Reviewer's report

Title: Addressing identification bias in the design and analysis of cluster-randomized pragmatic trials: a case study

Version: 1 Date: 29 Oct 2019

Reviewer: Beatriz Goulao

Reviewer's report:

Thank you for your responses to my comments. I think the article is now easier to read and the framework presented helps understand the steps taken to address potential bias.

I still think the article could be improved in terms of being more focused. For example, the introduction focuses on several aspects related to pragmatic trials that could potentially be presented in a more succinct way. It has a paragraph describing the case study, however this info appears again in the methods section. Specifically, the information towards the end of page 6 reappears in page 9 which is unnecessary. Also in this section, the authors state there is a possibility of selection bias (page 6, line 27). Before that and after, the authors consistently refer to identification bias, not selection bias. I would suggest being consistent across the article. If you meant to distinguish those types of biases (selection vs identification) that needs to be explicitly stated and explained. From reading the paper, they sound like the same bias in this context.

I think it would be helpful to add some information about usual ways to avoid identification bias in cluster trials and why they can't be used here. The authors have done that in different parts of the paper, but I think a few sentences on this in the introduction it would make for a stronger case regarding the importance of the topic and what the paper brings that is new.

It is still unclear why the intervention might lead to a change in the process by which the data gets recorded - I would make this more clear. The way the intervention has been described it looks like it would be more likely for patients to be diagnosed because of the nurse assessment? How is that related to the process by which the data gets recorded?

Framework (methods)

Step 0 - identifying susceptibility to identification bias

I agree this is an important first step to consider and might help other researchers think through their trials and this possibility. However I found the writing confusing in particular the second condition presented. When would the study objective not be estimating treatment effect among an identified population?
The authors state note that the analytical sample is not the same as the sample being offered care from the PROUD nurse. I don't understand this. Would that not depend on which analytical sample you decide to use according to the framework?

Step 2 - Choice of unbiased analytic sample

The authors mention a "power analysis" but there is no detail on how this was done. In the results section, it becomes clear that two approaches (one through formulae and another using simulation) were implemented. This needs to stated in the methods section with details on which formulae was used, what software and the details about the simulation conducted. More information is also needed on what the "high sensitivity" and "high specificity" scenarios actually mean in the methods (and not the results).

Results

Figure 2 - are these results calculated using a formulae or simulation? This needs to be explicit.

Level of interest
Please indicate how interesting you found the manuscript:

An article whose findings are important to those with closely related research interests

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

Acceptable

Quality of figures
All images and figures within the manuscript should be genuine i.e. without evidence of manipulation. No specific feature within an image may be enhanced, obscured, moved, removed, or introduced. If you have concerns about the veracity of the figures you should choose the first option below.

Statistical review
Is it essential that this manuscript is seen by an expert statistician? If so, please give your reasons in your report.

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 organisation 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 have no competing interests to declare

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

Were you mentored through this peer review?

No