

Use of the stepped wedge design cannot be recommended: A critical appraisal and comparison with the classic cluster randomized controlled trial design

Citation for published version (APA):

Kotz, D., Spigt, M., Arts, I. C. W., Crutzen, R., & Viechtbauer, W. (2012). Use of the stepped wedge design cannot be recommended: A critical appraisal and comparison with the classic cluster randomized controlled trial design. *Journal of Clinical Epidemiology*, 65(12), 1249-1252. <https://doi.org/10.1016/j.jclinepi.2012.06.004>

Document status and date:

Published: 01/12/2012

DOI:

[10.1016/j.jclinepi.2012.06.004](https://doi.org/10.1016/j.jclinepi.2012.06.004)

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.

Download date: 20 Aug. 2022

VARIANCE AND DISSENT

Use of the stepped wedge design cannot be recommended: A critical appraisal and comparison with the classic cluster randomized controlled trial design

Daniel Kotz^{a,*}, Mark Spigt^a, Ilja C.W. Arts^{b,c}, Rik Crutzen^d, Wolfgang Viechtbauer^e

^aDepartment of General Practice, CAPHRI School for Public Health and Primary Care, Maastricht University, PO Box 616, 6200 MD Maastricht, The Netherlands

^bDepartment of Epidemiology, CAPHRI School for Public Health and Primary Care, Maastricht University, PO Box 616 Maastricht, The Netherlands

^cDepartment of Epidemiology CARIM School for Cardiovascular Diseases, Maastricht University, PO Box 616 Maastricht, The Netherlands

^dDepartment of Health Promotion, CAPHRI School for Public Health and Primary Care, Maastricht University, PO Box 616 Maastricht, The Netherlands

^eDepartment of Psychiatry and Psychology, MHeNS School for Mental Health and Neuroscience, Maastricht University, PO Box 616 Maastricht, The Netherlands

Accepted 11 June 2012; Published online 8 September 2012

In a conventional randomized controlled trial (RCT), participants are randomly allocated to either an intervention or a control/comparison condition. However, for certain types of interventions, it is not valid to randomize participants at an individual level because contamination between the intervention and control groups can occur. In those situations, an alternative approach is to randomize entire groups of participants (clusters) to the two conditions. Studies using this type of design are typically called cluster RCTs.

A special form of this approach is the stepped wedge design, in which the intervention is administered in a random sequence to all clusters over a number of time periods (Fig. 1). Each time point at which a new cluster receives the intervention is called a “step,” and all clusters will have received the intervention after the final step. Before and after each step, the dependent variable of interest is measured in all clusters. The analysis of data from such a design involves the comparison of the intervention wedge (gray-shaded cells in Fig. 1) with the control wedge (non-shaded cells).

In a recent review of the literature, Mdege et al. [1] identified 25 studies using the stepped wedge design, of which about half were published between 1987 and 2006 and half from 2007 onward. The authors of this review attributed the growing use of the stepped wedge design to a heightened awareness of its existence among researchers and an “increased acknowledgement of the usefulness of the design in evaluating the effectiveness of interventions in routine

practice” [1]. But how useful is the design really? In this commentary, we critically assess the most important arguments that are made in favor of this design by various authors [1,2] and weigh these against its potential disadvantages.

1. Argument 1: all participants receive the intervention

Because the intervention is rolled out sequentially to all clusters of participants (Fig. 1), all participants will eventually receive the intervention in the stepped wedge design. Thereby, none of the participants are excluded from receiving an intervention that is believed to be effective. This is an advantage from an ethical point of view in a situation in which an intervention is tested with the prior belief that it will do more good than harm.

Counterargument: it is also possible to roll out an intervention to all individuals in a classic cluster RCT

In a classic cluster RCT, the clusters are randomized to either the intervention or the control condition. All clusters start at the same time with the treatment to which they were allocated (Fig. 2). Unless additional follow-up measurements are planned, the study concludes with the posttreatment assessment in all clusters. However, before the start of the study, one could easily agree to deliver the intervention to the control clusters after the final data collection so that all participants will eventually receive the intervention. Alternatively, one could consider making the implementation of the intervention in the control clusters contingent on the results of the study (i.e., only if the findings corroborate the prior belief that the intervention is beneficial).

* Corresponding author. Tel.: +31-43-38-82893; fax: +31-43-36-19344.

E-mail address: d.kotz@maastrichtuniversity.nl or <http://www.daniel-kotz.de> (D. Kotz).

		Step 1	Step 2	Step 3	Step 4	Step 5	Step 6	
Cluster	A	O	X	O	O	O	O	
	B	O	O	X	O	O	O	
	C	O	O	O	X	O	O	
	D	O	O	O	O	X	O	
	E	O	O	O	O	O	X	O
	F	O	O	O	O	O	O	X

Fig. 1. The stepped wedge design with six steps. The letter O indicates measurements of the dependent variable, and the letter X indicates implementation of the intervention. The letters A–F indicate six clusters. The gray shading indicates that the cluster has received the intervention.

2. Argument 2: sequential implementation of the intervention is advantageous

During the conduct of a stepped wedge design, the intervention is implemented sequentially, which may be necessary because of logistical, practical, and financial constraints. For example, if the intervention involves extensive training of caregivers, it may not be possible to implement the intervention simultaneously in multiple clusters if resources are limited. A sequential implementation of the intervention also allows studying underlying trends or control for time effects.

Counterargument: it is also possible to implement an intervention sequentially in a classic cluster RCT

It is true that it may be necessary to implement an intervention sequentially for various reasons. However, sequential implementation is also possible (and often done) in a classic cluster RCT. Fig. 3 illustrates an example in which clusters A, B, and C start with the intervention one after the other. In this example, the time between the preintervention

		Pre	Post
Cluster	A	O	X
	B	O	X
	C	O	X
	D	O	O
	E	O	O
	F	O	O

Fig. 2. The classic cluster randomized controlled trial design with three intervention clusters (A–C), three control clusters (D–F), and optional implementation of the intervention in control clusters after the conclusion of the study. The letter O indicates measurements of the dependent variable, and the letter X indicates implementation of the intervention. The letters A–F indicate six clusters. The gray shading indicates that the cluster has received the intervention.

(baseline) and the postintervention measurement of the dependent variable is kept constant across clusters and the starting point of the measurements is similar for pairs of intervention and control clusters (to keep potential time effects constant across intervention and control clusters). However, variations of this design are also possible: For example, it is also possible to collect data intermittently in the clusters to study the underlying time trends (Fig. 4).

3. Argument 3: stepped wedge designs require fewer clusters compared with classic cluster RCTs

The stepped wedge design may require fewer clusters than classic cluster RCTs. The statistical power is higher because the intervention effect is estimated not only by between-cluster comparisons as in the parallel group design but also by within-cluster comparisons.

Counterargument: the increased power of the stepped wedge design is because of the large number of measurements of the dependent variable and it is not an inherent characteristic of the design

In the stepped wedge design, data are collected from each cluster at each step. Therefore, a large number of measurements are made before and after the clusters start the

Cluster	A	O	X	O			
	B		O	X	O		
	C			O	X	O	
	D	O		O			X
	E		O		O		X
	F			O		O	X

Fig. 3. The cluster randomized design with three intervention clusters (A–C), three control clusters (D–F), sequential implementation of the intervention, and optional implementation of the intervention in control clusters after the conclusion of the study. The letter O indicates measurements of the dependent variable, and the letter X indicates implementation of the intervention. The letters A–F indicate six clusters. The gray shading indicates that the cluster has received the intervention.

Cluster	A	O	X	O	O	O	O	
	B	O		O	X	O	O	
	C	O		O		O	X	O
	D	O		O		O		X
	E	O		O		O		X
	F	O		O		O		X

Fig. 4. The cluster randomized design with three intervention clusters (A–C), three control clusters (D–F), sequential implementation of the intervention, intermittent measurements of the dependent variable, and optional implementation of the intervention in control clusters after the conclusion of the study. The letter O indicates measurements of the dependent variable, and the letter X indicates implementation of the intervention. The letters AeF indicate six clusters. The gray shading indicates that the cluster has received the intervention.

intervention. Assuming S denotes the number of steps, the stepped wedge design requires $(S + 1)/2$ times as many measurements when compared with a classic cluster RCT (e.g., $(6 + 1)/2 = 3.5$ times as many measurements in Fig. 1 compared with Fig. 2). Clearly, the increased amount of data obtained in this way leads to more precise estimates of treatment effects and hence increased power. However, based on the findings of Hussey and Hughes [3], we can conclude that the power advantage of the stepped wedge design derives primarily from this increase in the amount of data collected and not from the increase in randomization steps. Hence, one could essentially obtain the same amount of power by simply adding an equivalent number of pre- and postintervention measurements in the classic cluster RCT. Moreover, the study by Pearson et al. [4], which has been cited by Mdege et al. as supporting the claim that the stepped wedge design has higher power, provides no statistical evidence with respect to this claim. In fact, Pearson et al. [4] only provide an informal comparison between the implementation of a particular intervention using the stepped wedge design vs. a nonrandomized pretest–posttest design.

4. Argument 4: the stepped wedge design allows an intervention to be improved during the study

Because the intervention is rolled out sequentially over time in the stepped wedge design, it is possible to improve or refine the intervention or its delivery where necessary with each successive implementation step.

Counterargument: improvement of an intervention would also be possible in a classic cluster RCT in which clusters are sequentially initiated

As explained in response to argument 2, sequential implementation of an intervention across clusters is also possible in a classic cluster RCT. Hence, it would also be possible to improve an intervention or its delivery before it is initiated in the next cluster. However, whether it is desirable to adapt the intervention during the runtime of a trial

is highly debatable. Because trials are usually not powered to draw firm conclusions on time trends, this will almost certainly lower the validity of the results.

5. Disadvantages of the stepped wedge design

As we argued, the proclaimed arguments in favor of the stepped wedge design do not provide important advantages of this design compared with the classic cluster RCT. Furthermore, we believe that the stepped wedge design has a number of important disadvantages that must be taken into account when considering the use of this design:

1. A trial using the stepped wedge design will take much longer to conduct than a classic cluster RCT. In the worst-case scenario, the duration of a stepped wedge design will be the duration of a classic cluster RCT multiplied by the number of steps. Our example shows that a study using the stepped wedge design with six clusters and six steps (Fig. 1) can take up to six times longer than a classic cluster RCT (Fig. 2). This has important consequences for the feasibility of studies using the stepped wedge design. Furthermore, this design limitation makes it more difficult to study interventions that require a long-term follow-up to estimate their effect.
2. The repeated measurements of the dependent variable put a heavy burden on patients, caregivers, and researchers. Furthermore, it may be difficult to explain to the patients and caregivers involved why so many measurements have to be taken while the intervention has not even started. As a consequence, the quality of data collected in a stepped wedge design may suffer [5].
3. There is an increased risk of contamination and attrition in patients from a cluster in a stepped wedge design that is supposed to receive the intervention at one of the later steps. Especially in a situation where an intervention is believed to be superior to control (which has been stated as an argument in favor of the stepped wedge design), both patients and caregivers will probably want to receive or deliver the intervention as soon as possible. In clusters that are starting at a later time point (which may be several months or years later, depending on the duration of each step), there is an increased risk of control patients dropping into the intervention or withdrawing from the study entirely. For example, caregivers may start to deliver the intervention in patients who would benefit from the superior intervention. Similarly, patients may seek the superior intervention in one of the other clusters (which often are other health care centers). Of course, such problems may also occur in a classic cluster RCT in which an intervention is rolled out to patients over a longer period of time.

4. The most important disadvantage, in our opinion, is that an intervention is implemented in all clusters of the stepped wedge design, whereas it has not yet been proven to be effective. If the intervention turns out to be ineffective, it would mean that large numbers of patients received a useless intervention. In a classic cluster RCT, it is also possible to roll out the intervention to all patients from the control group, but this can be done after the intervention has been proven to be effective in the intervention group. Furthermore, it is possible to judge the effectiveness of the intervention much faster with the use of a classic cluster RCT. Consequently, patients from the control group of such a trial will receive a proven effective intervention at an earlier stage than if they had participated in a trial of the same intervention but evaluated using the stepped wedge design.

6. Conclusion

The use of the stepped wedge design is increasing. Although this design may seem attractive to patients, caregivers, and researchers, we have argued that the proclaimed advantages of the stepped wedge design are no real advantages. Furthermore, the stepped wedge design has several

important disadvantages compared with the classic cluster RCT, such as increased burden of measurements, extended trial duration, and excessive administration of an unproven intervention. Taking these considerations into account, the use of the stepped wedge design cannot be recommended.

Acknowledgments

This commentary resulted from a meeting of the Journal Club Mesch, which is a discussion group of epidemiologists, statisticians, and primary care researchers from Maastricht University Medical Centre (www.journalclubmesch.nl).

References

- [1] Mdege ND, Man MS, Taylor (Nee Brown) CA, Torgerson DJ. Systematic review of stepped wedge cluster randomized trials shows that design is particularly used to evaluate interventions during routine implementation. *J Clin Epidemiol* 2011;64:936–48.
- [2] Brown CA, Lilford RJ. The stepped wedge trial design: a systematic review. *BMC Med Res Methodol* 2006;6:54.
- [3] Hussey MA, Hughes JP. Design and analysis of stepped wedge cluster randomized trials. *Contemp Clin Trials* 2007;28:182–91.
- [4] Pearson D, Torgerson D, McDougall C, Bowles R. Parable of two agencies, one of which randomizes. *Ann Am Acad Pol Soc Sci* 2010;628:11–29.
- [5] Lynn P. *Methods of longitudinal surveys*. John Wiley & Sons Ltd; 2009.