Reviewer’s report

Title: Challenges of one-year longitudinal follow-up of a prospective, observational cohort study using an anonymised database: recommendations for trainee research collaboratives

Version: 0 Date: 09 Sep 2018

Reviewer: Segun Bello

Reviewer's report:

Major comments

- I cannot understand the justification for excluding patients who died within 30 days of index surgery from Surgery. Authors should as a minimum emphasize the included population in the abstract and the methods section i.e. patients were included provided they were alive 30 days post index surgery. Otherwise, the paper is totally misleading.

- The second paragraph of 'outcome measures' is difficult to link anywhere in the manuscript. What message are the authors trying to pass across? It is not linked to the primary and secondary outcomes. It is confusing.

- The use of the term 'registered' might be reconsidered as it appears to depict willingness to participate which is not in line with the language in this manuscript. Suggest use 'were enrolled' alongside reasons for non-enrolment

- Is it possible to give a background description of REDCap system in the methods?

Minor comments

- Abstract line 51-59: 'factors associated with increased likelihood of achieving increased likelihood of achieving ≥95% data completeness were total number of patients to be followed-up, and central versus local storage of patient identifiers (72.5% vs 48.0%, respectively, p=0.006)' here two factors were listed but only one set of numbers (p-value) was provided. It is thus confusing which factor this belong to. Authors should add the number or at least the p-value for the other factor to make it more explicit

- What is the difference between data completion rate and follow-up rates. Authors should introduce this earlier in the introduction and why it is important major

- Methods line 17 'studies' should read 'student'

- Outcomes collected at one year were listed but no justification for these outcomes were apparent in the manuscript. Authors should clarify in the methods section why these outcomes are important.
- Line 1 statistical analysis, delete 'were'

- Statistical analysis; ………. Fisher's exact modification if group sizes were less than five…. Replace 'group sizes' with 'expected cell counts'

- Under the section follow-up rates, what is the rationale for including patients from 47 centres that were not enrolled in the study? These patients are not part of the denominator and should not be added in the calculation of percentages.

- Last sentence under characteristics of patients followed-up at one-year, '….. there were no significant different…….' Should read '……..there was no significant difference….'

- Data completeness need to be defined in the methods

- Investigator feedback survey; how many were contacted and what was the response rate?

- Discussion line 4; '……..there was no evidence of significant systematic bias in patients being followed-up' delete 'significant. It's either there is evidence or there is no evidence

- Discussion last line of first paragraph; 'validation' should read 'validating'

- Discussion (page 14) Locate this sentence 'In the OAKS study, there were no significant differences in the patient-level demographics, operative indications or ASA grades between the group that underwent one-year follow-up and those that did not' What population is referred to here? Does this also include those patients who died during the first 30 days post-surgery?

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 am able to assess the statistics

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

Acceptable

**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.

None declared

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