COMMENTARY

Improving efficiency in cluster-randomized study design and implementation: taking advantage of a crossover

Nicholas G Reich[1]
Aaron M Milstone[2,3]

[1]Division of Biostatistics and Epidemiology, School of Public Health and Health Sciences, UMass-Amherst, Amherst, MA, USA; [2]Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA; [3]Department of Pediatrics, Johns Hopkins University, Baltimore, MD, USA

Abstract: While individually randomized trials have long provided the gold standard of clinical evidence, the use of cluster-randomized trials in biomedical and social scientific research has expanded rapidly in recent years. In certain settings, randomizing by group or cluster can provide distinct advantages over individual randomization. However, a central challenge for cluster-randomized trials is ensuring that the study arms are balanced across important participant characteristics. One method to combat imbalance between study arms is to incorporate a crossover into the study design. In this design, every cluster is observed under each treatment condition, in a randomly assigned sequence. We provide a concrete example of how incorporating a crossover into a cluster-randomized study can improve balance between arms and increase statistical efficiency of a trial. However, a crossover design cannot always be effectively implemented. This commentary illustrates the potential benefits and discusses the challenges and disadvantages to incorporating a crossover in a cluster-randomized study design.

Keywords: cluster-randomized clinical trials, crossover design, controlled comparisons, statistical power, balance, efficiency

Introduction

We are in the midst of a marked national movement toward clinical effectiveness research supported by over US$1 billion of federal funds in the United States alone. As scientists work to broaden our understanding of treatments that create measurable improvement in patient outcomes, they must have the best analytical tools at hand. Individually, randomized trials have long served as the holy grail of scientific evidence. In certain settings, however, persuasive arguments can be made that randomizing by group or cluster (eg, a hospital ward, village, or school) has distinct advantages over individual randomization. These advantages include, but are not limited to, increased generalizability and improved cost-effectiveness.[1,2] In this commentary, we discuss strategies that can reduce the risk of confounding and bias in cluster-randomized studies. In particular, we focus on cluster-randomized studies that include a crossover. Such studies, referred to as cluster-randomized crossover studies, have been employed in several recent prominent cluster-randomized studies[3–7] and can provide a cost-effective method to assess the clinical effectiveness of interventions for important health outcomes.

Evaluating interventions in a "real world setting" is ideal for testing an intervention's effectiveness. Cluster-randomization enables this more readily than individual randomization, because many interventions may benefit a group and not just an individual. In a standard cluster-randomized study, the treatment or intervention is randomly assigned to a group of people, such as a hospital ward, a village, or

Correspondence: Nicholas G Reich
Division of Biostatistics and Epidemiology, School of Public Health and Health Sciences UMass-Amherst, 715 North Pleasant Street, Amherst, MA 01003, USA
Tel +1 413 545 4534
Email nick@schoolph.umass.edu

a school. The treatment effect can be estimated by comparing outcomes in groups that received the treatment to those that did not. Large-scale cluster-randomized studies are routinely carried out in medical, public health, behavioral, and educational research.

Despite the growing use of cluster-randomization in recent years, there remain significant barriers to effectively implementing the study design. Cluster-randomization is most effective with large numbers of clusters to balance potentially confounding characteristics in the treatment and control arms of the study. Unfortunately, regulatory challenges and financial constraints limit large-scale, multisite studies. Incorporating a crossover into a cluster-randomized study can improve logistical and statistical efficiency. Cluster-randomized crossover study design is not a new concept, but the distinctions between it and a standard cluster-randomized trial are not widely understood, and the pros and cons of each design are not well characterized in the literature.

What is balance, and why is it important for cluster-randomized trials?

In individually randomized studies, each participant is randomly assigned to a treatment group. Randomization is a very powerful tool. In studies with enough participants, randomization virtually ensures that the arms of a study will have similar profiles of all measured and unmeasured characteristics. This similarity between arms of a study is commonly referred to as balance. When an individually randomized study has enough participants and an intention-to-treat analysis is performed, good balance between arms is guaranteed. However, in cluster-randomized trials, achieving good balance often requires quite a bit more attention.

Many cluster-randomized trials do not have enough clusters to guarantee that the treatment arms will be balanced. As an extreme example, consider a trial that has ten large hospitals as clusters in a study looking at patient outcomes. If five hospitals serve major urban centers and the other five serve rural areas, some possible randomizations would by chance create gross imbalance in important demographic or baseline characteristics. If the one arm of the study contained all five urban hospitals, and the other arm all the rural hospitals, the treatment effect would be confounded with several covariates. It would be virtually impossible to ensure that the reported treatment effect would be free from bias resulting from this confounding.

Why consider a crossover?

The cluster-randomized crossover design retains the advantages of a cluster-randomized design while leveraging its unique design to minimize imbalance, a common pitfall of cluster-randomized studies. In a cluster-randomized crossover design, investigators take advantage of the fact that often a cluster will be its own best control. In this type of study, clusters are randomly assigned to an initial intervention for a period of time, and then, after a washout period, they are observed under another intervention. Unlike with matching and stratification, to achieve balance in this setting, the researcher would not need to pre-specify the two or three "important" variables to group on, as most covariates will naturally be controlled for by the crossover. Methods for analyzing cluster-randomized crossover studies have been studied in some detail.[8,9] It is important to note that many studies may not be eligible to implement a crossover study due to the requirement that the intervention must be able to be "washed out" and not have carryover effects between study periods. Because it can be difficult to recognize the presence of and make adjustments for carryover effects, researchers must be thoughtful when considering a crossover in the study design.

A recent example

Climo et al incorporated a crossover into their recent cluster-randomized trial that investigated the effectiveness of daily bathing with antibacterial solution to reduce the risk of infection with multidrug-resistant organisms and the incidence of hospital-acquired bloodstream infections.[7] This study analyzed data on 7,727 patients from nine clusters. Five of the clusters implemented the intervention for 6 months and then switched to control for 6 months. Four clusters did the reverse. The authors found a significant lower risk of the primary outcomes when comparing the units on intervention to those on control.

In many situations, a crossover can enable fewer clusters to be included in the study, while maintaining adequate statistical power.[10] We conducted a power analysis for Climo et al's original study design under two separate scenarios: with and without a crossover. Taking data from the paper, we used the average number of patients enrolled per cluster per month, the mean length of stay, and the observed baseline rates to estimate power for hypothetical future trials similar to Climo et al's. For trials without a crossover, we assumed that each cluster was observed for 12 months. For trials with a crossover, we assumed that each cluster was observed for 6 months on treatment and 6 months off treatment.

In this way, we ensured that the number of hypothetical individuals enrolled remained constant across all of our comparisons. Since standard formulas do not exist to calculate power for cluster-randomized crossover trials, we used the clusterPower package for R to simulate power.[10,11]

After looking at crossover and non-crossover trials across a range of number of clusters, we found that a non-crossover version of this study would need an additional 40 clusters to maintain the same statistical power that an equivalent crossover trial achieves (see Figure 1). Additionally, the cluster-randomized crossover design achieves 50% better power than the standard cluster-randomized design with 12 clusters (the original number in Climo et al, before dropouts). These results confirm that for a trial like Climo et al's, implementing a crossover was a good choice. Significantly, what makes the crossover a good choice here is that there are a small number of clusters and some variability in the outcomes between clusters. Our R code used to generate these results is freely available at https://github.com/nickreich/crossover.

Current practices

In cluster-randomized trials without a crossover, investigators typically use one of several techniques to carefully restrict the randomization procedure. If properly implemented, such methods can simultaneously preserve some balance between arms of the study while maintaining the integrity of the randomization process. Two common techniques – matching and stratification – rely on grouping the clusters together based on a small number of possible confounding variables and then ensuring that some from each group are assigned to each arm of the study.[12,13] Additionally, restricted or constrained randomization schemes can be used to identify an acceptable randomization procedure.[14]

The advantages of these techniques are that they can greatly increase the balance of the study while still randomly assigning the intervention. However, implementing a matched or stratified randomization can be difficult. For example, the strata or matched pairs need to be based on a small number of variables, the choice of which will rarely be obvious or easily made. (Although Imai et al do recommend always matching on cluster size.[13]) Furthermore, guidelines on whether or how to include the matching or stratification in the analysis are unclear and lack consensus in the scientific literature.

For the specific situation when all clusters must end up following an intervention for political or logistical reasons, the "stepped-wedge" design has also been promoted as an alternative design in cluster-randomized trials.[15–17] This design has been shown to have some gains in efficiency over standard cluster-randomization but may not be robust when the secular trend for an outcome is changing over time.[18]

When can a crossover design improve efficiency?

There are important benefits that can be derived from adding a crossover to a cluster-randomized design. Primarily, a crossover enables fewer clusters to be included in the study. In standard analysis of cluster-randomized studies,
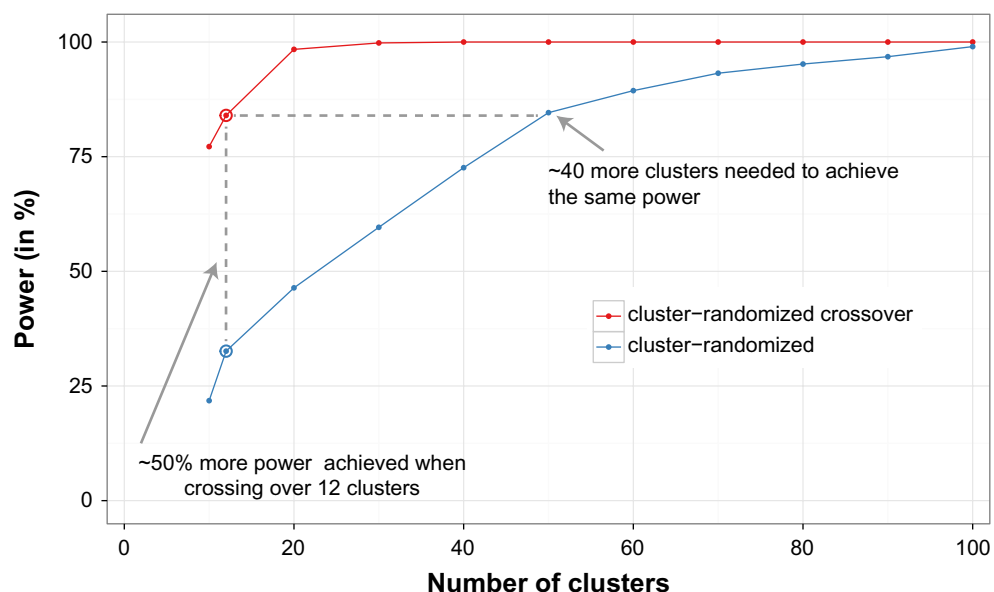

Figure 1 Estimated power for cluster-randomized crossover and standard cluster-randomized trials. These estimates were computed using simulation via the clusterPower package in R.

comparisons are made between clusters in the treatment and control groups, assuming an even distribution of characteristics between clusters. The crossover design can limit the impact of imbalance of characteristics between clusters (for example, clusters with sicker patients happened to be in one treatment arm of a study) because the treatment effect is compared within a given cluster during the two periods. A crossover can be an especially useful tool when the number of available clusters is limited. Additionally, as shown in the example above and in Reich et al,[10] a cluster-randomized crossover study can have more power to detect a given treatment effect than a cluster-randomized non-crossover study with the same number of participants. The analysis of crossover studies controls for within-group variation, providing a more efficient statistical comparison that analysis of non-crossover studies. More research is needed to identify explicitly the circumstances where the crossover design provides the greatest boost in efficiency.

While a crossover design has many benefits, it cannot be used for all studies. Crossover studies cannot be conducted if receiving the intervention in the first period of study would contaminate observations in the second. For example, an education campaign to improve medication compliance could not be crossed over because once a group received that education there would not be a way to "erase" the effects of the intervention during the control period. However, if the intervention does not have lingering effects, such as in the example above, then the crossover design could be implemented. Furthermore, a crossover might only be cost-effective when the per-cluster costs of enrolling a cluster are large or when there are few available clusters.

An additional consideration for cluster-randomized crossover studies is whether there would be any overlap of study participants across the different treatment periods. It is common to account for correlation within a given cluster across time periods of a study.[10] However, additional analytical tools (eg, a multi-level random effects model) may be needed to account for correlation within the same individual between multiple periods of study.

Conclusion

The cluster-randomized crossover design has been in use for more than a decade, but it remains an underutilized tool. While the cluster-randomized crossover design may not be the right choice for every setting, it should be considered alongside other design strategies (such as matching, stratification, restricted randomization, stepped-wedge,

or baseline comparisons) to ensure efficient implementation of cluster-randomized trials.

When considering whether to use a crossover, the three central issues to consider are:

1. Can the intervention be successfully "washed out"?
   Effects of an intervention may linger after the first period and contaminate the second period. This could either be due to continued treatment of the same patients or to fundamental changes in a clinic such as behavioral changes in health care workers due to an intervention. If a possible effect of an intervention cannot be "washed out," then a crossover cannot be effectively implemented.
2. Can the study period be prolonged?
   Since a crossover study may take up to twice as long as a study without the crossover, there must not be limitations on funding or participation during this time frame.
3. Will a crossover improve the efficiency and statistical power of the study?
   In many cases, adding a crossover to a cluster-randomized study design can markedly improve the efficiency of the study.[10] This means that a smaller effect size could be detected or fewer participants and/or clusters would be needed.

These issues are virtually the same criteria used for determining whether to conduct a standard crossover study. However, more research is needed to expand our understanding of how to best incorporate a crossover design into the practical implementation of cluster-randomized studies. Because there are significant differences between cluster-randomized and cluster-randomized crossover trials, the CONSORT statement on best practices for cluster-randomized trials should provide adequate guidance for reporting or analysis of a cluster-randomized crossover study.[19] A number of influential studies on patient-centered outcomes have incorporated a crossover design,[3–7] and this approach should be considered by researchers in the early design stages of future cluster-randomized trials.

Acknowledgments

Disclosure

The authors report no conflicts of interest in this work.

References

1. Platt R, Takvorian SU, Septimus E, et al. Cluster randomized trials in comparative effectiveness research: randomizing hospitals to test methods for prevention of healthcare-associated infections. *Med Care*. 2010;48(6):S52.
2. Perencevich EN, Lautenbach E. Infection Prevention and comparative effectiveness research. *JAMA*. 2011;305(14):1482–1483.
3. Jeyaratnam D, Whitty CJM, Phillips K, et al. Impact of rapid screening tests on acquisition of meticillin resistant Staphylococcus aureus: cluster randomised crossover trial. *BMJ*. 2008;336(7650):927–930.
4. de Smet AMGA, Kluytmans JAJW, Cooper BS, et al. Decontamination of the digestive tract and oropharynx in ICU patients. *N Engl J Med*. 2009;360(1):20–31.
5. Hejblum G, Chalumeau-Lemoine L, Ioos V, et al. Comparison of routine and on-demand prescription of chest radiographs in mechanically ventilated adults: a multicentre, cluster-randomised, two-period crossover study. *Lancet*. 2009;374(9702):1687–1693.
6. Milstone AM, Elward A, Song X, et al. Daily chlorhexidine bathing to reduce bacteraemia in critically ill children: a multicentre, cluster-randomised, crossover trial. *Lancet*. 2013;381(9872):1099–1106.
7. Climo MW, Yokoe DS, Warren DK, et al. Effect of daily chlorhexidine bathing on hospital-acquired infection. *N Engl J Med*. 2013;368(6):533–542.
8. Turner RM, White IR, Croudace T. Analysis of cluster randomized cross-over trial data: a comparison of methods. *Stat Med*. 2007;26(2):274–289.
9. Parienti JJ, Kuss O. Cluster-crossover design: a method for limiting clusters level effect in community-intervention studies. *Contemp Clin Trials*. 2007;28(3):316–323.
10. Reich NG, Myers JA, Obeng D, Milstone AM, Perl TM. Empirical power and sample size calculations for cluster-randomized and cluster-randomized crossover studies. *PLoS One*. 2012;7(4):e35564.
11. Reich NG, Obeng D. clusterPower: power calculations for cluster-randomized and cluster-randomized crossover trials (v0.4-2). 2013. Available from: http://cran.r-project.org/web/packages/clusterPower/index.html.
12. Donner A, Donald A. Analysis of data arising from a stratified design with the cluster as unit of randomization. *Stat Med*. 1987;6(1):43–52.
13. Imai K, King G, Nall C. The essential role of pair matching in cluster-randomized experiments, with application to the Mexican universal health insurance evaluation. *Stat Sci*. 2009;24(1):29–53.
14. Moulton LH. Covariate-based constrained randomization of group-randomized trials. *Clin Trials*. 2004;1(3):297–305.
15. The Gambia Hepatitis Intervention Study. The Gambia Hepatitis Study Group. *Cancer Res*. 1987;47(21):5782–5787.
16. Brown CA, Lilford RJ. The stepped wedge trial design: a systematic review. *BMC Med Res Methodol*. 2006;6:54.
17. Hussey MA, Hughes JP. Design and analysis of stepped wedge cluster randomized trials. *Contemp Clin Trials*. 2007;28(2):182–191.
18. Woertman W, de Hoop E, Moerbeek M, Zuidema SU, Gerritsen DL, Teerenstra S. Stepped wedge designs could reduce the required sample size in cluster randomized trials. *J Clin Epidemiol*. 2013;66(7):752–758.
19. Campbell MK, Piaggio G, Elbourne DR, Altman DG; CONSORT group. Consort 2010 statement: extension to cluster randomised trials. *BMJ*. 2012;345:e5661.