Author’s response to reviews

Title: Quality assessment of systematic reviews on total hip or knee arthroplasty using mod-AMSTAR

Authors:

Jing Li (lijing68@hotmail.com)
Xinyu Wu (18081901368@163.com)
Huan Sun (sun911027@163.com)
Xiaqin Zhou (18380199260@163.com)
Ji Wang (wangjimc@126.com)

Version: 2 Date: 25 Oct 2017

Author’s response to reviews:

Dears Editors,

Thank you very much for your letter and advice. We have revised the manuscript, and would like to re-submit it for your consideration. We have addressed the comments raised by the reviewers, and the amendments are highlighted in revised mode in the revised manuscript. In addition, we have checked and made some refinements of expressions again this time. Point by point responses to the reviewers’ comments are listed below.

We would like to express our sincere thanks to the reviewers for the constructive and positive comments.

Replies to Dawid Pieper, MPH:

- you clarified that there was an a priori developed protocol. Could make it available for reasons of transparency?
Reply-we have added an appendix for our protocol. As our study was not eligible for registered in PROSPERO platform, we wrote a protocol before the study and kept by ourselves as a guidance for the study process.

- I find your definition of a SR sill a bit too vague. It does not allow for a clear operationalization. For example, you can face meta-analyses without mentioning a literature search. Would that fit your inclusion criteria? What Need also to be more precise what is meant with "reproducible methods". Do the authors Need to Report all or one search strategy(ies) in full?

Reply-Sorry to make you confused. The purpose of our study is to assess the quality of systematic review or meta-analysis. As long as the author claimed the study as systematic review or meta-analysis, we included in our study. As to whether or not the study met the criteria of a qualified systematic review or meta-analysis, we used the AMSTAR to assess it.

- delete "We restricted language to English and Chinese" from the method section. This is already mentioned in the sentence before

Reply-thanks for your advice, we have deleted the sentence you mentioned.

- was data extraction also performed independently? What happened in case of disagreements?

Reply-Yes, the data extraction was performed independently, and we discussed or consult the third author if we met disagreements. We further revised it and make it clearer in study text (Page 4, paragraph 2, line 3-4).

- your idea of mAMSTAR seems to be very similar to R-AMSTAR

Reply-yes, at the beginning of our study, we have comprehensively conducted a search on the use of AMSTAR scale, R-AMSTAR (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2948145/pdf/TODENTJ-4-84.pdf) was one of the reference we referred to. We have cited it in our revised manuscript (page 4, paragraph 3, line1-2).

- I think it is more informative to state the 25th and 75th percentile when reporting the IQR instead of calculating the difference of them.
Reply-in consideration of your suggestion, we have modified the calculating difference to the 25th and 75th percentile.

- you elaborated on the (Regression) coefficient. It was clear this coefficient was from the Regression but it is not entirely clear what it is. I guess it is an Odds Ratio? Please clarify and state how do you Interpret the finding of 0.38

Reply-this coefficient was not Odds Ratio and was from the linear regression equation \( y = 0.3835x + 5.6986 \), but rounded with two digits after the decimal point. It means when the Journal impact factor increased 1 the mAMSTAR score increased 0.38

- you state in the discussion section now: "The overall methodological quality of SRs on THA was better than that..." Does this only refer to THA? What about TKA?

Reply-it’s our fault to lose TKA, we have applied TKA in the corresponding part in our manuscript as:

The overall methodological quality of SRs on THA and TKA was better than that…

-ACROBAT-NRSI is already outdated. Please refer to the ROBINS tool (https://sites.google.com/site/riskofbiastool/welcome/home)

Reply-thanks for your suggestion, we have read the article carefully and revised manuscript correspondingly. (Page 7, paragraph 6, line 5 to line 11)

- consider to add two sentences about risk of bias assessment of RCTs in the Quality of Primary studies section as all Reviews included RCTs, but you make a Point for other study designs

Reply- Thanks. We added following sentences in the section: The Cochrane Collaboration recommends a specific tool for assessing risk of bias of RCT, addressing seven specific domains, namely sequence generation, allocation concealment, blinding of participants and personnel, blinding of outcome assessment, incomplete outcome data, selective outcome reporting and ‘other issues’ that do not fit into these categories. (Page 7, paragraph 5, line8-12)
- when making the Point that applying mAMSTAR was challenging due to bad reporting, I would like to see some results of this in the manuscript. This could easily be done by referring to table 1 when focusing on the cannot answer category.

Reply-thanks for your advice, we have revised the corresponding part in Strength and Limitations. (Page 8, paragraph 6, line 7-9)

- what I think is still a Major concern is that you need more to Elaborate on the limitations. you modified AMSTAR, while AMSTAR is known to be reliable and valid. I think this statement might not hold true in your case as you made some extensive modifications. Please also note, that the score of AMSTAR has never been validated in any paper. It is mentioned in the external Validation paper you are referring to, but it has not undergone any test for validity or reliability. Problems with the scoring are also present in your manuscript as simply summing up scores implies weighting items. However, this weighting is not based any empirical evidence.

Reply-thanks for your helpful advice, we have revised the manuscript in the Strength and Limitations part. (Page 9, paragraph 1, line 5-7)

Supplement:

Data extraction table: "number of participants in SR" - do you mean number of authors?

Reply-Yes, to make this expression clearer, we modified this sentence to “number of authors in SR”

Replies to Al Mayhew:

1. There are instances when the authors need to justify their decision. For example, why only search from 2014 and 2015? Surely, there are SRs done in years before or after.

At the beginning of this study, we considered to include studies cover the last 5 years. However, we find the number of included studies was too large, thus we choose the last two years. We think the more recent published SR could better reflect current situations.

AM Sept 25, 2017: Is it the number of studies or your resources to do the work? I would put in a sentence in discussion saying "We only included studies published in 2014 or 2015 due to lack of resources. We understand that this may be a bias, as we would expect that the quality of more recent studies is likely higher than older studies"
2. Were there other criteria used to define systematic reviews other than words "systematic review" or "meta-analysis"? For example, a minimum number of databases searched?

We adopted the definition of “systematic review” from Cochrane handbook for systematic reviews of interventions in version 5.1.0, namely, a systematic review attempts to collate all empirical evidence that fits pre-specified eligibility criteria in order to answer a specific research question. It uses explicit, systematic methods that are selected with a view to minimizing bias, thus providing more reliable findings from which conclusions can be drawn and decisions made.

The key characteristics of a systematic review are:

a) A clearly stated set of objectives with pre-defined eligibility criteria for studies;

b) An explicit, reproducible methodology;

c) A systematic search that attempts to identify all studies that would meet the eligibility criteria;

d) An assessment of the validity of the findings of the included studies, for example through the assessment of risk of bias; and

e) A systematic presentation, and synthesis, of the characteristics and findings of the included studies.

We did not find compulsive requirement on the number of databases searched for this definition. We noticed that searching for at least two electronic sources was one of the requirements of AMSTAR items (item 3. Was a comprehensive literature search performed? from http://www.ncbi.nlm.nih.gov/pubmed/17302989).
AM Sept 25, 2017 Adopting this definition is fine, but you need to state this in the text "We used the definition of a systematic review from the Cochrane Handbook (Reference)."

All of your included reviews met all of the five criteria? I think it would be very helpful to describe eligibility using a PICO format for included reviews.

Reply-Thanks for your advice, we have re-stated the Inclusion and exclusion criteria part. As the purpose of this study was to assess the methodological quality of systematic reviews, we included the study once the authors claimed their study as systematic review so that we could assess whether it met the criteria of a real systematic review as required by AMSTAR and Cochrane Collaboration, and found the flaws of author claimed systematic review. So to make the inclusion and exclusion criteria of our study more clear, we re-wrote as following: We included all the author claimed systematic reviews or meta-analyses which focused on the effects and safety of procedures and prostheses in primary THA or TKA published in English or Chinese from 2014 to 2015. (Page 3, paragraph 4, line 5-8)

3. There are a lot of included studies. Did you give any thought to looking for SRs in other languages? Even if not to evaluate them, but just to have an understanding of the data you are missing. Would be good to justify.

When we scanned the title and abstract of the retrieved literature, we found 21 studies with full-text published in languages other than English and Chinese, but none of them were eligible for our study. We listed the references as follows:

(List removed)

We hope you can understand that in China it’s difficult for us to search studies in other language out of International databases and Chinese data bases.
There are 21 in this list but you only list 3 in Prisma figure. Can you clarify this? Did the non-English/non-Chinese papers come up in the English searches? It is acceptable in my opinion to exclude them but you need to be transparent about numbers (again, you can justify it because of limited resources).

Reply-the non-English/non-Chinese papers did come up in the English searches because they all had English titles and abstracts, and indexed in the English database. We only listed 3 in PRISMA figure because others were excluded when we screened the title and abstracts.

4. I believe that you did not clearly list all Chinese databases. "Like." implies you are not listing them all.

I am sorry to cause this misunderstanding, actually, we searched four Chinese databases and described in abstract, Figure 1 (the flow chart) and Appendix 1 (search strategies). To make it clear and more exact we deleted the word “like” in the revised manuscript (page 4, paragraph 2, line 3).

AM Sept 25, 2017 - Improved, thank you. Can you write out titles of databases in full? Reply-thanks for your advice, we have wrote out the full name of searched databases both in our manuscript and Appendix 1. (page 3, paragraph 5, line 3-5)

5. Did you develop the mAMSTAR or has it been used before? Any data on validity or reliability?

The contents of mAMSTAR are exactly the same as the original AMSTAR except for refining the items that include more than one content in a item. For example item 5 [Was a list of studies (included and excluded) provided?] was refined into 5.1(Was a list of included studies provided?) and 5.2 (Was a list of excluded studies provided?). We assessed the inter-rater agreement between two reviewers and the Cohen Kappa (κ) was provided in our study.

AM Sept 25, 2017 -I do not think the contents are the same for AMSTAR and mAMSTAR. By using mAMSTAR, you are able to provide partial scores (<1) on each item. My understanding is that a review had to include all the criteria to get a point in AMSTAR (e.g., both included and excluded studies in your example above.) The mAMSTAR score will almost always exceed the AMSTAR score; some items will get a partial score with mAMSTAR (e.g., 0.25, 0.67) but
would have received a score of 0 on AMSTAR as they did not meet all the criteria to get a point. This could lead to substantial differences between AMSTAR and mAMSTAR scores.

I was initially concerned that mAMSTAR had not been used before. However, I found it in a Cochrane review (Pollock et al, Interventions for improving upper limb function after stroke). You should definitely cite this review as having used mAMSTAR before. Any data on reliability/validity?

Reply-

(1) Thanks for your advice. We agree with you for the overall score of mAMSTAR can be higher than that of the original AMSTAR, and we did not test reliability and validity of the mAMSTAR. We have discussed it in revised manuscript, Strength and Limitations part. (Page 9, paragraph 1, line 5-12)

(2) When making our mAMSTAR scale, we referred to Pollock’s mAMSTAR, this article has been cited in our revised manuscript. But Pollock’s mAMSTAR was not completely in conformity with our scale, for example, we kept original Item 1 (Was an ‘a priori’ design provided?) in our mAMSTAR, but Pollock’s study extend this Item into four part (1.1 Were review subjects clearly defined? 1.2 Were review interventions described? 1.3 Were review comparisons specified? 1.4 Were review outcomes specified?), study from Pollock may over-interpret AMSTAR. (Page 4, paragraph 3, line1-2)

6. I do not understand the scoring of the mAMSTAR.

Our study did not change the total score by equally divided the score of each item to all its sub-items. For example, the item 5 was refined into 2 sub-items with each sub-item 0.5 point.

AM Sept 25, 2017 I looked at the table of mAMSTAR scores and the mAMSTAR is much clearer.

My preference would be to assess the studies using the original AMSTAR and then group them into the three categories: low (0-3), moderate (4-7) or high (8-11) and conduct and report all your comparisons using these three categories, rather than by using the AMSTAR or mAMSTAR score. I think this data would be much easier to present and for the reader to follow. I realize this means you would have to redo the analyses, but I think it would be a clearer presentation of the results. The challenge is that you will have a small number of studies in the high category but I think you could report how the reviews in the high category differ from those in the low category in text if the statistical comparisons were not useful due to small numbers.
If you choose not to change the analysis and continue with the mAMSTAR, then I feel strongly that you need to specify that the mAMSTAR is a new tool and the tool itself may influence your findings. You should discuss the strong likelihood of all reviews scoring higher using mAMSTAR than AMSTAR. You should also conduct a thorough literature search to demonstrate previous use of mAMSTAR other than Pollock et al and any supporting evidence of mAMSTAR.

Reply-

(1) Thanks for your advice; we quite agree with you that the mAMSTAR score will usually exceed the AMSTAR score. Due to time restriction, we will not change the analysis and continue with the mAMSTAR, but we discussed the limitation of using it and the possibility of higher total score in mAMSTAR in our revised manuscript. (Page 9, paragraph 1, line 5-12)

(2) We checked the use of mAMSTAR in Ovid-MEDLINE, Ovid-EMBASE, PUBMED, the Cochrane Library and all the Chinese databases, but only found Pollock’s study adopt this tool. This study has been cited in the revised manuscript.

7. In your original search, you found 10 times as many studies in English compared to Chinese. But in your included studies, there were only 3 times as many English studies as Chinese. I know that translating exact content of search strategies is difficult. I think this is an important point to discuss.

Although we used the same search words for both English databases and Chinese databases, it seems that the corresponding searching strategy is more sensitive in searching English databases than Chinese databases, resulting 10 times as many studies in English compared to Chinese. However, the number of studies not eligible for inclusion from English databases (1134) was higher than that from Chinese databases (115), resulting there were only 3 times as many English studies as Chinese.

AM Sept 25, 2017: I agree, but I wondered if you should discuss this in the text.

Reply-following your advice, we have discussed this at the beginning of the discussion. (Page 6, paragraph 2)
8. I like AMSTAR. But I think we have to be careful we don't analyze it in too much detail. I am happy to see that you reported the median score. You mention in your quality assessment that the "quality of the reviews was graded as high (8-11), medium (4-7) and low (0-3) quality," but you only report this once. I think you could have compared English and Chinese papers, year of publication using only high, medium or low. I know most of studies are in moderate category, but is it the same for Chinese and English papers? I think breaking down by individual item is too much analysis. It is also difficult to understand given my point #8 above.

We conducted a comparison between Chinese and English papers, year of publication using high, medium and low with Mann-Whitney U test, however, they were not statistically significant. Detailed information about two variables was displayed below:

(removed table)

AM Sept 25, 2017 See points number 7 and 8

Reply- We have discussed it in No.7 and No.8

9. 16 is a lot of variables to evaluate for bibliographic characteristics. How did you choose the characteristics?

We choose these characteristics mainly based on previous studies, which had conducted a quality assessment of systematic reviews and identify some factors that may affect the quality of systematic reviews. The explanation has been added in Methods section (page 4, paragraph 3, line 3 to 4).

AM Sept 25, 2017 Sixteen is still a lot of variables to analyze but I understand your rationale. I think it will be clearer if you categorize AMSTAR into low, moderate and high as described above.

Reply-We think using score in mAMSTAR being more informative, meanwhile only univariable analyses in each variable were conducted in our study, which cannot be affected by the number of variables. Meanwhile we checked table 3 and found we wrote wrong number of variables, we have modified it in our revised manuscript.
10. I think your conclusion that authors should consider AMSTAR and PRISMA when publishing is right. But, we don’t know if a review with a higher AMSTAR is more likely to change practice. It would be good to discuss that.

Thanks for your advice. The score of AMSTAR can only reflect the methodological quality of the systematic review, namely the internal validity. So a review with a higher AMSTAR score had more valid results. However, whether a review could change practice, we must also consider the clinical importance of the results and the generalizability of the review.

AM Sept 25, 2017 Why not mention this in the discussion? Your decision.

Reply-thanks for your advice, we have added this point to the conclusion. (page 9, paragraph 2, line 4-7)

11. Overall, I think you need to justify your decisions and methods more and spend less time on individual mAMSTAR details.

Thanks for your professional suggestion, we have extensively revised our methods and discussion part, we hope you are satisfied with our revision.


Reply-

(1) Most Items in original AMSTAR tool are multi-faceted, which can complicate the rating process, meanwhile important information can be neglected in the original tool. The newly published AMSTAR 2 (http://www.bmj.com/content/bmj/358/bmj.j4008.full.pdf) also noticed this problem, it kept and given more detailed coverage in two domains of original AMSTAR tool (split item 2 duplicate study selection and data extraction in to two domains and split Item 11 the possible influence of funding into primary studies and the review itself), which was exactly what we already considered in our mAMSTAR. Most items in AMSTAR were similar as those two Items. Thus, we refined AMSTAR items to simple univariable items. We have discussed it in our revised Strength and Limitations part. (Page 8, paragraph 5, line 4-8)
(2) We have explained in Introduction part that osteoarthritis (OA) is one of the ten most disabling diseases, while THA and TKA are the ultimate treatments of OA. Those measurements are both most commonly used treatment in total joint arthroplasty. The SRs in those two measurements was on the rise, but the study quality was not clear. In consideration of this, we tend to conduct a methodological quality assessment in Both TKA and THA.

(3) The characteristics were ones claimed to influence the methodological quality of SR from previous studies, we have explained it in our Study selection and data extraction part. (page 4, paragraph 2, line 4-6). Meanwhile, we checked our variables in table 3 and found we wrote wrong number of variables, we have modified it in our revised manuscript.

(4) Since we conducted a search on databases both in English and Chinese, we tend to compare the methodological quality and see the differences of flaws of SR in two different languages, and provide specific suggestions for Chinese researchers in this field.

12. I think it would also be good to include someone on your author team who has strong English writing skills; there are a lot of grammar mistakes.

   After rereading our manuscript sentence by sentence, some refinements of expressions have been made through cooperation of our authors and with the help of our colleagues who are good at English writing.

   AM Sept 25, 2017 The English writing still needs to be reviewed and corrected. There are many grammatical errors.

   Reply-thanks for your advice, we checked the article again and corrected some of the expressions, we hope this time you are satisfied with our work.

   AM Sept 25, 2017 Additional comments:

13. The strengths and weaknesses of this paper are not well described in the discussion.

   Reply-thanks for your advice, we have further thoroughly revised the Strength and Limitations part, we hope you are satisfied with our revision.
14. In Discussion, you state that "Another was that we refined the AMSTAR scale so that we could more accurately find the methodological flaws of included reviews."

This is true only if the reader looks at the full mAMSTAR table with each individual item. The overall mAMSTAR may be misleading.

Reply-in consideration of your idea, we have discussed it in our revised manuscript. (Page 9, paragraph 1, line 8-12)

Most items in AMSTAR was multi-faceted, those items can get one point only when they met all the criteria, however, mAMSTAR can get partial score when at least one criteria fulfilled. mAMSTAR thus hold strong possibility to get higher score than the original AMSTAR. When compare methodological quality with other fields using AMSTAR, the use of total mAMSTAR score should be interpreted with caution.

15. The discussion also includes the following sentence: "First, we only included reviews published in English and Chinese, bias could be introduced in reviews if positive findings are more likely to report in an international, English-language journal whereas negative findings are published in a local journal."

This is one area where I would be less specific. I would say something like: First, we only included reviews published in English and Chinese, bias could be introduced in reviews if positive findings are more likely to report in an international, English-language journal whereas negative findings are published in a local journal, and studies published in these two languages may differ from studies in other languages, especially with the high number of English and Chinese

Reply-thanks for your helpful advice, we have adjusted this description. (Page 8, paragraph 6, line 1-6)

We hope that the revised version of the manuscript is now acceptable for publication in your journal.

I look forward to hearing from you soon.
With best wishes,

Yours sincerely,

Xinyu Wu

Corresponding author:

Name: Jing Li

E-mail: lijing68@hotmail.com