



Review

What type of cluster randomized trial for which setting?

Karla Hemming^{a,*}, Andrew Copas^b, Andrew Forbes^c, Jessica Kasza^c^a Institute of Applied Health Research, University of Birmingham, Birmingham, UK^b MRC Clinical Trials Unit at University College London, London, UK^c School of Public Health and Preventive Medicine, Monash University, Victoria, Australia

ARTICLE INFO

Article History:

Received 16 August 2023

Accepted 8 January 2024

Available online 20 February 2024

ABSTRACT

The cluster randomized trial allows a randomized evaluation when it is either not possible to randomize the individual or randomizing individuals would put the trial at high risk of contamination across treatment arms. There are many variations of the cluster randomized design, including the parallel design with or without baseline measures, the cluster randomized cross-over design, the stepped-wedge cluster randomized design, and more recently-developed variants such as the batched stepped-wedge design and the staircase design. Once it has been clearly established that there is a need for cluster randomization, one ever important question is which form the cluster design should take. If a design in which time is split into multiple trial periods is to be adopted (e.g. as in a stepped-wedge), researchers must decide whether the same participants should be measured in multiple trial periods (cohort sampling); or if different participants should be measured in each period (continual recruitment or cross-sectional sampling). Here we outline the different possible options and weigh up the pros and cons of the different design choices, which revolve around statistical efficiency, study logistics and the assumptions required.

© 2024 The Author(s). Published by Elsevier Masson SAS. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)

Background

Cluster randomized trials randomize entire clusters of participants to intervention conditions. This is in contrast to individually randomized trials, where individuals themselves are randomized. In cluster randomized trials, clusters could be hospitals, units or wards within hospitals, entire geographical regions, schools, families, etc. Careful consideration and justification are required before opting for the use of cluster randomization as it increases the required sample size, increases risk of bias and has specific ethical considerations [1,2]. Cluster randomization is necessary when the intervention is applied at the cluster level, but outcomes are measured on the individuals within the cluster, or when individual randomization would lead to unacceptable levels of contamination (i.e. the treatment applied to one individual would influence the treatment or outcome of another individual) [1].

In the standard two-arm cluster randomized design clusters are randomized in a 1:1 allocation ratio to either intervention or control condition [Fig. 1a]. In Fig. 1 the clusters are represented as being allocated to “sequences” which are more commonly referred to as “arms” in parallel designs. There are many different variations of cluster randomized trials, many of which divide the trial up into

DOI of original article: <http://dx.doi.org/10.1016/j.jeph.2024.202198>, <http://dx.doi.org/10.1016/j.jeph.2024.202199>, <http://dx.doi.org/10.1016/j.jeph.2024.202196>, <http://dx.doi.org/10.1016/j.jeph.2024.202197>.

* Corresponding author.

E-mail address: k.hemming@bham.ac.uk (K. Hemming).

<https://doi.org/10.1016/j.jeph.2024.202195>

2950-4333/© 2024 The Author(s). Published by Elsevier Masson SAS. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)

multiple periods [3,4]. Some of these variations are similar to the sorts of variations available when using individual randomization, while others are unique to cluster randomization. The parallel cluster trial with a baseline period is perhaps the simplest of the variations [Fig. 1b]. Similar to the individually-randomized trial, this design simply requires a baseline measure of the outcome from participants before randomization, in addition to the measure taken after randomization, although typically in cluster trials different participants are measured in the two periods (see below). Another variation is to switch between control and intervention conditions potentially multiple times over multiple time periods (often referred to as sequences of treatments, hence the use of the term sequence) [Fig. 1c]. Such a design has an individually-randomized analogue in individual-level crossover trials, where individual patients cross between different treatments, with randomization to the sequence (order) of treatments received. In the case of randomizing clusters this is known as the cluster randomized cross-over design (sometimes referred to by the acronym CRXO) – with clusters rather than participants crossing between intervention conditions. Similar to individually randomized designs, this might simply consist of a single switch between treatment and control; or randomization to a sequence of switches – so that at set times clusters transition between receiving the intervention or control condition [5]. In the stepped-wedge cluster randomized trial (SW-CRT), clusters are randomized to a sequence which

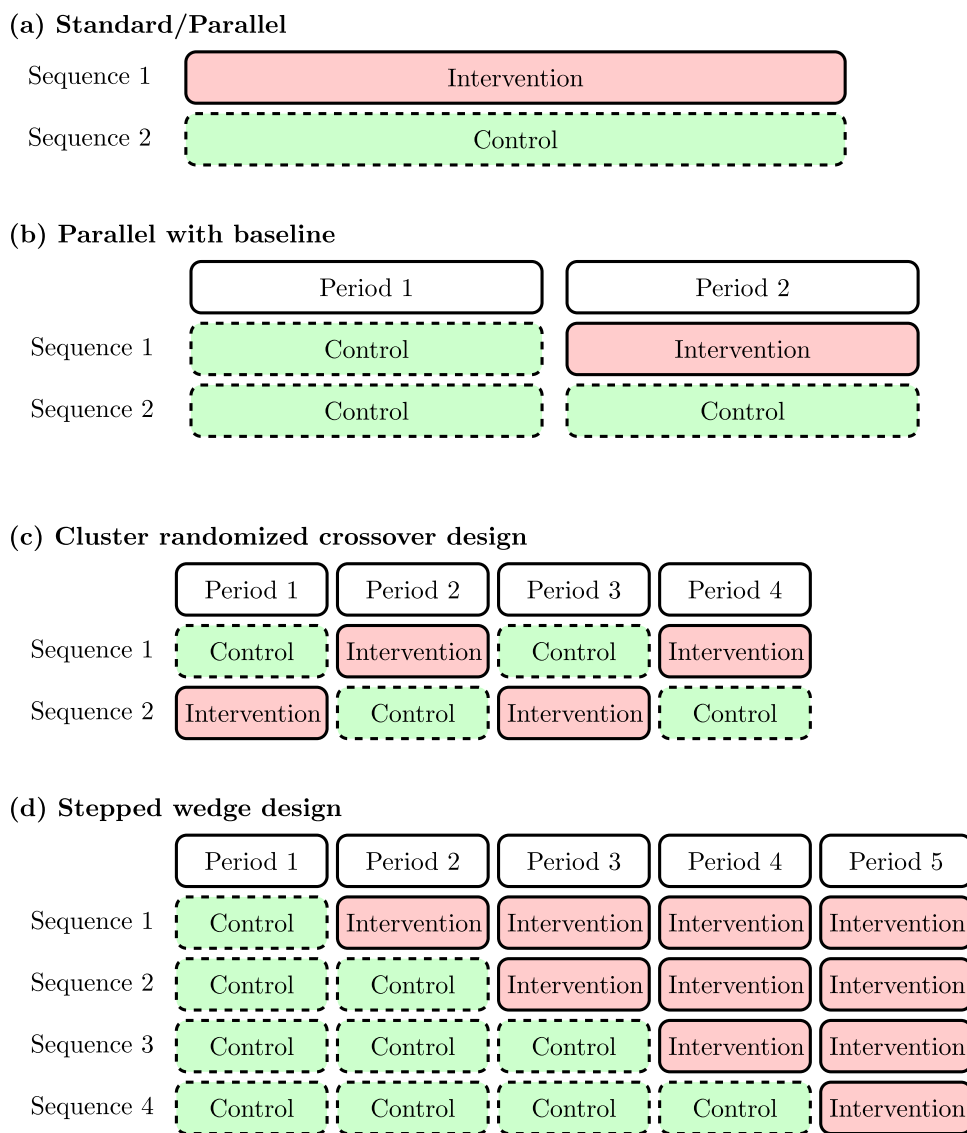


Fig. 1. A schematic representation of different types of cluster randomized trials.

dictates the period of transition from control to intervention condition – again a design in which time is split into a number of periods – but where every cluster ultimately receives the intervention condition [Fig. 1d]. Variations of the SW-CRT include the batched stepped-wedge design (a series of SW-CRTs run in batches), and the staircase design [6,7].

Citation analyses show an increasing adoption of these variations to the parallel CRT [8,9]. However, careful consideration is needed when considering which of these variations of cluster randomized designs to choose. Important considerations include the required sample size (number of clusters, number of time periods, number of participants per cluster) – often referred to as statistical efficiency; how many clusters are exposed to the intervention condition; as well as risk of bias (including carryover effects or within cluster contamination, identification and recruitment biases, and biases due to secular trends); and whether to measure different or the same participants in each of the different time periods. Our objective here is to provide some guidance on how to choose between these different designs.

Choices around different sampling structures

If a design in which time is split into multiple trial periods is to be adopted, the same or different participants might be measured in the

different trial periods. If clusters are large and data collection induces costs then researchers also need to decide whether to measure all individuals in clusters or just a random sample in each. Where the same participants are measured in each trial period this is known as closed cohort sampling. Alternative sampling structures where participants provide only one measurement throughout the entire trial are possible. A cross-sectional sampling structure would occur when samples of different participants are taken from each cluster in each period. A continuous recruitment structure involves different participants entering the trial during each period of time [10]. Alternatively, in some situations some participants may provide measurements in more than one period, and others only once, known as an open cohort sampling structure [11,12].

For some trials the setting determines how participants will be measured. For example, in the LUSTRUM trial participants enter the trial on diagnosis of chlamydia at a clinic (a trial cluster) and their participation in the trial ends soon after [13]. In this setting recruitment is inevitably continuous and, unless an individual is diagnosed more than once, different participants will be measured in different periods. In other settings researchers will however need to make a choice of sampling structure. For example, in a school trial with measurements taken initially in year 7, options include follow up one year later by measuring the same children (now year 8) or different

children in the same school year (new children, year 7). Here we may prefer to measure the same children again since this allows us to directly measure within-participant change and could increase power to estimate intervention effects. On the other hand, this sampling structure may be more susceptible to bias if, for example, the repeated assessment itself induces a change in behavior, or representativeness declines due to attrition (children leaving the school between school years) [14].

In other settings researchers will however need to make a choice between a cohort or cross-sectional sampling approach. For example, suppose we wish to evaluate an intervention through a baseline and an endline survey. If entire clusters are exposed to the intervention and clusters are open cohorts (e.g., communities) with a gradual rate of people leaving and joining clusters (i.e., 'churn') then the two sampling approaches target two slightly different estimands: the cohort approach targets those exposed from the start (but is subject to drop-out from people moving away) whilst the cross-sectional approach targets the broader cluster effect (i.e., the 'culture change') with a mixture of individual exposure durations. In settings where the 'churn' is low then the two estimands will be similar, and the choice of sampling approach is often based on logistical issues or statistical efficiency.

When to use the cluster randomized cross-over (CRXO)?

We start by considering when to use the cluster randomized cross-over design. This design features first in our list because it can be very statistically efficient – by incorporating cross-overs the CRXO design allows each cluster to act as its own control, and resultingly can recuperate some of the losses in efficiency brought about by cluster randomization. Indeed, if correlations between outcomes is expected to decay smoothly over time, the it has been shown that increasing the number of crossovers while maintaining the trial duration can increase the efficiency of the CRXO design [5]. However, the CRXO design is often not a reasonable choice when for example the intervention involves any change in human behavior such as education [15]. In addition, in settings where the intervention cannot be completely removed, the cross-over design can put the study at risk of carryover effects (that is, the control observations can become exposed to the intervention condition). Although steps can be put into place to mitigate the biases which arise from carryover effects – such as washout periods – these can both increase the duration and complexity of the study and leave doubts about whether the intervention truly brought about any observed change. Rather, the CRXO design should only be used when the intervention can be removed after roll-out and the cluster can return to its pre-trial state.

The PEPTIC trial is an example of a cluster randomized cross-over trial, including 50 clusters (including >26,000 participants) switching between two treatments in common use for stress ulcer prophylaxis in intensive care patients receiving invasive mechanical ventilation [16]. The crossover design was considered appropriate because the switching between interventions was straightforward and different patients in each period mitigated against any possible carryover effects from the treatment administered first.

When to use a design with baseline measures?

When cross-over is not possible or would likely put the study at risk of bias due to carryover effects, uni-directional cross-over designs can also increase statistical efficiency over the simple parallel design. The CRT with baseline design is particularly useful when the outcomes are ascertained from routinely collected data (and so have limited cost or delay implications). Again, the magnitude of the increase in statistical precision depends on within cluster

correlations, strengths of correlations between cluster-periods and cluster size [17]. It is also worthy to note that the cluster trial with a baseline measure will never be as statistically efficient as the CRXO design. An example of a CRT with a baseline period that uses cohort sampling is a CRT in which 10 clusters were randomized to either an unconditional or conditional cash transfer program to determine the effects on child health – here the same participants were measured before and after randomization [18]. An example of a CRT with a baseline period which uses cross-sectional sampling, is where schools were randomly allocated to control or a school-based mindfulness program, to determine the effect on well-being – here different children were measured pre and post randomization [19].

When to use the SW-CRT?

The stepped-wedge CRT can also be a pragmatic and statistically efficient design choice [20,21]. In the standard stepped-wedge CRT, all clusters must start the trial at the same time, and all must follow the pre-specified (randomized) intervention roll-out schedule. The knowledge that all clusters will eventually receive the intervention can enhance stakeholder engagement and participation in a SW-CRT. However, this appeal must be balanced against the knowledge that at the time of implementing the study it will be unknown if the intervention is effective [22]. The SW-CRT can also have an appeal where rationing or staggering the roll-out of the intervention is needed, for example because there are insufficient resources to roll-out the intervention to all clusters simultaneously (but this can achieve this under parallel CRT [4]).

This design has had a large uptake in recent years, but given the difficulties that occur with adherence to the strict scheduling of this design, it is often not the most appropriate design choice [23]. In addition, because it is not possible to increase the duration of the study once it has started, the design can also be at risk of under-recruitment issues [24]. Further, the stepped-wedge design induces confounding between the intervention and time. This is because the observations taken under the control condition are collected systematically earlier in calendar time compared to those under the intervention condition [Fig. 1d]. To mitigate this, it is necessary to adjust for time effects at the analysis stage – but these adjustments require assumptions (which cannot be verified) such as the effect of time is linear, or that the effect of time is same across all clusters.

So, whilst statistical efficiency can appear to increase under the SW-CRT design (again, the increase depends on strength of correlation, size of clusters) [25], and it can have a certain amount of pragmatic appeal, it is a design that should be adopted with caution and recognition of the assumptions made about time trends. Consider the SW-CRT across 121 hospitals in 10 different countries, evaluating a bundle of interventions for acute cerebral hemorrhage, including observations from more than 10,000 participants [26]. In this trial standard adjustment for calendar time was implemented in analysis, making the assumption that secular trends in the outcome functional recovery were identical across all 10 countries. An assumption of a common secular trend across 10 countries might be questionable. If this assumption was not tenable the resulting effect estimate might be biased. Including homogenous clusters, such as only clusters from within the same country, where calendar time effects (i.e. secular trends) are more likely to be similar would likely make assumptions about common trends more tenable [27].

While in the standard stepped-wedge design, all of the clusters must be ready to initiate the study at the same point in time, the batched stepped-wedge design relaxes this requirement to a degree because the trial is conducted in "batches" of clusters each in a 'mini' stepped-wedge design [6,23]. Further, the standard stepped-wedge design requires that all clusters participate in the trial for the entire trial duration; alternatives that do not require this include the dog-leg design and the staircase design [28,7].

Summary

There is a wide variety of types of CRTs. We recommend parallel CRTs as the preferred choice for settings where there are a large number of clusters. There are however more pragmatic design choices that can lead to increased study power, but there is no universal solution—and many of these alternative designs come with their own logistical issues or increased risk of bias, or analytical assumptions in statistical models. When designs with multiple periods are adopted there are different choices to make about sampling structures at the level of the individual. Whilst the selected sampling structure will often be dictated by the trial setting or practical constraints, it does have important ramifications. These different choices around designs will lead to different analytical [29] and sample size requirements [30].

There are also other factors we have not considered but which will nonetheless be important. For example, pilot trials and feasibility work can help determine the appropriateness of some of the design choices. There are also other design choices, not considered here, especially when there is interest to consider more than one intervention, however design complexity and associated assumptions can increase significantly – but we are mindful that whilst factorial designs and adaptive designs can make a very important contribution, they often can increase the required sample size and thus be infeasible. There are also other assumptions inherent in cluster trials we have not considered, such as whether intervention effects vary with time since roll-out of the intervention [31], or vary across clusters and in particular by size of cluster [32].

Declaration of Competing Interest

The authors declare there is no conflict of interests.

References

- Giraudeau B, Weijer C, Eldridge SM, Hemming K, Taljaard M. Why and when should we cluster randomize? *J Epidemiol Popul Health* 2024;72(1):202197.
- Caille A, Billot L, Kasza J. Practical and methodological challenges when conducting a cluster randomized trial: Examples and recommendations. *J Epidemiol Popul Health* 2024;72(1):202199.
- Cornfield J. Randomization by group: a formal analysis. *Am J Epidemiol* 1978;108(2):100–2 PMID: 707470. doi: 10.1093/oxfordjournals.aje.a112592.
- 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(2):181–96 Epub 2014 Oct 24. PMID: 25346484; PMCID: PMC4286109. doi: 10.1002/sim.6325.
- Grantham KL, Kasza J, Heritier S, Hemming K, Litton E, Forbes AB. How many times should a cluster randomized crossover trial cross over? *Stat Med* 2019;38(25):5021–33 Epub 2019 Sep 1. PMID: 31475383. doi: 10.1002/sim.8349.
- Kasza J, Bowden R, Hooper R, Forbes AB. The batched stepped wedge design: a design robust to delays in cluster recruitment. *Stat Med* 2022;41(18):3627–41 Epub 2022 May 21. PMID: 35596691; PMCID: PMC9541502. doi: 10.1002/sim.9438.
- Grantham KL, Forbes AB, Hooper R, Kasza J. The staircase cluster randomised trial design: A pragmatic alternative to the stepped wedge. *Stat Methods Med Res* 2024;31(1):24–41. pub 2023 Nov 30. PMID: 38031417; PMCID: PMC10863363. doi: 10.1177/09622802231202364.
- Murray DM. Influential methods reports for group-randomized trials and related designs. *Clin Trials* 2022;19(4):353–62 Epub 2022 Jan 6. PMID: 34991379; PMCID: PMC9256846. doi: 10.1177/17407745211063423.
- Offorha BC, Walters SJ, Jacques RM. Statistical analysis of publicly funded cluster randomised controlled trials: a review of the National Institute for Health Research Journals Library. *Trials* 2022;23(1):115. PMID: 35120567; PMCID: PMC8817506. doi: 10.1186/s13063-022-06025-1.
- Hooper R, Copas A. Stepped wedge trials with continuous recruitment require new ways of thinking. *J Clin Epidemiol* 2019;116:161–6 Epub 2019 Jun 11. PMID: 31272885. doi: 10.1016/j.jclinepi.2019.05.037.
- Copas A, 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.
- Kasza J, Hooper R, Copas A, Forbes AB. Sample size and power calculations for open cohort longitudinal cluster randomised trials. *Stat Med* 2020;39:1871–83.
- Estcourt CS, Stirrup O, Copas A, Low N, Mapp F, Saunders J, Mercer CH, Flowers P, Roberts T, Howarth AR, Owusu MW, Symonds M, Nandwani R, Ogwulu C, Brice S, Johnson AM, Althaus CL, Williams E, Comer-Schwartz A, Tostevin A, Cassell JA. Accelerated partner therapy contact tracing for people with chlamydia (LUSTRUM): a crossover cluster-randomised controlled trial. *Lancet Public Health* 2022;7(10):e853–65 PMID: 36182235. doi: 10.1016/S2468-2667(22)00204-3.
- Atienza AA, King AC. Community-based health intervention trials: an overview of methodological issues. *Epidemiol Rev* 2002;24(1):72–9 PMID: 12119859. doi: 10.1093/epirev/24.1.72.
- Goldstein CE, Giraudeau B, Weijer C, et al. When and how should we cluster and cross over: methodological and ethical issues. *Can J Anesth/ J Can Anesth* 2018;65:760–5. doi: 10.1007/s12630-018-1131-1.
- Young PJ, Bagshaw SM, Forbes AB, Nichol AD, Wright SE, Bailey M, Bellomo R, Beasley R, Brickell K, Eastwood GM, Gattas DJ, van Haren F, Litton E, Mackle DM, McArthur CJ, McGuinness SP, Mouncey PR, Navarra L, Oppenorth D, Pilcher D, Saxena MK, Webb SA, Wiley D, Rowan KM. Effect of stress ulcer prophylaxis with proton pump inhibitors vs histamine-2 receptor blockers on in-hospital mortality among ICU patients receiving invasive mechanical ventilation: the PEPTIC randomized clinical trial. *JAMA* 2020;323(7):616–26 PMID: 31950977; PMCID: PMC7029750. doi: 10.1001/jama.2019.22190.
- Copas AJ, Hooper R. Cluster randomised trials with different numbers of measurements at baseline and endline: sample size and optimal allocation. *Clin Trials* 2020;17(1):69–76 Epub 2019 Oct 3. PMID: 31580144; PMCID: PMC7334046. doi: 10.1177/1740774519873888.
- Robertson L, Mushati P, Eaton JW, Dumba L, Mavise G, Makoni J, Schumacher C, Crea T, Monasch R, Sherr L, Garnett GP, Nyamukapa C, Gregson S. Effects of unconditional and conditional cash transfers on child health and development in Zimbabwe: a cluster-randomised trial. *Lancet* 2013;381(9874):1283–92 Epub 2013 Feb 27. PMID: 23453283; PMCID: PMC3627205. doi: 10.1016/S0140-6736(12)62168-0.
- Kuyken W, Ball S, Crane C, Ganguli P, Jones B, Montero-Marín J, Nuthall E, Raja A, Taylor L, Tudor K, Viner RM, Allwood M, Aukland L, Dunning D, Casey T, Dalrymple N, De Wilde K, Farley ER, Harper J, Kappelmann N, Kempnich M, Lord L, Medlicott E, Palmer L, Petit A, Philips A, Pryor-Nitsch I, Radley L, Sonley A, Shackelford J, Tickell A, Blakemore SJ, Team TM, Ukoumunne OC, Greenberg MT, Ford T, Dalgleish T, Byford S, Williams JMG. Effectiveness and cost-effectiveness of universal school-based mindfulness training compared with normal school provision in reducing risk of mental health problems and promoting well-being in adolescence: the MYRIAD cluster randomised controlled trial. *Evid Based Ment Health* 2022;25(3):99–109 Epub ahead of print. PMID: 35820992; PMCID: PMC9340028. doi: 10.1136/ebmental-2021-300396.
- Hemming K, Taljaard M. Reflection on modern methods: when is a stepped-wedge cluster randomised trial a good study design choice? *Int J Epidemiol* 2020;49(3):1043–52 PMID: 32386407; PMCID: PMC7394949. doi: 10.1093/ije/dyaa077.
- 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(3):245–50 Epub 2020 Jun 16. PMID: 32546592; PMCID: PMC7907557. doi: 10.1136/bmjqs-2020-011620.
- Binik A. Delaying and withholding interventions: ethics and the stepped wedge trial. *J Med Ethics* 2019;45:662–7 - Pub Med.
- Caille A, Taljaard M, Le Vilain-Abraham F, Le Moigne A, Copas AJ, Tubach F, Dechartres A. Recruitment and implementation challenges were common in stepped-wedge cluster randomized trials: results from a methodological review. *J Clin Epidemiol* 2022 Aug;148:93–103 Epub 2022 Apr 26. PMID: 35483552. doi: 10.1016/j.jclinepi.2022.04.024.
- Eichner FA, Groenwold RHH, Grobbee DE, Oude Rengerink K. Systematic review showed that stepped-wedge cluster randomized trials often did not reach their planned sample size. *J Clin Epidemiol* 2019;107:89–100 Epub 2018 Nov 17. PMID: 30458261. doi: 10.1016/j.jclinepi.2018.11.013.
- 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 Epub 2016 Jan 7. PMID: 26748662; PMCID: PMC4949721. doi: 10.1002/sim.6850.
- Ma L, Hu X, Song L, Chen X, Ouyang M, Billot L, Li Q, Malavera A, Li X, Muñoz-Venturelli P, de Silva A, Thang NH, Wahab KW, Pandian LD, Wasay M, Pontes-Neto OM, Abanto C, Arauz A, Shi H, Tang G, Zhu S, She X, Liu L, Sakamoto Y, You S, Han Q, Crutzen B, Cheung E, Li Y, Wang X, Chen C, Liu F, Zhao Y, Li H, Liu Y, Jiang Y, Chen L, Wu B, Liu M, Xu J, You C, Anderson CS. INTERACT3 investigators. The third intensive care bundle with blood pressure reduction in acute cerebral haemorrhage trial (INTERACT3): an international, stepped wedge cluster randomised controlled trial. *Lancet* 2023;402(10395):27–40 Epub 2023 May 25. Erratum in: *Lancet*. 2023 Jul 15;402(10397):184. PMID: 37245517. doi: 10.1016/S0140-6736(23)00806-1.
- Campbell MJ, Hemming K, Taljaard M. The stepped wedge cluster randomised trial: what it is and when it should be used. *Med J Aust* 2019;210(6):253–4 e1 Epub 2019 Feb 13. PMID: 30761546. doi: 10.5694/mja2.50018.
- Hooper R, Bourke L. Cluster randomised trials with repeated cross sections: alternatives to parallel group designs. *BMJ* 2015;350:h2925.
- Billot L, Copas A, Leyrat C, Forbes A, Turner EL. How should a cluster randomized trial be analyzed? *J Epidemiol Popul Health* 2024;72(1):202196.
- Leyrat C, Eldridge S, Taljaard M, Hemming K. Practical considerations for sample size calculation for cluster randomized trials. *J Epidemiol Popul Health* 2024;72(1):202198.
- Kenny A, Voldal EC, Xia F, Heagerty PJ, Hughes JP. Analysis of stepped wedge cluster randomized trials in the presence of a time-varying treatment effect. *Stat Med* 2022;41(22):4311–39 Epub 2022 Jun 30. PMID: 35774016; PMCID: PMC9481733. doi: 10.1002/sim.9511.
- Seaman S, Pavlou M, Copas A. Review of methods for handling confounding by cluster and informative cluster size in clustered data. *Stat Med* 2014;33(30):5371–87 Epub 2014 Aug 4. PMID: 25087978; PMCID: PMC4320764. doi: 10.1002/sim.6277.