

Planning a cluster randomized controlled trial Methodological issues

Christie, J., O'Halloran, P., & Stevenson, M. (2009). Planning a cluster randomized controlled trial Methodological issues. *Nursing Research*, 58, 128-134.

Published in:
Nursing Research

Queen's University Belfast - Research Portal:
[Link to publication record in Queen's University Belfast Research Portal](#)

General rights

Copyright for the publications made accessible via the Queen's University Belfast Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The Research Portal is Queen's institutional repository that provides access to Queen's research output. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact openaccess@qub.ac.uk.

Planning a Cluster Randomized Controlled Trial

Methodological Issues

Janice Christie ▼ Peter O'Halloran ▼ Mike Stevenson

Editor's Note

Additional information provided by the authors expanding this article is on the Editor's Web site at <http://www.nursing-research-editor.com>.

- ▶ **Background:** The standard approach in a randomized controlled trial (RCT) is to randomize individuals to intervention and control groups. Yet, nursing and other health interventions are often implemented at the levels of health service organizational unit or geographical area. It may be more appropriate to conduct a cluster RCT. However, cluster randomization requires consideration of a number of important issues.
- ▶ **Objective:** The objective of this study was to show how critical issues in relation to design and analysis can be addressed.
- ▶ **Approach:** Two cluster RCTs conducted by the authors are used as examples. Guidance on the conduct and reporting of cluster RCTs is also offered.
- ▶ **Results:** A rationale for choosing this design was provided, and issues in relation to study design, calculation of sample size, and statistical analysis were clarified. A decision tree and checklist are provided to guide researchers through essential steps in conducting a cluster RCT.
- ▶ **Discussion:** Cluster RCTs present special challenges in relation to design, conduct, and analysis. Nevertheless, they are an appropriate and potentially powerful tool for nursing research. With careful attention to the issues addressed in this article, researchers can use this approach successfully.
- ▶ **Key Words:** cluster randomized trial · multilevel model · nested structure

Although many nurse-patient interactions are one-to-one encounters, nursing is essentially a group activity. Most nurses work with groups of colleagues, for example, as part of a ward team in a hospital in Northern Ireland or as a network of community nurses in rural Romania. Similarly, nurses care for people who live or work in groups, perhaps students from a school in Idaho or members of a mountain tribe in Thailand. Conse-

quently, when carrying out nursing research, whether focusing on outcomes for nurses themselves or their patients, it is important to recognize that participants are often linked through membership of a group, and hence, any data collected potentially are clustered. Clustered data are collected from people who are members of a group and who may be presumed, by virtue of the fact that they are members of that group, to have a greater similarity to those

within the group than that to individuals outside it. These similarities may be evident to the observer or unknown. Data collected from such clusters should be distinguished from data that are collected from unrelated individuals and grouped by a researcher on the basis of common characteristics (e.g., age and gender) for the purposes of subgroup analysis. Early researchers often did not address clustering effects either because of computation difficulties or because such effects were not recognized. In recently conducted studies, however, it is more likely that clustered data are controlled for (Bland, 2004).

In this article, the importance of taking into account the clustered nature of the data obtained from groups in planning and analyzing randomized controlled trials (RCTs) is considered. A variant of randomized trials, sometimes called *group randomized* or *community randomized* trials but more

Janice Christie, PhD, MA, RN, RSCPHN, PGCE, is Teaching Fellow, School of Nursing and Midwifery; Peter O'Halloran, PhD, MSc, RN, RNT, is Lecturer, School of Nursing and Midwifery; and Mike Stevenson, BSc, FSS, PGCHET, is Senior Lecturer in Medical Statistics, School of Medicine and Dentistry, Queen's University Belfast, Belfast, Northern Ireland, United Kingdom.

commonly known as *cluster randomized trials*, will be discussed. Here, groups of people rather than individuals are randomized to intervention and control groups. Also identified are the major methodological issues that need to be addressed when planning and undertaking a cluster RCT. This is illustrated through discussion of two trials, one an evaluation of a policy to provide hip protectors (O'Halloran et al., 2004) and the other an evaluation of home visitation schedules offered to first-time mothers by health visitors (United Kingdom-based specialist community public health nurses; Christie, 2005).

Reasons for Randomizing by Cluster

The standard approach in an RCT is to randomize individuals, and, as discussed later, randomizing by cluster raises a number of important design implications. So why use this approach to test an intervention? Such studies may be justified on a number of grounds (Medical Research Council, 2002). First, it may be that the intervention itself is designed to be delivered to groups of people rather than to individuals—for example, nurse-facilitated smoking cessation support groups. Alternatively, the intervention might be a treatment that requires health professionals to change their behavior to impact patient outcomes—for example, where nurses provide an educational intervention for patients with diabetes. It would be difficult for such nurses to avoid using the intervention techniques on patients designated as controls. Randomizing by cluster may also be used if there is a risk of contamination, where individuals randomized to receive an intervention, within a group, may influence others within the group. For example, patients attending a primary care practice who are offered breast screening may talk to other control-group-allocated patients attached to the practice who may then ask for screening themselves. A cluster RCT may be justified if the intervention involves supplying equipment to a particular set of units (clinics, hospitals, etc.). In this scenario, it may be cheaper or more convenient to randomize by unit rather than to supply every unit with the equipment.

Randomizing by cluster can also be appropriate for pragmatic trials in which the effectiveness of an intervention in routine clinical practice is mea-

sured. This method typically requires that interventions are offered to all who are likely to benefit rather than to randomly assigned individuals within a particular patient subgrouping. It should be noted that pragmatic trials are different from explanatory trials which measure the efficacy of an intervention under ideal conditions, often using carefully selected participants in a research environment (Roland & Torgerson, 1998).

Effects of Cluster Randomization

Randomization of individuals in a traditional or nonclustered RCT is carried out in an attempt to ensure that there is a probabilistically equal distribution of relevant individual characteristics between intervention and control groups. Standard statistical tests are suitable for this design because they are based on the assumption that the characteristics of individuals (and therefore their outcomes) are independent of one another. With a cluster RCT design, clusters (rather than individuals) are randomized to treatment and control groups. Consequently, statistical independence cannot be assumed because geographical areas and organizations tend to contain individuals who are more similar to each other than individuals within other areas or organizations. The causes of these group similarities will be discussed later, with examples to help nurse researchers identify potential clusters in their study population.

There are a number of reasons clustering effects exist within study populations, including choices of individuals, the impact of factors that directly influence the whole cluster, and indirect effects of the cluster's social environment. For example, older people or their families may select nursing care homes on the basis of cost, care ethos, or perceived desirability of the institution. In consequence, individual choice may result in residents within the home being more similar in status or aspiration than that of individuals in other establishments. The effect of a common cluster-level influence is evident when maternal well-being outcomes vary systematically between the case loads of community nurses. This is because community nurses may have differing approaches to providing care (Gomby, 1999).

Outcomes may also vary between clusters for reasons related to the wider social environment. For example, com-

munity nurses may be allocated families to visit on the basis of geographical catchment areas. Within these areas, residents and hence potential clients are likely to share socioeconomic status and environmental influences on their health status. If such patients are randomized according to geographical area, the variation in health between clusters (area) is likely to be greater than the health variation within clusters (Ukoumunne, Guillford, Chinn, Sterne, & Burney, 1999). Once potential clusters have been identified, the researcher will need to consider how these clustering effects will influence the design and analysis of a cluster RCT.

Cluster Randomization Design Issues

The unit of analysis within a cluster randomized trial can be either the cluster or the individuals within clusters. When analyzing individual-level data, the researcher must take into account the lack of statistical independence of data collected within each cluster. In this case, the researcher needs to statistically account for clustering effects at the power calculation and data analysis stages of a cluster RCT. Ukoumunne et al. (1999) recommend that researchers address a number of issues when designing a cluster RCT, including taking the cluster design into account when calculating the sample size, allowing for the number and size of included clusters, considering the use of stratification of clusters, and taking account of the cluster design when analyzing the results of the study.

At the outset, the cluster design must be taken into account when estimating the sample size. Standard sample size calculations are based on the assumption that the responses of individuals within clusters are independent. This assumption is unwarranted in a cluster RCT. In comparison with an RCT randomized at the level of the individual, a cluster RCT with the same sample size has a reduced power to detect an intervention effect, thus increasing the risk of concluding that there is not an effect when in reality there is (a Type II error). Therefore, it is essential that an accurate sample size is calculated before the study by allowing for the effect of a cluster randomized design (Figure 1). A cluster randomized trial will always require a larger sample size than that of a comparative noncluster

trial (Kerry & Bland, 1998). This will have cost and workload implications, and therefore calculation of the prospective sample size should be a prelude to decisions regarding the feasibility of a cluster randomized study.

The introduction of a covariate that explains (and hence reduces) between-cluster variation will also increase the power of a study (Raudenbush, 1997). For example, in the study by Christie (2005), maternal well-being variables were measured pretest and posttest for control and intervention groups, and data regarding potential confounding variables such as nurses' years of experience and qualifications were collected and analyzed. In concordance with the CONSolidated standards of Reporting Trials statement (CONSORT) (Begg et al., 1996)—a set of standards endorsed by the International Committee of Medical Journal Editors (2007) designed to alleviate the problems arising from inadequate reporting of RCTs—it is good practice to specify the assumptions used when estimating the size of cluster and within-cluster samples in the trial report (Campbell, Elbourne, & Altman, 2004).

In addition to the overall sample size, thought should be given to the number of clusters recruited. Consider two studies with the same number of participants; a study with a larger number of clusters and fewer individuals within clusters will be better able to distinguish intervention effects than does the one that has fewer clusters but larger numbers of individuals within clusters. It should be noted, however, that recruiting a large number of small clusters may become counterproductive, especially if there is a large amount of individual variation within the cluster. In this case, there will be large standard errors within the cluster (i.e., statistical estimates of intraclass variation will be unstable), leading to greater uncertainty in sample size calculations (Ukoumunne et al., 1999).

Researchers should also consider the use of stratification of clusters. In stratified randomization, recruited clusters are assigned to groups according to cluster-level characteristics that are thought to affect individual outcomes. Clusters are then randomized within strata to ensure a more even distribution of cluster-level characteristics between intervention and control groups. By systematically randomizing

The sample size calculated for randomization at the level of the individual must be multiplied by a factor known as *the design effect* to give a sample size with the same power as a study using individual randomization. To calculate the design effect, first estimate the intraclass correlation (ICC). If individuals within a particular cluster are no more likely to have similar outcomes than those in other clusters, then the ICC will be 0. If all individuals in a cluster have the same outcome, then the ICC will be 1. The larger the ICC coefficient, the greater the design effect, and, hence, a greater sample size will be required to match the power of a study randomized by individual. For standard regression multilevel models the ICC can be estimated by using variance that is shared within group clusters (level 2 variance) and variance that is shared by individuals in those clusters (level 1 variance) in the following equation:

$$\rho = \frac{\sigma_{u0}^2}{\sigma_{u0}^2 + \sigma_e^2}$$

Where ρ (rho) is the ICC, σ_{u0}^2 is random level 2 variance, and σ_e^2 is random level 1 variance. Random level 2 or 1 variance values can be obtained from previous research studies that have used the same outcomes and design as the study planned [e.g. Reading, Harvey, and Mclean (2000)]. Where there is no previous study that presents cluster variance or ICC values, these can be estimated from a pilot study or computer simulation based on previous research findings.

The steps in estimating the ICC, design effect, and sample size will be illustrated from a study carried out by O'Halloran et al. (2004). The aim of the study was to evaluate the effectiveness of a policy of making hip protectors available free of charge in nursing homes. This was a cluster randomized trial, so homes (not individuals) would be randomized to intervention or control groups. The main outcome measure was the rate of hip fractures in intervention and control homes.

The first step was to estimate a sample size for a study where individuals and not clusters were randomized to treatment and control groups. This estimation was based on the rates of hip fracture, and reductions in rates, observed in two trials of hip protectors in Scandinavian nursing homes (Ekman, Mallmin, Michaelsson, & Ljunghall, 1997; Lauritzen, Petersen, & Lund, 1993). Residents in the non-intervention groups of these studies experienced hip fractures at a mean rate of 5.6 per 100 residents over 12 months. Reductions in fracture rates of between 55 and 65 percent were observed in the intervention groups. The expected rate of hip fracture in the non-intervention limb was estimated to be the mean of the rates observed in the Scandinavian studies (a fracture rate of 8.4% in residents over an 18 month period). The power calculation [using *Epi Info* statistical software (CDC, 1999)] indicated that a study randomized by individual would require 716 residents in the intervention group and 1,432 in the control group (a total of 2,148), for 80% power at the 0.05 level of significance to detect a 40% reduction in the rate of hip fracture in the intervention group over an 18-month period.

The second step was to estimate the ICC. As there were no previous studies giving ICCs for similar trials, the ICC was obtained by simulations of frequencies of fractures in homes (Fleiss, 1981). The mean number of residents per home (cluster) was estimated as 36. With 36 residents and a rate of 8.4%, the expected number of fractures was approximately 3. The simulations consisted of calculating ρ for samples with varying numbers of fractures from 0-7, such that each sample had a mean of approximately 3. The values of ρ ranged from approx 0 to 0.03. A conservative approach was adopted and a $\rho = 0.02$ was selected as the basis for the power calculation for the cluster RCT study. (The actual ICC was calculated on completion of the study: average number of participants = 33 per cluster; $\rho = 0.0156$.) After estimating the ICC, the third step was to calculate the design effect using the formula: $Deff = 1 + (n_0 - 1)\rho$ [Where $Deff$ is the design effect, n_0 is the average number of individuals per cluster, and ρ (rho) is the intraclass correlation coefficient (Ukoumunne et al., 1999)]. So, in this case, the calculation for the design effect was: $1 + (36 - 1)0.02 = 1.7$

Once the design effect was calculated, the fourth step was to calculate the sample size. This was the original sample required for individual randomization multiplied by the design effect: $2148 \times 1.7 = 3652$

FIGURE 1. Estimating sample size in a cluster randomized trial. CDC = Centers for Disease Control and Prevention.

within relatively homogeneous strata, the amount of variation within the sample is reduced, making inferential statistical testing more efficient. For example, in the study undertaken by O'Halloran et al. (2004), homes that agreed to participate were allocated randomly to either intervention or control groups in a 1:2 ratio using block

randomization within strata. Strata were determined by the organizational characteristics of the homes thought to affect the risk of hip fracture or the rate of adherence to the use of hip protectors (type, size, client category, and affiliation of the home).

Ukoumunne et al. (1999) recommend that researchers take the cluster

design into account when analyzing the results of the study. The limitations that affect sample size calculations must also be allowed in individual-level analyses. Within a cluster randomized trial, a cluster of individuals rather than individuals themselves are randomized to intervention or control groups. As individuals within such clusters share a measure of common variance, the statistical assumption of independence within the sample is violated. If a researcher conducts a traditional *t* test or an analysis of variance or analysis of covariance type of analysis to test for the effect of treatment on individuals without considering the clustered nature of the data, the results are likely to be inaccurate as standard errors would be underestimated. This is because the between-cluster variance would be averaged across the sample, and this may produce confidence intervals that are narrower and *p* values that are smaller than their actual value. In consequence, because the data have been inappropriately analyzed, it could be inferred falsely that there is a difference between the control and intervention groups when in fact there is none: a Type I error (Snijders & Bosker, 1999). In this scenario, the researcher may assume that the intervention has had an effect when in reality there has been no effect. Figure 2 considers data analysis options for adjusting for clustering effects.

Planning a Cluster RCT

Many of the challenges faced by researchers planning cluster RCTs will also be confronted by those conducting trials randomized by individual. However, there are some issues that pertain to cluster trials in particular (Campbell et al., 2004). All of these issues should be addressed in published reports of the trial. This will not only allow readers to appraise the research critically but also be of use to those planning similar cluster RCTs in the future. Planning and reporting issues pertaining specifically to cluster RCTs and those in common with all RCTs are summarized in a table on the Editor's Web site at <http://www.nursing-research-editor.com> to help guide nurse researchers through the process.

Given the special challenges entailed in conducting a cluster RCT, the first

Three statistical approaches can be used to test the effect of a treatment within a cluster randomized trial (Ukoununne et al., 1999). A researcher can calculate a summary measure for each cluster and compare clusters using standard techniques on the basis of this measure (Donner & Klar, 2000). This cluster-based analysis is not statistically efficient as each cluster of individuals only produces one data value, and when there is a wide cluster variation each summary value needs to be weighted. A second method is to perform statistical analysis of individual level data (using standard techniques such as *t*-tests) which can be adjusted to correct for the design effect, thereby correcting standard errors (Ukoununne et al., 1999).

A third alternative is multilevel modelling (also known as hierarchical linear modelling or random effects modelling). This approach is preferred because of its flexibility: any standard statistical regression based on the General Linear Model (e.g., logistic and Poisson regression for binary and count data, respectively), can be modified to estimate random variation additionally associated with both individuals and clusters. The technique also allows adoption of a model-fitting analysis strategy in which progressive additions to a basic model specification (such as explanatory covariates or random or contextual effects) can be evaluated (Kreft & De Leeuw, 1998). Outcome variation attributable to individuals and clusters can be estimated through either a random intercept or random slope multilevel model. A multilevel random intercept model will provide corrected values of treatment and control group effects in addition to basic information regarding the amount of unexplained outcome variation attributable to individuals and clusters within a data set. However, the researcher can gain additional information regarding the nature of cluster variation (contextual effects) through estimation of random slope models (see Park & Lake, (2005) for further information).

The results of a multilevel analysis will be illustrated from a cluster randomized controlled trial carried out by Christie (2005). Christie (2005) studied the effect of home visitation on first-time family outcomes. Public health nurses (and hence the cluster of families they visited) were allocated randomly to treatment or control groups. Low-risk mothers who agreed to participate in the study received home visits either weekly for 6 weeks (treatment group) or once (control group) during the first 8 weeks following the birth of their first infant. In addition to a range of other family outcomes the researcher established whether these low-risk families had been identified as having any health need by 8 weeks postpartum.

A conventional logistic regression was used to estimate if there was a difference between the identification of control and intervention families' health needs. The result was an odds ratio of 2.095 (95% confidence interval 1.209 to 3.63 and *p* = .008). This result indicated that more families who were visited weekly had been identified with a family health need than those families who were visited once between 2-8 weeks postbirth.

The effect of the intervention was then estimated employing a random intercept model using the statistical package MLwiN (Rashbash et al., 2005). The cluster adjusted odds ratio was found to be 1.94 (95% confidence interval 0.191 to 1.29, *p* = .0912). This analysis indicated that after taking public health nurse clustering effects into consideration, there was no statistically significant difference between control and intervention families in relation to the identification of family health needs. The ICC for this multilevel model analysis was 0.29. This indicated that 29% of the total variance in assessed family need was attributable to differences between individual public health nurses' care. These differences could not be explained by only considering the number of home visits that families had received.

These findings demonstrate the principle that analysing clustered data using conventional statistical tests, without accounting for the effects of clustering, may result in making a type I error (concluding that there is a difference between treatment/control study participants when none exists). Making such a type I error, through incorrect analysis, is most likely to occur when there is a large ICC, a small number of clusters or there is a large number of participants per cluster (Bland, 2004).

FIGURE 2. Analyzing the results of a cluster randomized trial. ICC = intraclass correlation coefficient.

task is to provide a rationale for using a cluster randomized design. A decision tree to help researchers determine whether a cluster design is the optimum approach to their trial is provided (Figure 3). As indicated in Figure 3, if an RCT is not an appropriate methodology for a particular research question or situation, then nurse researchers can also address the impact of clustering effects when using survey and quasi-experimental approaches (Ukoununne et al., 1999). When a cluster RCT is the chosen approach, then it is necessary to specify whether interventions, objectives, hypotheses, and primary and secondary outcome measures per-

tain to the level of the individual, the cluster, or both. Similarly, researchers should determine whether to analyze data at the level of the individual, the cluster, or both. For example, Christie (2005) hypothesized that more postpartum home visits would impact both the individual mother and public health nurse outcomes measured at the level of the case load for each nurse. Thus, in this study, both individual- and cluster-level data were collected and analyzed.

Researchers should specify how the intraclass correlation will be calculated for each primary group-level outcome. The *intraclass correlation*

coefficient (ICC) is “the proportion of the true total variation in the outcome that can be attributed to differences between clusters” (Ukoumunne et al., 1999, p. 22). Thus, an ICC gives an indication of the shared unexplained variance within a cluster (see Figure 1). Then, potential confounding variables should be identified at both the individual and cluster level, and the method of measurement should be specified. Researchers will need to estimate the likely size of clusters as this will affect the calculation of the ICC and therefore the final sample size, together with the number of clusters to be included (studies in which fewer than four clusters are randomized to each group are unlikely to yield conclusive results). It will be evident from this discussion that the ICC is an estimate with a degree of uncertainty around the result. It is reasonable to take this into account by estimating ICC confidence limits and obtaining the corresponding confidence limits for the estimated sample size (Ukoumunne et al., 1999).

Eligibility criteria must be defined not only for individuals but also for clusters. For example, in the cluster RCT by O’Halloran et al. (2004), where the clusters were nursing homes, “all homes registered with the Registration and Inspection Unit (RIU) of the Eastern Health and Social Services Board (EHSSB), Northern Ireland, to offer residential or nursing care to the old and infirm (O&I), and the elderly mentally infirm (EMI)” were eligible for the study (p. 583).

When randomizing to intervention and control groups, the researcher also needs to consider if they will use simple cluster randomization, block randomization, stratification, or matching of clusters. Although both the participants and researchers ideally should be unaware (blind to) who is allocated to treatment or control conditions, this is hard to achieve in cluster randomized trials where often behavioral interventions are being studied (Donner & Klar, 2000).

Researchers naturally will plan which statistical methods will be used to compare groups for primary outcomes and for prespecified additional analyses, such as subgroup analyses. However, purchase of statistical software and associated statistical training may also be required. Although multilevel mod-

Planning and reporting issues pertaining specifically to cluster RCTs and those in common with all RCTs are summarized in a table on the Editor’s Web site.

els can be estimated through many types of statistical software packages, some dedicated programs are available such as MLwiN (Rashbash, Steele, Browne, & Prosser, 2005) and Hierarchical Linear and Nonlinear Modeling (HLM; Raudenbush, Bryk, & Congdon, 2008).

As noted earlier, a cluster RCT is a useful approach to testing the effec-

tiveness of an intervention in routine clinical practice (a pragmatic trial). With this in mind, consideration should be given to intention-to-treat analysis. With this approach, all participants are included in the arm to which they were allocated whether or not they received (or completed) the intervention given to that arm. This gives an estimate of the effect of treatment unbiased by the loss of participants as it includes in the analysis patients who did not get the intended treatment or who deviated from the trial protocol. Thus, it more closely reflects a real-world situation (Heritier, Gebiski, & Keech, 2003).

A first step to achieve an intention-to-treat analysis is to obtain complete data on all randomized participants (Lachin, 2000). Where possible, systems should be considered that will follow up nonresponders. In the study by Christie (2005), all nonrespondents received follow-up by telephone and letter. From a group of 295 mothers who agreed to take part, 3 participants withdrew and 12 did not respond to follow-up. Maternal outcomes were

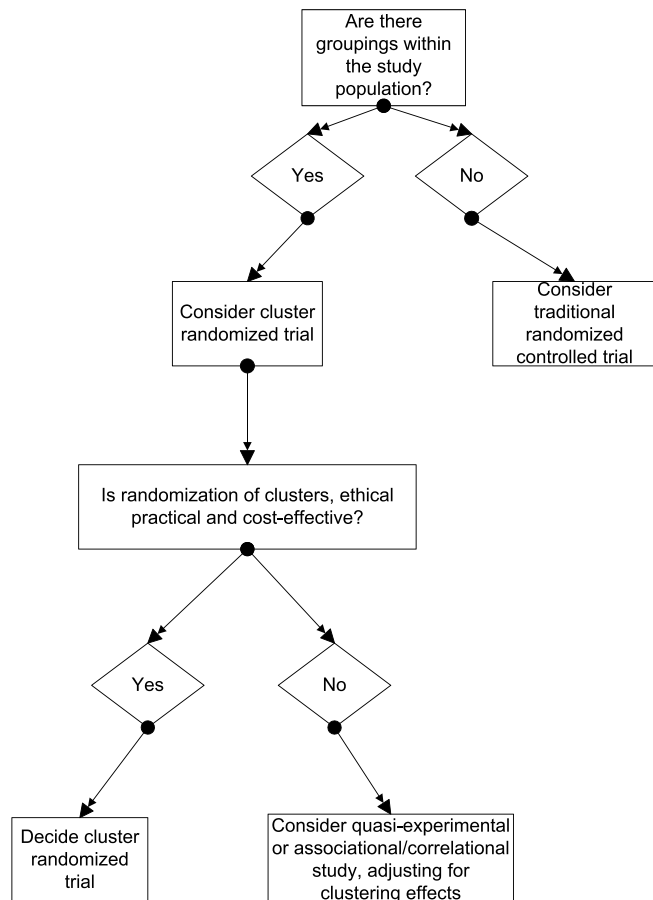


FIGURE 3. Cluster randomized trial decision tree.

analyzed according to which public health nurse group (intervention or control) the parents' health visitor was allocated to initially, even if the mothers did not actually get the intended number of home visits (treatment).

As with all RCTs, trial registration should be considered. Currently, the International Committee of Medical Journal Editors has a policy of requiring researchers to record trial details on an accepted trial registry before patient recruitment into an RCT (Laine et al., 2007). Failure to register a trial may result in the study being rejected by certain journals (De Angelis et al., 2004).

Ethical Issues for Cluster RCTs

Cluster RCTs present special challenges in relation to consent and representation (Medical Research Council, 2002). Because cluster RCTs occur at the level of the group or community, it can be difficult to provide for individual choice within the group when the intervention is implemented. For example, if a health promotion initiative is implemented at the community level, it would be impractical to obtain individual consent from all members of the community and impossible to guarantee that a nonconsenting member of the community would not be exposed to the intervention. When planning a cluster RCT, the first task is to determine whether the intervention is received by a whole cluster together, without the possibility of individual choice, or one where participants can decide individually (Donner & Klar, 2000). If individual consent cannot be obtained, then the researcher must make special efforts to ensure that the interests of individual participants are monitored and guarded. Generally, this means identifying and gaining the approval of the guardians of the participants' interests and the gatekeepers of access to patient groups.

Once the trial is underway, there must be an identified representative individual or body with the authority and ability to monitor and represent the interests of the clusters. Institutional review boards will expect informed individual consent where possible, even in cluster RCTs; however, the Office for Human Research Protections guidance to institutional review boards indicates circumstances where studies may be approved, although they alter standard

consenting processes (Department of Health and Human Services, 2005). These special circumstances include minimal risk to the well-being, rights, or welfare of participants and where it is impractical to carry out the research using standard consenting procedures. The statement further recommends that information should be given to research participants after the research.

Summary

Nursing interventions are often implemented across groups of people within healthcare organizations or geographical areas. Consequently, RCTs evaluating nursing interventions may be more appropriately randomized at the level of the group rather than at the level of the individual. However, cluster RCTs are typically more complex and expensive than nonclustered designs.

Researchers must consider a number of methodological issues when designing a cluster RCT and should determine whether interventions, objectives, hypotheses, and primary and secondary outcome measures pertain to the level of the individual, the cluster, or both and analyze accordingly. They should also specify the potential confounding variables at both individual and cluster level, eligibility criteria for clusters and their likely size, the approach to randomization, and whether the analysis will be by intention to treat. Multilevel modeling is a suitable approach to statistical analysis for cluster RCTs as it can be used to estimate treatment effects and, in addition, outcome variation between and within clusters.

Cluster RCTs make many demands on research design, conduct, and analysis. The decision tree, checklist, and worked examples provided will help guide nurses who are planning a cluster RCT. With careful attention to the issues addressed in this article, researchers can avoid methodological pitfalls and use this approach successfully. ▀

Accepted for publication July 10, 2008.

The authors thank Professor Jean Orr and Mrs. Marianne Moutray, head and associate head of the School of Nursing and Midwifery, Queen's University Belfast, respectively, for their help and support in undertaking this article. Acknowledgement is also due to the Research and Development Office, Depart-

ment of Health, Social Services and Public Safety in Northern Ireland which supported and funded the original research upon which this article is based.

Corresponding author: Janice Christie, PhD, MA, RN, RSCPHN, PGCE, School of Nursing and Midwifery, MBC Building, Queen's University Belfast, 97 Lisburn Road, Belfast BT9 7BL, Northern Ireland, United Kingdom (e-mail: j.christie@qub.ac.uk).

References

- Begg, C., Cho, M., Eastwood, S., Horton, R., Moher, D., Olkin, I., et al. (1996). Improving the quality of reporting of randomised controlled trials. The CONSORT statement. *Journal of the American Medical Association*, 276(8), 637–639.
- Bland, M. (2004). Cluster randomised trials in the medical literature: Two bibliometric surveys. *BMC Medical Research Methodology*, 4(21).
- Campbell, M. K., Elbourne, D. R., & Altman, D. G. (2004). CONSORT statement: Extension to cluster randomised trials. *British Medical Journal*, 328, 702–708.
- CDC. (1999). *Epi info* (Version 6). Retrieved August 24, 2007, from <http://www.cdc.gov/epiinfo/epiinfo.htm>
- Christie, J. (2005). An evaluation of health visitors' postpartum home visits to first-time mothers. A mixed method study, including a cluster randomised trial evaluating the effect of health visitor postpartum home visits to first-time mothers. Unpublished doctoral thesis, University of Ulster, Jordanstown, Northern Ireland, United Kingdom.
- De Angelis, C., Drazen, J. M., Frizelle, F. A., Haug, C., Hoey, J., Horton, R., et al. (2004). Clinical trial registration: A statement from the International Committee of Medical Journal Editors. *Annals of Internal Medicine*, 141(6), 477–478.
- Department of Health and Human Services. (2005). *Title 45—public welfare, Part 46—protection of human subjects*. Retrieved April 18, 2008, from <http://www.hhs.gov/ohrp/humansubjects/guidance/45cfr46.htm#46.116>
- Donner, A., & Klar, N. (2000). *Design and analysis of cluster randomised trials in health care*. London: Arnold.
- Ekman, A., Mallmin, H., Michaelsson, K., & Ljunghall, S. (1997). External hip protectors to prevent osteoporotic hip fractures. *Lancet*, 350(9077), 563–564.
- Fleiss, J. L. (1981). *Statistical methods for rates and proportions*. New York: Wiley.
- Gomby, D. S. (1999). Understanding evaluations of home visitation programs. *Future of Children*, 9(1), 27–43.
- Heritier, S. R., Gebiski, V. J., & Keech, A. C. (2003). Inclusion of patients in clinical trial analysis: The intention to treat

- principle. *Medical Journal of Australia*, 179, 438–440.
- International Committee of Medical Journal Editors. (2007). *Uniform requirements for manuscripts submitted to biomedical journals: Writing and editing for biomedical publication*. Retrieved April 15, 2008, from <http://www.icmje.org/index.html>
- Kerry, S. M., & Bland, M. (1998). Statistics notes: Sample size in cluster randomisation. *British Medical Journal*, 316(549), 1455–1460.
- Kreft, I., & De Leeuw, J. (1998). *Introducing multilevel modeling*. London: Sage.
- Lachin, J. M. (2000). Statistical considerations in the intent-to-treat principle. *Controlled Clinical Trials*, 21, 167–189.
- Laine, C., Horton, R., DeAngelis, C. D., Drazen, J. M., Frizelle, F. A., Godlee, F., et al. (2007). Clinical trial registration: Looking back and moving ahead. *JAMA*, 298(1), 93–94.
- Lauritzen, J. B., Petersen, M. M., & Lund, B. (1993). Effect of external hip protectors on hip fractures. *Lancet*, 341(8836), 11–13.
- Medical Research Council. (2002). *Cluster randomised trials methodological and ethical considerations*. London: Author.
- O'Halloran, P. D., Cran, G. W., Beringer, T. R. O., Kernohan, G. W., O'Neill, C., Orr, J., et al. (2004). A cluster randomised controlled trial to evaluate a policy of making hip protectors available to residents of nursing homes. *Age and Ageing*, 33, 582–588.
- Park, S., & Lake, E. T. (2005). Multi-level modeling of a clustered continuous outcome. *Nursing Research*, 54(6), 406–413.
- Rashbash, J., Steele, F., Browne, W., & Prosser, B. (2005). *MLwiN* (Version 2). Retrieved October 31, 2008, from <http://www.cmm.bristol.ac.uk>
- Raudenbush, S., Bryk, A., & Congdon, R. (2008). *HLM* (Version 6). Retrieved October 31, 2008, from <http://www.ssicentral.com/hlm/index.html>
- Raudenbush, S. W. (1997). Statistical analysis and optimal design for cluster randomised trials. *Psychological Methods*, 2(2), 173–185.
- Reading, R., Harvey, I., & Mclean, M. (2000). Cluster randomised trials in maternal and child health: Implications for power and sample size. *Archives of Disease in Childhood*, 82, 79–83.
- Roland, M., & Torgerson, D. J. (1998). Understanding controlled trials: What are pragmatic trials? *British Medical Journal*, 7127(316), 285.
- Snijders, T., & Bosker, R. (1999). *Multilevel analysis, an introduction to basic and advanced multilevel modeling*. London: Sage.
- Ukoununne, O. C., Guilliford, M. C., Chinn, S., Sterne, J. A. C., & Burney, P. G. J. (1999). Methods for evaluating area-wide and organisation-based interventions in health and health care: A systematic review. *Health Technology Assessment (Winchester)*, 3, iii–92.