Response to Keriel-Gascou et al.: Higher efficiency and other alleged advantages are not inherent to the stepped wedge design

Citation for published version (APA):

Document status and date:
Published: 01/07/2014

DOI:
10.1016/j.jclinepi.2014.02.015

Document Version:
Publisher's PDF, also known as Version of record

Document license:
Taverne

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:
www.umlib.nl/taverne-license

Take down policy
If you believe that this document breaches copyright please contact us at:
repository@maastrichtuniversity.nl
providing details and we will investigate your claim.
In conclusion, as we have mentioned previously in situations where interventions are being rolled out because it is believed that they do more good than harm or where there are limited resources that prevent the delivery of the intervention at different locations simultaneously, the stepped wedge design can be used to evaluate the intervention while delivering it to all intended participants. Woertman et al. [8] have shown that the stepped wedge design could reduce the sample size in cluster randomized trials; however, this is not universal and depends on several factors including the intracluster correlation coefficient and the number of steps. Although the stepped wedge design might be favorable in some circumstances, its use should be limited to cases where the alternative designs will lead to a less rigorous evaluation of the intervention.

Noreen D. Mdege*
Mona Kanaan
Department of Health Sciences
University of York
Heslington
York YO10 5DD
United Kingdom
*Corresponding author. Department of Health Sciences
University of York, Heslington
York YO10 5DD, United Kingdom
Tel.: +44-1904321836; fax: +44-1904321838
E-mail address: Noreen.mdege@york.ac.uk (N.D. Mdege)

References


Response to Keriel-Gascou et al.: Higher efficiency and other alleged advantages are not inherent to the stepped wedge design

In reply:

We thank Keriel-Gascou et al. [1] for their letter in response to our discussion papers regarding the stepped wedge design (SWD) [2,3]. A number of arguments were raised in the letter to which we would like to provide a point-by-point response.

Argument: The SWD is preferable to a classic cluster randomized trial (CRT) as it is far more efficient.

Our response: SWDs have been alleged to have higher power or efficiency by a number of authors [4–6]. The issue was also raised in a previous letter [7] and discussed in our response [8]. We still consider such claims to be an oversimplification and certainly do not agree that SWDs have inherently more power than classic CRTs. Hemming and Girling [9], in response to the article by Woertman et al. [6], provide examples that illustrate how the total sample size (or the total number of measurements taken) may be higher or lower in an SWD compared with a classic CRT, depending on cluster size and the intracluster correlation coefficient (see Ref. [10] for the authors’ response). Therefore, we do not agree with claims that SWDs are inherently more powerful or more efficient than classic CRTs. A single example (as the one provided in Ref. [9]) is sufficient to invalidate such a generalized claim. Whether an SWD or CRT is preferable in any particular context with respect to efficiency is a more complex issue and requires the consideration of a variety of factors, including the number of clusters available, number of subjects/measurements that can be taken per cluster, costs of recruiting a subject, costs of taking an individual measurement, and costs of implementing the intervention in each additional cluster.

Argument: Because an SWD requires fewer clusters than a classic CRT, clusters will tend to be larger, which typically also implies a higher number of professionals per cluster. As a result, the risk of having clusters with no inclusions is reduced and intercluster contamination is less likely to occur.

Our response: A small number of large clusters can also be disadvantageous. For example, the consequences of a cluster terminating its participation (eg, because of a decision at the management level) are much more severe than when one of many smaller clusters drops out. In addition, the risk of interperiod contamination (ie, between participants who have already received the intervention and those still waiting for it) [11] may increase. Finally, we would like to point out that whether intercluster contamination is an issue has much more to do with geographical aspects and catchment sizes (and potential overlap thereof) than simply the size of the clusters.

Argument: The SWD is advantageous because the intervention is introduced sequentially.

http://dx.doi.org/10.1016/j.jclinepi.2014.02.014
Our response: As we have described in our discussion paper [2], sequential introduction of the intervention is also possible in the context of a classic CRT.

Argument: The lower number of clusters required by an SWD “can improve group comparability in terms of population characteristics.”

Our response: If we understand the authors’ statement correctly, they claim that, if one were to compare an SWD with a small number of large clusters with a somehow comparable classic CRT (e.g., involving the same total number of subjects) with a large number of small clusters, the SWD would be advantageous, as subjects receiving the intervention vs. those not receiving the intervention would tend to be more similar to each other (in terms of characteristics, such as age, gender, and so on). Note that in a longitudinal SWD (in which the same subjects within each cluster are followed over time), the characteristics of the two groups must in fact be exactly identical (because each subject is part of both groups), which is unrelated to the number of clusters or the cluster size. Therefore, it appears that this argument is made in the context of a cross-sectional SWD (where new subjects are sampled within each cluster with each step). Regardless, the authors have not provided a proof of or reference for this claim, so it is difficult for us to verify this. However, with respect to the classic CRT, it is clear that, if one had to choose between a CRT with two clusters à 100 subjects (i.e., one cluster treated, the other untreated) or 100 clusters à two subjects, the latter would be preferable (with only two clusters, any differences with respect to the primary outcome would be completely confounded by any differences in subject characteristics). In a cross-sectional SWD, cluster characteristics will indeed remain the same over the various measurement occasions within each cluster, but each step involves the sampling of new subjects (some of which are treated and some are not), which contributes to differences in subject characteristics between the two groups. Consequently, a comparison between the SWD and CRT would also need to take into consideration the number of steps (besides the number of clusters and the cluster size). A more detailed examination of this issue is beyond the scope of this letter, so we regard this claim as unproven and the issue as unsettled for now.

Argument: The high burden of repeated measurements can be avoided in an SWD by sampling new subjects at each measurement occasion within clusters.

Our response: We agree that a cross-sectional SWD can be used to avoid the repeated measurements within individual subjects. However, clusters are then still required to repeatedly recruit new subjects and take repeated measurements. This puts a high burden on the clusters themselves (except when the data can be routinely collected).

Argument: In case the intervention is actually ineffective, the same numbers of subjects (i.e., half) receive the ineffective intervention in the classic CRT as in the SWD when sampling new subjects at each measurement occasion.

Our response: We agree that this is true for the cross-sectional (but not the longitudinal) SWD.

Argument: Professionals are motivated to continue their participation because they will eventually receive the intervention.

Our response: The promise of receiving the intervention could indeed increase motivation of professionals to continue study participation. However, clusters that are randomized to receive the intervention later may also lose interest in continuing participation because of the long delay. In addition, as we described in our discussion paper [2], optional implementation of the intervention in control clusters after the conclusion of the study can be easily added to a classic CRT (at which point the effectiveness of the intervention has already been established).

Argument: In the Information for Participating Actively (InPAct) study, the use of the SWD results in a reduced sample size.

Our response: With only eight clusters, the SWD is indeed more efficient in this example. As we demonstrated earlier, that is not inherently true for all SWDs in comparison to CRTs. In fact, as the authors themselves show, for a larger number of clusters, the CRT would be more efficient.

Conclusion: Although we agree that the SWD may offer some advantages over the classic CRT under certain circumstances (higher power, fewer clusters of a larger size), those advantages also need to be weighed against a number of potential disadvantages (can also have lower power, more severe consequences of a larger cluster dropping out, potential for interperiod contamination, high burden on subjects and/or clusters, decreasing motivation of clusters to continue participation). Therefore, claims as to the inherent superiority of the SWD are certainly false.

Wolfgang Viechtbauer*
Department of Psychiatry and Psychology
MHeNS School for Mental Health and Neuroscience
Maastricht University
PO Box 616, 6200 MD
Maastricht, The Netherlands

Daniel Kotz
Department of Family Medicine
CAPHRI School for Public Health and Primary Care
Maastricht University
PO Box 616, 6200 MD
Maastricht, The Netherlands

Mark Spigt
Department of Family Medicine
CAPHRI School for Public Health and Primary Care
Maastricht University
PO Box 616, 6200 MD
Maastricht, The Netherlands
A call for revising the strengthening the reporting of observational studies in epidemiology statement to include ecologic studies

The strengthening the reporting of observational studies in epidemiology (STROBE) statement that contains recommendations for accurate and complete reporting of 3 main observational study designs that is cohort, case-control, and cross-sectional designs published in 2008 [1]. Two extensions of the STROBE statement soon after released to cover genetic association studies [2] and molecular epidemiology studies [3]. However, one important type of observational studies that is ecologic studies is overlooked in this statement and its extensions.

Ecologic study is one of the oldest study designs that looks for the associations between disease incidence and variables of interest at aggregate, global, or environmental level. Missing individual data on the joint distribution of two or more variables within each group is the motive for focusing on the comparison on aggregate level rather than individual level [4].

Regardless of its numerous applications and advantages, ecologic design inherits a number of severe limitations, which may consist of “ecologic fallacy”, “within-group misclassification”, “insufficient data”, “troubles in controlling the confounders”, “collinearity”, “migration across groups”, and “temporal uncertainty” [5]. Therefore, it needs a robust design and a vigorous reporting.

Unfortunately, the results of a rather new bibliometric review study revealed that substantial number of cross-sectional ecologic studies apply untrustworthy methods or embrace statistical mistakes. For instance, 64% of studies under investigation improperly adjusted their outcomes for age or sex. Furthermore, 31% of studies failed to present important information when arguing the ecologic character of their design and so forth [6].

As a result, it is highly suggested that the STROBE statement should be revised or have another extension to provide an up-to-date set of guidelines that standardize reporting on ecologic studies.