OPEN ACCESS

Grand rounds in methodology: improving the design of staggered implementation cluster randomised trials

Samuel I Watson ,[1] Richard Hooper[2]

[1]Department of Applied Health Sciences, University of Birmingham, Birmingham, UK
[2]Wolfson Institute of Population Health, Queen Mary University of London, London, UK

Correspondence to
Prof Samuel I Watson;
s.i.watson@bham.ac.uk

Received 9 May 2025
Accepted 1 July 2025

ABSTRACT

The stepped-wedge cluster randomised trial is a popular design in implementation and health services research. All clusters, such as clinics or hospitals, start in the control state, and gradually switch over to treatment in a random order until all clusters have received the intervention. The design allows for the incorporation of an experiment into the gradual roll-out of an intervention across clusters. However, the traditional stepped-wedge layout may not be the best choice in many scenarios. In this article, we discuss modifications to the stepped-wedge design that maintain a staggered roll-out, but which may improve some key characteristics. We consider improving the timing of implementation periods, reducing the volume of data collection and allowing for the recruitment of clusters over the course of the trial.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ The stepped-wedge cluster trial is a useful design in implementation science, in which all clusters receive the intervention incrementally; however, it has several potential downsides that can present ethical, financial or practical hurdles to its use.

WHAT THIS STUDY ADDS

⇒ This article describes modifications to the stepped-wedge design that can improve power, reduce the burden on study participants and facilitate recruitment in different circumstances.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ Improvements and modifications to cluster trial designs that address the specific circumstances of an evaluation can improve the feasibility and acceptability of a design, while reducing costs.

THE SITUATION IN PRACTICE

The stepped-wedge cluster trial has become a mainstay of implementation science research.[1–3] This study design features longitudinal observations of clusters, such as hospitals, clinics or care homes. Traditionally, all clusters begin the trial in the control state and incrementally switch over to the intervention state until all clusters have received the intervention.[4 5] All clusters provide data in all periods of the study. The staggered nature of the intervention's implementation over time results in the 'step' of 'stepped-wedge'. Figure 1a illustrates the classic stepped-wedge design with six sequences (ie, trial 'arms').

The stepped-wedge cluster trial has two main benefits that make it an ideal tool for evaluation of health service, public health and community care interventions. First, it can be more statistically efficient than a parallel-groups design in some circumstances, requiring fewer clusters. Second, it may help overcome constraints on research resources that prevent half of the clusters in the evaluation from receiving the intervention simultaneously.[3 5 6] The design is often used as a way of incorporating evaluation into planned roll-outs of service delivery changes and interventions. In addition, many investigators choose this design because they want all the clusters eventually to receive the intervention.[3] They see the stepped-wedge design as a way to incentivise participation in the trial, since intervention receipt is guaranteed. For example, the Placental growth fActor Repeat sampling for Reduction of adverse perinatal Outcomes in women with suspecTed pre-eclampsia (PARROT) trials in the UK and Ireland evaluated whether testing for placental growth

Check for updates

To cite: Watson SI, Hooper R. *BMJ Qual Saf* Epub ahead of print: [please include Day Month Year]. doi:10.1136/bmjqs-2025-018697

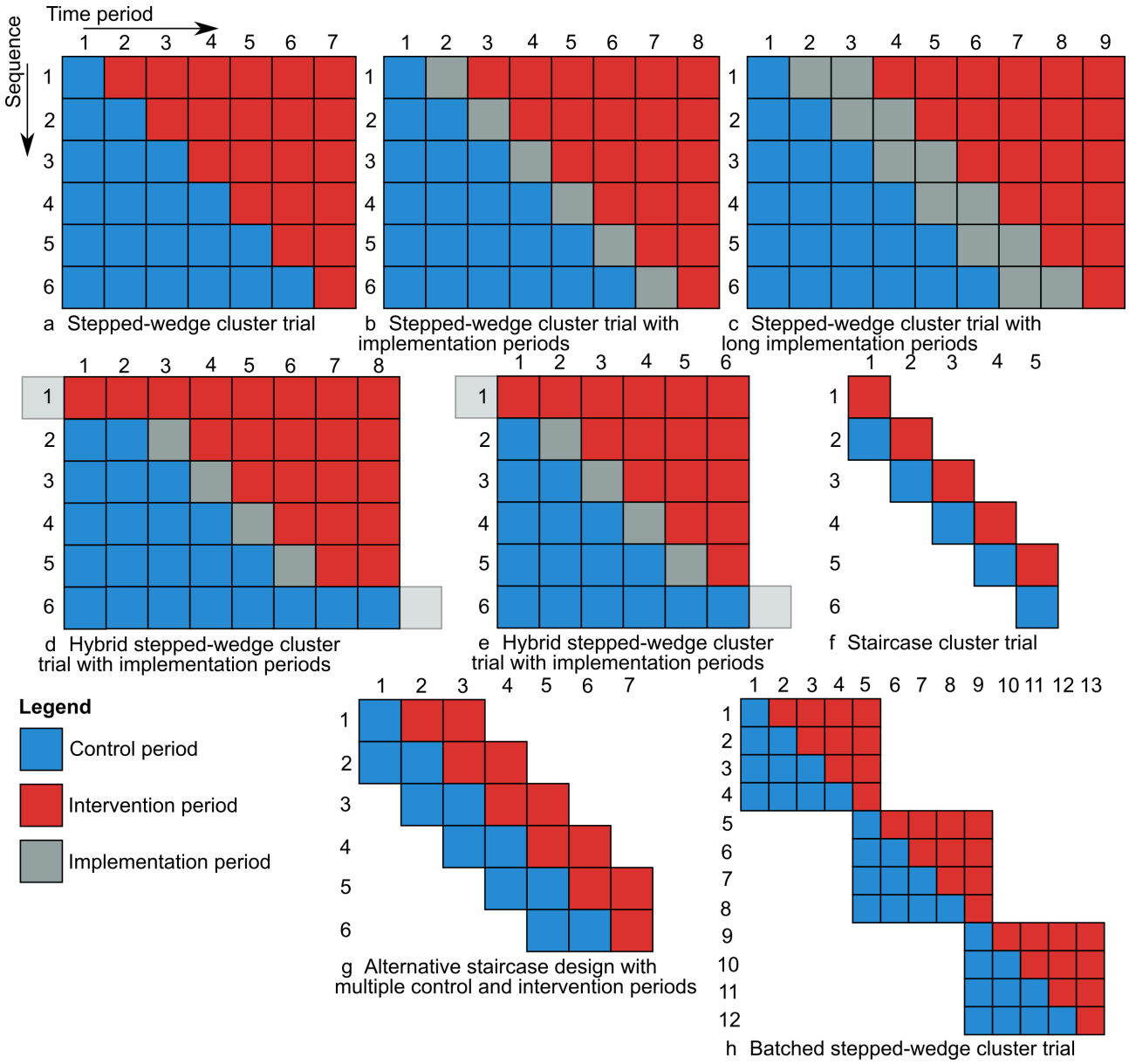

Figure 1 Examples of staggered implementation cluster randomised trial designs. Within each sub-figure, each row of the design is a trial sequence to which clusters are randomly allocated, each column of the design is a time period and the colour of each cell describes the intervention status for each sequence and time period. The designs are: (a) stepped-wedge cluster trial; (b) stepped-wedge cluster trial with implementation periods; (c) stepped-wedge cluster trial with long implementation periods; (d) Hybrid stepped-wedge cluster trial with implementation periods; (e) Alternative hybrid stepped-wedge cluster trial with implementation periods; (f) Staircase cluster trial; (g) Alternative staircase design with multiple control and intervention periods; (h) Batched stepped-wedge cluster trial.

factor in women with suspected pre-eclampsia reduced time to diagnosis and improved maternal and perinatal outcomes.[7 8] The studies report using a stepped-wedge design 'to increase the social acceptability of the trial to the hospitals'. But the incentive argument in favour of a stepped-wedge design over a simpler, parallel-groups design is moot.[9] Not every individual participating in a stepped-wedge trial receives the intervention (only half of the women in the PARROT trials did). Some clusters still have to wait for the intervention. And even in a parallel-groups design, we could still offer the intervention to all clusters at the end of the evaluation (a wait-list control).

The number of articles per year with 'stepped-wedge' in the title indexed on PubMed has increased from 25 in 2014 to 208 in 2024. There are reviews of relevant methodology,[10 11] a wide range of sample size calculators[12] and an extension to the Consolidated Standards of Reporting Trials statement.[13] However, the design has several characteristics that can undermine its use in practice. First, many health service interventions require implementation periods for training, infrastructure or to prevent contamination between control and treatment.[4] These implementation periods between control and treatment states can significantly reduce the overall power of a stepped-wedge design. Second,

the simplest forms of stepped-wedge designs are inefficient from a statistical power perspective. Third, some clusters must wait in the control state for a long time, withholding the intervention from patients while also collecting data. This can present an ethical, financial or practical hurdle for healthcare providers. Fourth, the stepped-wedge design requires us to recruit and randomly allocate all clusters before the beginning of the trial. Recruitment of clusters can take time, which may limit the time available to conduct the study.

Our aim in this article is to discuss alternatives to the stepped-wedge that have similar features, particularly that all clusters cross over in one direction and that the roll-out is staggered. The modifications and variations we describe have several desirable properties without sacrificing (much) efficiency and while permitting experimental evaluation alongside intervention roll-outs. We examine several designs and methods that address the four concerns outlined above. For some of these designs, there are only a small number of applied examples in the literature so far, since many of them have emerged from relatively recent methodological innovations. However, we identify relevant trials where possible.

IMPLEMENTATION PERIODS

Implementation periods are required when clusters cannot typically switch over from control to intervention states instantly. Instead, clinical staff may require training, infrastructure may need improving or new systems integrated (figure 1b, c).[4] For example, 3-month implementation periods were used in a stepped-wedge trial of integrating a mental health programme into chronic care clinics in Malawi. The implementation periods were used to train clinical staff in the delivery of group therapy and new care pathways.[14] Some trial designs will also require a 'closure period' after identification of control participants has stopped but before the implementation of the intervention begins, to allow the last control participants time to complete their exposure to routine care without that being contaminated by the intervention.[15] Together, the closure period and implementation period form a transition period between the periods of identification of control participants and identification of intervention participants. To align with other published work, and for simplicity, we will continue to use the term 'implementation period' for both.

The inclusion of implementation periods can significantly reduce the power of the design relative to an equivalently sized trial with no implementation periods.[9] Table 1 provides a comparison of the power of the designs in figure 1 for a continuous outcome and one cluster per sequence and a standardised effect size of 0.35. When the intracluster correlation coefficient (ICC) is relatively high, the power in this scenario drops from 80% to 66% between figure 1a and b, respectively, and then to 49% for figure 1c (table 1). Where a study may be underpowered, the standard response will be to increase the sample size. However, by understanding why we lose power, we can identify some small modifications to the design, which can recover most of the power and not require much change in the sample size.

Table 1 Comparison of power for the designs in figure 1 for a standardised effect size of 0.35, one cluster per sequence, a cluster-period size of 20, cross-sectional sampling and a cluster autocorrelation coefficient of 0.8 with a block exchangeable correlation structure

Design	ICC	Power (%)
(a)	0.05	80
	0.01	91
(b)	0.05	66
	0.01	78
(c)	0.05	49
	0.01	60
(d)	0.05	79
	0.01	93
(e)	0.05	68
	0.01	86
(f)	0.05	59
	0.01	67
(g)	0.05	70
	0.01	86

ICC, intraclass correlation coefficient.

STATISTICAL EFFICIENCY

There have been several recent methodological studies looking at which cluster trial designs achieve the greatest power for a given sample size. These are referred to as the 'optimal' or most 'efficient' designs.[16–18] For staggered implementation designs, including the stepped-wedge, a key finding is that not all cells (ie, cluster periods) contribute the same amount of 'information' to the estimated treatment effect. This means that the consequences of losing a cell on the power of the design is not the same for each cell; indeed, some cells can be removed from the full stepped-wedge design with almost no loss in power.

In a stepped-wedge design, the most informative cells are those on the main diagonal where clusters cross over from control to treatment.[16] Extending this finding, we can observe that the most informative time periods, where we should place the greatest amount of budget and effort, are those with contemporaneous, randomised comparisons of intervention and control participants.

In figure 1a, the most informative time periods are 2–6. The designs with implementation periods have fewer time periods with contemporaneous comparisons: four out of eight in figure 1b and only three out of nine in figure 1c. Indeed, figure 1c effectively has three baseline and three endline periods, and so most

of the observations in the study contribute the equivalent of a before-and-after type design. However, if we were to shift the implementation periods to just before and just after the study in the first and last sequences, and so maintaining a staggered roll-out, as shown in figure 1d, we can recover almost all the power lost due to incorporating implementation periods (table 1). Such a design is often called a 'hybrid' design as it incorporates sequences where clusters remain in the same treatment condition (control or intervention) throughout the evaluation (effectively a parallel-groups design), as well as sequences that incrementally cross clusters over in a staggered fashion (a traditional stepped-wedge design).[18] Indeed, if we drop the first and last time periods, as shown in figure 1e, the resulting design has comparable power to the original stepped-wedge design with implementation periods (figure 1b) with 25% fewer observations and time periods.

EXPOSURE TIME EFFECTS AND STAIRCASE DESIGNS

The full or 'complete' stepped-wedge design requires data collection in every period in every cluster (except implementation periods). For larger designs, running over long periods of time, this data collection can represent a significant burden on clusters and patients where the data are not abstracted from routine sources. The design also requires collecting control data for an extended interval in some clusters, which could arguably test their patience when they have agreed to participate in an evaluation of the new intervention. (Note that whether we collect this control data or not, there will be some clusters in a stepped-wedge design that have to wait a long time before they get the intervention, and the ethical challenges of this will remain).[19] This problem could be exacerbated in designs that include cohorts of patients whose inclusion in the study requires their identification as having a treatment need, but who must wait to receive treatment. For example, trials of interventions for mental healthcare will require screening patients to collect data on the presence and severity of a condition, which may not have been collected in standard practice. In a stepped-wedge setting, though, these patients may not then receive any intervention until much later in the study.

There is growing interest in the 'staircase' cluster trial design, which includes only the cells on the main diagonal (figure 1f, g), and other 'incomplete' designs where data are not collected in every period in every cluster.[20] Staircase designs are often the most efficient way to allocate experimental resources in these settings.[21] If the number of observations can be increased on the main diagonal of the designs, then one generally requires significantly less data to achieve the same power relative to a stepped-wedge design with equal size cluster periods. In addition to the cost-efficiency argument, these designs also only require clusters to *actively* participate in the trial for a small number of periods, which can reduce the burden on patients and clinics. They also focus attention on the effect of the intervention immediately after its introduction—useful if this is when the intervention effect is at its strongest, but not if the benefits of the intervention take a while to achieve.

The Data-driven Quality Improvement in Primary care study[22] was a cluster randomised trial evaluating an education-based intervention to reduce high-risk prescribing in primary care practices in Scotland. 34 practices were randomly allocated to one of 10 trial sequences, which formed a staircase design. Outcome data were collected at six time points pre-intervention and six post-intervention, with staggered start dates to form a staircase similar to figure 1g. A trial of the Extension for Community Health Outcomes model[23] to improve clinical practice and knowledge around autism screening and comorbidity management also adopted a staircase design. Ten medical centres were randomly allocated to one of five trial sequences with four total data collection time periods (two control and two intervention). This design was justified by the need to efficiently distribute data collection resources across the clusters during the study.

CONTINUOUS CLUSTER RECRUITMENT AND BATCHED DESIGNS

For parallel cluster trials, as with individual-level trials, it is straightforward to randomly allocate participants and clusters as they are recruited into the study, even after the study has started. However, the staggered implementation trial designs covered in this article require the recruitment and randomisation of all clusters before the trial starts. This requirement could delay already lengthy trials and still present an issue for clusters who may have to wait for long periods of time before starting their participation in the trial. It is possible, instead, to randomly allocate clusters while continuing to recruit new clusters. To maintain the benefits of randomisation, we still need to randomly allocate groups or 'batches' of clusters at once, so that the order of treatment initiation is not determined by recruitment date alone.[24]

A batched stepped-wedge design is illustrated in figure 1h.[24] Smaller groups of clusters are randomly allocated within smaller stepped-wedge designs. When the next group (or batch) of clusters is ready, they are then randomly allocated within the next substudy (which might or might not overlap with previous substudies), and so forth. One may view this design also as an incomplete stepped-wedge design that uses a restricted randomisation mechanism. The substudies need not be full stepped-wedge designs; batched randomisation could be applied to staircase or other designs with differing start dates. As with any other type of restricted randomisation, one must ensure appropriate

adjustment for the factors determining the randomisation in any analysis. Specifically, one must adjust for batch but potentially also the interaction between time and batch if there is overlap.[24]

The European Society of Coloproctology (ESCP) Safe-anastomosis Programme in Colorectal Surgery (EAGLE) study used a batched, incomplete stepped-wedge design to evaluate an educational intervention to reduce anastomotic leak after right colectomy.[25] The clusters were surgical units, and they were randomised in a series of batches, roughly every 2 months, provided at least 18 clusters were ready to be randomised at this point (batches varied in size). Each batch then participated in a substudy with a three-sequence staircase design. The authors noted that the batched design allowed for sequential entry of the clusters and staggered start times, which allowed them to manage a global study that was interrupted at several points by the COVID-19 pandemic. The SCANPatient trial will use a similar design to evaluate the effects of a new reporting template on the diagnosis of pancreatic ductal adenocarcinoma.[26] The trial will use three 'batches' of 10–12 hospitals randomly allocated to a stepped-wedge substudy with five sequences each.

CONCLUSIONS

We have described several stepped-wedge variants in this article that may alleviate some of the issues presented by a traditional stepped-wedge trial. The approach to analysis for these designs is the same as for the full stepped-wedge,[27] with the same considerations for allowing for clustering and correlation,[27] small sample biases[28] and appropriate adjustment for the restricted randomisation scheme in the case of batched designs. Several studies have highlighted the greater efficiency and better value for money that might be leveraged by considering staircase and other incomplete designs,[16] at least in situations where the primary data are not routinely collected.[29]

The cluster trial can be highly flexible in its design, especially when data can be collected longitudinally. Many practitioners, though, are only familiar with 'classic' interpretations of these designs. We have aimed to highlight that these designs need not be a Procrustean bed. The timing of data collection and intervention roll-out can be designed to suit often complex real-world circumstances and facilitate the inclusion of rigorous evaluation into the roll-out of interventions across health and social care.

Contributors SIW conceived the article and wrote the first draft. SIW and RH contributed to writing and reviewing all subsequent drafts. SIW is the guarantor.

Funding This study was funded by National Institute for Health and Care Research (NIHR 204294, NIHR158242, NIHR200132), Medical Research Council (MR/V038591/1).

Competing interests None declared.

Patient consent for publication Not applicable.

Ethics approval Not applicable.

Provenance and peer review Commissioned; externally peer reviewed.

Data availability statement There are no data in this work.

ORCID iD

Samuel I Watson http://orcid.org/0000-0002-8972-769X

REFERENCES

1 Varghese E, Briola A, Kennel T, *et al*. A systematic review of stepped wedge cluster randomized trials in high impact journals: assessing the design, rationale, and analysis. *J Clin Epidemiol* 2025;178:111622.
2 Brown CA, Lilford RJ. The stepped wedge trial design: a systematic review. *BMC Med Res Methodol* 2006;6:54.
3 Beard E, Lewis JJ, Copas A, *et al*. Stepped wedge randomised controlled trials: systematic review of studies published between 2010 and 2014. *Trials* 2015;16:353.
4 Hemming K, Haines TP, Chilton PJ, *et al*. The stepped wedge cluster randomised trial: rationale, design, analysis, and reporting. *BMJ* 2015;350:h391.
5 Hooper R. Key concepts in clinical epidemiology: Stepped wedge trials. *J Clin Epidemiol* 2021;137:159–62.
6 Hooper R, Eldridge SM. Cutting edge or blunt instrument: how to decide if a stepped wedge design is right for you. *BMJ Qual Saf* 2021;30:245–50.
7 Hayes-Ryan D, Khashan AS, Hemming K, *et al*. Placental growth factor in assessment of women with suspected pre-eclampsia to reduce maternal morbidity: a stepped wedge cluster randomised control trial (PARROT Ireland). *BMJ* 2021;374:n1857.
8 Duhig KE, Myers J, Seed PT, *et al*. Placental growth factor testing to assess women with suspected pre-eclampsia: a multicentre, pragmatic, stepped-wedge cluster-randomised controlled trial. *The Lancet* 2019;393:1807–18.
9 Westgate PM, Nigam SR, Shoben AB. Reconsidering stepped wedge cluster randomized trial designs with implementation periods: Fewer sequences or the parallel-group design with baseline and implementation periods are potentially more efficient. *Clin Trials* 2024;21:710–22.
10 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.
11 Hemming K, Eldridge S, Forbes G, *et al*. How to design efficient cluster randomised trials. *BMJ* 2017;358:j3064.
12 Ouyang Y, Li F, Preisser JS, *et al*. Sample size calculators for planning stepped-wedge cluster randomized trials: a review and comparison. *Int J Epidemiol* 2022;51:2000–13.
13 Hemming K, Taljaard M, McKenzie JE, *et al*. Reporting of stepped wedge cluster randomised trials: extension of the CONSORT 2010 statement with explanation and elaboration. *BMJ* 2018;363:k1614.
14 McBain RK, Mwale O, Mpinga K, *et al*. Effectiveness, cost-effectiveness, and positive externalities of integrated chronic care for adults with major depressive disorder in Malawi

(IC3D): a stepped-wedge, cluster-randomised, controlled trial. *The Lancet* 2024;404:1823–34.
15 Hooper R, Copas A. Stepped wedge trials with continuous recruitment require new ways of thinking. *J Clin Epidemiol* 2019;116:161–6.
16 Watson SI, Girling A, Hemming K. Optimal study designs for cluster randomised trials: An overview of methods and results. *Stat Methods Med Res* 2023;32:2135–57.
17 Kasza J, Forbes AB. Information content of cluster-period cells in stepped wedge trials. *Biometrics* 2019;75:144–52.
18 Girling AJ, Hemming K. Statistical efficiency and optimal design for stepped cluster studies under linear mixed effects models. *Stat Med* 2016;35:2149–66.
19 Binik A. Delaying and withholding interventions: ethics and the stepped wedge trial. *J Med Ethics* 2019;45:662–7.
20 Grantham KL, Forbes AB, Hooper R, *et al*. The staircase cluster randomised trial design: A pragmatic alternative to the stepped wedge. *Stat Methods Med Res* 2024;33:24–41.
21 Girling AJ, Watson SI. Optimal Allocation of Observations in Stepped-Wedge and Other Cluster Studies With Correlated Cluster-Period Effects. *Stat Med* 2025;44:e70100.
22 Dreischulte T, Donnan P, Grant A, *et al*. Safer Prescribing--A Trial of Education, Informatics, and Financial Incentives. *N Engl J Med* 2016;374:1053–64.
23 Mazurek MO, Parker RA, Chan J, *et al*. Effectiveness of the Extension for Community Health Outcomes Model as Applied to Primary Care for Autism. *JAMA Pediatr* 2020;174:e196306.
24 Kasza J, Bowden R, Hooper R, *et al*. The batched stepped wedge design: A design robust to delays in cluster recruitment. *Stat Med* 2022;41:3627–41.
25 Venn ML, Knowles CH, Li E, *et al*. Implementation of a batched stepped wedge trial evaluating a quality improvement intervention for surgical teams to reduce anastomotic leak after right colectomy. *Trials* 2023;24:329.
26 Li L, Kasza J, Recasens A, *et al*. SCANPatient: study protocol for a multi-centre, batched, stepped wedge, comparative effectiveness, randomised clinical trial of synoptic reporting of computerised tomography (CT) scans assessing cancers of the pancreas. *Trials* 2024;25:388.
27 Li F, Hughes JP, Hemming K, *et al*. Mixed-effects models for the design and analysis of stepped wedge cluster randomized trials: An overview. *Stat Methods Med Res* 2021;30:612–39.
28 Thompson JA, Hemming K, Forbes A, *et al*. Comparison of small-sample standard-error corrections for generalised estimating equations in stepped wedge cluster randomised trials with a binary outcome: A simulation study. *Stat Methods Med Res* 2021;30:425–39.
29 Rezaei-Darzi E, Grantham KL, Forbes AB, *et al*. The impact of iterative removal of low-information cluster-period cells from a stepped wedge design. *BMC Med Res Methodol* 2023;23:160.