Reviewer's report

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

Version: 0 Date: 04 Sep 2019

Reviewer: Beatriz Goulao

Reviewer's report:

The article is interesting and raises an important source of bias in cluster randomised trials explaining the process the authors to minimise it. However, I think it could be more helpful if it was more succinct, leaving some details about the main trial out and referring to the protocol, and the decision making process was more clearly structured. I would also prefer if the authors used fewer acronyms in their text, which made the reading harder.

Specific comments

- It is unclear who will identify the potential trial participants and why the possible solution of having an independent person doing that is not considered in order to minimise selection bias.

- If the main reason for differences in case detection are related to having the intervention available, then should the offer of a delayed intervention be considered?

- Both table 1 and 2 are helpful, but I could not understand why scenarios where everyone in the practice is included were even being considered. If the intervention is targeting a certain disease, why would you consider offering it to everyone? What if participants without the disease accepted the intervention? Is it ethical to offer an intervention to people that do not have the condition?

- I don't understand if the analytical sample is the sample being included in the trial or the sample analysed? These are not necessarily the same due, for example, to drop-out.

- After table 2, scenarios are brought up again (page 15 of the manuscript), but they now use numbers instead of the letters used in the table which is confusing. The possibility of identifying potential participants with the disease in advance by using specific variables is presented in the text but not in the scenarios table 2. Why is that?

- I thought table 3 was not essential and should have been an appendix, whereas presenting the different power results would have been helpful as part of the main text.

- Some of the detail about the sample size calculation, ie assuming the clustering would influence the different scenarios in the same way and therefore was not initially considered should be in the methods. The methods should outline the decision rules used, ie scenarios were generated
and we used power calculations to decide on the best design to use in terms of pool of eligible individuals.

- Page 18/19 outlines potential strategies to include post-randomisation identified individuals in the analysis. Even though the considerations are interesting, these are not results per se and feel more like a statistical analysis plan. I am not sure they belong in the results section and should perhaps be excluded with some details in the methods and/or discussion, ie deciding the best way to design the trial is just one part of the problem, this is what we plan to do next.

**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:

Not suitable for publication unless extensively edited

**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 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 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