Ribeiro DC, Milosavljevic S, Abbott JH. Sample Size Estimation for Cluster Randomized Controlled Trials. Musculoskelet Sci Pract. 2018;34:108-11.

Title: Sample size estimation for cluster randomized controlled trials

Daniel Cury Ribeiro, PT, PhD – School of Physiotherapy, Centre for Health, Activity and Rehabilitation Research, University of Otago

Stephan Milosavljevic, PT, PhD – School of Physical Therapy, College of Medicine, University of Saskatchewan

J. Haxby Abbott, DPT, PhD, FNZCP – Department of Surgical Sciences, Dunedin School of Medicine, University of Otago

Corresponding author:
Daniel Cury Ribeiro
325 Great King Street, Dunedin, New Zealand
Centre for Health, Activity and Rehabilitation Research – School of Physiotherapy – University of Otago
Email: Daniel.ribeiro@otago.ac.nz

Reference: Ribeiro, D.C.; Milosavljevic, S.; Abbott, J.H. (2018). Sample size estimation for cluster randomized controlled trials. Musculoskeletal Science and Practice, 34, 102-111

DOI: https://doi.org/10.1016/j.msksp.2017.10.002

Abstract

Cluster randomized controlled trials (cRCTs) are commonly used by clinical researchers. The advantages of cRCTs include preventing treatment contamination, enhancing administrative efficiency, convenience, external validity, ethical considerations, and likelihood of increased compliance by participants. However, when designing a cRCT, clinical researchers are faced with challenges, such as cluster units that may not have an equal number of participants within each. In this Technical Note, we discuss approaches for estimating the sample size, while taking into account unequal cluster sizes, and strategies for optimizing the design of cluster trials.

Key words: cluster randomized trial, sample size, unequal cluster size, and design effect.

Introduction

Randomized controlled trials (RCTs) are the gold standard design for experimental studies, as they can reduce many of the risks of bias that threaten clinical trials.[3,18] However, there remain other risks of bias that random allocation alone does not address.[19] Treatment contamination is one example, and can occur when treatment providers or participants learn what the 'other' group have been doing, and begin to blend that into their allocated intervention, thus 'contaminating' it. This corrupts the internal validity of the study, weakens its ability to detect between-groups differences, resulting in falsely concluding the trial treatment does not have a significant effect, when in truth it does (Type II error). When treatment contamination is a risk, a cluster RCT (cRCT) is recommended.[3,6,18,19]

In cRCT, randomization is done at the level of the study sites, centres, clinics or clinicians.[18] All participants attending that site or clinician are automatically in that "cluster", and receive the intervention allocated to the cluster.[3] This reduces the likelihood of contact with the clinicians or participants of the 'other' group. Other advantages of a cRCT include enhanced administrative efficiency, convenience, increased external validity, ethical considerations, and likelihood of increased compliance by participants.[3,6,18] On the other hand, cRCT design reduces the statistical efficiency of the trial,[7,8,17] adds complexity to the statistical approach for estimating the sample size and analysing the main findings.[3,18] Obtaining a robust estimate of the required sample size is crucial for conducting a trial that is statistically sound and financially feasible.[18,20]

This technical note aims to discuss factors that affect sample size estimation of a cRCT, and present different approaches to estimate the sample size when designing a two-arm, cRCT with a continuous outcome measure. Numerous factors need to be taken into account when designing and estimating the sample size of a cRCT (Table 1). Below, we present an overview of each of these factors, and the effect of these on planning and estimating sample size of a cRCT.

Table 1. Effect of different factors on the design and sample size of cluster randomized controlled trials.

Factor	Effect on required sample size
Design effect	Cluster RCTs add sampling error compared with standard RCTs, therefore are statistically less efficient and require a larger sample size
Number of clusters	The greater the number of clusters, the smaller the required sample size
Size of clusters	The fewer participants per cluster, the smaller the required sample size
Intracluster correlation coefficient (ICC)	The smaller the ICC, the smaller the required sample size
Allocation ratio	Equal allocation ratio requires smaller sample size
Attrition	Researchers may consider accounting for individual or cluster drop-outs
Baseline measurements	Including covariates into the analysis increases statistical power, reducing the required sample size
Outcome measure	The type of outcome measure (i.e., binary, continuous, count, ordinal, time-to-event and rate) dictates the formula used to estimate the sample size of the trial.

The design effect

Cluster RCTs are statistically less efficient than normal RCTs, due to the problem of variance inflation, caused by the fact that participants within a cluster unit are dependent, which increases sampling error in this type of trial.[4,20] To account for the statistical inefficiency of cRCTs, a larger sample size is usually required, when compared to a standard RCT.

There are different methods for estimating sample size of cRCT.[3,20] A common and simple approach to estimate sample size for a cluster trial is to multiply the estimated sample size of a standard RCT by a factor, referred to as the "design effect" (DE) (Equation 1). Inflating the sample size of a standard trial by DE increases the statistical power of the cRCT.[5]

Equation 1

$$DE = 1 + (n - 1) \times \rho \quad (1)$$

Where:

DE = design effect;

n = cluster size (i.e. number of participants per cluster);

ρ = intracluster correlation coefficient;

The DE is a function of cluster size and the intracluster correlation coefficient (ICC). The ICC measures the degree of similarity of clustered data,[18] and takes into account how much the variance differs within and between-clusters.[10] Therefore, the larger the ICC or the cluster size, the larger the DE. The impact of cluster size and ICC on the DE is illustrated in Figure 1.

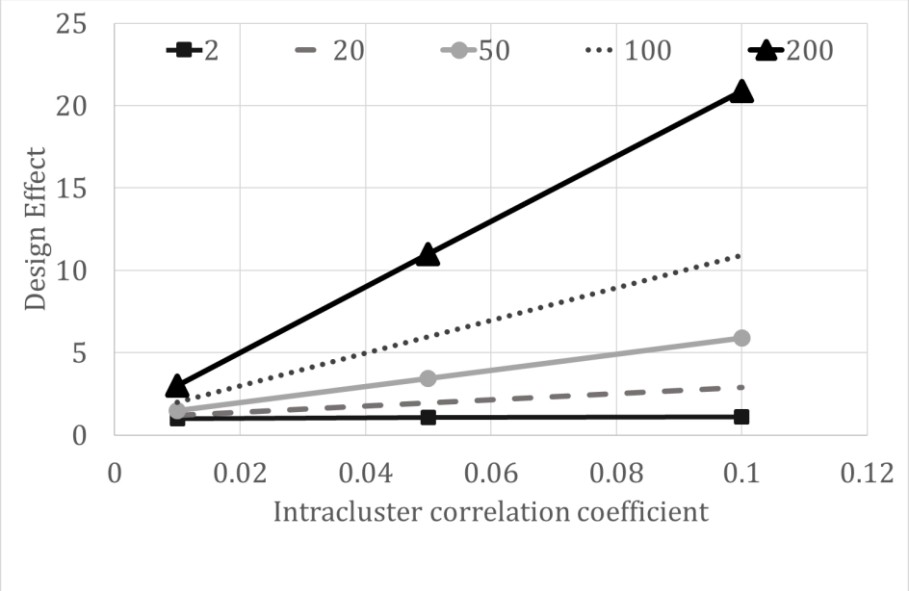

Figure 1. Relationship between intracluster correlation coefficient (x axis), the Design Effect (y axis) and the cluster size.
Black line with square = cluster size of 2 participants per cluster; Grey dashed line = cluster size of 20 participants per cluster; Grey line with circle = cluster size of 50 participants per cluster; Dotted black line = cluster size of 100 participants per cluster; black line with triangle = cluster size of 200 participants per cluster.

The total number of participants (considering a two-arm trial, with equal allocation) for a cRCT is defined by Equation (2):

Equation 2

$$SS_{cluster\ RCT} = SS_{standard\ RCT} \times \text{DE} \quad (2)$$

Where:

$SS_{cluster\ RCT}$ = total sample size in a cluster RCT;

$SS_{standard\ RCT}$ = total sample size in a standard RCT;

DE = design effect, from (1).

The number of participants required per group in a standard RCT can be readily calculated using trusted online resources (e.g. Sample Size Calculator)[11], and is defined by Equation (3):

Equation 3

$$n_{standard\ RCT} = \frac{2\sigma^2\left(Z_{1-\propto/2} - Z_{1-\beta}\right)^2}{\Delta^2} \quad 3)$$

Where:

Z = the x'th percentage point of the standard normal distribution;

Δ = clinically important difference between groups for the primary outcome measure;

σ^2 = variance of primary outcome measure;

α = significance level;

β = power;

$n_{individual\ RCT}$ = sample size per group.

While the approach described above is simple to implement, it assumes clusters with similar size. Recruitment for clinical trials is usually a challenge, and so most cluster trials tend to end up with unequal cluster sizes.[7] This reduces the statistical power of the trial.[8] Therefore, it is recommended that researchers adopt Equation (2) for estimating, *a priori*, the sample size of a cRCT with unequal cluster sizes, so that the trial will not be underpowered. Guidance on how to estimate the sample size requirement in a cRCT with unequal cluster sizes is provided below – but first, some considerations with regard to the number and size of clusters, and variability of the outcome between clusters.

Number of clusters

When estimating the sample size, researchers need to determine the number of clusters and the number of participants per cluster.[20] Trials should avoid having too few clusters. Using a small number of clusters increases the required sample size (Table 1), because of variance inflation. The greater the number of clusters, the closer to a normal distribution data will be.[18] Adding an extra cluster is an effective way to increase the power of a trial.[6] However, adding an extra cluster to a trial will likely increase costs and logistical challenges, as it will involve recruiting a relatively large number of participants in order to match the size of the other clusters in the trial.[3]

There are cases where the number of clusters is fixed due to geographical or logistic issues.[3,9] In these cases, assuming that the size of clusters is equal, the number of clusters (defined *a priori*) will be appropriate as long as it is larger than the product of the number of required participants and the estimated ICC.[9]

Size of clusters

The size of each cluster impacts on the statistical power of a trial, as it impacts on the variability of the outcome measure.[20] The fewer participants per cluster, the smaller the required sample size (Table 1, Figure 1), however a greater number of clusters then becomes necessary. That occurs because the larger the cluster size, the larger the DE (Equation 1).

In trials with unequal cluster sizes, the larger the size difference between clusters, the larger the sample required to achieve the same statistical power for a certain alpha (Figure 1). That happens because the outcome measure estimates from smaller clusters will be less precise than those from larger clusters, decreasing the statistical power.[13] Therefore, cRCTs with unequal cluster sizes need to account for that loss on statistical power when estimating the sample size.

Intracluster correlation coefficient

The ICC is an estimate of how much variability is present among clustered data, and equals to the between-cluster variability divided by the sum of within- and between-cluster variability (Equation 4).[10] The smaller the ICC, the more precise the outcome measure sampling, and the smaller the required sample size (Table 1).[3]

Equation 4

$$\rho = \frac{\sigma^2_{between}}{\sigma^2_{between} + \sigma^2_{within}} \quad (4)$$

Where:

ρ = intracluster correlation coefficient;

$\sigma^2_{between}$ = between-cluster variance for the outcome measure;

σ^2_{within} = within-cluster variance for the outcome measure.

Outcome measurements are likely to be more similar within a cluster than between clusters,[18] impacting on the independence of measurements and causing variance inflation.[6,18] Therefore, it is necessary to adjust the sample size to compensate for the loss in statistical power.

A strategy to reduce the ICC consists of including baseline measurements as covariates when determining the outcome measure for each cluster.[3,18] This results in reduced variance, as the change in scores between baseline and follow-up are generally less variable than are follow-up scores alone.

Ideally, researchers should conduct a pilot or feasibility trial[16] prior to conducting the full cluster trial.[1] Conducting a pilot or feasibility trial allows estimating the variability of the outcome measure and, most importantly in the case where cRCT is planned, it allows estimating the variability of the ICC.[3]

Sample size estimation for cluster RCT with unequal cluster sizes

The majority of trials tend to have unequal cluster sizes. Hence, researchers should consider cluster size variability when estimating the sample size.[8,18] The challenge is to know the variability of the cluster size prior to starting data collection.[18] Equation (5) can be used for estimating the design effect for unequal cluster sizes.[8] This approach is considered conservative when planning individual-level analysis.[8]

Equation 5

$$DE_{unequal} = 1 + [(1 + cv^2) \times \bar{x} - 1]\rho \quad (5)$$

Where:

cv = coefficient of variation of cluster size;

$\bar{x}$ = mean cluster size;

ρ = intracluster correlation coefficient.

When the coefficient of variation (cv) of cluster size is unknown, it can be estimated by entering the likely minimum and maximum cluster size. First, the standard deviation is estimated by Equation (6), and then the cv is estimated by Equation (7):

Equations 6 and 7

$$sd = \frac{CS_{Range}}{4} \quad (6)$$

$$cv = \frac{sd}{\bar{m}} \quad (7)$$

Where:

sd = standard deviation of cluster size;

CS_{Range} = cluster size range;

cv = coefficient of variation of cluster size.

The sample size for the cluster RCT is then determined (Equation 8)[8,15]:

Equation 8

$$SS_{cluster\ RCT} = SS_{standard\ RCT} \times DE_{unequal} \quad (8)$$

Where:

$SS_{cluster\ RCT}$ = total sample size for cluster RCT;

$SS_{standard\ RCT}$= total sample size for standard RCT;

$DE_{unequal}$= design effect for unequal cluster size.

Planned statistical analysis

The statistical analysis of cRCTs may be undertaken by conducting cluster- or individual-level analysis.[8,12] When conducting cluster-level analysis, outcomes are calculated for each cluster.[8] In this case, the average of individual measurements is calculated within each cluster, and a linear regression model may be used.[20] If clusters have the same size and there are no covariates, the approach is the same as conducting individual-level analysis.[14,20] Cluster-level analysis reduces the data, and uses as input in the statistical analysis only one observation per cluster.[8,18] While simple,[18] this approach is not appropriate in cases where clusters do not have the same size[18], which occurs in the majority of cases.[7] In trials with variable cluster size, individual-level analysis should be planned[18], this is considered more efficient than cluster-level analysis weighted by cluster size.[12] This is commonly undertaken using linear mixed effects model that include cluster as random effects, or alternatively a generalised estimating equation (where the within-cluster correlation is modelled).[18]

Outcome measures

The formula used to estimate the sample size of a cRCT depends on the type of outcome measure (i.e. continuous, binary, count, ordinal, time-to-event, and rate). Musculoskeletal clinical researchers commonly use continuous outcome measures (e.g. VAS, disability scores). Thus, we are focusing on continuous outcome measures only. For more information on this topic, the reader is referred to the literature.[3,18]

Allocation ratio

Equal allocation ratio is the most efficient option requiring the smallest sample size. An unequal allocation ratio may be considered where issues limit the number of participants that can be recruited to a particular arm of the trial, such as cost of an intervention, limited number of

participants available, or difficulty recruiting people into a trial where there is a 50% chance of being allocated to an 'unattractive' group (e.g. no treatment).[3,18] In such cases, consult a biostatistician for assistance with sample size calculations.

Attrition

In a standard RCT, participants may withdraw from the trial. In a cRCT, individual participants within a cluster or the whole cluster (less common) may withdraw. Attrition problems are likely to be higher in cRCTs where there can be lack of researcher contact with individual participants.[3] Researchers may include a drop-out rate for individuals or clusters into the sample size estimation to account for this withdrawal effect (e.g. to account for an estimated 20% dropout, it is necessary to add 25% more participants or two or more additional clusters into the trial, respectively).[18]

Baseline measurements and other covariates

Adding baseline measurements and other covariates in the analysis increases the statistical power. This is particularly useful to minimize the impact of smaller than planned sample size or higher than anticipated drop-out. Adding covariates reduces the between-cluster variability[3], however adding covariates have no direct effect on *a priori* sample size calculations.

Measurements may be taken at individual or cluster level. These measurements may include demographic characteristics (e.g. age, body mass index), and baseline measurements of the primary outcome measure.[18] It is important that the covariate is associated with the primary outcome measure. For example, quadriceps muscle strength has been used in a trial on the management of knee osteoarthritis;[2] and duration of symptoms in a trial on the management shoulder disorders.[21] Measuring covariates may increase the cost of the trial; on the other hand, it will lead to a smaller sample size. If it is simple and inexpensive to gather additional information, then researchers should seriously consider this option. However, only useful additional data should be collected to minimize participant burden. It is possible to estimate and compare the costs of adding covariate measurements or increasing the sample size[18]. More details on the planning and sample size estimation of cRCTs (with different designs and outcome measures) can also be found in the literature.[3,18]

Ribeiro DC, Milosavljevic S, Abbott JH. Sample Size Estimation for Cluster Randomized Controlled Trials. Musculoskelet Sci Pract. 2018;34:108-11.

Final considerations

Cluster clinical trials have design strengths, but also carry challenges that need to be considered at the design stage by clinical researchers. We have discussed factors influencing the design and sample size estimation of cRCTs, and presented two approaches to estimate the sample size: one ideal for trials with balanced and another for trials with unequal cluster sizes.

Acknowledgement

The research team received financial support from Health Research Council Emerging Researcher First Grant (15/527) and Lottery Health Research (AP353007).

References

1. Abbott JH. The distinction between randomized clinical trials (RCTs) and preliminary feasibility and pilot studies: what they are and are not. *The Journal of orthopaedic and sports physical therapy*. 2014;44:555-558.

2. Abbott JH, Robertson MC, Chapple C, Pinto D, Wright AA, Leon de la Barra S, Baxter GD, Theis JC, Campbell AJ, team MOAT. Manual therapy, exercise therapy, or both, in addition to usual care, for osteoarthritis of the hip or knee: a randomized controlled trial. 1: clinical effectiveness. *Osteoarthritis and cartilage / OARS, Osteoarthritis Research Society*. 2013;21:525-534.

3. Campbell MJ, Walters SJ. *How to Design, Analyse and Report Cluster Randomised Trials in Medicine and Health Related Research*. 1. Wiley; 2014.

4. Cornfield J. Randomization by group: a formal analysis. *Am J Epidemiol*. 1978;108:100-102.

5. Donner A, Birkett N, Buck C. Randomization by cluster. Sample size requirements and analysis. *Am J Epidemiol*. 1981;114:906-914.

6. Donner A, Klar N. Pitfalls of and controversies in cluster randomization trials. *Am J Public Health*. 2004;94:416-422.

7. Eldridge SM, Ashby D, Feder GS, Rudnicka AR, Ukoumunne OC. Lessons for cluster randomized trials in the twenty-first century: a systematic review of trials in primary care. *Clin Trials*. 2004;1:80-90.

8. Eldridge SM, Ashby D, Kerry S. Sample size for cluster randomized trials: effect of coefficient of variation of cluster size and analysis method. *International journal of epidemiology*. 2006;35:1292-1300.

9. Hemming K, Girling AJ, Sitch AJ, Marsh J, Lilford RJ. Sample size calculations for cluster randomised controlled trials with a fixed number of clusters. *BMC Med Res Methodol*. 2011;11:102.

10. Killip S, Mahfoud Z, Pearce K. What is an intracluster correlation coefficient? Crucial concepts for primary care researchers. *Ann Fam Med*. 2004;2:204-208.

11. Kohn MA, Jarret MS, Senyak J. Sample Size Calculators. Available at: http://www.sample-size.net/sample-size-means/. Accessed 09/12/2016, 2016.

12. Lake S, Kammann E, Klar N, Betensky R. Sample size re-estimation in cluster randomization trials. *Stat Med*. 2002;21:1337-1350.

13. Lauer SA, Kleinman KP, Reich NG. The effect of cluster size variability on statistical power in cluster-randomized trials. *PLoS One*. 2015;10:e0119074.

14. Moerbeek M, van Breukelen GJ, Berger MP. A comparison between traditional methods and multilevel regression for the analysis of multicenter intervention studies. *Journal of clinical epidemiology*. 2003;56:341-350.

Ribeiro DC, Milosavljevic S, Abbott JH. Sample Size Estimation for Cluster Randomized Controlled Trials. Musculoskelet Sci Pract. 2018;34:108-11.

15. Ribeiro DC, Milosavljevic S, Abbott JH. Effectiveness of a lumbopelvic monitor and feedback device to change postural behaviour: a protocol for the ELF cluster randomised controlled trial. *BMJ Open*. 2017;7:e015568.

16. Ribeiro DC, Sole G, Abbott JH, Milosavljevic S. The effectiveness of a lumbopelvic monitor and feedback device to change postural behavior: a feasibility randomized controlled trial. *The Journal of orthopaedic and sports physical therapy*. 2014;44:702-711.

17. Rotondi M, Donner A. Sample Size Estimation in Cluster Randomized Trials: An Evidence-Based Perspective. *Computational Statistics and Data Analysis*. 2011;56:1174-1187.

18. Rutterford C, Copas A, Eldridge S. Methods for sample size determination in cluster randomized trials. *International journal of epidemiology*. 2015;44:1051-1067.

19. Torgerson DJ. Contamination in trials: is cluster randomisation the answer? *Bmj*. 2001;322:355-357.

20. van Breukelen GJ, Candel MJ. Calculating sample sizes for cluster randomized trials: we can keep it simple and efficient! *Journal of clinical epidemiology*. 2012;65:1212-1218.

21. Yiasemides R, Halaki M, Cathers I, Ginn KA. Does Passive Mobilization of Shoulder Region Joints Provide Additional Benefit Over Advice and Exercise Alone for People Who Have Shoulder Pain and Minimal Movement Restriction? A Randomized Controlled Trial. *Physical therapy*. 2011;91:178-189.