Journal of Clinical Epidemiology 137 (2021) 159–162

Journal of Clinical Epidemiology

ORIGINAL ARTICLE

Key concepts in clinical epidemiology: Stepped wedge trials

Richard Hooper*

Institute of Population Health Sciences, Queen Mary University of London, London, UK

Received 16 March 2021; Received in revised form 10 April 2021; Accepted 13 April 2021; Available online 20 April 2021

Abstract

A stepped wedge trial evaluates an intervention that is implemented over a number of time periods according to a staggered timetable. Stepped wedge trials are usually cluster randomized, the intervention being delivered at some geographical, service or other cluster level. There is considerable variety in the design and conduct of stepped wedge trials in practice. The analysis of a stepped wedge trial often assumes that the effect of the intervention is maintained at a constant level once it has been implemented. It is important when estimating this effect to adjust for a period effect or underlying secular trend, since time is confounded with intervention, and to account for the clustering of outcomes. The advantage often cited for a stepped wedge design is that every cluster ends up getting the intervention, though in any trial design we can offer the intervention preferentially to control clusters after the trial has finished. The real advantage of a stepped wedge design is likely to be practicality or statistical efficiency.

Keywords: Stepped wedge; Clinical trial; Cluster randomised; Crossover

1. Introduction

A stepped wedge trial is a kind of crossover trial – that is, one in which you randomize not to one treatment condition or another, but to one *sequence* of treatment conditions or another over a succession of time periods [1]. In a stepped wedge trial, the crossover is unidirectional: a sequence may switch from treatment condition B to treatment condition A, but never from A back to B (Figs. 1). Condition B might be the current standard of care, while A might be an innovation that is straightforward enough to introduce, but much harder (practically or politically) to take away – an intervention that changes practice or is difficult to unlearn, or that policy has decreed will eventually be rolled out to everyone [2] .Think of "B" for "Before" and "A" for "After".

Where an intervention is to be rolled out as a policy change it might seem that the moment for evaluation has passed us by, but as long as the implementation is not instantaneous we still have an opportunity to conduct a randomized trial [2]. Randomizing to unidirectional sequences amounts to staggering the implementation of the intervention according to a predetermined, randomized schedule. Stepped wedge trials are nearly always randomized in clusters: that is, an entire administrative region, hospital catchment area, general practice list, or other grouping of individual participants is randomized to a particular sequence. Randomizing in clusters is statistically inefficient, but suits health service or policy interventions that are delivered at an organizational or regional level.

2. Variations on a theme

Fig. 1(a) illustrates a typical stepped wedge design, with four randomized sequences and five time periods. In Fig. 1(a), all the sequences begin in the control condition and end in the intervention condition, and in each period another sequence crosses over to the intervention. This pattern, though common, is not set in stone: Figs. 1(c-f) illustrate alternative stepped wedge schemes. Some stepped wedge trials are designed to be "incomplete", *i.e.* not to use data from every period in every cluster – see the examples in Fig. 1(d) and (f).

There is considerable variety in the design and conduct of stepped wedge trials in practice [3]. The "time periods" of a stepped wedge trial may variously represent a series of repeated cross-sectional slices through clusters, a continuous period of recruitment or identification demarcated by the dates of intervention implementation in different sequences, or multiple follow-up assessments of a single cohort of individuals recruited at the beginning of the trial

Conflict of interest: None

* Corresponding author. Yvonne Carter Building, 58 Turner Street, London E1 2AB Tel: 020 7882 7324, Fax: 020 7882 2552.

E-mail address: r.l.hooper@qmul.ac.uk.

https://doi.org/10.1016/j.jclinepi.2021.04.003

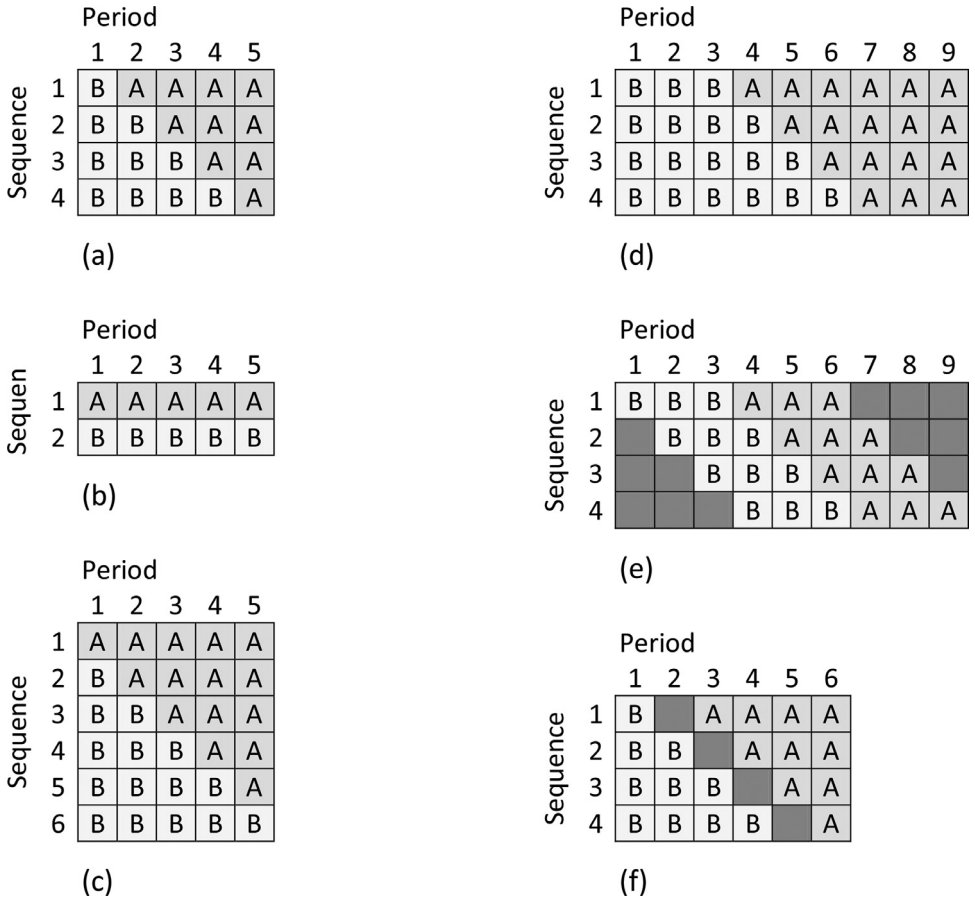

Fig. 1. Unidirectional crossover designs. (a) A typical stepped wedge design with four sequences and five periods: each sequence begins in the control condition (B) and ends in the intervention condition (A), and in each period another sequence crosses over to the intervention. The pattern of 'A's forms a stepped wedge shape, giving this family of designs its name. (b) A two-sequence design over five periods with no cross-over (though it still adheres, strictly, to the unidirectional principle that a sequence may switch from treatment condition B to treatment condition A, but never from A back to B). (c) A stepped wedge design that combines sequences from (A) and (B). (d) An 11-period, four-sequence design where each sequence spends at least three periods in the control condition at the beginning, and three periods in the intervention condition at the end. (e) An incomplete version of (D) where data are collected only in the three periods immediately before and the three periods immediately after cross-over (this might be to reduce the burden or cost of data collection, for example). (f) An incomplete design that leaves a transition period between the last control period and the first intervention period (this might be to allow time for the intervention to be implemented).

(Figs. 2) [4]. In each case a number of assessments can be taken from each cluster in each period. Each sequence could have several clusters allocated to it, or just one, depending on the design. CONSORT guidance is available to help with reporting stepped wedge trials [5].

3. Example

The Devon Active Villages Evaluation (DAVE) evaluated a community-level intervention aimed at increasing physical activity [6]. 128 rural UK villages were randomized to four different schedules for implementing the intervention, in a stepped pattern, as in Fig. 1(a). The villagers' physical activity levels were surveyed at five fixed time-points throughout the trial: on each occasion a random sample were selected from each village to receive the postal survey. This is an example of a repeated cross-section stepped wedge trial Fig. 2(a).

4. Analysis

How you analyze a stepped wedge trial depends on how much complexity you want to model. The analysis often assumes, for simplicity, that the effect of the intervention – the difference in expected outcomes between clusters in the intervention and control conditions – is maintained at a constant level once the intervention has been implemented. It is important when estimating this effect to adjust for a period effect or underlying secular trend, since time and intervention are confounded (the unidirectional cross-over means that as time goes on, more of the clusters are in the intervention condition). It is also vital in any cluster randomized or longitudinal design to account for the clustering of outcomes.

One way to do all of this is using a multi-level model [7] , with fixed effects of period and intervention, the latter depending on cluster and period. Be aware that when the

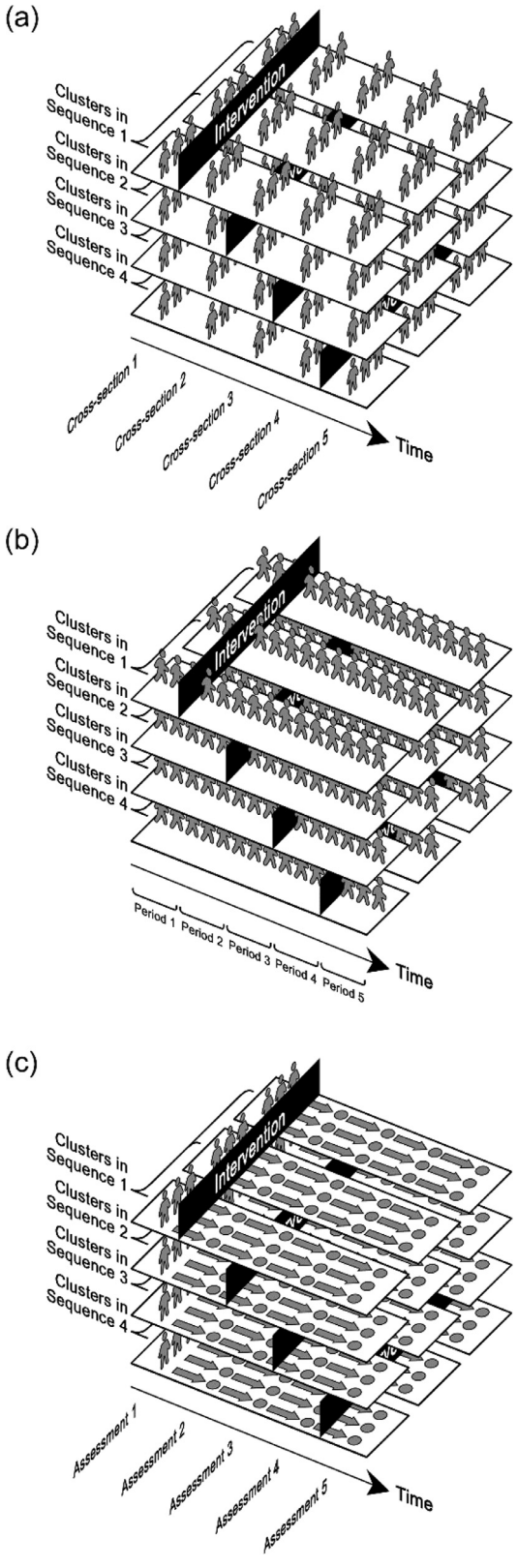

Fig. 2. Three ways of sampling individuals in different periods of a cluster randomized stepped wedge trial: (a) repeated cross-sections through the clusters; (b) continuous recruitment demarcated by the dates of intervention implementation in different sequences; (c) The same cohort assessed at multiple follow-up times.

number of clusters is small, multi-level models may inflate the type I error rate (the chance of finding a statistically significant intervention effect when none exists) unless a correction is applied [8]. A small number of clusters also increases the risk that a key cluster characteristic will be confounded with sequence: all the rural clusters might be randomized to cross to the intervention early in the trial, and all the urban clusters late, for example.

5. Advantages

The staggered schedule of a stepped wedge trial mimics a natural (non-experimental) implementation process, and this pragmatic feel may make the design attractive to triallists and participating centres [9] .Still, the stepped wedge may not be the best choice of trial design on purely theoretical grounds. Why use Fig. 1(a) in preference to Fig. 1(b), for example? The advantage often cited for a stepped wedge design is that every cluster ends up getting the intervention, but even in Fig. 1(b) we could offer the intervention preferentially to the control clusters after the trial has finished. Ethical arguments are flawed: if it is unethical to randomize some clusters to a trial arm that does not receive the intervention, then it is also unethical to ask some clusters to wait for it in a stepped wedge design [9].

The real advantages of staggering the intervention are twofold. First, efficiency: Figure 1(a) may, in some circumstances, need fewer clusters to achieve the same precision and statistical power than Fig. 1(b) (this arises, roughly speaking, when the cluster size is large or the clustering of outcomes is relatively strong), though in fact the hybrid of the two designs – Fig. 1(c) – may be more efficient than either of them [10]. Second, practicality: there may not be "enough" of the intervention to deliver to half the clusters simultaneously (think of vaccine availability, for example). Further discussion can be found in the references and reading list that follow.

References

[1] Hussey MA, Hughes JP. Design and analysis of stepped wedge cluster randomized trials. Contemp Clin Trials 2007;28(2):182–91.
[2] Mdege ND, Man MS, 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] Beard E, Lewis JJ, Copas A, Davey C, Osrin D, Baio G, et al. Stepped wedge randomized controlled trials: systematic review of studies published between 2010 and 2014. Trials 2015;16:353.
[4] 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.
[5] 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.
[6] Solomon E, Rees T, Ukoumunne OC, Hillsdon M. The Devon Active Villages Evaluation (DAVE) trial: study protocol of a stepped

wedge cluster randomised trial of a community-level physical activity intervention in rural southwest England. BMC Public Health 2012;12:581.

[7] Li F, Hughes JP, Hemming K, Taljaard M, Melnick ER, Heagerty PJ. Mixed-effects models for the design and analysis of stepped wedge cluster randomized trials: an overview. Stat Methods Med Res 2020. doi:10.1177/0962280220932962.

[8] Kahan BC, Forbes G, Ali Y, Jairath V, Bremner S, Harhay MO, et al. Increased risk of type I errors in cluster randomised trials with small or medium numbers of clusters: a review, reanalysis, and simulation study. Trials 2016;17:438.

[9] Prost A, Binik A, Abubakar I, Roy A, De Allegri M, Mouchoux C, et al. Logistic, ethical, and political dimensions of stepped wedge trials: critical review and case studies. Trials 2015;16:351.

[10] Girling AJ, Hemming K. Statistical efficiency and optimal design for stepped cluster studies under linear mixed effects models. Stat Med 2016;35(13):2149–66.

Further reading

Hargreaves JR, Copas AJ, Beard E, Osrin D, Lewis JJ, Davey C, et al. Five questions to consider before conducting a stepped wedge trial. Trials 2015;16:350 (Reflections on motivation, design, analysis, sample size, and reporting.).

Binik A. Delaying and withholding interventions: ethics and the stepped wedge trial. J Med Ethics 2019;45:662–7 (More on the ethical arguments for and against stepped wedge designs.).

Hooper R, Eldridge SM. Cutting edge or blunt instrument: how to decide if a stepped wedge design is right for you. BMJ Quality & Safety 2021;30:245–50 (A questioning approach to the use of stepped wedge trials, with a worked example looking at efficiency.).

Hemming K, Kasza J, Hooper R, Forbes A, Taljaard M. A tutorial on sample size calculation for multiple-period cluster randomized parallel, cross-over and stepped-wedge trials using the Shiny CRT Calculator. Int J Epidemiol 2020;49(3):979–95 (How to go about calculating sample size for a stepped wedge trial.).