

PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (<http://bmjopen.bmj.com/site/about/resources/checklist.pdf>) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	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, Carole; Thirugnanasampanthar, Sai Surabi; Kanji, Salmaan; Ferri, Nicola; Thabet, Pierre; Pieper, Dawid; Tasnim, Sara; Nelson, Harrison; Reid, Emma; Zhang, Jia He (Janet); Kalkat, Banveer; Chi, Yuan; Thompson, Jacqueline; Abdoulrezzak, Reema; Zheng, Di Wen (Wendy); Pangka, Lindy; Wang, Dian (Xin Ran); Safavi, Parisa; Sooch, Anmol; Kang, Kevin; Whitelaw, Sera; Tricco, Andrea

VERSION 1 – REVIEW

REVIEWER	Kilicoglu, Halil University of Illinois System
REVIEW RETURNED	28-Aug-2021

GENERAL COMMENTS	<p>A study protocol to replicate Jadad algorithm in a sample of overviews is presented. The goal is to describe how overviews address discordant findings in results across multiple systematic reviews (SR) and evaluate Jadad algorithm for utility, efficiency, and comprehensiveness. 30 overviews will be considered. Since Jadad algorithm is sometimes vague, the authors clarify how they will interpret various aspects of the algorithm based on an iterative process, which is a strength of the paper. Stages of the study (screening, Jadad piloting, data extraction, etc.) are described in detail. I don't have a major issue with the protocol. The study may help with making Jadad algorithm more precise and lead to its more consistent use. The weakness of the paper is the writing and organization, which should be improved for clarity and tightened up to be acceptable.</p> <p>Specific comments:</p> <ul style="list-style-type: none">- More context for Jadad use can be provided. How commonly is it used? What kind of inconsistencies occur in its application?- Figures are useful. The narrative, especially in Section 3.7, on the other hand can be hard to follow. Several steps seem to apply to overviews, SRs, and RCTs and it is easy to get confused about what a particular step applies to.- "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" I am unable to follow.- Stage 2 screening bullets can be more consistent (e.g., "Includes...." and "Uses...."
-------------------------	---

	<ul style="list-style-type: none"> - "We will use Google translate to interpret non-English studies for screening and assessment." Given that Google Translate may have significant issues, this may be problematic. - "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" Could 'serious' be subjective? - "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" It does not make much sense to me. 'primary outcome' -> 'primary intervention'? - "Two authors will extract the primary intervention and outcome, and disagreements will be discussed until consensus is reached." It was unclear to me whether this is for SRs. - "SRs with meta-analysis of RCTs" This is repeated several times, is it necessary to qualify with "with meta-analysis of RCTs"? - More details on AMSTAR, AMSTAR 2, and ROBIS can be provided. - ""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 I2, tau2, or chi2)" It sounds ungrammatical. - "Reporting only Steps E1, E2 or E3 is not considered a systematic approach to evidence synthesis." However "E1 only" higher than "E2 and E3". Why? - "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. " Confusing sentence and possibly ungrammatical. - ""we considered the Jadad scale and allocation concealment to be the most common scale between 1995 and 2009".. "allocation concealment" does not seem to make sense. - In 3.9.2, it does not seem necessary to repeat the definitions of direction of effect, etc. from Jadad. - In Funding: "No grants or funding were successful for this important study." sounds inappropriate. - Figure 1 caption mentions step A and B. But these are not in the figure. Is this an error? <p>Minor:</p> <ul style="list-style-type: none"> - overview authors assessment -> authors' - discordant or concordant discordance assessments?? - for continuous data" -> " seems spurious.
--	--

REVIEWER	Murad, M. Hassan Mayo Clinic
REVIEW RETURNED	30-Oct-2021

GENERAL COMMENTS	<ol style="list-style-type: none"> 1. Considering that the Jadad scale was published 24 years ago and was not used that much, wouldn't it be more fruitful to, instead of trying to validate it, to go ahead and do content analysis and propose a new algorithm? I think such content analysis should be the output of this work. 2. The Jadad article specifically dealt with conflicting reviews. However, most "overviews of reviews" are not done for that purpose. Most of them are done to address multiple outcomes not addressed in single reviews, or are done to provide an overarching
-------------------------	--

	<p>summary of a topic, or to identify individual studies for re-analysis (see Cochrane definition of overviews for example). Overviews encounter the challenge of choosing between two or more very similar systematic reviews (that may not be conflicting), and they have to choose between them. In this case, you can use AMSTAR criteria for example, or another, but again, the goal of overviews is not to compare conflicting reviews.</p> <p>3. Some of the text and figures (e.g., figure 5) explicitly use the term "RCT." Wouldn't all this apply to reviews of non RCTs?</p> <p>4. Note that many studies have shown that restricting literature searches to English did not make a difference in terms of systematic reviews conclusions. Also, it is a standard in some large systematic review programs, like the AHRQ EPC Program, to search for non English studies only to understand their distribution, but then exclude them.</p> <p>5. Having 22 authors on the protocol paper despite the fact that the sample of included overviews is going to be very small (30) requires some justification.</p>
--	--

REVIEWER	Iansavichene, Alla London Health Sciences Centre
REVIEW RETURNED	10-Nov-2021

GENERAL COMMENTS	<p>This is a well written and presented protocol. The strength of the study protocol is that it attempts to address an important topic that has not been covered before to evaluate and replicate Jadad algorithm assessments of discordance in SRs.</p> <p>Please find comments and suggestions for improvement below.</p> <p>1. Page 6, Line 26: For consistency and accuracy, having specified Cochrane Database of Systematic Reviews (CDSR) [Page 9, Line 21], it would be helpful to replace 'Cochrane Database', listed as source searched in the Abstract, with the specific database name e. g., Cochrane Database of Systematic Reviews. The Cochrane Library is a collection of databases and CDSR is only one of them.</p> <p>2. Page 6, Line 26 AND Page 9, Line 20-21 For the transparency, it might be helpful to indicate interface through which other databases were searched similar to what was reported for Medline (e.g., via Ovid, Wiley InterScience, etc.)</p> <p>3. Page 9, Line 39-40 AND Page 9, Line 20-21 For clarity for the reader, since it's indicated [Page 9, Line 39-40] inclusion of only Ovid MEDLINE database into the search update vs. 3 databases utilized for the original search [Page 9, Line 20-21], it might be helpful to provide rationale. Also, the CDSR is an appropriate choice, but the NLM https://www.nlm.nih.gov/pubs/techbull/jf02/jf02_cochrane.html has been indexing the CDSR for MEDLINE since 2000, so considerable number of the Cochrane Reviews indexed in the Cochrane Library's CDSR database are available within MEDLINE. Searching the MEDLINE (Ovid) and Epistemonikos would've been sufficient, unless authors found added value of including CDSR. Would authors consider including searching Epistemonikos in addition to MEDLINE (Ovid) for the search update?</p> <p>4. Page 9, Line 38-45 For accuracy of reporting and so that other researchers can replicate the search following the same algorithm, it could help to report or refer to the original search strategy (e. g., https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6186052/) used to derive the database with 1218 overviews. The reported in the protocol search string does not account for truncated variations of</p>
-------------------------	--

	<p>search terms in the title or abstract. Also, if you use a search tag the way it is reported (e.g., overlap.tiab) the search will not return any results as each field name (e.g., .ti) needs to be separated by a comma (e.g., .ti,ab) after the field name, and a period after the field name at the end cannot be omitted [e.g., .ti,ab. as in the following (overlap\$ or discrepan\$ or discordan\$ or differenc\$ or conflicting\$.ti,ab. or Jadad\$.ab.)] in Ovid Medline. Was the search strategy re-typed and, thus, introduced some inaccuracies? It is unclear if the reported search was carried out with a guidance from a trained medical/health librarian and/or peer reviewed.</p> <p>5. Page 9, Line 38-45</p> <p>Again, for accuracy. It is difficult to accurately assess the sensitivity of a reported search strategy, but authors might consider expanding the range of search terms to include specific names of quality assessment tools: AMSTAR, GRADE, ROBIS among others (e. g., (inconsisten\$ or (methodological\$ adj3 (quality or concern\$)) or (AMSTAR\$ and quality\$) or primary studies\$.tw,kf.) for the concept used (e. g., (overlap\$ or discrepan\$, etc.) to cover other ways in which it could be described.</p>
--	--

VERSION 1 – AUTHOR RESPONSE

Reviewer 1: Dr. Halil Kilicoglu, University of Illinois System

Please note that we have renamed overviews “discordant reviews” to distinguish them from overviews, which do more than assess discordance across conflicting systematic reviews, and may only include concordant reviews.

- More context for Jadad use can be provided. How commonly is it used? What kind of inconsistencies occur in its application?

Thank you very much for your feedback. There were no other studies identified that provided any of this information. No methodological investigations or replications of the Jadad algorithm itself were identified, and therefore we are not aware of any study to which we could directly compare our study results. We are only aware of studies that applied the Jadad algorithm for choosing the best systematic review. Studies 17-19 are examples of studies using the Jadad algorithm.

- Figures are useful. The narrative, especially in Section 3.7, on the other hand can be hard to follow. Several steps seem to apply to overviews, SRs, and RCTs and it is easy to get confused about what a particular step applies to.

As a rule, 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.

We have read through this section to add clarity. If the reviewer has further suggestions, they would be very welcome.

- ""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" I am unable to follow.

We have made the edit to line 189-191: "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."

- Stage 2 screening bullets can be more consistent (e.g., "Includes...." and "Uses....")

Thank you. We have made the edit in lines 199-201

- "We will use Google translate to interpret non-English studies for screening and assessment." Given that Google Translate may have significant issues, this may be problematic.

Several authors, including myself, are fluent in several languages. I can speak English, French, German, Spanish and a battering of other languages less well. Other authors are fluent in Mandarin, and Italian. Using Google translate will help with translation but will not solely relied upon. We believe this is the best way to minimise the potential for publication bias.

- "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" Could 'serious' be subjective?

Thank you for the comment. We will simply choose the first outcome mentioned in the manuscript, if no designated primary outcome is mentioned. We have therefore amended the sentence to: "If the primary outcome is not found in any of these sections, we will extract the first outcome mentioned in the manuscript [26, 27]." (Lines 226-228)

- "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" It does not make much sense to me. 'primary outcome' -> 'primary intervention'?

In overviews of reviews, it may be the case that multiple interventions are investigated. Some interventions may not examine the primary outcome, but a secondary outcome.

We therefore simplified and reworded this section to: "The primary intervention was selected according to the one related to the primary outcome. If this was unclear, we chose the first intervention highlighted in the title or abstract"

- "Two authors will extract the primary intervention and outcome, and disagreements will be discussed until consensus is reached." It was unclear to me whether this is for SRs.

We have edited the text in lines 234-37 to the following:

"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. "

- "SRs with meta-analysis of RCTs" This is repeated several times, is it necessary to qualify with "with meta-analysis of RCTs"?

We have deleted one of 5 instances (in the heading for 3.9) of "with meta-analysis of RCTs."

- More details on AMSTAR, AMSTAR 2, and ROBIS can be provided.

We have added the following sentences as further information:

“AMSTAR [30] and the updated AMSTAR 2 [31] are tools to assess methodological quality (i.e. quality of conduct and reporting) and ROBIS [32] 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).”

- ""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 I2, tau2, or chi2)" It sounds ungrammatical.

Thank you for the comment. We have edited to "This 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 I2, tau2, or chi2)"

- "Reporting only Steps E1, E2 or E3 is not considered a systematic approach to evidence synthesis." However, "E1 only" higher than "E2 and E3". Why?

Reviews that meet all three: Step E1 (data extraction methods), E2 (clinical heterogeneity methods), and E3, are highest in our hierarchy (random or fixed effect model; statistical heterogeneity methods). Ideally, we want all three to be reported by reviews, and it is true that they should not be considered equal. However, we had to select a hierarchy in the case when all three were not reported.

- "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. " Confusing sentence and possibly ungrammatical.

Edited to: For reviews dated 2009 or earlier, we considered the Jadad [34] and Schulz [35] scales to be the most common scales used between 1995 and 2011."

- ""we considered the Jadad scale and allocation concealment to be the most common scale between 1995 and 2009".. "allocation concealment" does not seem to make sense.

Edited to: Thus, after discussion with authors, we applied a decision rule that for reviews dates 2009 and earlier, we considered the Jadad [34] and Schulz [35] scales to be the most common scale between 1995 and 2009, in addition to other tools being considered on a case-by-case basis.

- In 3.9.2, it does not seem necessary to repeat the definitions of direction of effect, etc. from Jadad.

Agreed, we have deleted that section

- In Funding: "No grants or funding were successful for this important study." sounds inappropriate.

The sentence was reworded to: "we did not obtain any specific funding for this study"

- Figure 1 caption mentions step A and B. But these are not in the figure. Is this an error?

We have added "Step A and B are not shown in this figure; however,..." to clarify.

Minor:

- overview authors assessment -> authors'

Thank you. Edits were made to Lines 582 and 590.

- discordant or concordant discordance assessments??

Edit made

- for continuous data" -> " seems spurious.

Edit made

Reviewer 2 Dr. M. Hassan Murad, Mayo Clinic

1. Considering that the Jadad scale was published 24 years ago and was not used that much, wouldn't it be more fruitful to, instead of trying to validate it, to go ahead and do content analysis and propose a new algorithm? I think such content analysis should be the output of this work.

Thank you for your comment. A study of content analysis is being planned for the first study (objective 1, study 1; line 136). We have edited it accordingly.

2. The Jadad article specifically dealt with conflicting reviews. However, most "overviews of reviews" are not done for that purpose. Most of them are done to address multiple outcomes not addressed in single reviews, or are done to provide an overarching summary of a topic, or to identify individual studies for re-analysis (see Cochrane definition of overviews for example). Overviews encounter the challenge of choosing between two or more very similar systematic reviews (that may not be conflicting), and they have to choose between them. In this case, you can use AMSTAR criteria for example, or another, but again, the goal of overviews is not to compare conflicting reviews.

Thank you. We renamed "overviews" in our text to "discordant reviews" to distinguish them from overviews, which do more than assess discordance across conflicting systematic reviews. We also restructured the background section to explain this change.

3. Some of the text and figures (e.g., figure 5) explicitly use the term "RCT." Wouldn't all this apply to reviews of non RCTs?

The Jadad algorithm clearly states that only systematic reviews with meta-analysis of RCTs are to be included and assessed.

4. Note that many studies have shown that restricting literature searches to English did not make a difference in terms of systematic reviews conclusions. Also, it is a standard in some large systematic review programs, like the AHRQ EPC Program, to search for non English studies only to understand their distribution, but then exclude them.

Yes, we are aware of the research stating that searching the non-English literature did not make a difference in terms of systematic review conclusions. While the potential impact of studies published in languages other than English in a meta-analysis may be minimal, it is difficult to predict in which cases this exclusion may bias a systematic review. Since searching the non-English literature is not hard to do in our group as we have a handful of authors that speak different languages, we will include studies in all languages.

5. Having 22 authors on the protocol paper despite the fact that the sample of included overviews is going to be very small (30) requires some justification.

While we appreciate the reviewer's perspective, we can confirm that all authors comply with the ICMJE standards for authorship. Each author has made substantial contributions to the conception or design of the work through weekly study meetings to interpret the Jadad algorithm; wrote parts of the manuscript and designed the figures, and substantively revised the protocol manuscript.

This is a multi-year and multi-study project. Study 2, for example, has 120 data points to collect for each study requiring reflection and consultation. A strong collaboration makes the project and studies more robust.

Reviewer: 3

Mrs. Alla Iansavichene, London Health Sciences Centre

1. Page 6, Line 26:

For consistency and accuracy, having specified Cochrane Database of Systematic Reviews (CDSR) [Page 9, Line 21], it would be helpful to replace 'Cochrane Database', listed as source searched in the Abstract, with the specific database name e. g., Cochrane Database of Systematic Reviews. The Cochrane Library is a collection of databases and CDSR is only one of them.

Thank you for your comments. This has been done.

2. Page 6, Line 26 AND Page 9, Line 20-21

For the transparency, it might be helpful to indicate interface through which other databases were searched similar to what was reported for Medline (e.g., via Ovid, Wiley InterScience, etc.)

We made the following amendment: "We searched the CDSR through the website interface using the filter for Cochrane reviews." (Line 161)

Epistemonikos does not have an interface, and we searched the CDSR through their website: <https://www.cochranelibrary.com/advanced-search> using the filter for Cochrane reviews.

3. Page 9, Line 39-40 AND Page 9, Line 20-21

For clarity for the reader, since it's indicated [Page 9, Line 39-40] inclusion of only Ovid MEDLINE database into the search update vs. 3 databases utilized for the original search [Page 9, Line 20-21], it might be helpful to provide rationale. Also, the CDSR is an appropriate choice, but the NLM https://www.nlm.nih.gov/pubs/techbull/jf02/jf02_cochrane.html has been indexing the CDSR for MEDLINE since 2000, so considerable number of the Cochrane Reviews indexed in the Cochrane Library's CDSR database are available within MEDLINE. Searching the MEDLINE (Ovid) and Epistemonikos would've been sufficient, unless authors found added value of including CDSR. Would authors consider including searching Epistemonikos in addition to MEDLINE (Ovid) for the search update?

Thank you very much for your comment. Since the aim of our methods study is to replicate the Jadad algorithm assessments, we also do not think updating our search would make a difference to the robustness of the results. Moreover, most of the studies were from Medline, and as this is an unfunded study we did not update all databases.

Furthermore, the protocol was submitted to BMJ Open on June 10th, and we received this peer review feedback on Nov 11th. So while we agree with your suggestion to include Epistemonikos in the update, we are unable to update the search since the protocol took so long to get back to us.

We have thus included this limitation in the highlights section, and in the "strengths and limitations section" of the discussion. "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." (Lines 764-766)

4. Page 9, Line 38-45

For accuracy of reporting and so that other researchers can replicate the search following the same algorithm, it could help to report or refer to the original search strategy (e. g., <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6186052/>) used to derive the database with 1218 overviews. The reported in the protocol search string does not account for truncated variations of search terms in the title or abstract. Also, if you use a search tag the way it is reported (e.g., overlap.tiab) the search will not return any results as each field name (e.g., .ti) needs to be separated by a comma (e.g., .ti,ab) after the field name, and a period after the field name at the end cannot be omitted [e.g., .ti,ab. as in the following (overlap\$ or discrepan\$ or discordan\$ or differenc\$ or conflicting\$.ti,ab. or Jadad\$.ab.)] in Ovid Medline. Was the search strategy re-typed and, thus, introduced some inaccuracies? It is unclear if the reported search was carried out with a guidance from a trained medical/health librarian and/or peer reviewed.

Thank you for your comment, we have referenced the validated search strategy for overviews as a reference in the text. We have also added the following as a rationale: "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 [22]. As a rationale, we believe this combination would retrieve an equal proportion of overviews.”

Ovid MEDLINE search using the following search string:

Thank you. We have made the following edits: ("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.)

However, since we have completed our search and data extraction, we will not be conducting a search update.

We did not use the services of an information specialist. We have listed these as limitations as well.

5. Page 9, Line 38-45

Again, for accuracy. It is difficult to accurately assess the sensitivity of a reported search strategy, but authors might consider expanding the range of search terms to include specific names of quality assessment tools: AMSTAR, GRADE, ROBIS among others (e. g., (inconsisten\$ or (methodological\$ adj3 (quality or concern\$)) or (AMSTAR\$ and quality\$) or primary studies\$.tw,kf.) for the concept used (e. g., (overlap\$ or discrepan\$, etc.) to cover other ways in which it could be described.

Thank you for your suggestions. The Jadad algorithm does not promote any quality assessment tool used. We believe that if we search for the terms Jadad in the abstract or as a keyword in the full text, we will have captured all articles either using the Jadad scale to assess the quality of RCTs, or the Jadad algorithm to assess discordance across SRs. This would minimise the number of irrelevant citations we would need to screen.

VERSION 2 – REVIEW

REVIEWER	Murad, M. Hassan Mayo Clinic
REVIEW RETURNED	01-Dec-2021
GENERAL COMMENTS	adequate response
REVIEWER	Iansavichene, Alla London Health Sciences Centre
REVIEW RETURNED	06-Dec-2021
GENERAL COMMENTS	Reviewer 3's comments have been fully addressed in the revised manuscript. Authors have revised each relevant section and explained problems that were pointed out as limitation of the study in both section 5.3 and the strengths and limitations highlights. Thank you!