Author's response to reviews

Title: Designing a stepped wedge trial: three main designs, carry-over effects and randomisation approaches

Authors:

Andrew J Copas (a.copas@ucl.ac.uk)
James J Lewis (James.Lewis@lshtm.ac.uk)
Jennifer A Thompson (Jennifer.Thompson@lshtm.ac.uk)
Calum Davey (calum.davey@lshtm.ac.uk)
Gianluca Baio (g.baio@ucl.ac.uk)
James Hargreaves (James.Hargreaves@lshtm.ac.uk)

Version: 3 Date: 9 June 2015

Author's response to reviews:

Response to Reviewers

We thank Dr Kanaan for her thoughtful review and now list our responses to each of her comments.

Reviewer's report:

This manuscript is one of a series of articles under review looking at issues related to the stepped wedge trial (SWT) design. It identifies three main designs; however, its main focus is on closed and open cohorts designs. It also discusses potential carry-over effects and randomisation approaches. This manuscript extends the recent attempt of Hemming et al (2015) to establish a framework for SWT designs.

Major Compulsory Revisions

1. The paper focuses mainly on the published SWTs post the systematic review of Mdege et al (2011). However, given that the manuscript attempts to add to the general framework of reporting of SWTs, it would be of interest to include those as well in the results.

Response

We carefully considered this suggestion but we feel that our article is primarily a methodological article that sets out design principles for SWTs, rather than a review article. Whilst we do report the number of trials that follow each of 3 main designs on the basis of a recent review (see response to next comment), we do not feel that this quantification is central to the paper, and we are also aware that the relative frequency with which each design is used may change over time. In the future there may be further designs used but we feel our article sets a framework that can be extended naturally. For these reasons rather than extend the allocation of published trials to particular designs further into the past we have now changed the abstract and main text in other instances to be clear that

...
the any statements we make about the relative of frequency of designs are based on recent trials.

2. On Page 3 under Background the authors mention that most “SWT described in the literature do not follow this particular design”, this most probably depends on the time frame the authors investigate. They cite one of the other articles in this series of under review articles but it would be beneficial if further details are also given here. Also, it would be valuable to provide the distribution of the type of SWT designs of all the studies included in the previous two systematic reviews and the systematic review that is currently under review.

Response
Please see response to comment 1. The reason we are able to make the assertion that most SWT described recently in the literature do not follow this particular design is because, in the terminology we later develop, the continuous recruitment short exposure design and closed cohort design are distinct from that particular design and these together represent most recent trials. Specifically the continuous recruitment short exposure design involves data collection in continuous time and the closed cohort design involves repeated observation of the same individuals. Open cohort trials may or may not follow the particular design we describe on page 3 depending on the measurement scheme. In our article we do report the number of trials that follow each design (13 continuous recruitment short exposure, 11 closed cohort, 11 open cohort, 2 ‘other) and have now added to page 5 that the review article examined 37 trials. On page 7 we refer back to the introduction and state that: “In the introduction we mentioned that the design literature has focussed mainly on designs where measurements are obtained cross-sectionally at pre-defined discrete time points [2-5]. Now that we have outlined a range of possible SWT designs, we see that amongst recent trials the design literature mainly addresses two special cases: the open cohort design with only a very small proportion of participants sampled at each time point (so that participants are measured at most once), and the repeated closed cohort design of Williams and colleagues [12].”

3. Figure 1 does not seem to add further value to the text and it’s also hard to follow why the different labels are where they are. For example, the duration of the step, (c), does it refer to the whole period after the introduction of the intervention? Would this notation also apply to the continuous recruitment with short exposure design? Also, it’s not clear why (a) is placed in that particular cell.

Response
We would prefer to retain the figure as we feel it has some value, particularly in clarifying that some trials involve data collection either whilst all clusters are in the control condition and/or when all clusters are in the intervention condition. However we thank the reviewer for alerting us that the figure is unclear. We have heavily revised the figure, and feel the figure now also helps the reader understand our terminology, which has changed slightly from the previous version.

4. On Page 5 under “Continuous recruitment with short exposure”, the authors
mention thirteen trials that use this design. It would be interesting to quantify the total number of trials that were included in the review; the review article needs to be cited here as well. As mentioned above, it would be interesting to report how many trials in total across the three systematic reviews that have used this design and how many have used the other two designs, possibly, using the typology detailed on Page 8. Also, how many SWTs will not fit the proposed typology?

Response

Please see response to comments 1 & 2. Two of the recent trials considered did not follow any of the three main designs, but they can still be described using the typology in terms of how individuals start exposure, exposure duration and measurement.

Minor Revisions

5. The numbering of the figures does not match the results section.
Response:

Sorry for this error, we hope we will be able to control the Figure numbers displayed on resubmission

6. Page 8, last sentence replace “measurement” with “measurements”.
Response:

This text has been modified

7. Page 10, define BHOMA on first appearance.
Response:

Corrected

We thank Dr Porcher for his thoughtful review and now list our responses to each of his comments.

General comments

The paper belongs to a series of articles on stepped wedge trials (SWT) all considered together for publication in Trials (references 1, 13, 14 and 15). Overall, the reviewer found it sometimes difficult to follow the many references to these other manuscripts (that were not available). Perhaps fewer references to these works would make this paper more self-contained and easier to read.

Response: we have made some changes but using references to other papers has helped to keep the paper length controlled, and will assist the reader of the complete series

Major Compulsory Revisions

1. After reading the paper, it appears that it contains useful information. Nonetheless, it is very narrative, and it is difficult to have a clear idea of the aims and methods. For instance, it is unclear whether it is a result of a systematic review, a non-systematic review, or simply a collection of topics covered by the experience of the authors. It is also not so clear whether the paper primarily aims
at helping the design or the reporting of SWT. This should be clarified, and the paper revised accordingly.

2. The different topics covered are all useful, but their succession lacks a clear progression or transitions. We are left without knowing if all these topics are the most important ones when designing a SWT (but then the issue on incomplete and unbalanced designs is too short to be really helpful), or just some topics the authors develop among others.

3. Related to my previous point, the issue of incomplete designs should be covered more in-depth, with pros and cons. We even learn more in the discussion, which is strange.

Response to comments 1-3.

We thank the reviewer for pointing out this weakness in the paper. We have now clarified in the introduction and methods sections that the topics addressed in the paper are firstly how individuals participate in the trial and the implications of this for the possibility of carry-over effects, and secondly guidance on four key design choices in planning a SWT: randomisation method, choice of step length and number of steps, collection of data before and after the rollout period, and whether the design will be complete or incomplete. We leave the other key design choice, that of sample size, to our companion paper in the series. We address reporting of the design of a SWT, principally in terms of how individuals participate. We have addressed the issue of incomplete designs in more depth than before.

4. The case studies are only briefly described in terms of main design, and then quickly referred to when tackling different issues such as carry-over. A more in-depth analysis of the case studies may help illustrating the authors’ take-home message on the different points they cover.

Response:

We have extended our description of each case study when tackling each of the design choices issues which are key to the design of a SWT, including now any lag in the implementation of the intervention (see response to comment 7 from Referee 3). We have however retained the format of first describing the case studies in terms of main design (individual exposure and measurement) because we feel that exposure and measurement are difficult to understand without examples.

5. The issues concerning analysis of SWT are described in another paper of the series. But this manuscript refers to analyses several times (e.g. page 10, end of the first paragraph; pages 11-12 on mixed-effects models …) A short introduction on how SWT are analysed would therefore be helpful.

Response:

We considered this comment but, partly because the paper is already long, we have instead decided to resolve this problem by reducing the referral to analysis methods in this paper. We now describe the choice of whether to collect outcome data before or after the rollout period in more general terms.
6. The manuscript would be improved if the authors could make a table with the three main designs, the other issues, and summarize the advantages and disadvantages of each.

Response:
We considered this and are prepared to add a table if requested, but felt on balance this was not justified. We felt the three main issues that affect the three designs differently are the likelihood of carry over effects, the information that can be used in stratified or restricted randomisation, and when a long step duration might be chosen (or alternatively an analysis approach or incomplete design chosen to avoid this). We did not feel that these issues deserved a dedicated table and felt the issues for each design are described clearly in the text.

Minor Essential Revisions
1. Many statements lack precision. For instance, page 11, “[the gain in efficiency] is probably greater when the baseline and subsequent values are taken from the same individuals”. Does the “probably” mean the authors are unsure of their statement? It would be better to be more specific or to give a reference.

Response:
We have tried to be more precise in the text. The particular instance mentioned by the reviewer we have dropped in this revised version as we felt the issue was not especially important and in order to manage the overall length of the paper whilst adding new text.

2. The numbering of figures in the text does not fit the numbered figures at the end of the manuscript.

Response:
Sorry for this error, we hope we will be able to control the Figure numbers displayed on resubmission

3. Page 6, 3 lines from the bottom, isn’t a “time point” missing?

Response:
Thank you, corrected

4. The paragraph on the need to check for balance throughout the trial is either too short (and then the point would deserve more details) or too long (and thus can be omitted, since the issue is covered in reference 15).

Response:
Thank you, we agree this comment is too short to be helpful and have removed the sentence.

Discretionary Revisions
1. The issues raised by data collection in case study no. 3 (in particular during the 21 months after rollout) may be further developed to better show why this would not be advisable.
Response:
We have revised all the text on data collection before or after the rollout period, referring to the case study.

We thank Dr Giraudeau for his thoughtful review and now list our responses to each of his comments.

Reviewer’s report:
This paper on the stepped-wedge design is of great interest and up-to-date, due to the increasing use of this design. It is clearly written and of importance, especially the proposed typology of SWT.

I would have the following questions/comments:

1) In some cases, for some clusters, there is no control period (cf example 3b). This means that some clusters directly start with the intervention period, maybe because the intervention has already been implemented in these clusters, i.e. prior the study. Can authors comment more on this point? Is this truly acceptable? Should it be avoided? What are the consequences (for instance, there is no longer any randomization for these clusters)?

Response:
In Figure 3b we see that for case study 2 some clusters begin the intervention at the start of the trial, but they are allocated to do so at random and so randomisation is preserved for the complete trial. In the case that some clusters have already implemented the intervention we would recommend conducting the SWT in other clusters. We have added some text to the description of case study 2 to clarify that this is still a randomised SWT.

2) Open cohort. Authors evoke the possibility for some individuals to change between trial clusters. This may also occur in the closed cohort design (even though participants are followed from the beginning to the end of the trial), and should be acknowledged. This comes down to be a switch from one cluster to another one, and, in the end, anytime there is some participant follow-up such a situation may be encountered.

Response:
We have added some text to page 6 where we describe the closed cohort design to highlight that some changing between clusters could potentially occur here also “…and participate from start till end, typically without any changing clusters”.

3) I feel a bit ill at ease by the fact that authors propose a typology, and, at the same time, acknowledge that 2 trials in their review adopted a different design.

Response:
I hope we can reassure the reviewer by pointing out that although 2 trials did not follow one of our three most commonly conducted designs, these trials can still be described in terms of our terminology. See also response to comment 4 from Referee 1.
4) It would be of interest to know whether the authors succeeded in using their TDM description for all of the SWT they identified in their review.

Response:
We have not explicitly described all trials in the review in terms of our proposed TDM approach, and for some trials there may not be enough detail provided in the report for this to be done. We have however looked in some detail at many in order to review other aspects of the trial and are unaware of any trials which are well described but for which the TDM approach could not be applied.

5) Regarding randomization, it would be of interest to discuss stratification and restriction in view of the typology previously defined. Surely these possibilities are not relevant for all the types of SWT.

Response:
We agree, noting primarily that covariates of individual participants cannot be used in a continuous recruitment design, whereas in an open or closed cohort it could be possible to identify participants and collect information from them before randomisation occurs.

6) I would also suggest discussing collection of outcome data before and after the rollout period in view of the type of SWT. Thus, I suspect that for a SWT with continuous recruitment with short exposure, the issues discussed by the authors no longer hold.

Response:
We feel that the issues are broadly the same, so have not drawn attention to the distinction between the different types of SWT here. We do however believe that collection of outcome data before or after rollout will be more informative in a closed cohort trial as the measurements will be on the same individuals and so likely to be more highly correlated over time.

7) I think we miss some discussion regarding the different interventions assessed in SWT. Thus, on the one hand, in case study 2 (introducing free school breakfasts), the intervention is such that there is no “learning curve”: the day the intervention starts, the day it applies (except, I agree, that pupils have to become familiar with this new opportunity)? On the other hand, in case study 3, the intervention implies staff training, with some associated learning curve. This surely has impact on the way the SWT is planned and on the way it is analyzed. I would be interest in having some comments on this issue, namely in view of the review performed by the authors.

Response:
Thank you for alerting us to this gap in the paper. We have expanded the section on duration and number of steps to reflect the potential impact here of whether there is a lag period during which the intervention is either partially implemented or has not yet affected the outcome in individuals. We have also addressed this issue in the expended section on incomplete designs. The impact on analysis is addressed in our companion paper.