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)
Xiaoqin Zhou (18380199260@163.com)
Ji Wang (wangjimc@126.com)

Version: 1  Date: 25 Aug 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 made some refinements of expressions through cooperation of our authors and with the help of our colleagues who are good at English writing 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:

-It is unclear which coefficient is meant, please be more precise

Reply 1: This coefficient is the regression coefficient based on the linear regression test between journal impact factor and mAMSTAR score of each included studies. The regression equation
was displayed in Figure 2. The coefficient is rounded to decimal point two. We have corrected the word “coefficient” to “regression coefficient” in the Abstract (page 2, Results, line 10) in the revised version, and it has been explained in method section (page 6, paragraph 1, line 8 to 9).

-I don’t understand the conclusion. The whole results section is about methodological quality. So why do you refer to PRISMA (reporting quality) in the conclusions?

Reply 2:The AMSTAR appraisal process was difficult to implement when the reporting quality was poor, especially some important components of systematic review are not reported, so it is recommended that PRISMA should be followed. But in order to emphasize the theme, we modified several sentence in abstract (page 2, conclusion).

Introduction:


Reply 3:Thank you for your recommendation, we read and compared this reference to reference 1 carefully and found that the content of the study was more update to reflect the current situation. Thus we decided to adopt the reference you recommended and made a relevant revision of the first sentence of the introduction.

-"From 2005 to 2015, there was..." this needs to be referenced

Reply 4:The relevant reference has been attached to the Introduction (page 3, paragraph 2, line 6) in the revised version. While we found the word “systematic review” was not proper, thus we corrected it to “meta-analysis” in Introduction (page 3, paragraph 2, line 5).

-it is important that you already make a differentiation between methodological quality and reporting quality in the introduction (and all forthcoming sections)

Reply 5:The statement of two kinds of quality has been added in background (page 3, paragraph 1, line 6 to 9)
do not use mAMSTAR in the introduction section, it should only be introduced in the methods section

Reply 6: According to your instructions, we have changed the word mAMSTAR to AMSTAR in Introduction section (page 2, paragraph 2, line 7).

Methods:

-please state there this review has been planned a pri/registered

Reply 7: Our study focused on the methodological assessment of published systematic reviews which is not eligible for registration in PROSPERO (Registration are limited to systematic reviews of the effects of interventions and strategies to prevent, diagnose, treat, and monitor health conditions, for which there is a health related outcome). However, we had a plan before the assessment. The statement of prior plan has been added in Methods (page 3, paragraph 3) in the revised version.

-please provide a definition of a SR/MA you have used

Reply 8:The definition has been added in Methods (page 4, paragraph 4, line 1 to line 4) in the revised version. Systematic review was defined as a type of literature reviews that critically appraise and formally synthesize the best existing evidence to provide a statement of conclusion that answer specific clinical questions. Meta-analysis is the use of statistical methods to summarize the results of independent studies (Glass 1976). By combining information from all relevant studies, meta-analyses can provide more precise estimates of the effects of health care than those derived from the individual studies included within a review (http://handbook-5-1.cochrane.org/).

-it is not clear whether you have only focused on intervention reviews. This is important as AMSTAR has been originally developed for SRs or RCTs

Reply 9:Yes the reviews we included were intervention reviews, the systematic reviews or meta-analyses we included was focused on the effects and safety of procedures and prostheses in primary THA or TKA. We explained it in Methods (page 4, paragraph 1, line 1 to 3)
what is the rationale for only including the last two years (i.e. 2014-2015). This is quite uncommon, why not the last 5 years?

Reply 10: 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 big, thus we choose the last two years. We think the more recent published SR could better reflect current situations.

the literature search was not conducted from January 2014 to December 2015, this is only time span for inclusion of your articles

Reply 11: Sorry to make confusion. We searched Medline and Embase from Jan. 2014 to Dec. 2015, but Cochrane database and Chinese databases on Sept. 2016 and only included reviews published from Jan. 2014 to Dec. 2015. To make clear, the flow chart has been revised.

as you refer to the PRISMA guideline, I would recommend that you follow it themselves. For example, it is not reported how the study selection was performed

Reply 12: The study selection has been added in Methods section (page 4, paragraph 3, line 1 to line 3). Meanwhile, we have further read the PRISMA guideline carefully and try to make our study meet the requirement of this guideline.

"AMSTAR is a freely accessible validated tool for assessing the methodological quality of MA"
- please revise this statement to SR instead of MA

Reply 13: Thank you for your amendment, the correction has been made in Methods section (page 4, paragraph 4, line 2).

-when referring to AMSTAR please make sure that you cite the original papers. Reference 15 where you refer to the summary score is not an original AMSTAR paper. I believe that the summary score was not mentioned in the AMSTAR development paper by Sheal et al. 2007 at all.

Reply 14: Thanks for your careful checking of the references. We read the reference again and found this reference provides no empirical evidence to advice on the appropriate weighting for each item when calculating a total score. But we found the study measure on validation of AMSTAR mentioned the sum of all items in AMSTAR scale and stated that the total score was calculated by summing one point for each “yes” and no point for “no”, “can’t answer”, resulting
In summary, scores from 0 to 11 ([https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2131785/pdf/pone.0001350.pdf](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2131785/pdf/pone.0001350.pdf)). Correction has been made in Methods section (page 4, paragraph 4, line 6 to 7).

- You state that "Our study did not change the total score and the methodological quality of the reviews was graded as high (8-11), medium (4-7) and low (0-3) quality" but it is unclear how this was calculated.

Reply 15: 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 2 was refined into 4 sub-items with each sub-item 0.25 point.

- You are talking about Chinese journals and English journals. I guess you mean publications in that language.

Reply 16: We do mean reviews published in Chinese and English. We have revised it in Methods section (page 5, paragraph 1, line 3 to 4).

- You need to expand a bit more on the statistical analysis.

Reply 17: Correction has been made in the revised version (page 5, paragraph 1).

Results:

- Inter-rater reliability instead of reliability, and please be more concrete on you measure, do you mean Cohens kappa?

Reply 18: Thanks for your professional advice, the correction has been made in the revised version (page 4, paragraph 5, line 1).

- Some more descriptive characteristics would be helpful for the reader. In particular, are all reviews interventional?

Reply 19: According to your suggestions, we have displayed more descriptive characteristics in Results (page 6, paragraph 3) in revised version.
you state "The overall median score for all 63 included reviews was 6.366±1.225". However, the ± implicates that this is a standard deviation. But this makes only sense together with a mean (not median) score!

Reply 20: In consideration of your suggestion, we had consulted the statistician and found it not proper, thus we corrected it in Results section (page 5, paragraph 3, line 1 to 2), meanwhile we adjusted the relevant data in Table 3.

- "In general, included reviews displayed better" -->better than what?

Reply 21: We deemed items of coincidence rate over 80% as displayed better, and coincidence rate less than 20% as poor, we explained it in Results section (page 5, paragraph 4, line 1 and line 5).

-how did you derive the 16 bibliographical characteristics? please ammend in the method section

Reply 22: The 16 bibliographical characteristics were derived from previous studies that suggested these characteristics may influence the quality of systematic reviews. Correction has been made in Methods section (page 4, paragraph 3, line 3 to 4).

-how did you search for the impact factor? Is it the IF of a given year or of the corresponding year of publication?

Reply 23: It is the IF of corresponding year of publication and was found in the Journal Citation Reports of Web of Science (https://jcr.incites.thomsonreuters.com/JCRJournalHomeAction.action?SID=B1-qgrlANuFaPlx2FWkSHaWSQ4Zec0qQbg8CQ-18x2dQF6rd29tnrV6ZPQtUZbfUAx3Dx3D2Qx2BhBYtpPoCkP1UavKmzAQx3Dx3D-9vvmzcndpRgQCGPd1c2qPQx3Dx3D-wx2BJQh9GKVmtdJw3700KssQx3Dx3D&refineString=null&SrcApp=IC2LS&timeSpan=null &Init=Yes&wsid=Q2C56X7OFJdt81wt9il), impact factors of Chinese journal was found in CNKI (http://navi.cnki.net/KNavi/Journal.html) which was a Chinese databases. Explanation has been added in Methods (page 4, paragraph 3, line5 to 8) in the revised version.
Discussion:

-please put your results into the context of other results investigating methodological quality of SRs. I don’t think the results of TKA/THA are worse or better than in other fields.

Reply 24: As your suggestion, we compared our results with the methodological quality of SRs in other fields. Correction has been made in appropriate part of the discussion.

-you are talking about KEY quality item - how do you define this? providing the list of excluded primary studies is not a key quality item, in my opinion

Reply 25: I agree with you that some of the items in the AMSTAR seem not so relevant to the methodological quality such as providing a list of included and excluded studies (item 5). So it is inappropriate to talk about KEY quality item. We have revised the expression.

-in general, I think the discussion (that is well structured) can much more refer to published studies investigating the methodological quality of SRs in other fields.

Reply 26: We added more information by comparing our results with the methodological quality of SRs in other fields.

-I would be more cautious on recommending a tool for observational studies such as the NOS. The authors did not investigate this. Furthermore, the NOS has also been criticized for content (https://www.ncbi.nlm.nih.gov/pubmed/20652370), and inter-rater reliability (https://www.ncbi.nlm.nih.gov/pubmed/20652370)

Reply 27: Thanks for your professional advice. Correction has made in Discussion (page 8, paragraph 3) in the revised version.

-one of the main results refers to "Reviews included literatures with English and other languages and reviews declared conflicts of interest displayed better than who did not report (7.00 vs 5.71, P=0.04; 7.37 versus 5.67, p=0.033)." However, the authors give no explanation for this finding. This leaves me with the question whether this is a relevant finding or just due to chance?

Reply 28: The explanation has been made in Discussion part (page 9, paragraph 1 to 2) as following.
- I would refrain from stating that AMSTAR has been proven to be reliable and valid as you have revised the original AMSTAR heavily!

Correction has been made in our revised manuscript.

Reply 29: In fact, the contents of mAMSTAR are exactly the same as the original AMSTAR except for refining the items that include more than one content in an 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?).

-the section limitations and strengths needs a thorough revision

Reply 30: Thanks for your suggestions, the Strengths and Limitations part has been revised thoroughly.

Although I am not a native speaker, I feel that the manuscript would benefit from language editing by a native English speaker

Reply 31: 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.

Replies to Al Mayhew:

1. The issue is a very important one and understanding the quality of systematic reviews (SRs) on TKA and THA is important

Reply: Thank you for agreeing with our work.

2. The research question is not well defined in the background. It would be helpful if the authors stated, "The purpose of this study was to...." The statement is included in the abstract, but should be expanded upon in the background.

Reply: Thanks. We have revised the background based on your suggestion (page 3, paragraph 2, line 7-10).
3. 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.

Reply: 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.

4. 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?

Reply: 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).

5. 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.

Reply: 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:


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.

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

Reply: 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).
7. Did you develop the mAMSTAR or has it been used before? Any data on validity or reliability?

Reply: 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.

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

Reply: 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.

9. 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.

Reply: 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.

10. 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.
Reply: 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:

<table>
<thead>
<tr>
<th>Bibliographical Characteristic</th>
<th>High</th>
<th>Medium</th>
<th>Low</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Published Language</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chinese</td>
<td>0</td>
<td>16</td>
<td>0</td>
<td>0.424</td>
</tr>
<tr>
<td>English</td>
<td>4</td>
<td>43</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Published Year</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2014</td>
<td>1</td>
<td>29</td>
<td>1</td>
<td>0.187</td>
</tr>
<tr>
<td>2015</td>
<td>3</td>
<td>29</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

11. Why do you think 58/63 studies were "moderate quality"?

Reply: The methodological quality of the reviews was graded as high (8-11), medium (4-7) and low (0-3). Since those 58 studies’ mAMSTAR scores ranged from 4 to 8 (not include point 8), which means they are moderate quality.

12. 16 is a lot of variables to evaluate for bibliographic characteristics. How did you choose the characteristics?

Reply: 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).
13. Studies with negative findings are not the same as grey literature. Very important point.

Reply: Thanks for your professional suggestion; the correction has been made in Discussion section (page 7, paragraph 3).

14. I had a look at your list of included studies. I thought that most of them would be about THA vs non surgery; or TKA vs non surgery. At first I was surprised, many of them are about one surfacing technique versus another or one surgical approach versus another. Now I understand why you found so many studies in 2 years of publication. I wonder if the inclusion criteria are too broad.

Reply: We included all the systematic reviews as long as the reviews followed the Cochrane’s definition of systematic review to understand the methodological quality of SR in this field.

15. 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.

Reply: 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.

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

Reply: Thanks for your professional suggestion, we have extensively revised our methods and discussion part, we hope you are satisfied with our revision.
17. 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.

Reply: 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.

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