Reviewer's report

Title: Stepped wedge randomised controlled trials: systematic review of studies published between 2010 and 2014

Version: 2 Date: 18 March 2015

Reviewer: Noreen D Mdege

Reviewer's report:

I thank the Editors for the opportunity to review this manuscript. I found it very interesting to read and it is well written. I have a number of comments that could help improve the manuscript:

Major Compulsory Revisions

Methods:
One author reviewed the titles and abstracts. Why one and not at least two? This also has to be picked up in the limitations section in terms of the implications of this on the findings.

Categorization of interventions using the behaviour change wheel and coding of behaviour change techniques: Who performed this and how was it done? Was it done independently in pairs? What was the level of agreement within pairs? How were disagreements resolved? If not done independently in pairs- what are the implications on the findings?

Discussion:
There are a number of interesting issues that were raised in some parts of this manuscript that should have been picked up in more detail in the discussion section. Some of them have been highlighted already. Here are others:

- I would have expected more discussion on the controversies around the use of this design. This is very important because even with such controversies and concerns around reasons for using the design, these reasons seem not to have changed and the design seems to even be getting more popular. In their article Kotz et al actually argue that the use of this design should not be recommended; there was a response to that of course which I think would be relevant in this discussion (Mdege et al 2012. There are some circumstances where the stepped-wedge cluster randomized trial is preferable to the alternative: no randomized trial at all. Response to the commentary by Kotz and colleagues. JCE, 65(12):1253-4). It is important for the author to discuss their findings within the context of this current debate, and what the way forward could be.

- Avoiding disappointment effects is one of the reasons cited for using this design. Is this reason supported by any empirical evidence in any of the articles? How does this fit with what the experience has been with cluster randomised trials in general?
- ‘Cannot implement everywhere at same time’ is referenced for a number of studies in Table 1. Did the studies actually say this because with any other experimental design you would not implement everywhere at same time. I would have expected something like ‘cannot implement in many clusters at same time’ or something along those lines.

Minor Essential Revisions

Results:
Table 1: The column study start date is confusing, particularly for research articles. Please use publication date so it is easy to relate to your time restriction of 2010-2014 and your reference list.

Sample size calculation: Currently in read as if all studies, regardless of whether they were full articles, conference abstracts or trial registrations, were assessed on reporting sample size calculation. This is okay for results papers and protocols, but I do not think assessing this is relevant for conference abstracts or trial registrations.

How many studies accounted for the stepped wedge design? Out of these how many used Hussey and Hughes' approach?

General comments:
There are also a number (very few though) of typos in the manuscript.
I hope the authors will find these comments useful.

**Level of interest:** An article of outstanding merit and interest in its field

**Quality of written English:** Acceptable

**Statistical review:** No, the manuscript does not need to be seen by a statistician.

**Declaration of competing interests:**
I declare that I have no competing interests