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

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 trial2 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.
Several prominent examples of this trial design have been published in recent years.\textsuperscript{8–12} Although some trials have evaluated interventions at the cluster level (e.g., electronic reminders integrated into the electronic healthcare record\textsuperscript{13} and enhanced versus conventional cleaning strategies in critical care\textsuperscript{14}), 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,\textsuperscript{15,17} patient level decontamination strategies,\textsuperscript{18,19} antiseptic treatments,\textsuperscript{20,22} different types of heparin solutions,\textsuperscript{23} gastric ulcer prophylaxis,\textsuperscript{11} and different types of saline solutions.\textsuperscript{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.\textsuperscript{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.\textsuperscript{25} In the cluster randomised crossover design, clusters cross between treatment and control conditions.\textsuperscript{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).\textsuperscript{24,29} 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 (e.g., 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.\textsuperscript{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.\textsuperscript{31} 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.\textsuperscript{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.\textsuperscript{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
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, 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.
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).

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 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

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

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

<table>
<thead>
<tr>
<th>Row</th>
<th>Characteristic</th>
<th>Parallel cluster randomised trial</th>
<th>Two period CRXO</th>
<th>Six period CRXO</th>
<th>24 period CRXO</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>C</td>
<td>T</td>
<td>C</td>
<td>T</td>
</tr>
<tr>
<td>1</td>
<td>No of clusters</td>
<td>10</td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Teaching hospital: clusters</td>
<td>0§</td>
<td>10 (100) §</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Influenza outbreak: clusters</td>
<td>10 (100)‡</td>
<td>10 (100)‡</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Quality assurance: clusters</td>
<td>0§</td>
<td>10 (100)§</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>No of cluster periods</td>
<td>—</td>
<td>—</td>
<td>240‡</td>
<td>240‡</td>
</tr>
<tr>
<td>6</td>
<td>Teaching hospital: cluster periods</td>
<td>—</td>
<td>—</td>
<td>120 (50)‡</td>
<td>120 (50)‡</td>
</tr>
<tr>
<td>7</td>
<td>Influenza outbreak: cluster periods†</td>
<td>—</td>
<td>—</td>
<td>60 (25)‡</td>
<td>60 (25)‡</td>
</tr>
<tr>
<td>8</td>
<td>Quality assurance: cluster periods</td>
<td>—</td>
<td>—</td>
<td>120 (50)§</td>
<td>0§</td>
</tr>
</tbody>
</table>

Data are number (%) unless stated otherwise. C=control; T=treatment; CRXO=cluster randomised, crossover trial.

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). 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.

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.
crossover trial without patient consent,50 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. 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.54 In cluster randomised trials, it has been argued that the use of a waiver of consent should be extended to include cluster level interventions.6 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 32 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.38 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 intrACLuster 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 intrACLuster correlation within periods and the intrACLuster 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 NHRI senior research fellowship (SRF-2017-1002).

Competing interests: All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_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 Cardialex, 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 planned have been explained.

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


Chalmers I. Why transition from alternation to randomisation in clinical trials was made. BMJ 1999;319:137-9. doi:10.1136/bmj.319.7211.1372


