



Use of multiple period, cluster randomised, crossover trial designs for comparative effectiveness research

Karla Hemming,¹ Monica Taljaard,^{2,3} Charles Weijer,⁴ Andrew B Forbes⁵

¹Institute of Applied Health Research, University of Birmingham, Birmingham B15 2TT, UK

²Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, ON, Canada

³School of Epidemiology and Public Health, University of Ottawa, Ottawa, ON, Canada

⁴Departments of Medicine, Epidemiology, and Biostatistics, and Philosophy, Western University, London, ON, Canada

⁵School of Public Health and Preventive Medicine, Monash University, Melbourne, VIC, Australia

Correspondence to: K Hemming k.hemming@bham.ac.uk (ORCID 0000-0002-2226-6550) Additional material is published online only. To view please visit the journal online.

Cite this as: *BMJ* 2020;371:m3800 <http://dx.doi.org/10.1136/bmj.m3800>

Accepted: 18 September 2020

Many treatments are adopted into clinical practice without a solid evidence base and might be used heterogeneously across settings. Rigorous randomised controlled trials are therefore needed to inform decisions about the comparative effectiveness of treatments in common use. The mainstay of comparative effectiveness research is pragmatic trial design, which emphasises broad eligibility criteria, simple logistics, routinely collected outcome data, and cost efficient designs. Although treatment differences at the individual level might be small, they can become important when aggregated across large populations. To detect these small differences, very large trials are often required. In multiple period, cluster randomised, crossover trials, the study design randomises clusters (eg, hospitals) to exposure to different interventions in a randomly determined order, and is an attractive design for comparative effectiveness research. The trial design can be highly statistically efficient, compared with other competing designs, and can have many logistical advantages. Several

prominent examples of this trial design have been published recently, yet practical guidance is lacking on how best to design these trials to ensure that they provide robust evidence. Some considerations include how to determine the frequency and number of crossovers, the importance of a time balanced design, how to determine the required sample size, and how to analyse appropriately. The justification for using this design (over a design randomising on the level of individual patients) also raises ethical concerns when used to evaluate individual level interventions without the prospective informed consent of individual participants. In this article, we outline the key methodological and ethical requirements needed for the robust design of multiple period, cluster randomised, crossover trials.

Many treatments are adopted into clinical practice without a solid evidence base and might be used heterogeneously across settings.^{1 2} Furthermore, differences between competing treatments are often small but can nevertheless be important at a population level. Large, pragmatic, randomised controlled trials are needed to inform clinical decision making.³⁻⁵ The multiple period, cluster randomised, crossover trial design is relatively new, in which clusters (such as wards or hospitals) cross between interventions multiple times, such that the treatment a patient receives is dictated by the treatment their cluster (eg, hospital) has been allocated to in the period they present for treatment. Treatments in each cluster period are therefore determined at random, and often in such a way that clusters alternate between the treatments under comparison (box 1 outlines an example trial^{2 6}). The design is statistically powerful and can also be resource efficient.⁷ In settings where the intervention can be implemented and withdrawn quickly, where concerns about the effect of an intervention persisting from one period to the next would be minimal, and where whole or representative clusters can be included, the design has potential to advance evidence based medicine.

SUMMARY POINTS

In multiple period, cluster randomised, crossover trials, clusters randomly cross between the treatment and control conditions many times over many periods

The trial design can be highly statistically efficient, compared with other competing designs, and can have many logistical advantages

In settings where the intervention can be implemented and withdrawn quickly, and where whole or representative clusters can be included, the design has potential to answer important questions on the comparative effectiveness of routinely used treatments

Risks of biases need to be minimised, and researchers should consider blinding of treatments, a time balanced design, and an analysis that allows for residual time confounding and temporal correlations

Several prominent examples of this trial design have been published in recent years.⁸⁻¹² Although some trials have evaluated interventions at the cluster level (eg, electronic reminders integrated into the electronic healthcare record¹³ and enhanced versus conventional cleaning strategies in critical care¹⁴), many have evaluated interventions delivered directly to individuals. Such interventions are often already used routinely, but limited evidence could exist to support a preference for one over the other. Examples include different antibiotic strategies in hospitals,¹⁵⁻¹⁷ patient level decontamination strategies,^{18 19} antiseptic treatments,²⁰⁻²² different types of heparin solutions,²³ gastric ulcer prophylaxis,¹¹ and different types of saline solutions.^{9 10}

In this paper, we consider the what, why, when, and how aspects of the trial design, by:

- Describing the characteristics of the cluster crossover design in general
- Considering advantages over designs for both conventional cluster randomised trials and individual patient randomised trials
- Describing the conditions needed for the design to be feasible
- Providing practical advice on implementation, as well as recommendations for analysis and sample size calculation.

What is the multiple period, cluster randomised, crossover trial design?

Randomised trials have two major types: the individual patient randomised trial and the cluster randomised trial. In individual patient randomised trials, patients are allocated to different arms, which might be treatment or control conditions or two or more active treatments

(for simplicity we refer to these as control and treatment conditions). In contrast, the cluster randomised trial randomises entire clusters.²⁴ The conventional form of the cluster randomised trial is the parallel design, which randomises clusters to either control or treatment conditions (fig 1A). In individual patient randomised trials, patients might be randomly allocated to receive both treatment and control conditions in a random order—in a crossover design.²⁵ In the cluster randomised crossover design, clusters cross between treatment and control conditions.²⁶⁻²⁸ Most commonly, cluster crossover designs have two periods such that each cluster spends one period in the treatment condition and one period in the control condition (perhaps interspersed with a washout period; fig 1B).^{26 29} In multiple period, cluster randomised, crossover trials, clusters cross between the treatment and control conditions many times over many periods, typically simply alternating between the conditions (fig 1C and fig 1D), although other ordering patterns (formally known as sequences) are possible (eg, treatment, control, control, and treatment in four periods).

Although the design might be used for comparing more than two treatments, here we focus on designs comparing two treatments, where effects are not expected to persist after the treatment is switched and where the effect of the intervention is not expected to change over the duration of the study. We also focus on designs where different people gradually present over time as eligible participants, participants are measured once in each period, and each participant mostly receives one of the two treatments, which would typically be the case in hospital based settings.³⁰ Here, we do not consider the special case of studies in one cluster.

Why use the cluster crossover design?

Increased statistical efficiency over conventional parallel cluster randomised trials

Cluster randomised trials require a larger sample size than the individual patient randomised design.²⁴ The increase in sample size is due to participants within a cluster tending to have more similar outcomes than participants across different clusters, and therefore provide less information than that of the same number of participants in separate clusters. In individual patient randomised trials, designs that randomise patients to cross between treatments are known to increase statistical precision.^{31 32} The cluster randomised crossover design, by crossing over interventions, offers the chance to recuperate some of the efficiency losses due to cluster randomisation. This increase in statistical efficiency arises because each cluster receives both treatment conditions thus allowing comparisons within clusters.³³ These comparisons within clusters have other advantages: even in studies with small number of clusters, they eliminate chance imbalances between intervention and control conditions on any time invariant, cluster level characteristics. Furthermore, increasing the number of periods can help eliminate chance imbalance on any time varying, cluster level characteristics. In

Box 1: Example of a multiple period, cluster randomised, crossover trial design

The Prevention of Arrhythmia Device Infection Trial (PADIT) had a multicentre, multiple period, cluster randomised, crossover design that compared different strategies of antimicrobial prophylaxis in high risk patients undergoing the commonly used procedure of cardiac rhythm device implantation in specialist treatment centres in Canada and the Netherlands.²⁶ The specialist treatment centres each adopted their own highly standardised operating procedures for prevention of device infection.² The trial compared two different types of antimicrobial prophylaxis. Under the control condition, patients received one dose of preoperative antibiotics; under the intervention condition, patients received a strategy of incremental antimicrobial prophylaxis.

The unit of randomisation was a treatment centre (cluster); and the trial included 28 centres. The trial lasted for 24 months, and during this period, the type of antimicrobial strategy used in the treatment centre alternated between the two strategies every six months. Randomisation was to one of four alternating sequences of treatment (T) and control (C): TCTC, TCCT, CTCT, CTTC. The primary outcome was hospital admission for device infection. The study included data on 12 842 procedures and used only routinely collected outcome data. Most centres approved trial conduct with waiver of consent so that patients were not aware they were participating in a trial, although some centres required patient consent for data collection.

The investigators claimed that their trial design allowed inferences to be generalised to the entire population of high risk patients, and can thus could inform decisions at the level of the healthcare provider as to which treatment strategy reduces the average number of device related infections in the population.

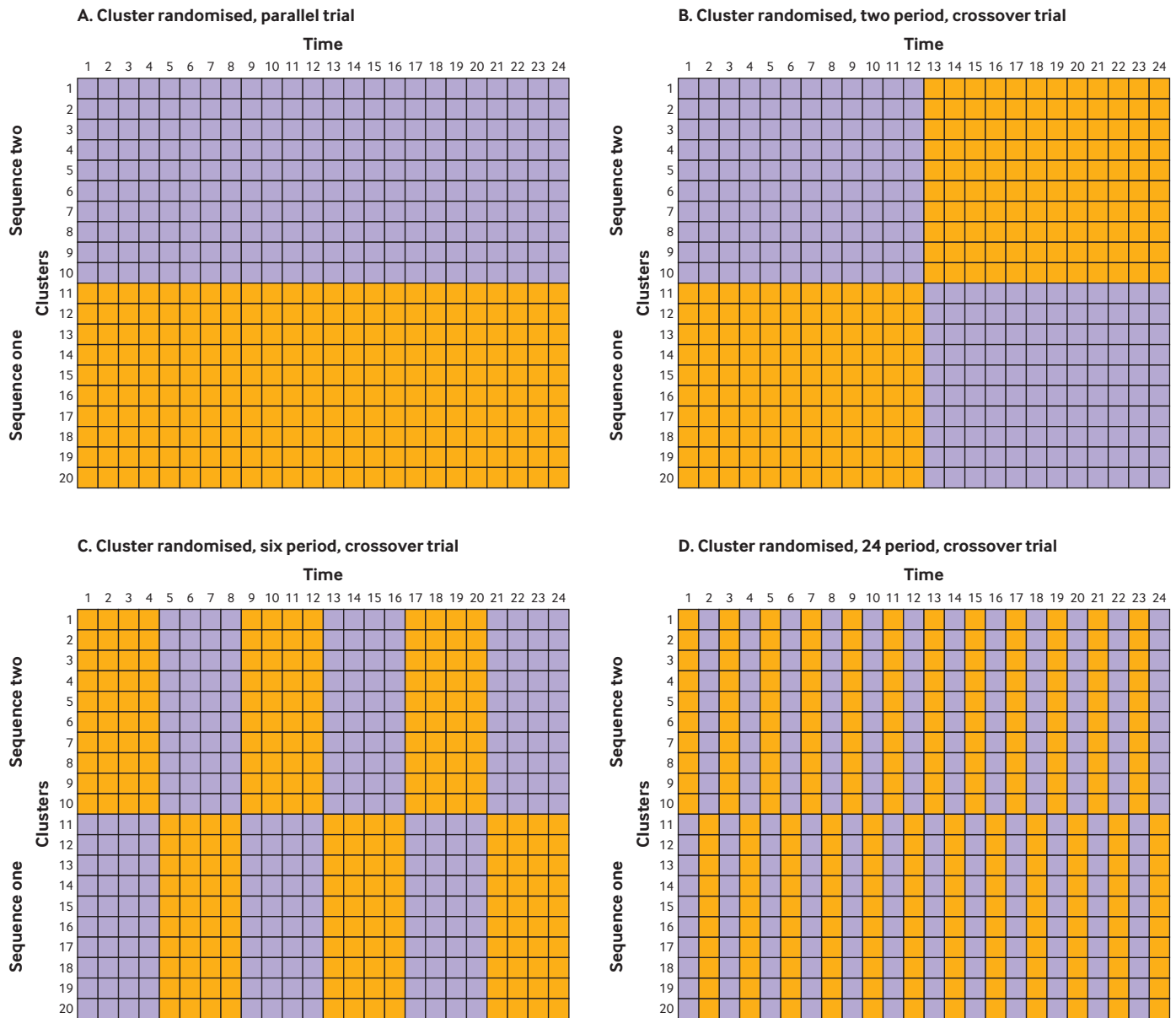


Fig 1 | Study designs of example cluster randomised trials over 24 months. Purple=exposure to control; orange=exposure to intervention. Example includes 20 clusters for illustrative purposes

a hypothetical example (box 2), we illustrate how chance imbalances might affect inferences in a cluster trial with a small number of clusters.

Reduced logistical complexity over individual patient randomised trials

Trials randomising individual patients can be resource intensive and costly,^{34 35} which could prevent socially important research from taking place. Cluster randomised designs, on the other hand, can be more resource efficient: only one type of intervention needs to be delivered in a cluster at any particular time, and so the complexity of the research infrastructure and associated costs can be reduced. Yet, parallel cluster randomised trials by nature are statistically inefficient, which makes them an impractical choice for detecting small but clinically important differences

in comparative effectiveness research. The cluster crossover design, by including multiple switches can, however, substantially reduce the required number of clusters, and thus can be a resource efficient choice.

When can the design for multiple period, cluster randomised crossover trials be used?

Treatments should be easily switched back and forth

Carry-over effects refer to the persistence of treatments into the next period (or even subsequent periods). The cluster crossover design will only be suitable when treatment conditions can be removed easily and when there is no carry-over. Consequently, the design will not be suitable for evaluating knowledge translation or behaviour change interventions, such as those that seek to encourage the uptake of evidence based practice.

Box 2: Hypothetical example of the multiple period, cluster randomised, crossover trial design

A parallel trial randomises 20 hospitals to treatment or control and they are observed for 24 months (fig S1A). The 10 hospitals allocated to the treatment might, by chance, be different from those allocated to the control (eg, they might all be teaching hospitals). Such chance imbalance is unlikely to occur in trials randomising many clusters, but in practice, cluster trials often include a small number of clusters.¹²

Table 1 illustrates an example of cluster level characteristics for different cluster randomised trials. In this example, teaching hospital status would be imbalanced across the treatment conditions (table row 2). However, in a two period design for a cluster randomised crossover trial (fig S1B), all 20 hospitals would be observed equally under both conditions for 12 months (table row 6).

Table 1 Example of cluster level characteristics in different cluster randomised trials*

Row	Characteristic	Parallel cluster randomised trial		Two period CRXO		Six period CRXO		24 period CRXO	
		C	T	C	T	C	T	C	T
1	No of clusters	10	10	—	—	—	—	—	—
2	Teaching hospital: clusters	0§	10 (100) §	—	—	—	—	—	—
3	Influenza outbreak: clusters	10 (100)‡	10 (100)‡	—	—	—	—	—	—
4	Quality assurance: clusters	0§	10 (100)§	—	—	—	—	—	—
5	No of cluster periods	—	—	240‡	240‡	240‡	240‡	240‡	240‡
6	Teaching hospital: cluster periods	—	—	120 (50)‡	120 (50)‡	120 (50)‡	120 (50)‡	120 (50)‡	120 (50)‡
7	Influenza outbreak: cluster periods‡	—	—	60 (25)‡	60 (25)‡	60 (25)‡	60 (25)‡	60 (25)‡	60 (25)‡
8	Quality assurance: cluster periods	—	—	120 (50)§	0§	40 (17)§	80 (33)§	60 (25)‡	60 (25)‡

Data are number (%) unless stated otherwise. C=control; T=treatment; CRXO=cluster randomised, crossover trial.
 *Figure S1 shows the designs and interaction of clusters, treatment, and cluster level characteristics (teaching hospital and exposure to quality assurance programme). Example assumes that each cluster includes same number of observations in each cluster period, and that teaching hospital status does not change over duration of study but that quality assurance status changes from 0 to 1 for clusters allocated to the second sequence.
 †Influenza outbreak is a binary indicator of whether the cluster or cluster period is observed under the influenza outbreak (second half of the design, all clusters).
 ‡Balanced characteristics across designs.
 §Imbalanced characteristics across designs.

However, a risk of imbalance remains due to temporal changes (eg, seasonal effects). Suppose an influenza outbreak (or some other communicable disease) occurs in the second half of the trial, and suppose it affects all hospitals equally, causing increased endpoint event rates. By adopting a design with an equal number of the opposite of each treatment sequence (eg, TCCT and CTTC, where T=treatment and C=control), the increased event rate in the second half of the study would occur equally in both treatment and control conditions (table row 7). The conventional parallel cluster randomised trial would show good balance on this characteristic because the parallel cluster randomised trial also induces a natural balance on time (table row 3).

More often temporal changes occur in some clusters and not others (eg, it is unlikely that all hospitals will be affected by exactly the same degree by any influenza outbreak). Suppose the 10 hospitals allocated to receive the intervention first (eg, those allocated to the second sequence), happen to be the 10 teaching hospitals, and they are coincidentally requested to participate in a quality assurance programme during the second period of the study (fig S1B). Exposure to the quality assurance programme shows imbalance under both the parallel and two period, cluster randomised, crossover design (table row 4). Yet, increasing the number of periods can mitigate these imbalances. For example, a six period and 24 period design (figs S1C and S1D) achieve greater balance on the exposure to the quality assurance programme. This is because the multiple crossovers increase the likelihood of balance on time varying, cluster level characteristics that affect clusters differently. With a large number of clusters, randomisation would prevent such chance imbalance.

Even in those studies where the intervention can be removed, washout periods might still be needed to allow for transition between treatment conditions. These washout periods will be important if an instantaneous transition between treatment conditions cannot be made or if participants have a prolonged exposure to the intervention. For example, an intensive care department might easily be able to withdraw a saline solution (the intervention) from use, seemingly obviating the need to include a washout period. However, a washout period (sometimes referred to as closure periods) might still be desirable to allow patients who have already started treatment to continue on the same treatment and to allow delayed assessment of outcomes (such as length of intensive care stay) before the cluster transitions to the next treatment condition.³⁶

Risk of recruitment bias should be low

The cluster crossover design is particularly attractive when whole clusters can be included—that is, without

the need to recruit participants by obtaining a waiver of consent.¹ Post-randomisation recruitment is a well known source of bias in cluster randomised trials, especially when participants and those individuals recruiting them into the trial cannot be blinded.^{37 38} In cluster crossover designs, any patient recruitment necessarily occurs after cluster randomisation; thus, the design is particularly vulnerable to this form of bias. While broad eligibility criteria and recruitment by someone blind to the treatment condition might mitigate these biases, such procedures can increase complexities.³⁹ The design has thus become particularly popular in settings with broad eligibility criteria and without individual participant recruitment.

How should the design be implemented?

Waivers of consent should only be used when ethically appropriate

Although it might be tempting to use the increased risks of bias as a motivation to implement a cluster

crossover trial without patient consent,⁴⁰ waivers of consent are generally not acceptable for individual level therapeutic interventions.⁴¹⁻⁴³ International ethics guidelines articulate a general requirement to obtain the informed consent of research participants.⁴⁴ A research ethics committee might approve a waiver of consent provided that three conditions are fulfilled: the study offers the prospect of substantial social value; if informed consent was required, the study would be impracticable or impossible to conduct; and study participation poses only minimal risk.⁴⁴ In cluster randomised trials, it has been argued that the use of a waiver of consent should be extended to include cluster level interventions.⁴⁵ This is because cluster members typically cannot avoid interventions delivered at the level of the cluster (eg, public health messages), which precludes the meaningful refusal of study participation.⁴⁵

Yet, recent literature has witnessed an attempt to expand the use of waiver of consent in cluster crossover designs to include drugs, such as benzodiazepines in cardiac anaesthesia and prophylactic antibiotics in cardiac pacemaker implantation.^{6 34} Investigators have framed these interventions as policy interventions (that is, cluster interventions that are not amenable to patient consent).^{34 40} However, drugs given as part of an institutional policy are not cluster level interventions, because they remain divisible at the individual level.^{35 46} Further work is needed to develop scientifically robust and ethically appropriate strategies (such as modified methods of consent) to recruit participants into pragmatic trials that evaluate routinely used medical interventions.⁴⁷

Increasing the number of treatment switches increases statistical efficiency but reduces practicality

Increasing the number of treatment switches (or crossovers), in a design of fixed duration, is likely to increase the statistical precision (that is, narrow the width of the confidence interval), at least under realistic correlation assumptions. However, gains could be tapered beyond eight crossovers.³³ Moreover, increasing the number of crossovers will increase the likelihood of balance across treatment and control conditions on any time varying, cluster level characteristics. Furthermore, increasing the number of clusters will increase the likelihood of balance across sequences on time invariant, cluster characteristics (this balance might be less critical because comparisons within clusters will cancel out any such imbalances). On the other hand, increasing the number of crossovers and clusters will likely increase the logistical complexity, and could increase costs and reduce adherence.

Alternating sequence increases statistical efficiency but induces predictability

Much of the theoretical appeal of the cluster crossover design stems from having a balance in control and treatment conditions in each time period and observing each cluster an equal number of periods in

each treatment condition (fig 1B, fig 1C, and fig 1D). In the technical literature on crossover designs, this is referred to as being uniform on periods and clusters,³¹ although for simplicity we refer to this as balance. First and foremost, designs that are time balanced, in theory and under certain conditions (such as equal cluster sizes and a large number of clusters), reduce the need to adjust for period effects. When the clusters alternate between the two conditions, the design can be shown empirically to have greater statistical precision (under certain conditions about the correlation structure; supplementary material 1). The Prevention of Arrhythmia Device Infection Trial (PADIT; box 1) used the sequences TCTC, TCCT, CTCT, and CTTC (where T=treatment and C=control, and where seven clusters were allocated to each of these four sequences). The design is time balanced (that is, in any given time period, 14 clusters are under the control condition and 14 are under the treatment condition) and uniform on clusters (that is, every cluster spends two periods in the control condition and two periods in the treatment condition), but might not be statistically optimal, because clusters allocated to sequences two and four do not alternate between treatment conditions.

However, while a design with alternating treatments creates a statistically powerful design, it induces predictability into the allocation. Predictability of the treatment sequence is likely to be inconsequential when the treatments are blinded.^{48 49} Or, for example, in evaluations of interventions in an emergency department, where the fact that the upcoming allocation is predictable is unlikely to delay treatment until the next period. Yet, in studies that include patient recruitment and where the treatment is not blinded (as is often the case in pragmatic trials), knowledge of an upcoming switch to a treatment perceived as preferable might impede recruitment and induce an imbalance of recruitment under different treatment conditions.

Consequently, study designs need to consider both statistical efficiency and risks of predictability. Allocating an equal number of clusters to dual pairs of sequences such that the second sequence in any pair is the inverse of the first, will help ensure a time balanced design. Sequences consisting of randomised permuted blocks (where each block contains an equal replication of treatment conditions) will help prevent predictability of upcoming assignment (fig S3). Statistically efficient alternating sequences should only be used where concerns over predictability are minimal.

Realised designs should retain properties of time balance

In practice, studies might not be amenable to recruiting all clusters simultaneously. This means that even in studies where clusters are randomised to alternating or dual sequences, if clusters commence these sequences at different points in time, the realised design might not necessarily contain an equal replication of clusters under control and treatment conditions in each period (that is, it will not be balanced on time). Supplementary figure S2 illustrates a realised design showing time

imbalance. Stratifying the trial into groups based on when clusters are able to commence participation (where each group contains alternating or dual sequences) can help promote balance on time within groups and allow the possibility of different clusters commencing the trial at different points in time.¹¹

Additionally, although it might be thought that the length of each observation period for each cluster should be tailored so that the same number of patients are assessed in each period, this creates an imbalance on time periods across clusters. Therefore, timing of the switches should be based on calendar time rather than patient volume, even if that results in different numbers of patients across clusters in each period. Again, stratifying the trial into groups (based on size of cluster) and randomising within groups can help promote balance.

Blinding of participants, providers, and outcome assessors should be considered carefully

An important consideration in a cluster crossover trial is who should be blinded to the allocated treatment. Different considerations might apply at the stages of identification or recruitment into the trial (if any) and during stages of intervention delivery and outcome assessment. Blinding of research staff or those individuals identifying or recruiting participants is considered above. Furthermore, blinding of outcome assessors should be considered when outcomes are subjective. The question of whether to disguise treatments from participants and providers (eg, treating physicians) after recruitment is more nuanced. An important consideration is that cluster crossover trials often have a pragmatic intent so that interest is usually in the total effect of an intervention under usual care conditions where patients and providers are aware of the treatment. Furthermore, blinding might be difficult or impossible for non-pharmacological interventions. However, blinding might nevertheless be useful, when decision makers are interested in the comparative effectiveness of treatments based purely on biological activity.^{50 51} On the other hand, even when participants and providers are blinded, the possibility remains that a physician with a strong preference for one or the other treatment might source treatments from elsewhere.

What are the added statistical complications?

Analysis should allow for both clustering and temporal correlations

Analysis must allow not only for intraclass correlations (as is essential in cluster randomised trials) but also for decay in the strength of correlations within clusters with increasing time separation between measurements. In two period, crossover designs, this analysis is achieved using random effects for clusters and random effects for cluster periods.^{26 52} This approach allows for non-independence of observations within clusters, such that correlations between observations in the same period can be larger than correlations between observations in

different periods. In designs with more than two periods, correlation structures have been proposed to be allowed to decay with increasing separation between measurements.^{53 54} One approach that can accommodate such correlation structures is with linear or generalised linear mixed models, with appropriate small sample corrections for trials with a small number of clusters.⁵⁵ However, bias can be induced if the correlation structure is misspecified.⁵⁶ Alternative approaches might be to use generalised estimating equations, which are robust to misspecification of the correlation structure when there are a large number of clusters.⁵⁷

Analysis should adjust for period effects

Residual confounding of period effects can occur in studies that are designed to be balanced on time, but for various reasons have some imbalance on time.⁵⁸ This can occur when clusters drop out, when cluster sizes vary, when clusters initiate a sequence late, or when the randomisation does not ensure a time balanced design. Adjusting for period effects in the analytical model can be achieved by including time as categorical variable or as linear or spline functions, typically under the assumption of a common secular trend across all clusters. Trials that are imbalanced on time and do not adjust for time effects or those that adjust for time effects using misspecified forms will provide biased estimates of treatment effects or standard errors, resulting in under or over coverage of their confidence intervals. The consequences of model misspecification of time effects are likely to be greater for trials with either a small number of clusters or small number of periods, and they become increasingly less consequential as the number of clusters or periods increase.

Sample size calculations do not always collapse into a simple design effect

Sample size calculations must also allow for correlations both within clusters and across time within clusters. In two period, crossover designs, these correlations are often categorised by the intraclass correlation within periods and the intraclass correlation between periods.^{59 60} Under this assumed correlation structure, sample sizes can easily be determined by inflating the sample size needed under individual randomisation by a design effect. However, these correlation structures are unlikely to be sufficient for cluster crossover designs with more than two periods: these designs are likely to lend themselves to a correlation structure that allows correlations to decay with increasing separation of measurements.^{53 54} Unfortunately, under these more complicated correlation models, sample size inflation cannot be characterised by simple design effects. However, analytical solutions are available; thus, the required number of clusters could be determined given a specified total cluster size and vice versa.⁶¹ These calculations require specification of correlation parameters, which are ideally estimated from relevant routinely collected data, from the same set of clusters, and over a similar time period. We provide an

illustrative sample size calculation for the PADIT trial (supplementary material 1) implemented using an RShiny app.⁶¹

Conclusion

The complexities and costs associated with conducting individual patient randomised trials have resulted in a lack of evidence about comparative effectiveness of routinely used interventions in medical practice. The multiple period, cluster randomised, crossover design has the potential to advance the evidence base for comparative effectiveness of low risk interventions in routine use. The unique appeal of the design lies in the combination of cluster implementation and crossover aspects. The cluster randomisation provides enhanced operational and logistical appeal while the crossover aspect enhances statistical power and reduces chance imbalances. Provided that the design includes a sufficient number of time periods and clusters, and is well balanced across time and clusters, it has the potential to provide robust and internally valid evaluation. When the design includes a representative sample of clusters, it is more likely to have the potential to provide generalisable evidence and allow evaluation of consistency of effects across settings or clusters.

Risks of bias associated with the design include identification and recruitment biases in studies where there is individual patient recruitment; as well as carry-over effects when used in the evaluation of interventions whose effects persist over periods of the study; or from deviations from intended treatments, which might be problematic in unblinded trials. The design is likely to be most appealing from internal and external validity perspectives when it includes whole clusters. For the evaluation of individual level interventions, the design will probably only be part of the toolkit when not seeking full individual patient consent is ethically appropriate.

Contributors: KH led the development of the idea. All the authors contributed to the development of the paper, made an intellectual contribution to the development of the ideas, commented on the draft version of the paper, and approved the final version of the paper. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Funding: This research was partly funded by the UK National Institute for Health Research (NIHR) Collaborations for Leadership in Applied Health Research and Care West Midlands initiative, and the Australian National Health and Medical Research Council (NHMRC; grant 1183303). KH is funded by an NIHR senior research fellowship (SRF-2017-10-002).

Competing interests: All authors have completed the ICMJE uniform disclosure form at www.icmje.org/doi_disclosure.pdf and declare: part support from the UK NIHR Collaborations for Leadership in Applied Health Research and Care West Midlands initiative and Australian NHMRC for the submitted work; CW receives consulting income from Cardialen, Eli Lilly, and Research Triangle Institute International; all other authors declare no support from any organisation for the submitted work no financial relationships with any organisations that might have an interest in the submitted work in the previous three years, no other relationships or activities that could appear to have influenced the submitted work.

The lead author affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned have been explained.

Provenance and peer review: Not commissioned; externally peer reviewed.

- Bellomo R, Forbes A, Akram M, Bailey M, Pilcher DV, Cooper DJ. Why we must cluster and cross over. *Crit Care Resusc* 2013;15:155-7.
- Connolly SJ, Philippon F, Longtin Y, et al. Randomized cluster crossover trials for reliable, efficient, comparative effectiveness testing: design of the Prevention of Arrhythmia Device Infection Trial (PADIT). *Can J Cardiol* 2013;29:652-8. doi:10.1016/j.cjca.2013.01.020
- Schwartz D, Lellouch J. Explanatory and pragmatic attitudes in therapeutic trials. *J Chronic Dis* 1967;20:637-48. doi:10.1016/0021-9681(67)90041-0
- Zwarenstein M, Treweek S. What kind of randomised trials do patients and clinicians need? *Evid Based Med* 2009;14:101-3. doi:10.1136/ebm.14.4.101
- Angus DC. Fusing randomized trials with big data: the key to self-learning health care systems? *JAMA* 2015;314:767-8. doi:10.1001/jama.2015.7762
- Krahn AD, Longtin Y, Philippon F, et al. Prevention of arrhythmia device infection trial: the PADIT trial. *J Am Coll Cardiol* 2018;72:3098-109. doi:10.1016/j.jacc.2018.09.068
- Hayes MJ, Morris GK, Hampton JR. Comparison of mobilization after two and nine days in uncomplicated myocardial infarction. *Br Med J* 1974;3:10-3. doi:10.1136/bmj.3.5922.10
- Young P, Bailey M, Beasley R, et al, SPLIT Investigators, ANZICS CTG. Effect of a buffered crystalloid solution vs saline on acute kidney injury among patients in the intensive care unit: the SPLIT randomized clinical trial. *JAMA* 2015;314:1701-10. doi:10.1001/jama.2015.12334
- Semler MW, Self WH, Wanderer JP, et al, SMART Investigators and the Pragmatic Critical Care Research Group. Balanced crystalloids versus saline in critically ill adults. *N Engl J Med* 2018;378:829-39. doi:10.1056/NEJMoa1711584
- Self WH, Semler MW, Wanderer JP, et al, SALT-ED Investigators. Balanced Crystalloids versus saline in noncritically ill adults. *N Engl J Med* 2018;378:819-28. doi:10.1056/NEJMoa1711586
- PEPTIC Investigators for the Australian and New Zealand Intensive Care Society Clinical Trials Group, Alberta Health Services Critical Care Strategic Clinical Network, Irish Critical Care Trials Group. 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;323:616-26.
- Arnup SJ, Forbes AB, Kahan BC, Morgan KE, McKenzie JE. The quality of reporting in cluster randomised crossover trials: proposal for reporting items and an assessment of reporting quality. *Trials* 2016;17:575. doi:10.1186/s13063-016-1685-6
- Lang E, Afilalo M, Vandal AC, et al. Impact of an electronic link between the emergency department and family physicians: a randomized controlled trial. *CMAJ* 2006;174:313-8. doi:10.1503/cmaj.050698
- Wilson AP, Smyth D, Moore G, et al. The impact of enhanced cleaning within the intensive care unit on contamination of the near-patient environment with hospital pathogens: a randomized crossover study in critical care units in two hospitals. *Crit Care Med* 2011;39:651-8. doi:10.1097/CCM.0b013e318206bc66
- Nijssen S, Fluit A, van de Vijver D, Top J, Willems R, Bonten MJ. Effects of reducing beta-lactam antibiotic pressure on intestinal colonization of antibiotic-resistant gram-negative bacteria. *Intensive Care Med* 2010;36:512-9. doi:10.1007/s00134-009-1714-y
- van Werkhoven CH, Postma DF, Oosterheert JJ, Bonten MJ. Antibiotic treatment of moderate-severe community-acquired pneumonia: design and rationale of a multicentre cluster-randomised cross-over trial. *Neth J Med* 2014;72:170-8.
- Postma DF, van Werkhoven CH, van Elden LJ, et al, CAP-START Study Group. Antibiotic treatment strategies for community-acquired pneumonia in adults. *N Engl J Med* 2015;372:1312-23. doi:10.1056/NEJMoa1406330
- de Smet AM, Kluytmans JA, Cooper BS, et al. Decontamination of the digestive tract and oropharynx in ICU patients. *N Engl J Med* 2009;360:20-31. doi:10.1056/NEJMoa0800394
- Wittekamp BH, Plantinga NL, Cooper BS, et al. Decontamination strategies and bloodstream infections with antibiotic-resistant microorganisms in ventilated patients: a randomized clinical trial. *JAMA* 2018;320:2087-98. doi:10.1001/jama.2018.13765
- Calfee DP, Farr BM. Comparison of four antiseptic preparations for skin in the prevention of contamination of percutaneously drawn blood cultures: a randomized trial. *J Clin Microbiol* 2002;40:1660-5. doi:10.1128/JCM.40.5.1660-1665.2002
- Parietti JJ, du Cheyron D, Ramakers M, et al, Members of the NACRE Study Group. Alcoholic povidone-iodine to prevent central venous catheter colonization: A randomized unit-crossover study. *Crit Care Med* 2004;32:708-13. doi:10.1097/01.CCM.0000115265.05604.7B
- Washer LL, Chenoweth C, Kim HW, et al. Blood culture contamination: a randomized trial evaluating the comparative effectiveness of 3 skin antiseptic interventions. *Infect Control Hosp Epidemiol* 2013;34:15-21. doi:10.1086/668777

- 23 Epperson EL. Efficacy of 0.9% sodium chloride injection with and without heparin for maintaining indwelling intermittent injection sites. *Clin Pharm* 1984;3:626-9.
- 24 Eldridge S, Kerry S. Introduction. In: *A practical guide to cluster randomised trials in health services research*. John Wiley and Sons, 2012. doi:10.1002/9781119966241.ch1
- 25 Jones B, Kenward MG. *Design and analysis of cross-over trials*. Chapman and Hall/CRC, 2014. doi:10.1201/b17537
- 26 Parienti JJ, Kuss O. Cluster-crossover design: a method for limiting clusters level effect in community-intervention studies. *Contemp Clin Trials* 2007;28:316-23. doi:10.1016/j.cct.2006.10.004
- 27 Forbes AB, Akram M, Pilcher D, Cooper J, Bellomo R. Cluster randomised crossover trials with binary data and unbalanced cluster sizes: application to studies of near-universal interventions in intensive care. *Clin Trials* 2015;12:34-44. doi:10.1177/1740774514559610
- 28 Amup SJ, McKenzie JE, Hemming K, Pilcher D, Forbes AB. Understanding the cluster randomised crossover design: a graphical illustration of the components of variation and a sample size tutorial. *Trials* 2017;18:381. doi:10.1186/s13063-017-2113-2
- 29 Amup SJ, Forbes AB, Kahan BC, Morgan KE, McKenzie JE. Appropriate statistical methods were infrequently used in cluster-randomized crossover trials. *J Clin Epidemiol* 2016;74:40-50. doi:10.1016/j.jclinepi.2015.11.013
- 30 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. doi:10.1186/s13063-015-0842-7
- 31 Matthews JN. Multi-period crossover trials. *Stat Methods Med Res* 1994;3:383-405. doi:10.1177/096228029400300405
- 32 Senn S. Cross-over trials in Statistics in Medicine: the first '25' years. *Stat Med* 2006;25:3430-42. doi:10.1002/sim.2706
- 33 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;38:5021-33. doi:10.1002/sim.8349
- 34 Spence J, Belley-Côté E, Lee SF, et al. The role of randomized cluster crossover trials for comparative effectiveness testing in anesthesia: design of the Benzodiazepine-Free Cardiac Anesthesia for Reduction in Postoperative Delirium (B-Free) trial. *Can J Anaesth* 2018;65:813-21. doi:10.1007/s12630-018-1130-2
- 35 Goldstein CE, Giraudeau B, Weijer C, Taljaard M. When and how should we cluster and cross over: methodological and ethical issues. *Can J Anaesth* 2018;65:760-5. doi:10.1007/s12630-018-1131-1
- 36 Hooper R, Copas A. Stepped wedge trials with continuous recruitment require new ways of thinking. *J Clin Epidemiol* 2019;116:161-6. doi:10.1016/j.jclinepi.2019.05.037
- 37 Puffer S, Torgerson D, Watson J. Evidence for risk of bias in cluster randomised trials: review of recent trials published in three general medical journals. *BMJ* 2003;327:785-9. doi:10.1136/bmj.327.7418.785
- 38 Bolzern J, Mnyama N, Bosanquet K, Torgerson DJ. A review of cluster randomized trials found statistical evidence of selection bias. *J Clin Epidemiol* 2018;99:106-12. doi:10.1016/j.jclinepi.2018.03.010
- 39 Eldridge S, Kerry S, Torgerson DJ. Bias in identifying and recruiting participants in cluster randomised trials: what can be done? *BMJ* 2009;339:b4006. doi:10.1136/bmj.b4006
- 40 Taljaard M, Goldstein CE, Giraudeau B, et al. Cluster over individual randomization: are study design choices appropriately justified? Review of a random sample of trials. *Clin Trials* 2020;17:253-63. doi:10.1177/1740774519896799
- 41 Edwards SJ, Brauholtz DA, Lilford RJ, Stevens AJ. Ethical issues in the design and conduct of cluster randomised controlled trials. *BMJ* 1999;318:1407-9. doi:10.1136/bmj.318.7195.1407
- 42 Medical Research Council. Cluster randomised trials: methodological and ethical considerations. MRC clinical trial series. 2002.
- 43 Sim J, Dawson A. Informed consent and cluster-randomized trials. *Am J Public Health* 2012;102:480-5. doi:10.2105/AJPH.2011.300389
- 44 Council for International Organizations of Medical Sciences (CIOMS). *International Ethical Guidelines for health-related research involving humans – in collaboration with the World Health Organization*. WHO, 2016.
- 45 Weijer C, Grimshaw JM, Eccles MP, et al. Ottawa Ethics of Cluster Randomized Trials Consensus Group. The Ottawa Statement on the Ethical Design and Conduct of Cluster Randomized Trials. *PLoS Med* 2012;9:e1001346. doi:10.1371/journal.pmed.1001346
- 46 Eldridge SM, Ashby D, Feder GS. Informed patient consent to participation in cluster randomized trials: an empirical exploration of trials in primary care. *Clin Trials* 2005;2:91-8. doi:10.1191/1740774505cn0700a
- 47 Kim SY, Miller FG. Informed consent for pragmatic trials--the integrated consent model. *N Engl J Med* 2014;370:769-72. doi:10.1056/NEJMHle1312508
- 48 Altman DG, Bland JM. Statistics notes. Treatment allocation in controlled trials: why randomise? *BMJ* 1999;318:1209. doi:10.1136/bmj.318.7192.1209
- 49 Chalmers I. Why transition from alternation to randomisation in clinical trials was made. *BMJ* 1999;319:1372. doi:10.1136/bmj.319.7221.1372
- 50 European Medicines Agency. ICH E9 (R1) addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials. 2020. https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e9-r1-addendum-estimands-sensitivity-analysis-clinical-trials-guideline-statistical-principles_en.pdf
- 51 Zwahlenstein M, Thorpe K, Treweek S, Loudon K. PRECIS-2 for retrospective assessment of RCTs in systematic reviews. *J Clin Epidemiol* 2020;S0895-4356(20)30743-5.
- 52 Turner RM, White IR, Croudace T, PIP Study Group. Analysis of cluster randomized cross-over trial data: a comparison of methods. *Stat Med* 2007;26:274-89. doi:10.1002/sim.2537
- 53 Kasza J, Hemming K, Hooper R, Matthews J, Forbes AB. Impact of non-uniform correlation structure on sample size and power in multiple-period cluster randomised trials. *Stat Methods Med Res* 2019;28:703-16. doi:10.1177/0962280217734981
- 54 Grantham KL, Kasza J, Heritier S, Hemming K, Forbes AB. Accounting for a decaying correlation structure in cluster randomized trials with continuous recruitment. *Stat Med* 2019;38:1918-34. doi:10.1002/sim.8089
- 55 Leyrat C, Morgan KE, Leurent B, Kahan BC. Cluster randomized trials with a small number of clusters: which analyses should be used? *Int J Epidemiol* 2018;47:321-31. doi:10.1093/ije/dyx169
- 56 Kasza J, Forbes AB. Inference for the treatment effect in multiple-period cluster randomised trials when random effect correlation structure is misspecified. *Stat Methods Med Res* 2019;28:3112-22. doi:10.1177/0962280218797151
- 57 Turner EL, Li F, Gallis JA, Prague M, Murray DM. Review of recent methodological developments in group-randomized trials: part 1-design. *Am J Public Health* 2017;107:907-15. doi:10.2105/AJPH.2017.303706
- 58 Morgan KE, Forbes AB, Keogh RH, Jairath V, Kahan BC. Choosing appropriate analysis methods for cluster randomised cross-over trials with a binary outcome. *Stat Med* 2017;36:318-33. doi:10.1002/sim.7137
- 59 Giraudeau B, Ravaud P, Donner A. Sample size calculation for cluster randomized cross-over trials. *Stat Med* 2008;27:5578-85. doi:10.1002/sim.3383
- 60 Hooper R, Bourke L. Cluster randomised trials with repeated cross sections: alternatives to parallel group designs. *BMJ* 2015;350:h2925. doi:10.1136/bmj.h2925
- 61 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;49:979-95. doi:10.1093/ije/dyz237

Web appendix: Supplementary material