Reviewer's report

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

Version: 1 Date: 25 Sep 2017

Reviewer: Al Mayhew

Reviewer's report:

1. I have added my comments to my original ones, I hope this is helpful. My new comments are preceded by "AM Sept 25, 2017"

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).

AM Sept 25, 2017 - Improved, thank you.

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.

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"

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.

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.

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.

AM Sept 25, 2017 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)

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).

AM Sept 25, 2017 - Improved, thank you. Can you write out titles of databases in full?

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.

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?

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.

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.

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.

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

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:

AM Sept 25, 2017 See points number 7 and 8

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.

AM Sept 25, 2017 Thank you, this is clear now.

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).
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.

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).

AM Sept 25, 2017 Thank you, this is clear now.

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.

AM Sept 25, 2017 Thank you, no revision required.

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.

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

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.

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

AM Sept 25, 2017 Additional comments:

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

19. 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.

20. 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

**Are the methods appropriate and well described?**
If not, please specify what is required in your comments to the authors.

Yes

**Does the work include the necessary controls?**
If not, please specify which controls are required in your comments to the authors.

Yes

**Are the conclusions drawn adequately supported by the data shown?**
If not, please explain in your comments to the authors.

Yes

**Are you able to assess any statistics in the manuscript or would you recommend an additional statistical review?**
If an additional statistical review is recommended, please specify what aspects require further assessment in your comments to the editors.

I recommend additional statistical review

**Quality of written English**
Please indicate the quality of language in the manuscript:

Not suitable for publication unless extensively edited
Declaration of competing interests
Please complete a declaration of competing interests, considering the following questions:

1. Have you in the past five years received reimbursements, fees, funding, or salary from an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

2. Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

3. Do you hold or are you currently applying for any patents relating to the content of the manuscript?

4. Have you received reimbursements, fees, funding, or salary from an organization that holds or has applied for patents relating to the content of the manuscript?

5. Do you have any other financial competing interests?

6. Do you have any non-financial competing interests in relation to this paper?

If you can answer no to all of the above, write 'I declare that I have no competing interests' below. If your reply is yes to any, please give details below.

'I declare that I have no competing interests'

I agree to the open peer review policy of the journal. I understand that my name will be included on my report to the authors and, if the manuscript is accepted for publication, my named report including any attachments I upload will be posted on the website along with the authors' responses. I agree for my report to be made available under an Open Access Creative Commons CC-BY license (http://creativecommons.org/licenses/by/4.0/). I understand that any comments which I do not wish to be included in my named report can be included as confidential comments to the editors, which will not be published.

I agree to the open peer review policy of the journal