# COMMENTARY

# Stepped wedge trials with continuous recruitment require new ways of thinking

Richard Hooper[a,*], Andrew Copas[b]

[a]*Centre for Primary Care & Public Health, Queen Mary University of London, London, UK*
[b]*MRC Clinical Trials Unit, University College London, London, UK*

## Abstract

**Objectives:** There is substantial variation in the design of stepped wedge trials. Many recruit participants continuously over time, although the methodological literature has tended not to differentiate closely between continuous recruitment and discrete sampling. We argue for a deeper understanding of the special features of stepped wedge trials with continuous recruitment.

**Study Design and Setting:** This is a commentary and informal review.

**Results:** We discuss the scheduling of recruitment and implementation in continuous time and how contamination might be avoided. We also offer some suggestions on reporting and terminology for stepped wedge trials with continuous recruitment and comment on issues for analysis.

**Conclusion:** Repeated cross-section and continuous recruitment stepped wedge trials are not the same thing. More work is needed to develop the theory and practice of stepped wedge designs with continuous recruitment. Thoughtful approaches to design and clarity of reporting are vital. © 2019 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

*Keywords:* Stepped wedge; Cluster randomized trials; Continuous recruitment; Contamination; Interrupted time series

## 1. Introduction

Stepped wedge trials are cluster randomized trials in which clusters are followed longitudinally, with different clusters randomized to cross from the control condition to the active intervention condition at different times [1]. This experimental approach mirrors the natural process of implementing an intervention over many sites, and is often associated with situations where a policy decision has already been made to roll out an intervention [2]. Stepped wedge trials are increasingly popular, with systematic reviews showing an exponential growth in the number of published reports and registered protocols [2–7]. A new CONSORT extension covering stepped wedge trials has been produced [8,9]. Methodologists have also embraced the topic: in the pages of the Journal of Clinical Epidemiology, a vigorous debate blossomed over the relative efficiency of stepped wedge and more traditional designs

[2,10–21]. This discussion illustrated, above all, that precision is key in discussing the unique features of stepped wedge trials: the devil is in the detail.

But the general definition of a stepped wedge trial given previously hides substantial variation in their design in practice, according to the plan for recruitment and the flow of individual participants through the study [22]. One common subtype involves the continuous recruitment of participants from each cluster over the same interval of time, with different clusters crossing over to the intervention at different points along this interval. For example, in their stepped wedge trial of placental growth factor testing to assess women with suspected pre-eclampsia, Duhig et al recruited participants continuously over a 16-month interval, as and when women presented with suspected pre-eclampsia at one of 11 participating maternity units [23]. This is in sharp contrast to the way data are collected for a trial such as the Devon Active Villages Evaluation by Solomon et al, in which participants were sampled from village communities in a series of discrete cross-sectional surveys [24].

The literature on stepped wedge trials has tended not to differentiate too closely between continuous recruitment

---

**What is new?**

**Key findings**
- Stepped wedge trials with continuous recruitment comprise an important category of stepped wedge trials.

- The literature on stepped wedge trials has tended not to differentiate closely between continuous recruitment and discrete sampling.

**What this adds to what was known?**
- Continuous time processes need special consideration at both the design and analysis stages.

- Careless timing of recruitment and implementation in continuous time can lead to contamination within a cluster.

**What is the implication and what should change now?**
- Repeated cross-section and continuous recruitment stepped wedge designs should be distinguished.

- The timing of recruitment and implementation should be clearly reported.

and discrete sampling. Here, we argue that a deeper understanding of the unique features of continuous recruitment processes will help researchers to identify risks of contamination and to make more appropriate design and analysis choices.

### 1.1. Continuous recruitment with short exposure

If participants remained exposed to a potential intervention in the cluster for a long period after their recruitment, then we could follow individuals longitudinally to see how their outcomes change when the cluster crosses from the control to the intervention. But more commonly participants who are recruited continuously are exposed only briefly, and each participant's primary outcome is assessed just once: either as a control or an intervention participant. In the typology of Copas et al, this is known as a continuous recruitment short exposure design [22].

For example, Poldervaart et al evaluated a policy which promoted the use of a scoring system to guide clinical decisions for patients with acute chest pain on arrival at hospital emergency departments [25]. Nine hospitals were randomized to change over to the intervention at staggered intervals over a recruitment period which ran from July 2013 to August 2014. The primary outcome, assessed for each individual participant, was a major cardiac event within 6 weeks of presentation at the hospital. Median

length of stay in the emergency department (exposure) was 4 hours [26].

A note on our terminology: the "exposure" period refers to the interval over which a participant is exposed to the possibility of receiving the active intervention if it were introduced at a cluster—either exposed directly, because the intervention is delivered at cluster level, or indirectly, through contact with other individuals in the same cluster who receive the intervention.

It is worth noting that in a continuous recruitment short exposure design, the outcome may be assessed some considerable time after the individual was recruited. The first ever stepped wedge trial, which evaluated the effectiveness of adding hepatitis B virus (HBV) vaccination to a routine program of childhood vaccination in The Gambia, provides an extreme example. Clusters were distinct geographical regions defined by the locations of 17 health centers. The schedule of vaccinations lasted for 9 months, but the window of opportunity for a child to be recruited into the trial, and for vaccination of a child to begin, was short (the first month after birth). If a child did not receive HBV vaccination at the first vaccination visit, they did not receive it at any subsequent visit. The primary outcome of hepatocellular carcinoma was not determined until 30 years later [27]. Even though participants might have lived in the same region of The Gambia for all of those 30 years, their direct "exposure" was short (1 month), because of the short window of opportunity for starting HBV vaccination. Thus participants could reliably be labeled as "control" or "intervention" depending on whether they were born before or after the crossover in their cluster. (The question of indirect exposure is less clear: all participants might have benefitted from some degree of herd immunity over the decades which followed the introduction of the HBV vaccination program to their communities.)

### 1.2. Contamination within a cluster

In many trials with continuous recruitment, it is not possible to cross seamlessly and instantaneously from recruitment under the control condition to recruitment under the active intervention condition. In a cluster in the process of crossing over from the control to the intervention, some participants may experience a mixture of the two, or something intermediate. The CONSORT extension to stepped wedge trials warns of the dangers of this kind of "contamination" of treatments [8]. To understand these issues in more detail (the devil is in the detail!), we must map out the timeline for crossover in a cluster. We start by defining terms, and then give examples.

In a stepped wedge trial, the crossover is unidirectional—from the control condition (usually routine care) to the active intervention, but never in reverse. With this kind of unidirectional crossover, there are (up to) four key phases in a cluster: (1) a period of recruitment under the control condition; (2) a period allowing for the

"closure" of the control recruitment period (we elaborate on this below); (3) a period of implementation of the active intervention; and (4) a period of recruitment under the active intervention.

The closure period (2) should be long enough that any participants who were recruited before the closure period begins will either have had their outcomes assessed or are no longer exposed by the end of the closure period. The closure period is therefore the shorter of the exposure period and the time from recruitment to outcome measurement. The implementation period (3) should be long enough for the intervention to be set up within a cluster and operating at full strength.

The closure period is often negligible (or at least short) because exposure is short. In Poldervaart's trial, the amount of time spent by each participant in the emergency department was a tiny fraction of the overall recruitment period. But this is not always the case. Berdal et al reported results from a stepped wedge trial of an experimental rehabilitation program for patients with rheumatic diseases who were admitted to rehabilitation centers in Norway [28,29]. The comparator was traditional rehabilitation, whose duration could vary between 1 and 4 weeks, and outcomes were not assessed until after a participant had been discharged. The closure period required for this trial would have been 4 weeks—long enough to ensure that all participants recruited under the control condition had left the center before the implementation of the intervention began.

The implementation period is also often assumed to be negligible. This is sometimes achieved in practice by delivering training materials to a cluster before the scheduled switch. In Berdal's trial, for example, a team of research intervention facilitators visited each center with study materials 2 weeks before the crossover was due. It is vital to understand whether intervention materials predelivered in this way could be accessed by health professionals or participants in a cluster before the official crossover. If they can, then the implementation period (as we define it here)

must be considered to begin from this point. Other trials have included explicit implementation periods in their design: Peden et al allowed a 5-week interval for their quality improvement intervention to be introduced in hospitals performing emergency laparotomies [30].

### 1.3. Avoiding contamination

To avoid contamination, we propose the following as a standard of good practice: that all recruitment be suspended for the combined duration of the closure period and implementation period, or at least that these participants be excluded from the primary comparison of control and intervention. This principle was followed in Peden's trial, for example. The combined closure and implementation period is sometimes referred to as a "transition period" between recruitment under the control and intervention conditions [1].

The closure period ensures that outcomes of "control" participants are not contaminated by the intervention, even while it is being set up. The implementation period ensures that "intervention" participants are never receiving a less-than-full-strength intervention—i.e., are not contaminated by a weaker form of the intervention.

### 1.4. Diagrams

The CONSORT extension recommends diagrams as a tool for understanding the risk of contamination in stepped wedge designs [8]. If the phases of control recruitment, closure, implementation, and intervention recruitment in a cluster are nonoverlapping (as we recommend), they can be shown schematically as a sequence of differently shaded or colored bars running along a time axis. The number of people recruited under the control (or intervention) condition in a given cluster will be the product of the recruitment rate and the length of the control (or intervention) recruitment period. In some situations, as noted previously, the closure or implementation period may be assumed to be negligible, meaning that a cluster can cross seamlessly
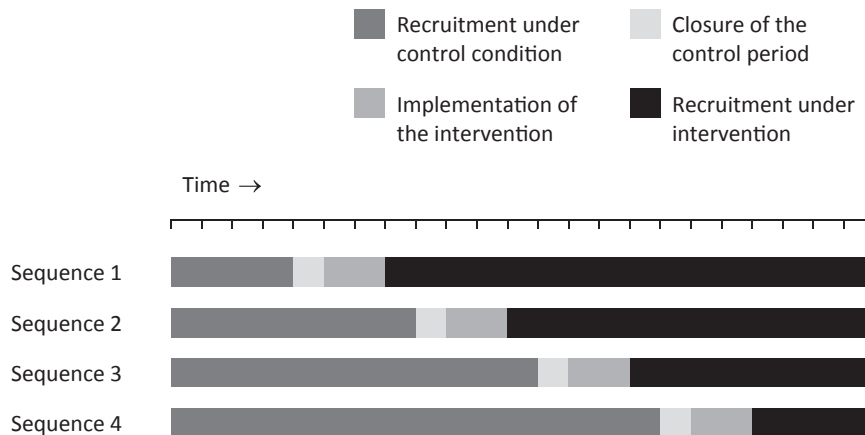


**Fig. 1.** Schematic representation of a stepped wedge trial with continuous recruitment, distinguishing four phases in each cluster: recruitment under the control condition, closure of the control recruitment period, implementation of the active intervention, and recruitment under the active intervention condition.

from control recruitment to intervention recruitment. When such an assumption is made, it should be justified. If a cluster is not required to include any control participants, then a closure period is unnecessary, and the cluster can begin in the implementation or intervention recruitment phase.

Figure 1 illustrates such a diagram for a complete stepped wedge trial design with clusters randomized to four different sequences. The example is a traditional stepped wedge, with crossover times regularly spaced across sequences, and all clusters starting in the control condition and finishing in the intervention condition. When recruitment is a continuous time process, it is worth bearing in mind that we have the freedom to choose crossover times anywhere along the continuous timeline, and the time selected for one cluster to cross over need not have any significance in another cluster. Nor is there any particular reason for the end of one transition to align with the beginning of another. As long as we have allowed sufficient time for closure and implementation, we can usefully begin recruitment again. Observe how different this is to the way time is conceived in the classic Hussey and Hughes model for stepped wedge trials, where time consists of discrete periods divided by cluster crossovers [31].

What every stepped wedge trial does require is a fine degree of control over which clusters make the transition to the active intervention and when. This is one of their enduring challenges.

### 1.5. Time as a continuous variable

The analysis of any stepped wedge trial must adjust appropriately for time, which will generally be confounded with treatment (as time goes on more and more clusters are in the intervention condition). Adjusting for time as a discrete factor, as in the Hussey and Hughes model, seems an inadequate approach to the analysis of a stepped wedge trial with continuous recruitment. An alternative might be to model time as a continuous variable with a linear or smoothly varying effect on outcome, rather like an interrupted time series analysis [32,33]. This kind of analysis choice feeds back to design and sample size calculation: the sample size needed for an analysis adjusting for a continuous, linear effect of time will not be the same as that needed if the adjustment is for discrete time [34].

The Hussey and Hughes model also assumes that the intracluster correlation—the correlation between the outcomes of two individuals from the same cluster—is the same for any pair of individuals. This is known as complete exchangeability. Other authors have attempted to generalize this to allow the intracluster correlation to grow weaker over time, while still treating time as discrete [35–37]. In these discrete-time models for intracluster correlation, there is still exchangeability within a given period. Again, as soon as we view recruitment as continuous, this approach becomes inadequate: two individuals recruited at either end of the "same" period will have much less to do with each other than two

individuals recruited just on either side of a "division" between periods. An approach where the intracluster correlation decays as a continuous function of the time separating the two individuals may be a better alternative [32]. Again, this impacts on design as well as analysis.

### 1.6. Terminology

Copas et al identified three distinct subtypes of stepped wedge trial: continuous recruitment short exposure designs, "open cohort" designs such as the Devon Active Villages Evaluation trial with its repeated survey samples, and "closed cohort" designs which follow the same individuals longitudinally [22]. Other authors blur the boundaries of open cohort and continuous recruitment designs to form a single category of "cross-sectional" stepped wedge trials, in which each participant is only assessed once [1,21]. This usage is misleading: "cross-sectional" conjures an image of a series of cross-sectional slices through time— something quite different to a continuous recruitment process. We suggest that the term "repeated cross-section" be restricted to trials involving discrete surveys of a cluster population. We prefer to use the term "continuous recruitment" for what it is. "Continuous recruitment short exposure" is a more specific term, but as noted previously, the exposure need not always be short.

### 2. Conclusions

Continuous recruitment stepped wedge trials are likely to continue to be popular with researchers. Depending on how such a trial is conducted, there can be a risk of contamination within a cluster, although some care and thought can help to address this. A diagram can help us to visualize the design and judge the risk more clearly.

We have proposed a diagram describing the timeline of recruitment and transition as planned in advance of a trial. In reality, individual clusters may deviate from the planned timeline. There could also be systematic deviations if, for example, the intervention takes longer to implement than anticipated. By careful monitoring of trial conduct, it may be possible to report the actual timelines experienced by clusters, as well as those planned, and to consider the consequent risks of contamination. Exactly what we mean in general by "intention-to-treat" or "as treated" analyses of a stepped wedge trial may be open to debate [8], but with clarity of reporting and reflection on the risks of bias we can start to explore such alternatives.

Although framed in terms of stepped wedge trials, the ideas presented here apply equally to simpler designs for cluster randomized trials with continuous recruitment, such as a parallel groups design with a baseline period. We have not touched on the question of optimal design: how should a schedule be chosen for transition and recruitment to minimize the total number of clusters or participants? This is a

fertile area for future research. Incorporating time as a continuous variable in the analysis of stepped wedge trials is also likely to be the focus of ongoing work. Sample size calculation and other design tools need development and expansion in parallel. A continuous timescale gives a stepped wedge trial something of the feel of an interrupted time series study, albeit a randomized one, and perhaps there is more to be learned about the analysis of continuous recruitment stepped wedge trials from a cross-fertilization with the interrupted time series literature.

# References

[1] Hemming K, Haines TP, Chilton PJ, Girling AJ, Lilford RJ. The stepped wedge cluster randomised trial: rationale, design, analysis and reporting. BMJ 2015;350:h391.

[2] Mdege ND, Man M-S, Taylor 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.

[3] Brown CA, Lilford RJ. The stepped wedge trial design: a systematic review. BMC Med Res Methodol 2006;6:54.

[4] Beard E, Copas A, Davey C, Osrin D, Baio G, Thompson JA, et al. Stepped wedge randomised controlled trials: systematic review of studies published between 2010 and 2014. Trials 2015;16:353.

[5] Martin J, Taljaard M, Girling A, Hemming K. Systematic review finds major deficiencies in sample size methodology and reporting for stepped-wedge cluster randomised trials. BMJ Open 2016;6: e010166.

[6] Barker D, McElduff P, D'Este C, Campbell MJ. Stepped wedge cluster randomised trials: a review of the statistical methodology used and available. BMC Med Res Methodol 2016;16:69.

[7] Grayling MJ, Wason JMS, Mander AP. Stepped wedge cluster randomized controlled trial designs: a review of reporting quality and design features. Trials 2017;18:33.

[8] Hemming K, Taljaard M, McKenzie JE, Hooper R, Copas A, Thompson JA, et al. Reporting of stepped wedge cluster randomised trials: extension of the CONSORT 2010 statement with explanation and elaboration. BMJ 2018;363:k1614.

[9] Hemming K, Carroll K, Thompson J, Forbes A, Taljaard M. Quality of stepped-wedge trial reporting can be reliably assessed using an updated CONSORT: crowd-sourcing systematic review. J Clin Epidemiol 2019;107:77−88.

[10] Kotz D, Spigt M, Arts ICW, Crutzen R, Viechtbauer W. Use of the stepped wedge design cannot be recommended: a critical appraisal and comparison with the classic cluster randomized controlled trial design. J Clin Epidemiol 2012;65:1249−52.

[11] Mdege MN, Man M-S, Taylor CA, Torgerson DJ. 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. J Clin Epidemiol 2012;65: 1253−4.

[12] Kotz D, Spigt M, Arts ICW, Crutzen R, Viechtbauer W. Researchers should convince policy makers to perform a classic cluster randomized controlled trial instead of a stepped wedge design when an intervention is rolled out. J Clin Epidemiol 2012;65:1255−6.

[13] Woertman W, de Hoop E, Moerbeek M, Zuidema SU, Gerritsen DL, Teerenstra S. Stepped wedge designs could reduce the required sample size in cluster randomized trials. J Clin Epidemiol 2013;66: 752−8.

[14] Hemming K, Girling A, Martin J, Bond SJ. Stepped wedge cluster randomized trials are efficient and provide a method of evaluation without which some interventions would not be evaluated. J Clin Epidemiol 2013;66:1058−9.

[15] Kotz D, Spigt M, Arts ICW, Crutzen R, Viechtbauer W. The stepped wedge design does not inherently have more power than a cluster randomized controlled trial. J Clin Epidemiol 2013;66:1059−60.

[16] Hemming K, Girling A. The efficiency of stepped wedge vs. cluster randomized trials: stepped wedge studies do not always require a smaller sample size. J Clin Epidemiol 2013;66:1427−8.

[17] De Hoop E, Woertman W, Teerenstra S. The stepped wedge cluster randomized trial always requires fewer clusters but not always fewer measurements, that is, participants than a parallel cluster randomized trial in a cross-sectional design. J Clin Epidemiol 2013;66:1428.

[18] Keriel-Gascou M, Buchet-Poyau K, Rabilloud M, Duclos A, Colin C. A stepped wedge cluster randomized trial is preferable for assessing complex health interventions. J Clin Epidemiol 2017;67:831−3.

[19] Mdege MD, Kanaan M. Response to Keriel-Gascou et al. Addressing assumptions on the stepped wedge randomized trial design. J Clin Epidemiol 2014;67:833−4.

[20] Weichtbauer W, Kotz D, Spigt M, Arts ICW, Crutzen R. Response to Keriel-Gascou et al.: higher efficiency and other alleged advantages are not inherent to the stepped wedge design. J Clin Epidemiol 2014;67:834−6.

[21] Zhou X, Liao X, Spiegelman D. "Cross-sectional" stepped wedge designs always reduce the required sample size when there is no time effect. J Clin Epidemiol 2017;83:108−9.

[22] Copas AJ, Lewis JJ, Thompson JA, Davey C, Baio G, Hargreaves JR. Designing a stepped wedge trial: three main designs, carry-over effects and randomisation approaches. Trials 2015;16:352.

[23] Duhig KE, Myers J, Seed PT, Sparkes J, Lowe J, Hunter RM, et al. Placental growth factor testing to assess women with suspected pre-eclampsia: a multicentre, pragmatic, stepped-wedge cluster-randomised controlled trial. Lancet 2019;393:1807−18.

[24] Solomon E, Rees T, Ukoumunne OC, Metcalf B, Hillsdon M. The Devon Active Villages Evaluation (DAVE) trial of a community-level physical activity intervention in rural south-west England: a stepped wedge cluster randomised controlled trial. Int J Behav Nutr Phys Act 2014;11:94.

[25] Poldervaart JM, Reitsma JB, Koffijberg H, Backus BE, Six AJ, Doevendans PA, et al. The impact of the HEART risk score in the early assessment of patients with acute chest pain: design of a stepped wedge, cluster randomised trial. BMC Cardiovasc Disord 2013;13:77.

[26] Poldervaart JM, Reitsma JB, Backus BE, Koffijberg H, Veldkamp RF, Ten Haaf ME, et al. Effect of using the HEART score in patients with chest pain in the emergency department: a stepped-wedge, cluster randomized trial. Ann Intern Med 2017;166:689−97.

[27] The Gambia Hepatitis Study Group. The Gambia hepatitis intervention study. Cancer Res 1987;47(21):5782−7.

[28] Kjeken I, Berdal G, Bø I, Dager T, Dingsør A, Hagfors J, et al. Evaluation of a structured goal planning and tailored follow-up programme in rehabilitation for patients with rheumatic diseases: protocol for a pragmatic, stepped-wedge cluster randomized trial. BMC Musculoskelet Disord 2014;15:153.

[29] Berdal Bø I, Dager TN, Dingsør A, Eppeland SG, Hagfors J, Hamnes B, et al. Structured goal planning and supportive telephone follow-up in rheumatology care: results from a pragmatic, stepped-wedge, cluster-randomized trial. Arthritis Care Res 2018;70(11):1576−86.

[30] Peden CJ, Stephens T, Martin G, Kahan BC, Thomson A, Rivett K, et al. Effectiveness of a national quality improvement programme to improve survival after emergency abdominal surgery (EPOCH): a stepped-wedge cluster-randomised trial. Lancet 2019;393:2213−21.

[31] Hussey MA, Hughes JP. Design and analysis of stepped wedge cluster randomized trials. Contemp Clin Trials 2007;28:182−91.

[32] Grantham K, Kasza J, Heritier S, Hemming K, Forbes A. Accounting for a decaying correlation structure in cluster randomised trials with continuous recruitment. Stat Med 2019;38:1918−34.

[33] Kontopantelis E, Doran T, Springate DA, Buchan I, Reeves D. Regression based quasi-experimental approach when randomisation is not an option: interrupted time series analysis. BMJ 2015;350: h2750.

34 Grantham KL, Forbes AB, Heritier S, Kasza J. Time parameterizations in cluster randomized trial planning. Am Stat 2019. https://doi.org/10.1080/00031305.2019.1623072.

[35] Hemming K, Lilford R, Girling AJ. Stepped-wedge cluster randomised controlled trials: a generic framework including parallel and multiple-level designs. Stat Med 2015;34:181–96.

[36] Hooper R, Bourke L. Cluster randomised trials with repeated cross sections: alternatives to parallel group designs. BMJ 2015;350:h2925.

[37] Kasza J, Hemming K, Hooper R, Matthews JNS, Forbes AB. Impact of non-uniform correlation structure on sample size and power in multiple-period cluster randomised trials. Stat Methods Med Res 2017;28:703–16.