.: **Introduction to ROBIS, a new tool to assess the risk of bias in a systematic review**. In: *23rd Cochrane Colloquium: 2015; Vienna, Austria*: John Wiley & Sons; 2015.
41. Viswanathan M, Patnode CD, Berkman ND, Bass EB, Chang S, Hartling L, Murad MH, Treadwell JR, Kane RL: **Assessing the risk of bias in systematic reviews of health care interventions**. *Methods guide for effectiveness and comparative effectiveness reviews [Internet]* 2017.
42. Jadad AR, Moher D, Klassen TP: **Guides for reading and interpreting systematic reviews: II. How did the authors find the studies and assess their quality?** *Archives of pediatrics & adolescent medicine* 1998, **152**(8):812-817.
43. Schulz KF, Chalmers I, Hayes RJ, Altman DG: **Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials**. *Jama* 1995, **273**(5):408-412.

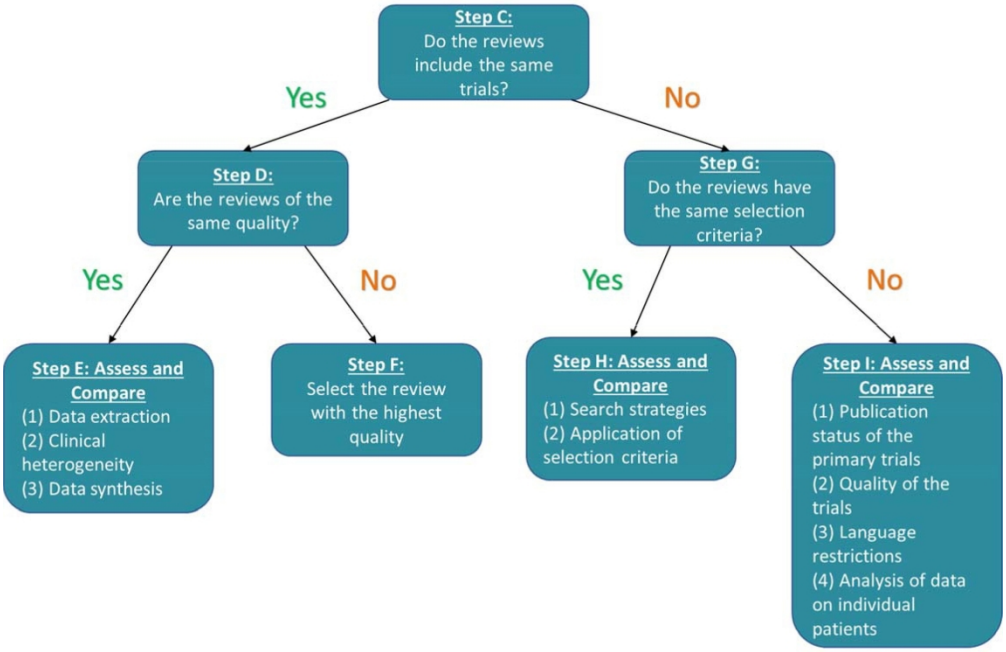

Figure 1. Jadad (1997) decision tree.

Figure 1

190x130mm (300 x 300 DPI)

Figure 2. Our Approach to Operationalizing Step E of the Jadad Algorithm.

Figure 2

164x208mm (300 x 300 DPI)

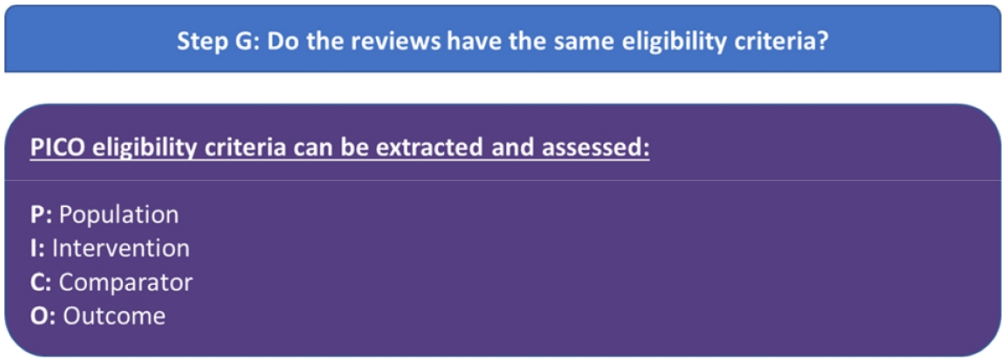

Figure 3. Our Approach to Operationalising Step G of the Jadad Algorithm

Figure 3

160x61mm (300 x 300 DPI)

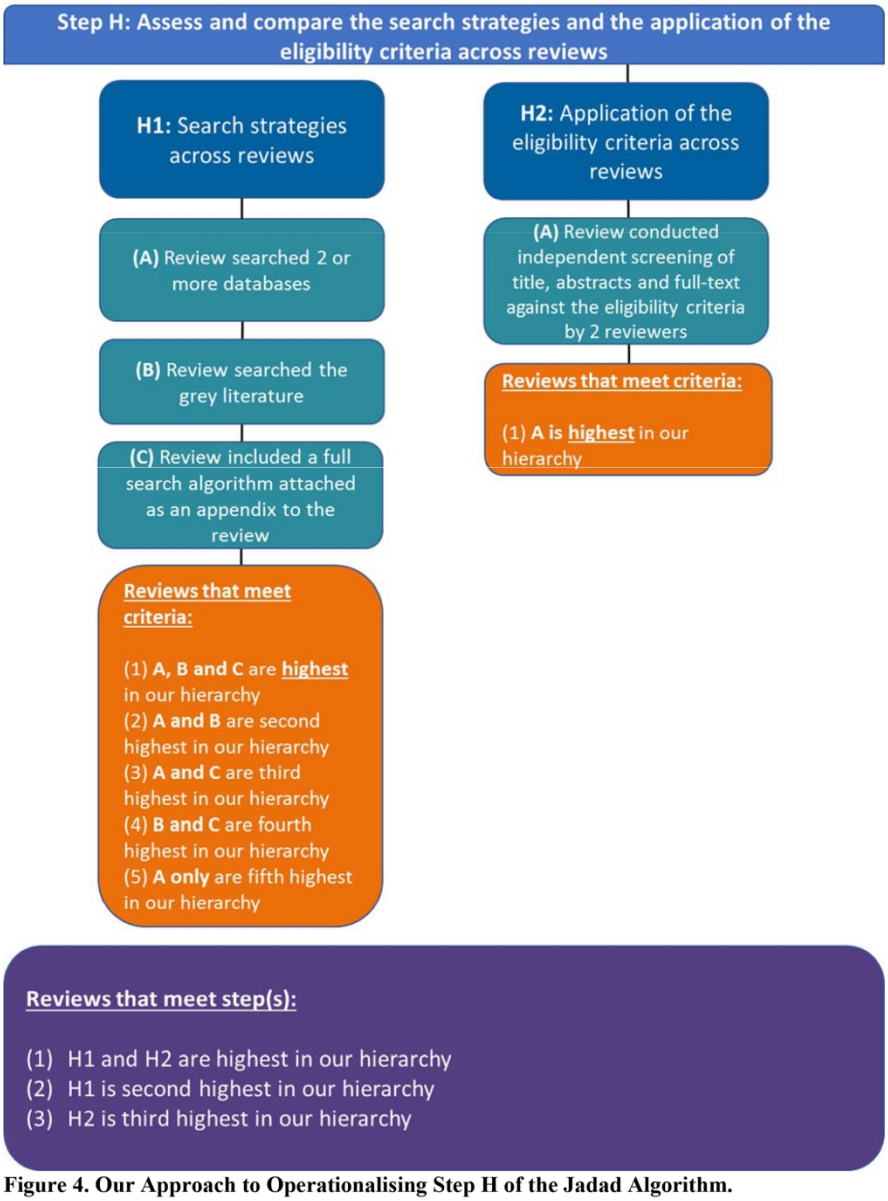

Figure 4. Our Approach to Operationalising Step H of the Jadad Algorithm.

Figure 4

163x222mm (300 x 300 DPI)

Figure 5. Our Approach to Operationalising Step I of the Jadad Algorithm.

Figure 5

175x220mm (300 x 300 DPI)