Author's response to reviews

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

Authors:

Emma Beard (e.beard@ucl.ac.uk)
James J Lewis (James.Lewis@lshtm.ac.uk)
Andrew J Copas (a.copas@ucl.ac.uk)
Calum Davey (calum.davey@lshtm.ac.uk)
David Osrin (d.osrin@ucl.ac.uk)
Gianluca Baio (g.baio@ucl.ac.uk)
Jennifer A Thompson (Jennifer.Thompson@lshtm.ac.uk)
Katherine L Fielding (Katherine.Fielding@lshtm.ac.uk)
Rumana Z Omar (r.omar@ucl.ac.uk)
Sam Ononge (ononge2006@yahoo.com)
James Hargreaves (James.Hargreaves@lshtm.ac.uk)
Audrey Prost (audrey.prost@ucl.ac.uk)

Version: 3 Date: 29 May 2015

Author's response to reviews: see over
Dear Professor Altman, Professor Furberg, and Professor Grimshaw

RE: 1677419149162212 - Stepped wedge randomised controlled trials: systematic review of studies published between 2010 and 2014

Thank you for yours and the reviewers’ comments on our manuscript titled “Stepped wedge randomised controlled trials: systematic review of studies published between 2010 and 2014”. We found them to be extremely helpful and have addressed the comments below and in the manuscript. We look forward to hearing from you.

Yours Sincerely

Emma and James (joint first authors), on behalf of all authors

Reviewer 1:

1. 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.
   
   We agree that it is advisable to always have at least two reviewers selecting the papers for a systematic review. EB, as part of an MSc project, re-ran the search and double checked that all eligible papers had been identified. This was previously not stated in the paper. We now include the following on page 6: “EB then re-ran the search to double check that all eligible papers had been identified between 1st of January 2010 and 14th May 2014”

2. 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?
   
   We have now included further details in the paper on the coding of behaviour change techniques and intervention functions. EB followed standardised protocol as specified previously (Michie et al, 2012). These details were excluded in the original submission due to word count. The following has been added on page 6: “EB coded all interventions using these nine functions. A subset of eight papers was also second-coded by a researcher familiar with the BCW, until 90% agreement was obtained. Any discrepancies were resolved through consensus discussions.”

3. 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. 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.

We were aware of these papers and the debates surrounding stepped wedge designs, and address them briefly on pages 4 and 5. These issues are discussed in more detail in the design paper and sample size papers which will make up the special issue of trials. We do not include more detail here as the aim was to systematically review trial content and we want to avoid repetition and substantial overlap where possible. However, we have now added the following sentence on page 5: “These issues are discussed in more detail in the other papers which make up this special issue of trials [15-17]”; and have added some debate to the discussion on page 10: “The reasons for using a stepped wedge design largely coincide with those reported previously: ethical, logistical and methodological [5, 4]. The potential impact of disappointment effects, whereby individuals not randomised to the treatment of choice fail to adhere or drop-out, was given by several studies as a reason for choosing the SWT design (Table 1). However, some authors argue that this is not an inherent feature of SWT, and that cluster RCTs can be extended to include a wait-list control group [13]. Thus the ethical argument that one should not withhold a potentially effective intervention from a group of individuals cannot form the sole justification for this trial design. It is possible that under certain circumstances, including the roll out of public health interventions, that a SWT would reduce required resources. One could easily envisage the situation of an intervention conducted by GPs, which would require one intervention trainer for a SWT (each GP is visited consecutively) and multiple for a cluster RCT (each GP is trained concurrently). SWT may also be suitable for optimising interventions, with the ability to modify content and delivery over time. However, the excess expense of this over factorial designs should be considered [64]. Finally, although it is possible under certain circumstances that the SWT is optimal in terms of power, due partly to the within- and between-cluster data, this is not always the case [17, 14].”

4. 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?

We have now added a paragraph to the discussion on page 10 which talks about disappointment effects in relation to SWT and cluster randomised trials: “The reasons for using a stepped wedge design largely coincide with those reported previously: ethical, logistical and methodological [5, 4]. The potential impact of disappointment effects, whereby individuals not randomised to the treatment of choice fail to adhere or drop-out, was given by several studies as a reason for choosing the SWT design (Table 1). However, some authors argue that this is not an inherent feature of SWT, and that cluster RCTs can be extended to include a wait-list control group [13]. Thus the ethical argument that one should not withhold a potentially effective intervention from a group of individuals cannot form the sole justification for this trial design.”
We could not find any empirical evidence in the cited articles for this worry of disappointment effects.

5. ‘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.

This was just the wording that we used. We meant in each cluster, and so have now changed ‘Cannot implement everywhere at same time’ to ‘cannot implement in many clusters at same time’ in Table 1.

6. 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.

We feel that it is useful to include the start date of the intervention, but have now also included the publication date for the intervention in the same column of table 1.

7. 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.

We agree that a number of the sections are not relevant for conference abstracts or trial registrations. In these instances the coder would report that the coding was not based on a full published protocol or paper and put N/A in the appropriate boxes. We have now included a sentence on page 6 which addresses this: “Additional, sections were collated by authors of the other papers in this special issue of trials [15, 17, 16]. For conference abstracts or trial registrations, a number of these sections were not relevant and were coded as ‘not applicable’.”

However, one trial registration did report their sample size calculation and we thought this should be mentioned. Hence, we have not excluded trial registrations or conference abstracts from this.

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

The answers to these questions are addressed in the sample size paper which makes up this special issue of Trials. The review addresses the general issues around reasons for the stepped wedge design and overall methodological issues. The sample size paper also addresses the method of Hussey and Hughes etc, so we thought it best this question was addressed there.

9. There are also a number (very few though) of typos in the manuscript.

Thank you. We have re-read the manuscript and made a few changes.

Reviewer 2:
1. Not clear if the literature search in the methods section applies to the "collated methodological" studies, or if these were collected on a more adhoc basis? Possibly not relevant to this article in the series, unless the methods are being described here and not in the other review articles?

   The term collated is slightly confusing. What we mean is that we reviewed papers which had written about the analysis of, design of, or sample size calculations for SWT. These studies were then summarised in the sample size, analysis and design papers, and also included in this special issue of Trials. We have now included two sentences on page 5 to make this clearer: “Some of these articles were identified through the formal literature search detailed above, others by checking the reference lists of identified articles. These are reviewed in the other publications of this special issue of trials [17-19].”

2. It might worth stating the year, 2007, that Hussey and Hughes, and Moulten et al were published in the sentence.

   Thank you for this suggestion, we have now include the year in the first sentence on page 4 which references these papers.

3. First sentence page 6 has extra full stop.

   Thank you for spotting this mistake, we have now corrected this (now page 8).

4. There are certain references I'd expect to see in the background that are missing if this was an over-arching background to the series, or are they stand-alone articles. For example, possibly worthwhile to cite Medege at al's response to Kotz [13] and Hemming and Girling's letter addressing lack of clarity in Woertman [10]. (Just suggestions). Mdege, N.D., Man, M.-S., Taylor nee Brown, C.A., Torgerson, D.J. 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 (2012) Journal of Clinical Epidemiology, 65 (12), pp. 1253-1254 Hemming K, Girling A. The efficiency of stepped wedge vs. cluster randomized trials: stepped wedge studies do not always require a smaller sample size. [Letter]. J Clin Epidemiol 2013;66:1428-8. [in this issue]Hemming's 2014 sample size paper is worth a mention even though it is Stata specific, very comprehensive (can be found at http://www.stata-journal.com/article.html?article=st0341)

   Thank you for these references. We have now included a couple of sentences on the Mdege et al response in page 5: “Mdege and colleagues have subsequently agreed with many of these arguments, however, they have also pointed out that although they hold for the evaluation of health care treatments, they do not generally hold for policy-type trials for which alternative is often no randomised trial [15].” We have also referenced the paper by Hemming and the STATA paper in the sample size paper.

5. Perhaps highlight in additional file 1 which items were relevant for this review on the form?

   We now refer to the exact sections of additional file 1 on page 5: “The current paper focuses on objectives 1-3 (Supplementary appendix 1 sections: 1.1-1.3, 2.1-2.3, 2.6-2.8, 3.1, 3.2, 3.5-3.9, 3.12-3.14, 5.1-5.3). Objectives 4-6 are considered in more detail in the other articles in this issue of Trials.”
6. Page 6 para 3, 7 studies said SWT had higher power, did you make any attempt to verify this claim, but perhaps more relevant for the Baio [59] paper?

_This is addressed in the sample size paper. We have now included a sentence on page 5 to refer to this: “These issues are discussed in more detail in the other papers which make up this special issue of trials [17-19].”_

7. How many studies were published results of Mdege or Brown protocols? Linked to this, is it possible to graphically summarise the increase over time (say articles and protocols per year) to re-enforce the increase over time?

_Four papers of published results were included as protocols in the Mdege et al review, one of which was included in the Brown et al. review as a protocol as well. We now include the following sentence on page 7: “Four of the published papers had previously been included as published protocols in the Mdege et al review [5].” We now include a figure (figure 2) which shows the rates of publications per month over the four years. Per month was used as year 2014 only contributed less than 5 months of data._

8. Page 8 para 1, the "two-way cross-over design" is mentioned for the first time here as an alternative to SWT or parallel groups trial, might it worth a sentence elaborating on this design? (http://www.systematicreviewsjournal.com/content/3/1/86 Andrew Forbes and collaborators have a protocol for a SR on these just as an aside).

_We now include a reference to this paper on page 9 and the following text “(a design which randomises half the clusters to intervention and half to control for the first half of the trial, at which point they switch condition until the end of the trial [63]).”_

9. I know of one SWT trial that I thought would be included but was not (although only a conference abstract and ISRCTN protocol available to describe it). Perhaps excluded for some reason? http://www.trialsjournal.com/content/14/S1/P142

http://www.isrctn.com/ISRCTN11640515

_Thank you for spotting this clinical trial. Our review did not identify this and so it was not included. This is spoken about as a potential limitation in the discussion (paragraph 2 of the discussion; page 9)._
There a multitude of possible classification systems out there. The BCW actually can be used for all types of interventions, including legislation, regulation, fiscal measures and marketing. Each of the nine functions can then be used to decide how the intervention will work e.g. training, coercion or education. The BCTs do not have to just be used to code behaviour change interventions. The reason for adopting this model, was that one of the authors had been trained in its use and was experienced in its application and thus aiding the reliability of the coding. It also provides a coherent framework. We have expanded the descriptions on pages 6 and 7 to make this clearer.

2. Asymmetric requirement to report reason for using, but not for not using SWD?
   Please see response to reviewers suggestions above. We have now included a more through overview of the pros and cons of adopting a SWT design. However, we do not go into large amounts of detail as this is addressed in the other papers which contribute to this special issue of trials. We do not view this “asymmetric” requirement as problematic, there is now such a diversity of trial designs that we do not expect an author to explain why they did not choose each design, rather to just justify their chosen design.

3. Only one reviewer to identify studies. Oh well, one vigilant reader is better than two sloppy ones.
   We should have elaborated in the first paper. Prior to the final submission, a second author double checked that all eligible papers had been identified since the 1st of January 2010 and 14th May 2014. We now refer to this on page 6: “One author (AP) reviewed the titles and abstracts of all identified research articles, conference abstracts, protocols, and trial registrations to decide on eligibility for full review. EB then re-ran the search to double check that all eligible papers had been identified between 1st of January 2010 and 14th May 2014”.

4. Usual to include some sort of PRISMA diagram, but perhaps not for a methodological review.
   We do provide PRISMA diagram (see Figure 1).

5. Graph of use vs. time?
   We have now included a graph of rate of publications versus time (see figure 2).

Editorial requests:

1. Please include the email addresses of all authors on the title page.
   We have now included the email addresses.

2. Please rename the Introduction, 'Background'.
   We have made this change.

3. Please include a Conclusions section as the last section of the text, after the Results and Discussion. This should state clearly the main conclusions of the research and give a clear explanation of their importance and relevance. Summary illustrations may be included.
We have now included a conclusion section which states: “This article aims to update previous systematic reviews on SWT, consider what interventions were tested and the rationale given for using a SWT. The popularity of stepped wedge trials was found to have increased since 2010, predominantly in high-income countries. However, many were poorly reported and thus there is a need for further guidance on the conduction and reporting of SWT.”

4. Please include your funding information in an Acknowledgements section at the end of the manuscript, before the reference list.
   We now include a separate funding section.

5. Please upload the figures as separate files via the online submission system. They should not be included within the main manuscript document.
   This has been done.

6. Please include a figure and title legend section after the reference list.
   This has been done.

7. For your additional file(s), please ensure that you list the following information after your reference section in your manuscript:

   Additional files:
   File name (e.g. Additional file 1)
   ? File format including the correct file extension for example .pdf, .xls, .txt, .pptx (including name and a URL of an appropriate viewer if format is unusual)
   ? Title of data
   ? Description of data

   Additional files should be named "Additional file 1" and so on and should be referenced explicitly by file name within the body of the article, e.g. 'An additional movie file shows this in more detail [see Additional file 1]'.

   This has been done.

8. For research articles, Trials requires the submission of a populated CONSORT checklist and flow diagram. If appropriate, please provide the flow diagram and checklist as additional
files (http://www.trialsjournal.com/authors/instructions/research#preparing-additional-files). A Word file of the checklist and flow diagram can be downloaded here: http://www.consort-statement.org/consort-statement/

This is a review article and so a CONSORT diagram would not be appropriate.