Reviewer’s report

Title: Reliability of patient-reported complications following hip or knee arthroplasty procedures

Version: 0 Date: 31 Jul 2018

Reviewer: Reviewer 2

Reviewer's report:

PEER REVIEWER COMMENTS: To view the full report from the academic peer reviewer, please see the attached file.

REVIEWER COMMENTS FROM REPORT: Overall impression: This is an important area of study. However the data collected, the design of the study and analysis plan and execution have major flaws. As a result, I am cautious with the conclusions drawn and would recommend that the authors reconsider the steps taken in designing the study to answer this question.

What the authors have done well: This is an interesting paper, most notably since various core outcome set recommendations, such as OMERACT, have stressed the need to collect data on adverse events and complications, with patient-reported measures and data collection approaches being the currently advised approach for registry findings. Through this, the paper is appropriate in its rationale and basis. Table 3 - this is very helpful and clearly presented. Thank you.

Ways it does not meet best practice: (A) As stated below, the recruitment of 3 surgeons who only reviewed patients at 6 weeks does not answer the over-arching question. (B) There are no form of veracity checking against the medical notes and therefore it is not clear what issue this has on the 'gold standard' comparisons (C) The analysis includes all complications and the analyses should be divided by severity of complication as I anticipate that this will provide very different outcomes.

REQUESTED REVISIONS:

Minor Issues
Abstract - there are numerous typing/spelling errors. For example…what is the 'medical les' is 'les' a typo? Specificity is incorrectly typed; non-significant differences in the Results Abstract section - please check and revise throughout.

Major Issues

Abstract - I am not sure the results section is sufficient. This should be revised when the paper is considered as a whole but I would anticipate a difference in reliability based on the type of complication reported. To analyze a PE with a superficial wound infection as equal complications feels somewhat inappropriate. I think the analysis should be stratified by severity of complication. This will change the abstract's results section.

Abstract - "Analyses performed categorizing patients by surgeon, joint operated and time between surgery and follow-up revealed non-significant differences between these groups" - I am not sure whether this level of detail is really needed in this paper. Answer the main question first and then consider this secondary questions. Ultimately the results are most probably underpowered therefore the non-significant differences may be attributed to type 2 statistical error and therefore erroneous. Please consider.

Abstract - Conclusion - this is the correct conclusion. The sensitivity detects how good a test is (PROMS in this instance) for detecting a complication when one exists. The specificity is how good a test is for detecting when one does not exist… the results (based on the data), indicate that patients are good at detecting the absence of a complication (0.98) but not good a reporting the presence (0.14). HOWEVER, I think the data should be analyzed by type of complication rather than complication per se. I suspect that this will radically change the results and interpretation. I would recommend the authors consider this, even if just by analyzing my major and minor complication. This is important for surveillance work and long-term monitoring. This will also help contextualize the recommendation on 'greater attention to the clarity of the questions asked' as we will have a much better idea on what questions should be asked then.

Background - Page 1 line 54 - Please provide some indication on what is 'relatively low' in complication rates. This is important to provide the reader with context.

- Page 2 line 12-15 - Please provide data on sensitivity and specificity so that the reader can start to get an ideas on what the comparative data is compared to this paper's results.
- Page 2 line 24 - The reasoning for doing this study is not clear. Please tell the reader explicitly why we need this analysis and what this analysis provides over previous studies [3-15 from the paper's reference list].

- Page 3 line 41 - Please provide a sample size calculation to provide a justification for this cohort of 'approximately 50' per surgeon. There is limited clarity on this and a risk of selection bias. How were the 6 surgeon's selected? This would be useful to know to aid the transparency on the cohort's identification.

- Page 3 line 58 - I do not think that the subgroup analyses are helpful. I think analyzing by joint (THA or TKA may be useful if the authors think that the pathway and process is different for these joints (they are not in my clinical experience from a surveillance perspective)). The time between surgery and follow-up should not be substantially different and if it is, I would suggest the authors consider whether this should be an exclusion criteria. It is hard to ascertain whether the 'attending surgeon' should be an important factor as we have insufficient information on them and the process to which they work. I think analyzing by type of complication would be much more valuable and would be more likely to provide differential results. Analyzing by re-admission, reoperation and other complication would be a start. I would also suggest re-analyzing by minor versus major complications. Page 3 lines 9 to 19 highlights a number of complications which the authors can make a clinical judgement on the severity of complication. This may be grouped into two or three different categorises. Please consider analyzing the PPV, NPV, sensitivity and specificity by such groupings. I think this will provide a more insightful outcome.

Results - Page 4 line 6 - I do not think that this is really a random sampling process at all. The researchers could select the time-scale, the surgeon's involved and potentially, as not clearly reported, the participants actually included. This needs clarifying.

Results - Page 4, line 7 - how can a participant (potentially 24) be included if they did not review a surgeon review which formed the basis of the comparison to PROM. Please consider the reporting and/or eligibility criteria.

Results - Page 4 Line 12 - I would strongly suggest excluding Surgeons A, B and E as the follow-up time and therefore recall error is potentially different to Surgeons C, D and F…this is a major confounder which massively impacts on the findings. Please re-consider this approach and identify alternative surgeons or include the cohort from Surgeon C, D and F to reach
sufficient participants to meet the sample size. Acknowledging this as Limitation 2 (Page 6, Line 60) is insufficient - something needed to be done to prevent this from happening and I think not recruiting from Surgeon A, B and E would be advantageous in this instance given the impact this could have on the study's conclusions.

Results - Page 4 line 30-40 - I am cautious about the interpretation of the results because of the grouping issues as stated above.

Results - Page 4 Line 40-47 - I am cautious about the interpretation until a power calculation has been presented, as stated previously.

Discussion - I am concerned about the interpretation because of the analysis approach adopted. Please find comments above.

Discussion - the interpretation is reasonable however I would recommend the authors reflect on the qualitative literature on reporting, expectations and outcomes which will help support their assumptions on patient recovery and complications in Page 5 and 6.

Discussion - Limitations (Page 6 Line 53-60) - I do not think the authors made a sufficient account for the major limitation in this study which is around the 'gold standard' comparison as 'surgeon records' (Page 3 line 53). This is major as there may be issues of poor reporting of medical complications, particularly minor complications which may have cause some discrepancy in reporting of complications. Furthermore, what is the risk of non-reporting complications to the surgeon's hospital but self-managing or seeking alternative medical care such as from a community physician or other orthopaedic surgeon? Again, this may provide a discrepancy which should be addressed in the paper's limitation section but also provides a reason for categorising analyses by different levels of severity of complication. Lesser complications may not require hospital consultation or may have greater risk of poor recall than more memorable major complications which need urgent medical attention.

Conclusion - I am cautious due to the methodological issues I have raised earlier. The authors need to re-format the data and then re-analyze. They may also need to half their sample due to
issues in their choice of surgeons. The authors may also need to consider how the data was collected. These are fundamental issues.

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

No

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

No

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

No

**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.

Not relevant to this manuscript

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

Needs some language corrections before being published

**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 organization 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 organization 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.

This reviewer has been recruited by a partner organization, Research Square. Reviewers with declared or apparent competing interests are not utilized for these reviews. This reviewer has agreed to publication of their comments online under a Creative Commons Attribution License attributed to Research Square and was paid a small honorarium for completing the review within a specified timeframe. Honoraria for reviews such as this are paid regardless of the reviewer recommendation.

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