Reviewer’s report

Title: A protocol for a cluster-randomized controlled trial testing an empowerment intervention to prevent sexual assault in upper primary school adolescents in the informal settlements of Nairobi, Kenya

Version: 0 Date: 03 Oct 2018

Reviewer: Daniel Tancredi

Reviewer's report:

Peer Review of PUBH-D-18-02455 "A protocol for a cluster-randomized controlled trial testing an engagement intervention to prevent sexual assault in upper primary school adolescents in the informal settlements of Nairobi, Kenya"

Summary: This is a clearly written description of a generally sound protocol for a sexual assault prevention intervention trial that aims to fill an important evidence gap in the literature, the need to develop and evaluate interventions for adolescents in urban settlements in low and middle-income countries. This reviewer has a small number of concerns for the authors to address.

Moderate concerns:

1. Page 8, lines 13-14. The authors report that schools that refuse to participate at baseline will be replaced with a different school, but then a couple of paragraphs later, in the "Randomization" section, they describe that randomization has already been performed. To preserve the statistical equivalence of the intervention and control schools, one would want to avoid randomizing schools until after ensuring that they would participate. Was that done? If so, then a simple revision would be needed, which would be to change the tense of the sentence in lines 13-14 to say that schools that refused to participate were excluded prior to randomization.

If schools can drop out after they have been randomized, then the authors will need to present an approach for dealing with that, including a sound rationale. This is particularly important given that matched pair design that the authors used. I recommend that the authors look closely at their reference #27, the 2009 Statistical Science paper by Imai et al, where the issue of drop-out clusters is extensively discussed, because the potential drop-out of a single cluster from a matched pair is something that can complicate the analysis. Replacing drop-out schools with other schools may not be desirable in regards to preserving the internal validity of the
randomized design. In any case, a rationale presented (and perhaps sensitivity analyses should be proposed) for dealing with drop-out schools.

2. Page 13, line 38. The authors should specify the regression framework that would be used for the instrumental variables analysis, picking one of the three options in Reference 36. I would guess that the authors intend to use the structural nested models framework, for example, but there are two other frameworks that could be used.

3. Page 14, line 48. The units of analysis in this study are participant surveys, but for any given outcome, it would appear that only male or only female participants will be involved, so the authors may wish to add this detail.

4. Page 13, line 55. Although a "difference-in-difference" estimand could potentially be estimated from a model that used a logistic-link function, using predictive margins for example, that would not be the natural thing to do. I suspect that the authors may have intended to say they would do, say, a between-group ratio of post:pre odds-ratios analysis. That's what would result from exponentiating the "time*arm" logistic regression coefficient to express it as a odds-ratio. Alternatively, the authors could retain the "difference-in-differences" phrasing, but it would need to be applied to the "log odds" of sexual assault, not the "rates", as described later in that sentence.

5. Page 14, line 45-46. The authors say that only complete-cases will be available for analysis because participants who exit the study will be lost, but that is questionable. Wouldn't they still have provided the baseline survey, which could potentially be analysed? Please revise this sentence to make it clear what the rationale is for dropping data from drop-outs. Also, please clarify whether unique identifiers for individuals are available for this study. I presume they are, but it's not clear.

Minor issues:

6. Page 6, line 28: Should "adolescents and urban settlements" be changed to "adolescents in urban settlements"? It seems from the nicely presented literature review that the gap is with respect to adolescents in urban settlements.

7. Page 9, line 27. The 40% decrease appears to be for the "odds", not the "odds-ratio". In other words, it appears the authors are talking about an odds-ratio of 0.60, which corresponds to the odds being reduced by 40%.

Suggested edits for style:

Page 8, line 28: Change "which" to "that".
Page 9, line 2: Change "among large schools" to "in large schools," (note addition of comma).

Page 11, line 1: Change "a ratio of approximately 1:15 students" to "a trainer:student ratio of 15".

**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.

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