Cluster randomized controlled trials

Suzann Puffer BSc,1 David J. Torgerson PhD2 and Judith Watson PhD3
1Research Assistant, 2Director, 3Research Fellow, York Trials Unit, Department of Health Sciences, University of York, York, UK

Correspondence
David J. Torgerson
York Trials Unit
Department of Health Sciences
Area 4
Seebohm Rowntree Building
University of York
York YO10 5DD
UK
E-mail: djt6@york.ac.uk

Keywords: cluster, controlled trial, methodology, randomized

Accepted for publication:
16 September 2004

Abstract
Cluster randomized controlled trial (RCT), in which groups or clusters of individuals rather than individuals themselves are randomized, are increasingly common. Indeed, for the evaluation of certain types of intervention (such as those used in health promotion and educational interventions) a cluster randomized trial is virtually the only valid approach. However, cluster trials are generally more difficult to design and execute than individually randomized studies, and some design features of a cluster trial may make it particularly vulnerable to a range of threats that can introduce bias. In this paper we discuss the issues that can lead to bias in cluster randomized trials and conclude with some suggestions for avoiding these problems.

Introduction
In most clinical trials participants are randomized as individuals to different treatments. However, sometimes individual allocation is not possible or desirable and groups of individuals are randomized instead. This is known as cluster or group randomization. In a cluster trial groups of participants (such as general practices or hospital wards) are the unit of randomization. Periods of time can also form a cluster. Thus, weeks of a year might be randomized so that an intervention might be implemented during 1 week and withdrawn during the control week. The use of cluster randomization, and appropriate statistical analysis, in the context of education was discussed as early as 1940 by Lindquist (Lindquist 1940). This is because many educational evaluations are within a naturally occurring clusters (e.g. classroom or school) and therefore the only feasible approach of undertaking a randomized controlled trial (RCT) is to use cluster allocation. Cluster trials are increasing in popularity among health service researchers (Donner & Klar 2000; Bland 2004). Indeed, health promotion trials among children are often undertaken within a school setting by randomizing whole schools to a novel health promotion curriculum, for example.

There are various reasons why a cluster design may be appropriate. For some types of intervention cluster randomization may be the only feasible approach. For example, if we wished to test the effectiveness of a set of clinical guidelines on patient outcome, we would randomize doctors to receive the guidelines or to act as controls. The effectiveness of the guidelines would be tested on the patients who are registered to those doctors. This approach is probably the only feasible and practicable way of evaluating such an intervention. An alternative approach would be to randomize individual patients to doctors who had received the guidelines and to those who had not. Whilst this approach is, in theory, feasible, in practice it would be difficult or even impossible to implement. Evaluating a vaccine or an intervention to prevent a communicable disease may also require cluster randomization to avoid the ‘herd’ effect. In a trial of communicable diseases the probability of an individ-
ual contracting the disease will be dependent upon the proportion of the surrounding population who are already immune to the disease. Individual randomization will underestimate the effectiveness of the vaccine because, in real life situations, a proportion of the population who remain unvaccinated will receive some protection from the bulk of the population who are vaccinated. Thus, the only sensible method of evaluating such interventions is through allocating people in clusters.

Cluster randomization is often used to avoid ‘contamination’ between those receiving the intervention and those who are not. For example, a randomized trial of an educational intervention to improve the diet and lifestyle of people at high risk of heart disease might use a cluster design to avoid the intervention group telling the control group about the treatment. If this did occur, then there is a danger of dilution bias resulting in a Type II error. To avoid dilution bias participants may be randomized by general practice on the basis that, compared with individual randomization, participants in the intervention practices are less likely to contaminate participants in the control practices. It is this need to avoid the possibility of contamination that has led to an increased use of cluster randomization.

Although RCTs are the most robust evaluative method, poorly conducted trials are susceptible to a number of factors that can bias their results. Methodological reviews of individually randomized trials have shown that rigorously conducted trials produce different effect estimates from those which are poorly conducted (Shulz et al. 1994; Kjaergaard et al. 2001). However, whilst cluster randomized trials are more complex to design and execute than individually randomized trials (Donner & Klar 2000), they have received relatively little attention. Thus, the implications of cluster randomization have not been well understood. Within this paper we attempt to highlight the particular difficulties with conducting cluster randomized trials and identify potential solutions to avoiding or minimizing these problems.

**Statistical issues**

Whilst cluster randomization may be advantageous for the evaluation of specific interventions, there are substantial drawbacks to the use of this design. Firstly, compared with an individually randomized trial testing the same hypothesis, cluster randomization requires a significantly larger sample size. This is because standard sample size calculations assume that outcomes between individuals are uncorrelated. However, when participants are randomized by cluster this assumption no longer holds. The greater the correlation between individuals within a cluster [known as the intraclass correlation coefficient (ICC)] the greater the number of participants required. The ICC is the proportion of the total variance of the outcome that can be explained by the variation between clusters. Typically, cluster trials need 50–100% more participants to achieve the same statistical power as an individually randomized trial: this is known as the design effect. For example, in a review of 36 cluster trials by Puffer and colleagues, only five papers reported the design effect, and these ranged from 21% to 200% with an average (median) of 120% (Puffer et al. 2003).

In general, cluster trials with fewer than five clusters per arm are inadvisable (Medical Research Council 2002). This is because there will be an inadequate number to balance out cluster level confounding. Randomization given sufficient numbers will result in factors that could affect participant outcome to be balanced across the groups. In a cluster trial there will be factors at the cluster level as well as the participant level that need to be balanced.

Secondly, the analysis of cluster trials is not straightforward. If the clustering is not taken into account and the data are treated as independent, the result may be $P$-values that are too small and confidence intervals that are too narrow (Bland 2004). Thus, leading to conclusions that may be false. This problem has been known for more than 60 years (Lindquist 1940). Despite this many cluster randomized trials still use statistical approaches that are appropriate only for individually randomized studies. Statistical methods that account for the clustering effect are necessary. These may take the form of relatively simple methods such as a two sample $t$-test, which instead of using individual patient values, compares cluster means (Kerry & Bland 1998). Alternatively, more complex approaches such as multi-level modelling may be required. The most appropriate analytical technique will depend on the
study design, the number of clusters and the number of individuals per cluster (Medical Research Council 2002).

**Selection bias**

The statistical drawbacks of using cluster randomization are relatively well documented within the literature (Murray 1998; Donner & Klar 2000), and are now therefore given at least some consideration in the majority of cluster trials published in major journals (Puffer et al. 2003; Bland 2004; Eldridge et al. 2004). However, cluster trials are prone to a series of other problems that can bias their results, and until fairly recently these issues have received insufficient attention (Puffer et al. 2003). These issues are discussed below.

In cluster trials potential bias can be introduced at two levels, the first of which is the cluster level. The randomization of clusters needs to be undertaken carefully and preferably independently. Otherwise it is possible for biased allocation to occur. That is, clusters are not allocated randomly and are instead allocated to a particular arm on the basis of reasons that might affect outcome. As has happened in individually randomized trials (Shulz 1995), it is theoretically possible for the allocation of clusters to be subverted. Furthermore, once clusters have been randomized it is important to retain the cluster in its allocated group to avoid the risk of attrition bias.

The second level at which bias can occur is after clusters have been allocated and when individual participants are recruited into the study. If the person recruiting participants has both knowledge of the clinical characteristics of the participants and of the allocation schedule, biased recruitment can occur. Subversion within individually randomized trials can occur when participants with poor prognostic characteristics are randomized so that they are more likely to enter the ‘unfavoured’ group (Shulz 1995; Kennedy & Grant 1997). Furthermore, evidence for the biasing effects of allocation foreknowledge has been shown on treatment effect sizes (Shulz et al. 1994; Kjaergaard et al. 2001). Consequently, those recruiting participants in individually randomized trials ought to be blind to the allocation schedule.

In cluster trials, however, it is not always possible to conceal treatment allocation. Thus, biased recruitment is often a possibility. Furthermore, selection bias can be introduced if participants withheld their consent for either treatment or data collection, particularly if this occurs differentially across the treatment groups. This is a well documented disadvantage of acquiring consent after randomization in individually randomized trials [known as Zelen’s Method (Zelen 1990)] because some refusal of treatment or data collection usually occurs (Altman et al. 1995). This is less problematic in non-Zelen designs, as participants are told in advance about the treatment options and if they decline to be exposed to one of the options they are not randomized.

Whatever the reasons for differences in recruitment, the consequences are potentially the same: selection bias is introduced and the trial results are unreliable. In a recent review Puffer and colleagues identified a sample of 36 cluster randomized trials published in three major medical journals, between 1997 and 2002 (Puffer et al. 2003). They found that 15 of these trials could have experienced bias in their recruitment of participants. Seven of these 15 trials showed some evidence of consenting differential numbers of participants or excluding participants in a selective fashion. Furthermore, whilst having no evidence of bias in the original published paper, one of the eight remaining trials was later subsequently found to have experienced recruitment bias (Jordhoy et al. 2002). Thus, it was found that 25% of cluster trials published in major clinical journals suffered potential selection bias.

A more recent review of 152 cluster randomized trials undertaken in primary care found that only eight (5%) had evidence for differential recruitment (Eldridge et al. 2004). However, unlike the work by Puffer and colleagues, in this review each trial was not carefully scrutinized to ascertain whether or not there was a problem of biased recruitment (Eldridge, personal communication).

Given that cluster trials are more difficult to undertake successfully than individually randomized studies we make the following suggestions that trialists might consider when designing such trials.

**Individual allocation**

As discussed above, one of the main reasons for using cluster randomization is to overcome the per-
ceived threat of contamination. However, whilst this threat may be real in some situations, in others there may be very little contamination. Indeed, in sample size terms it has been demonstrated that even when there are relatively high contamination rates (e.g. 20%) it may still be more efficient to randomize more patients in an individual trial and accept a diluted effect size (Slymen & Hovell 1997; Torgerson 2001). Thus, it is best to avoid using cluster trial methods if at all possible. However, this advice cannot always be adhered to, and for the robust evaluation of particular interventions cluster methods are required.

Prior identification of participants

In an attempt to avoid recruitment bias, in some instances it may be possible to identify participants before cluster allocation. For example, if we consider the evaluation of a school-based health promotion campaign, it would be possible to identify children within schools or individual classes before cluster allocation. The children and their parents are presented with the possible alternatives and are asked for their consent to participate. Once consent is obtained, schools or classes can be randomized to the different groups. Failure to identify children before cluster allocation has the potential to introduce bias. Again, however, prior identification is not always possible.

Independent recruitment

Where prior identification is not feasible, an ‘independent’ person needs to identify and recruit participants. Furthermore, this person should ideally be blind to group allocation. Allowing a person with foreknowledge of group allocation to recruit participants could introduce bias and should be avoided. An example of independent recruitment is provided from a trial by King and colleagues (King et al. 2002). In this trial general practitioners (GPs), in the intervention group, were trained to diagnose and treat depression. Clearly GPs in both groups were aware of their allocation and if they identified trial participants bias could result in one of two ways. First, because of the nature of the training, if it were effective, we might expect the GPs to actually improve patient identification, by identifying either more cases or patients with different clinical characteristics compared with GPs who did not have training. This process would lead to samples of participants in the control and treatment groups who were fundamentally different at baseline leading to bias. Alternatively, or in addition, some GPs may consciously or unconsciously ‘subvert’ the trial by recruiting participants with characteristics such that the study will be shown to be successful. To reduce the possibility of either of these problems King and colleagues used practice receptionists to identify and recruit participants. Receptionists from both groups of practices would have had identical training and therefore should recruit similar trial participants, which should avoid recruitment bias.

Reporting of cluster trials

Cluster trials must not only be designed and analysed appropriately but also reported in such a way that allows readers to assess trial quality and understand how the conclusions were reached. The reporting of trials in general is of poor quality (Huwiler-Muntener et al. 2002). However, since the introduction of the Consolidated Standards of Reporting Trials (CONSORT) statement in 1996 (Begg et al. 1996), and the revised version in 2001 (Moher et al. 2001), this issue has received much attention and improvements have been made. The CONSORT statement includes a checklist of items that should be included when reporting a trial, which are evidence-based whenever possible and are regularly reviewed (Altman et al. 2001). However, the original statement focused on the reporting of individually randomized trials and did not allow for the special methodological circumstances of cluster trials. Indeed a recent systematic review of cluster trials in primary care highlighted the problem of poor reporting and found no improvement in reporting of cluster trials between 1997 and 2000 (Eldridge et al. 2004). Recently, Campbell and colleagues have developed an extended version of the CONSORT statement to include cluster randomized trials (Campbell et al. 2004). The development of this statement will hopefully increase awareness of the implications of cluster randomization, improve the reporting of these trials and aid critical appraisals of trial quality.
Conclusions
Because of the difficulties of conducting cluster RCTs, they should be avoided unless individually randomized trials are inferior or practically impossible. Researchers should justify why an individually randomized trial is not appropriate and explain clearly why a cluster trial is preferable. The particular difficulties of using this design should be given sufficient attention from the initial planning stages of the trial. Furthermore, particular attention should be given to the reporting of the trial to ensure readers are able to assess the methodological quality.

References