

Sample size calculations for cluster randomised crossover trials in Australian and New Zealand intensive care research

Sarah J Arnup, Joanne E McKenzie,
David Pilcher, Rinaldo Bellomo and Andrew B Forbes

Designing randomised controlled trials (RCTs) in the intensive care setting can be challenging. A recent review of observed effect sizes in 38 mortality trials calculated an average effect size of 1.4%, in contrast to the average effect size of 10.1% that was hypothesised when the trials were planned.¹ Detection of such small intervention effects using individual randomisation, with reasonable levels of statistical power (eg, > 80%), can lead to designs in which it is not possible to recruit the required number of participants. Further, individual randomisation is often not feasible in intensive care settings for interventions such as infection control and “bundles of care”, which involve multiple practice changes simultaneously. This is because individual randomisation in intensive care settings involves a high risk of “contamination” between intervention arms. Individual randomisation is also not feasible for whole-of-intensive care unit processes, such as admission and discharge policies, because varying procedure by individual patient is impractical.

For evaluation of the above interventions, cluster randomised designs are generally preferred. In these designs, all patients in each ICU (cluster) receive the same intervention.² Cluster randomisation allows the intervention to be delivered as if it were standard operating practice and may increase the efficiency of data collection.³ However, cluster randomisation is likely to result in the need for more participants than individual randomisation. The statistical power in a cluster randomisation trial depends on the number of ICUs, the number of patients in each ICU, and the similarity of the outcome responses of patients within each ICU.² As the outcome responses of patients within an ICU become more similar to each other than to patients in different ICUs, (within-cluster correlation), cluster randomisation will further reduce the statistical power compared with individual randomisation.

An alternative to the parallel-group, cluster randomised design is the cluster randomised crossover (CRXO) design.^{4,5} Bellomo and colleagues³ have called for an increase in the use of the CRXO design to evaluate routinely used and low-risk interventions such as oxygen therapy, ulcer-prophylaxis therapy and intravenous fluids and nutrition in RCTs.

In a two-intervention, two-period CRXO design, each ICU receives the two interventions, but they receive them

ABSTRACT

Objective: The cluster randomised crossover (CRXO) design provides an opportunity to conduct randomised controlled trials to evaluate low risk interventions in the intensive care setting. Our aim is to provide a tutorial on how to perform a sample size calculation for a CRXO trial, focusing on the meaning of the elements required for the calculations, with application to intensive care trials.

Data sources: We use all-cause in-hospital mortality from the Australian and New Zealand Intensive Care Society Adult Patient Database clinical registry to illustrate the sample size calculations.

Methods: We show sample size calculations for a two-intervention, two 12-month period, cross-sectional CRXO trial. We provide the formulae, and examples of their use, to determine the number of intensive care units required to detect a risk ratio (RR) with a designated level of power between two interventions for trials in which the elements required for sample size calculations remain constant across all ICUs (unstratified design); and in which there are distinct groups (strata) of ICUs that differ importantly in the elements required for sample size calculations (stratified design).

Results: The CRXO design markedly reduces the sample size requirement compared with the parallel-group, cluster randomised design for the example cases. The stratified design further reduces the sample size requirement compared with the unstratified design.

Conclusions: The CRXO design enables the evaluation of routinely used interventions that can bring about small, but important, improvements in patient care in the intensive care setting.

 Crit Care Resusc 2018; 20 (2): 117-123

sequentially in two separate periods of time called “cluster-periods”. The order in which the interventions are delivered to each ICU is randomised to control for changes over time that are independent of the intervention (eg, policy changes), but which might have an impact on the trial outcomes. By comparing the interventions within each ICU, the ICU-specific component of variation is removed

from the estimate of the difference between interventions. Therefore, as for an individually randomised crossover trial, the inclusion of the crossover element has the benefit of reducing the required number of participants compared with a parallel-group design.⁶

The sample size calculation is a critical element in designing RCTs. To date, a tutorial on how to calculate the sample size for a CRXO trial, with a focus on designing intensive care trials, has not been available. Further, in practice, there are often distinct groups (strata) of ICUs that differ in the elements required for sample size calculation, such as mortality rates or within-cluster correlations. These elements need to be accounted for in the sample size calculations. The purpose of this article is to fill these gaps in the literature through the provision of examples of sample size calculations for a set of hypothetical intensive care trials.

Sample size considerations for CRXO trials

A sample size calculation for a CRXO trial needs to account for both the cluster randomisation and crossover aspects of the design. We now briefly discuss sample size and design elements. A detailed discussion by Arnup and colleagues is available elsewhere.⁷

Our focus is a two-intervention, two-period design in which different individuals are included in each period. Each ICU is randomised to receive intervention A for one period and then to cross over to intervention B for the second period, or vice versa. The basic sample size formula for the number of ICUs required to detect a risk ratio (RR) with power $1-\beta$ between two interventions, with probability α , is shown in Equation 1 (see Appendix I, online at cicm.org.au/journal.php).

Within-period correlation

When cluster randomisation is used, the responses of patients within the same ICUs are often more similar to each other than to patients in different ICUs.² This degree of similarity depends on how variable the response rate is across ICUs. In the CRXO design, the similarity of the responses from two patients within the same ICU, within the same period of time, is quantified by the within-ICU *within-period correlation* (WPC).

Between-period correlation

When a trial design with multiple time periods is used, we also need to consider the similarity of responses of patients within the same ICU, but in different periods of time. We quantify this similarity by the within-ICU *between-period correlation* (BPC). If the ICU environment is relatively stable, we might expect the responses of two patients from the

same ICU in different periods to be similar, but less so than if patients were from the same period, because of potential changes in the ICU environment over time. This typically results in the BPC being less than the WPC.

Parallel-group versus CRXO designs

To illustrate the potential benefit of the crossover aspect in the CRXO design, we compare the power to detect an RR of 0.80 (20% relative risk reduction) between two interventions in a parallel-group cluster randomised trial and a crossover cluster randomised trial, assuming a baseline risk of 10% and an ICU admission rate of 1000 patients per year. The sample size formula for the number of ICUs required to detect an RR with power $1-\beta$ between two interventions with probability α for a parallel-group cluster randomised trial is shown in Equation 2 (see Appendix I).

For a parallel-group, cluster randomised trial conducted over 24 months, the cluster size is 2000. For this example, we assume that the WPC = 0.02. For a CRXO trial conducted over 24 months, with two 12-month periods, the cluster-period size is 1000. We assume that the WPC = 0.02 and the BPC = 0.01. Given these assumptions, the required number of clusters for the parallel-group design is 67, and the required number of clusters for the crossover design is 38. Hence, by including a crossover, the required number of clusters has been reduced by 45%. The required number of patients for the parallel-group design is therefore $67 \times 2000 = 134\,000$; and the required number of patients for the crossover design is $38 \times 2 \times 1000 = 76\,000$. The inclusion of a crossover has reduced the required number of patients by 43%.

Effect of within-period correlation

To illustrate the effect of the WPC on the sample size requirements, we begin by discussing the extreme values of the WPC. When the WPC is 1, all patients within the same ICU in the same time period have the same response, so there is no gain in information from sampling more than one patient. When the WPC is 0, the responses of two patients in the same ICU are no more similar than two patients in different ICUs, so the information contained in the ICU is the same as the information that would be obtained from completely independent patients. Therefore, as the WPC increases from 0, the precision of the estimate of the intervention effect in each time period decreases. As a result, the power to detect a specified RR between interventions also decreases.

In the CRXO trial described in the previous section, the WPC was 0.02 (and the BPC was 0.01), and as a result, 38 ICUs (76 000 patients) were required to detect an RR of 0.80 with 80% power (Figure 1, grey solid curve). However, if the WPC is 0.03 (and the BPC remains 0.01), then 38 ICUs

will only achieve 54% power (Figure 1, black solid curve) to detect the same size effect. The effect of the WPC in a CRXO trial acts in the same way as the intracluster correlation coefficient in a parallel-group cluster randomised trial (as shown by Campbell and colleagues⁸).

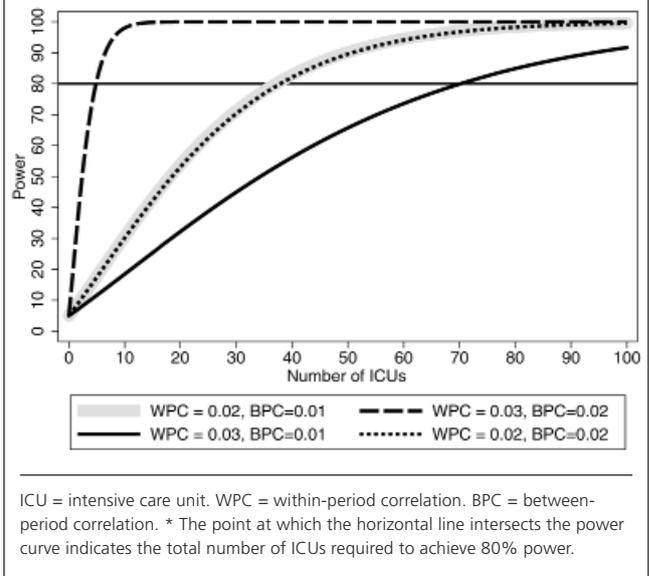
Effect of between-period correlation

To understand the effect of the BPC on the sample size requirement, we first note that the outcome of a specific patient depends on which ICU (cluster) the patient was admitted to, any time effects unique to that ICU, the intervention the patient receives, and factors idiosyncratic to the individual patient. In a CRXO design, we compare interventions A and B by comparing the responses of patients in the first period with the responses of patients in the second period (or equivalently responses in the second period with the first period), *within each ICU*. By performing these comparisons within each ICU, the component of the patient outcomes that is due to the ICU (cluster) that remains constant over time will cancel out from the comparison. The removal of this between-ICU variability enables us to obtain a more precise estimate of the intervention effect.

We now consider the extreme values of the BPC to demonstrate its effect. When the BPC is the same as the WPC, the responses of patients in two different periods are as similar as two responses from patients in the same period; this is the largest value that the BPC can take because the responses of two patients in different periods cannot be more similar than the responses from patients in the same period. In this scenario, all of the ICU-specific effect is removed from the comparison between intervention A and B. In contrast, when the BPC is zero, there is no similarity at all between patients in the same ICU in different periods and there is no value in using crossover design, because no part of the ICU-specific effect is removed from the comparison. Therefore, as the BPC becomes closer to the WPC, more of the ICU-specific effect is removed from the comparison between interventions A and B, and as a result, the power to detect a specified RR between interventions increases. (See Arnup and colleagues, 2017, for a detailed graphical explanation of this phenomenon.⁷)

As an example, when the BPC is half the value of the WPC (WPC = 0.02, BPC = 0.01), 38 ICUs are required to detect an RR of 0.80 with 80% power (Figure 1, grey solid curve). The same power can also be achieved if the absolute difference between the WPC and the BPC remains 0.01, but the values of the WPC and the BPC change; that is, WPC = 0.03, BPC = 0.02 (Figure 1, short-dashed curve).⁹ However, if the BPC takes the maximum value of equal to WPC (ie, WPC = BPC = 0.02), then 38 ICUs will achieve far greater power, close to 100% (Figure 1, long-dashed curve).

Figure 1. Power to detect a risk ratio of 0.8 (baseline risk, 10%; ICU size, 1000 patients/cluster-period) between interventions with varying within- and between-period correlations*



Effect of cluster size

As for parallel-group cluster randomised trials, power depends to a far greater extent on the number of ICUs than on the number of patients within each ICU.² For example, for a parallel-group cluster randomised trial designed to detect a RR of 0.80 (an absolute risk reduction from 10% to 8%) with an intracluster correlation of 0.01 and with 20 ICUs, doubling the size of the trial from 10 000 to 20 000 patients by increasing the ICU size from 500 to 1000 patients per period increases the power from 51% to 55%. In contrast, doubling the size of the trial from 20 to 40 ICUs with 500 patients per period increases the power to 80%. Therefore, to achieve the same level of power for a larger number of ICUs requires far fewer patients overall, compared with a smaller number of ICUs with more patients per ICU.

Period effects

Changes in the trial environment between the first period and the second period can lead to changes in the patient responses in the second period that are unrelated to the delivered intervention. Such “period effects” alter all patient responses in the second period by the same amount, independent of the intervention given in that period. In a balanced CRXO design, in which the same number of clusters receive intervention A and B in each period, any period effects are removed from the comparison between interventions A and B, and hence do not introduce bias to

the estimate of the intervention effect. However, when the design is not balanced, one needs to explicitly accommodate period effects into the analysis. The sample size formulae can be modified slightly to accommodate period effects, but that detail is beyond the scope of this article. Forbes and colleagues have discussed a relevant approach to this.⁹

Carry-over effects

A requirement of the design is that the effect of the intervention given in the first period does not carry over to affect the responses of patients in the second period. Otherwise, the estimated intervention effect may be biased. Carry-over can occur at the level of the individual patient, or at the level of the ICU environment. Including different patients in each period (ie, a cross-sectional design) removes the risk of carry-over at the level of the individual patient, but if the intervention administered in the first period can lead to changes in the behaviour of the health care team or in the ICU environment that persists into the second period, then the potential for carry-over exists. No statistical method can detect or remove carry-over effects from a two-intervention, two-period CRXO design, and therefore it is essential that the risk of carry-over is minimised by the trial design itself, for example by including an appropriate washout period.¹⁰

Stratification

ICUs can form strata with distinct characteristics (eg, diagnostic casemix; different ratios of high mortality risk emergency patients to elective surgical cases; different hospital protocols for delivery of care, such as infection control and admission and discharge policies; variation in the availability of therapeutic services within the hospital, such as interventional radiology, cardiac catheterisation and extracorporeal membrane oxygenation; or hospital location) which lead to variations between strata in the outcome measure and other elements required for sample size calculations, for example, mortality rate, ICU admission rate, and the WPC and BPC values, respectively. Failure to account for differences in these elements can lead an inappropriate sample size. An extension to the sample size formula to include multiple strata is given in Equation 3 (see Appendix I). We provide an illustration in Section 3.

Adjustments to sample size formula with small numbers of clusters

When the estimated number of ICUs is small (eg, fewer than 30), it is recommended than one additional ICU is included per intervention (two ICUs in total) in the sample size calculation.^{2,9}

Sample size examples

In this section we show how to perform sample size calculations for a hypothetical trial of interventions to reduce all-cause in-hospital mortality taking place over a total duration of 2 years (ie, a cluster period duration of 12 months). We begin with unstratified (ie, “conventional”) CRXO trials that include patients from only one stratum. We then extend to cases in which event rates, and hence clustering effects, vary between two strata: tertiary ICUs, and metropolitan and rural ICUs combined.

We obtain estimates of the WPC and BPC for all-cause in-hospital mortality, the annual number of ICU admissions, and baseline mortality rates from the Australian and New Zealand Intensive Care Society (ANZICS) Adult Patient Database (APD) clinical quality registry. The ANZICS-APD is a clinical quality registry, managed by the ANZICS Centre for Outcome and Resource Evaluation, which collects de-identified information on admissions to adult ICUs in Australia and New Zealand.¹¹ The methods used for data extraction and calculation of the correlations and annual number of admissions is shown in detail in Appendix II. A file with code to calculate power in the statistical package Stata, version 14 (StataCorp), is also provided.

The estimates calculated from the ANZICS-APD (Table 1) are used to design a two-intervention, two 12-month period CRXO trial, where all-cause in-hospital mortality in all admitted patients is the primary outcome. In Appendix III, we show the analogous calculation for a trial including a restricted patient group — those receiving mechanical ventilation. The formula for the sample size for the required number of ICUs is shown in Equation 1 in Appendix I.

Unstratified sample size calculation example

Suppose a CRXO trial is being planned to compare a buffered crystalloid fluid with saline in all patients requiring

Table 1. Number of ICUs meeting inclusion criteria,* and other ICU characteristics, by ICU stratum and combined across all strata

ICU stratum	Included ICUs (n)	Average annual admissions (n)	Harmonic mean annual admissions (n)	Mortality rate (%)	WPC	BPC
Tertiary	34	1356	1114	9.05%	0.006	0.005
Metropolitan and rural	54	638	553	6.65%	0.008	0.007
Combined	88	911	684	8.03%	0.009	0.006

ICU = intensive care unit. WPC = within-period correlation. BPC = between-period correlation. * See Appendix II, online at cicm.org.au/journal.php.

fluids. The effect to be detected is a 10% reduction in all-cause in-hospital mortality (RR = 0.90). The investigators wish to detect this effect with 90% power, assuming a 5% significance level.

The sample size formula requires values of the “harmonic mean” number of patients per ICU per 12-month period (see Appendix I for definition); the event rate of the outcome (in this example, mortality); and the WPC and BPC. Using the estimates of these values calculated from the APD data (see row labelled “combined” in Table 1) gives a sample size requirement of 105 ICUs (191 310 patients).

A case for stratification

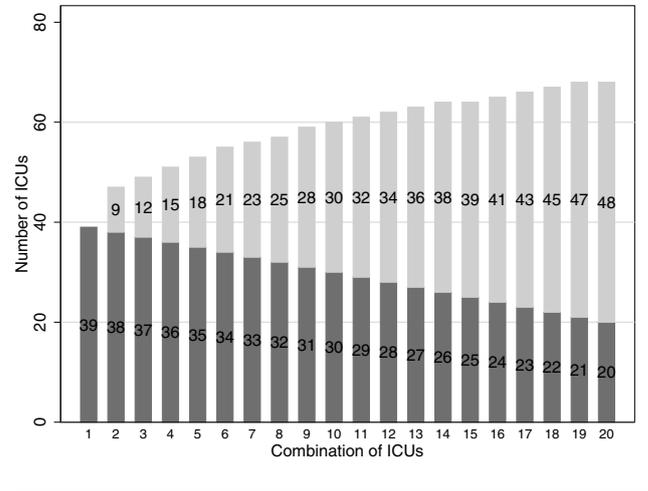
Rather than combining all ICUs together in a single calculation, as in Equation 1 (see Appendix I), we can include the differences between types of ICUs (strata) to reduce the overall sample size requirement. Table 1 shows that the number of annual admissions to tertiary ICUs is more than twice that of the metropolitan and rural ICUs; the mortality rate is highest in tertiary ICUs; and that there are small but important differences in WPC and BPC values across strata. The lower value of the WPC in the tertiary ICUs indicates less variability in mortality rates between tertiary ICUs compared with the metropolitan and rural ICUs. For WPC and BPC values, the differences between the tertiary and the metropolitan and rural ICU values are similar, suggesting that the variability of the ICU environment in these hospitals is similar over time.

Stratified sample size calculation example

The variance (or equivalently precision) of the specified intervention effect is a key element in all sample size calculations. For a stratified sample size calculation, this concept is extended to incorporate the different variances in each stratum, which arise from the different sample size elements in each stratum (eg, different mortality rates, WPC and BPC). Additionally, for a stratified sample size calculation, triallists need to decide how many ICUs (clusters) they will sample from each stratum. Different combinations of numbers across the strata will change the total required number of clusters.

Allowing for stratification can bring about important benefits in terms of reducing the number of required clusters in a CRXO design. We extend our previous example, allowing for the differences in the estimates of annual number of admissions, mortality rates and correlations (WPC and BPC) between the tertiary and the combined metropolitan and rural strata (see Appendix IV). As shown in Figure 2, depending on the combination of numbers of ICUs chosen from each stratum, the reduction in the required number of ICUs (clusters) and patients is *at least* 30%.

Figure 2. Sample size requirements for total number of ICUs, and number of ICUs from each of the strata, required to obtain 90% power to detect a risk ratio of 0.90 between interventions*



ICU = intensive care unit. * Each bar represents a different number (combination) of ICUs from each of the strata. Dark grey = tertiary ICUs. Light grey = metropolitan and rural ICUs.

Effect of stratification on sample size requirement

Stratification reduces the sample size requirement for a CRXO trial when ICUs can be differentiated by a known factor or set of factors (in this case, the type of hospital), and this factor strongly affects the outcome of interest in the trial (in this case mortality: mortality varies by type of hospital).

To understand the effect of stratification, first note that for any sample size calculation, the sample size will decrease when the intervention effect is estimated with more precision. In a CRXO trial, this occurs as the difference the WPC and the BPC decreases (See Section 2). In the ANZICS-APD data, there are substantial differences in mortality rates across the two strata, and as a result, the within-stratum variability in mortality rates (and hence the WPC within each stratum), is markedly smaller than for all ICUs combined (Table 1). Because the BPC remains relatively constant across the two strata, the difference between the WPC and the BPC is much smaller in the individual strata than in the combined data, and this serves to reduce the sample size.

Proportion of ICUs to select from each stratum

The next step is to determine the proportion of the total number of ICUs that are tertiary, with the remainder being metropolitan or rural. The tertiary ICUs offer considerable advantages: not only are they larger in size (patient numbers), but they also have a higher mortality rate (and hence a larger absolute intervention effect). As a result,

the smallest total required number of ICUs will be obtained when the trial is restricted to tertiary ICUs. However, if there is an insufficient number of tertiary ICUs to satisfy the sample size requirement, or for practical reasons or generalisability concerns, it may not be appropriate to restrict a trial to only tertiary ICUs. While the overall sample size requirement does increase as the fraction of tertiary ICUs is reduced, this may result in an unfeasible trial becoming feasible, as shown in the example in Appendix IV.

Discussion

We have discussed how to perform a sample size calculation for a two-period, two-intervention, cross-sectional CRXO trial in the intensive care setting. We have also provided estimates of the elements required to perform these calculations using the ANZIC-APD data. A review of the statistical methods used to determine the sample size for CRXO trials found that the methodology was frequently inadequate.¹² Inadequate methods may have been used because of lack of knowledge of the appropriate sample size methodology, limited availability of the elements needed to perform the sample size calculations or because of a lack of practical examples of how to implement a sample size formula. We have addressed these issues in this article.

We have provided a sample size formula to determine the number of ICUs required to detect a constant *relative* risk, rather than a constant *absolute* risk reduction, between two interventions. When considering multiple strata, each with their own baseline mortality (event) rate, evidence suggests that it is more plausible to expect a constant relative reduction rather than absolute reduction.¹³

When the sample size calculation accounts for multiple ICU strata, the same power can be achieved with differing numbers of ICUs from each stratum. For example, a trial can be designed with an equal contribution of ICUs from each stratum, or designed so that one stratum provides most of the required ICUs. In the ANZICS-APD, when the intervention is aimed at reducing all-cause mortality, the absolute minimum number of total ICUs will be obtained by including only tertiary ICUs. However, stratifying the sample size calculation provides flexibility to adjust the required number of ICUs from each stratum when the number of ICUs is limited by the availability of ICUs in the tertiary (or other) stratum.

Additional considerations

We have considered the two-intervention, two-period CRXO design with different patients in each period. Adaptations of this design, such as increasing the number of periods¹⁴ or including the same participants across periods,¹⁵ are possible, but they also increase the complexity of the sample

size calculation and analysis.¹⁶

The sample size calculation is sensitive to the difference between the WPC and the BPC, therefore both correlations should be chosen carefully.⁷ Considerations for choosing the WPC are similar to those for choosing an intracluster correlation for a parallel-group, cluster randomised trial.¹⁷⁻²⁰ Choosing the BPC is likely to rely on routinely collected data, pilot and feasibility data, or a reasoned best guess. However, in the absence of such data, recommendations of half the WPC²¹ and 0.8 of the WPC have been made.²²

The intended analysis should match the sample size methodology. One potentially appropriate analysis method to estimate the RR in a stratified CRXO trial is to use generalised estimating equations (GEE) for a binomially distributed outcome, with a logarithmic link and an exchangeable “working” correlation between individuals within an ICU. The ICU strata can be included as a covariate in the model, as can a term for period effects.

There is an active field of research investigating how well statistical analysis methods perform in cluster randomised trials with small numbers of clusters or low event rates. Our examinations displayed appropriate performance for an RR of 0.9 and two strata with event rates of 7% and 9%. However results from other research has shown that caution may need to be exercised with power formulae with fewer than 12 clusters²³ or when the event rates are 6% or lower.⁹

Conclusion

Sample size determination for CRXO trials requires the use of an appropriate sample size formula together with appropriate estimates of its component elements. We have provided the sample size formulae, estimates of the elements required by the formulae using ANZICS-APD data, and examples of how to determine the required sample size for unstratified and stratified CRXO trials in the intensive care setting.

Author details

Sarah J Arnup¹

Joanne E McKenzie¹

David Pilcher^{2,3,4}

Rinaldo Bellomo⁴

Andrew B Forbes¹

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

2 Centre for Outcome and Resource Evaluation, Australian and New Zealand Intensive Care Society, Melbourne, VIC, Australia.

3 Department of Intensive Care, The Alfred Hospital, Melbourne, VIC, Australia.

4 Australian and New Zealand Intensive Care Research Centre, School of Public Health and Preventive Medicine, Monash University, Melbourne, VIC, Australia.

Correspondence: andrew.forbes@monash.edu

References

- 1 Aberegg SK, Richards DR, O'Brien JM. Delta inflation: a bias in the design of randomized controlled trials in critical care medicine. *Crit Care* 2010; 14: R77.
- 2 Eldridge S, Kerry S. A practical guide to cluster randomised trials in health services research. Chichester, United Kingdom: John Wiley and Sons, 2012.
- 3 Bellomo R, Forbes A, Akram M, Bailey M, Pilcher DV, Cooper DJ. Why we must cluster and cross over. *Crit Care Resusc* 2013; 15: 155-157.
- 4 Turner RM, White IR, Croudace T. Analysis of cluster randomized cross-over trial data: a comparison of methods. *Stat Med* 2007; 26: 274-289.
- 5 Parienti JJ, Kuss O. Cluster-crossover design: a method for limiting clusters level effect in community-intervention studies. *Contemp Clin Trials* 2007; 28: 316-323.
- 6 Mills EJ, Chan AW, Wu P, Vail A, Guyatt GH, Altman DG. Design, analysis, and presentation of crossover trials. *Trials* 2009; 10: 27.
- 7 Arnup 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.
- 8 Campbell MK, Mollison J, Steen N, Grimshaw JM, Eccles M. Analysis of cluster randomized trials in primary care: a practical approach. *Fam Pract* 2000; 17: 192-196.
- 9 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.
- 10 Jones B, Kenwood MG. Design and analysis of cross-over trials. 2nd edition. London: Chapman and Hall/CRC, 2003.
- 11 Stow PJ, Hart GK, Hignett T, et al. Development and implementation of a high-quality clinical database: the Australian and New Zealand Intensive Care Society Adult Patient Database. *J Crit Care* 2006; 21: 133-141.
- 12 Arnup 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.
- 13 Deeks JJ. Issues in the selection of a summary statistic for meta-analysis of clinical trials with binary outcomes. *Stat Med* 2002; 21: 1575-1600.
- 14 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: 652-658.
- 15 Rietbergen C, Moerbeek M. The design of cluster randomized crossover trials. *J Educ Behav Stat* 2011; 36: 472-490.
- 16 Kasza J, Hemming K, Hooper R, et al. Impact of non-uniform correlation structure on sample size and power in multiple-period cluster randomised trials. *Stat Methods Med Res* 2017; Jan 1: 962280217734981. doi: 10.1177/0962280217734981 [Epub ahead of print].
- 17 Campbell MK, Fayers PM, Grimshaw JM. Determinants of the intraclass correlation coefficient in cluster randomized trials: the case of implementation research. *Clin Trials* 2005; 2: 99-107.
- 18 Gulliford MC, Adams G, Ukoumunne OC, Latinovic R, Chinn S, Campbell MJ. Intraclass correlation coefficient and outcome prevalence are associated in clustered binary data. *J Clin Epidemiol* 2005; 58: 246-251.
- 19 Gulliford MC, Ukoumunne OC, Chinn S. Components of variance and intraclass correlations for the design of community-based surveys and intervention studies: data from the Health Survey for England 1994. *Am J Epidemiol* 1999; 149: 876-883.
- 20 Adams G, Gulliford MC, Ukoumunne OC, Eldridge S, Chinn S, Campbell MJ. Patterns of intra-cluster correlation from primary care research to inform study design and analysis. *J Clin Epidemiol* 2004; 57: 785-794.
- 21 Donner A, Klar N, Zou G. Methods for the statistical analysis of binary data in split-cluster designs. *Biometrics* 2004; 60: 919-925.
- 22 Hooper R, Bourke L. Cluster randomised trials with repeated cross sections: alternatives to parallel group designs. *BMJ* 2015; 350: h2925.
- 23 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: 318-333. □