Research Master's Programme

Methods and Statistics in Social and Behavioral Sciences

Utrecht University, The Netherlands

MSc Thesis Charlotte Rietbergen

The Design of Cluster Randomized Crossover Trials

May 2009

Supervisor: Dr. Mirjam Moerbeek

Preferred journal of publication: Journal of Educational and Behavioral Statistics

Word count: 6104

The Design of Cluster Randomized Crossover Trials

Abstract

The inefficiency induced by between-cluster variation in cluster randomized trials can be reduced by implementing a crossover design. In a simple crossover trial each subject receives each treatment in random order. A powerful characteristic of this design is that each subject serves as its own control. In a cluster randomized crossover trial clusters of subjects are randomly allocated to a sequence of interventions. Under this design each subject is either included in only one of the treatment periods (crossover at cluster level) or in both periods (crossover at subject level). In this study the efficiency of both cluster randomized crossover trials relative to the cluster randomized trial without crossover is demonstrated. Furthermore, the optimal allocation of clusters and subjects given a fixed budget or desired power level is discussed.

Keywords: cluster randomized crossover design, statistical efficiency, optimal allocation

1 Introduction

For experimental trials in social research the data structure in the population of interest is often hierarchical; individual subjects are nested within clusters of subjects (for an introduction on the topic see Snijders and Bosker (1999), Hox (2002), Raudenbush and Bryk (2002) and Goldstein (2003)). For example, when a researcher wants to evaluate the effectiveness of a high school drug abuse prevention program, the population consists of pupils nested within classes or within schools. When one wants to examine the effect of a parenting intervention program on children's behavioral problems, the data structure consists of children nested within families. In a situation where a researcher is interested in burnout of employees in large businesses, the data structure comprises employees nested within companies.

Random allocation of subjects to intervention conditions could either be at individual level or at cluster level (Moerbeek et al., 2000). That is, either individual subjects are

randomly allocated to the conditions, or complete clusters of subjects are randomly assigned (Donner and Klar, 2000). Randomization of individuals is often problematic from a logistical point of view or due to contamination of treatment effects (Moerbeek, 2005) or high costs. Though cluster randomization enables the researcher to avoid those problems, it induces inefficiency due to between-cluster variation. The proportion of overall variation in response that is accounted for by between-cluster variation is expressed in the intra-cluster correlation coefficient (Donner and Klar, 2000). The degree of inefficiency increases with both the intra-cluster correlation and the average cluster size.

Many efforts have been made in order to compensate for the reduced efficiency in cluster randomized trials. Raudenbush (1997) and Moerbeek (2006) discussed the use of a pretreatment covariate to account for between- and within-cluster variability in order to increase efficiency of the cluster randomized trial. Murray (1998) discussed the procedure of matching or stratification of clusters prior to randomization to the different conditions. This approach will ensure that the conditions are reasonably comparable, at least on the matching or stratification factors (Murray, 1998), which is also a powerful strategy to protect against bias. The present study explores the adoption of a crossover design within a cluster randomized trial as a method to increase efficiency.

A crossover trial is defined as a trial in which subjects are given sequences of treatments with the object of studying the differences between individual treatments or sub-sequences of treatments (Senn, 2002; Jones and Kenward, 2003). The most common crossover design is the AB/BA crossover, where individual subjects are randomly allocated to a sequence of treatments A and B. Subjects in the first treatment sequence receive treatment A in the first period of the trial and B in the second period, whereas subjects in the second sequence receive both treatments in reverse order. In a German study (Widenhorn-Müller et al., 2008) into the effect of having breakfast on cognitive performance and mood in high school students, a crossover design with two periods and two treatment conditions was implemented as follows. Students were randomly assigned to two groups and were tested on two occasions. Students

in the first group received a standardized breakfast on the morning of the first testing day, while students in the second group received none. Seven days later on the morning of the second testing day, students from the second group received breakfast, while students from the first group did not receive breakfast. The attractive feature of this design, where each student is included in both treatment conditions subsequently, is elicited by each student serving as its own control. By studying the difference between the treatment conditions within students, between-subject variation is eliminated from the variance of the treatment effect estimator, which makes this design more powerful than a completely randomized trial (Brown, 1980).

In a cluster randomized crossover trial complete clusters of subjects are randomly assigned to sequences of interventions. For the example above this would mean that instead of randomization of students to one of the breakfast/non-breakfast sequences, complete school classes are allocated to the intervention sequences. Within the cluster randomized crossover design two subtypes can be distinguished; the first with crossover at cluster level and a second with crossover at subject level.

In a cluster randomized cluster crossover trial (CRCC) different subjects within clusters are recruited for each period of the trial. To translate this to the breakfast-example, this could come down to a situation where complete school classes are allocated to one of the breakfast/non-breakfast sequences, where a random half of each class is included in the first period and the other half in the second. This implies that the treatment effect is examined within clusters, which allows elimination of between-cluster variation from the total variance of the treatment effect estimator.

In a cluster randomized individual crossover trial (CRIC) the same subjects are included in both periods of the trial. So, for our example, students who receive breakfast on the first testing day of the trial are the same as those who do not receive breakfast on the second testing, and vice versa. This way the treatment effect is examined within persons within clusters. This design allows elimination of between-cluster as well as between-subject

variation from the variance of the treatment effect estimator.

Though cluster randomized trials are often used in social research, the crossover design within such a trial is not implemented as frequently as it is in medical research. The object of this paper is to stress the amount of efficiency gained by the adoption of the crossover design within a cluster randomized trial. The efficiency of the cluster randomized and ordinary AB/BA crossover design relative to the simple randomized trial is reviewed. Furthermore, the efficiency of both the CRCC and CRIC design relative to the cluster randomized trial is examined. This paper builds on previous papers by Giraudeau et al. (2008) and Harrison and Brady (2004) that provided a sample size formula for continuous outcomes under a cluster randomized crossover trial that is useful in the situation when crossover is at cluster level. The present paper discusses the optimal ratio between the number of clusters and cluster size given a fixed budget or desired power level for both cluster randomized crossover designs.

Though familiar to most readers, the statistical models for the simple randomized, cluster randomized and crossover trial, and the relative efficiency of those designs will be revisited in Section 2. An unambiguous notation of those models provides a solid base for building the more elaborate models for the CRCC and CRIC design in Section 3. In this section the efficiency of both cluster randomized crossover trials relative to the simple cluster randomized trial will be discussed as well. Illustrated by the breakfast-example, Section 4 provides formulae for the optimal allocation of clusters and subjects. Finally, a discussion of the findings can be found in Section 5.

2 Review of the Simple Randomized, Cluster Randomized and Crossover Design

The objective of building the statistical models for the CRCC and CRIC design based on the more familiar models, asks for a uniform use of subscripts and symbols. Therefore this section starts with an explanation of the subscripts used in the model notation of the designs

discussed in this paper.

In a simple randomized trial the arm of the trial is denoted by i, where $i = 1$ indicates the control arm of the trial and $i = 2$ the experimental arm. Subjects in both arms of the trial are indicated by subscript l, where $l = 1, ..., \frac{1}{2}n$ and n equals the total number of subjects included in the trial. In the model notation for cluster randomized trials the same subscripts i and l are used, but now the subjects are nested within clusters. Therefore, in a cluster randomized trial subjects within cluster are denoted by $l = 1, ..., n_1$ and n_1 equals the cluster size. Subscript j denotes the cluster, where $j = 1, ..., \frac{1}{2}n_2$, so that n_2 is the total number of clusters included in the trial. The total number of subjects in a cluster randomized trial equals $n_2 n_1 = n$.

For the crossover design $i = 1$ indicates the intervention sequence where subjects are included in the control condition in the first period of the trial and in the experimental condition in the second period. $i = 2$ indicates the inclusion of subjects in the intervention conditions in reverse order. Subjects in both arms of the trial are denoted by subscript l, where $l = 1, ..., \frac{1}{2}n$ and n equals the total number of subjects included in the trial. As a simple crossover design is assumed with two treatment periods, subscript k is introduced to indicate the period of the trial in which an observation is made. $k = 1$ indicates the first period of the trial and $k = 2$ the second. Since measurements are taken on each subject twice, those measurements are indicated by subscript m, where $m = 1, 2$. Notice that period and measurement occasion in the AB/BA example with only two measurements per subject as assumed here, will always coincide.

The meaning of the subscripts as used for the models of the simple and cluster randomized and crossover trial, will be used for both cluster randomized crossover trials whenever possible. The reader will be informed about deviations from the meaning of the notation in the relevant (sub)sections. For all designs under discussion a balanced design is assumed, so that an equal number of subjects within, if relevant, an equal number of clusters is included in each of the conditions.

2.1 The Simple Randomized and Cluster Randomized Design

The common experimental design used for the examination of the effect of an intervention is the simple randomized (SR) design. Within such a trial individual subjects are randomly assigned to either the control or experimental condition. For simplicity it is assumed observations at post-test are made on each subject only once. The well-known model for subject l in arm i is:

$$Y_{il} = \mu + \phi X_i + u_{il}$$

where μ equals the mean for subjects in the control condition. ϕ is the fixed treatment effect and X_i indicates the treatment received by the subject, where $X_1 = 0$ for subjects in the control group and $X_2 = 1$ for subjects in the experimental group. The independent random error u_{il} is normally distributed with mean zero and variance as displayed in Table 1. For this design σ_u^2 is equal to the total error variance σ^2.

As already explained in the introduction, when the data structure in the population of interest is hierarchical, in some situations it is necessary to use a cluster randomized (CR) design. Here complete clusters of subjects are randomly assigned to either the control or experimental arm of the trial. For this design a mixed-effects model is assumed with a fixed effect for treatment and random effects at cluster and subject level:

$$Y_{ijl} = \mu + \phi X_i + v_{ij} + u_{ijl}$$

where μ is the mean for subjects in the control condition. ϕ is the fixed treatment effect, with X_i indicating the treatment received by a subject as in a simple randomized trial. The random error is now split into two parts; u_{ijl} denotes the random effect at subject level and v_{ij} is added to the model to denote the random error at cluster level. Both are independent and normally distributed with mean zero and variances as displayed in Table 1. The sum of the two variance components equals the total error variance σ^2. The proportion of the

overall variation in scores accounted for by the between-cluster variability, the intra-cluster correlation ρ_2, is calculated as $(\sigma_v^2)/(\sigma_v^2 + \sigma_u^2)$.

—Table 1—

The efficiency of a design relative to another design is defined by the ratio of the variances of the treatment effect estimator under both designs. The variance formulae for the simple randomized and cluster randomized trial can be found in the right column of Table 1. When we take the ratio of the variance of the treatment effect estimator in a simple randomized trial and a cluster randomized trial, we find:

$$\frac{var(\hat{\phi}_{SR})}{var(\hat{\phi}_{CR})} = \frac{1}{1 + (n_1 - 1)\rho_2}$$

We recognize the well-known design effect or variance inflation factor $1 + (n_1 - 1)\rho_2$, that quantifies the effect of the clustering of subjects. In practice this means that the required number of individuals under a simple randomized trial should be multiplied by the variance inflation factor in a cluster randomized trial to obtain the same level of statistical power (Donner and Klar, 2000). The efficiency is plotted in the left hand graph of Figure 1 against the intra-cluster correlation for different cluster sizes. The plot shows that as the intra-cluster correlation increases, the efficiency of the simple randomized trial relative to the cluster randomized trial increases (at a faster rate for larger cluster sizes).

—Figure 1—

2.2 The Crossover Design

As the object of this paper is to examine to what extent the adoption of a crossover design within a cluster randomized trial enables the researcher to deal with the reduced efficiency of the latter, we now revisit the simple crossover (CO) design, where individual subjects are randomly assigned to a sequence of treatments. Subjects in the $i = 1$ arm of the trial are in the control condition at $k = 1$ and in the experimental condition at $k = 2$, and subjects in the $i = 2$ arm of the trial are in the experimental condition at $k = 1$ and in the control condition at $k = 2$. As the subjects in a crossover trial receive the treatments in different periods Senn (2002) describes the possible concern of *period by treatment interaction*, which is present when the effect of a treatment varies according to the period in which it is given. A special case of *period by treatment interaction* is *carry-over*, which occurs when the effect of a treatment given in a certain period lasts into a subsequent period. Adjusting for *carry-over* comes down to examining the treatment effect from observations taken in the period that is unaffected by any previous treatment regimens. In most cases this would come down to a simple parallel group trial and undoubtedly lead to a substantial decrease in effective sample size. One approach to prevent *carry-over* is by implementing a *wash-out period* between two subsequent study periods (as defined by Senn (2002)). In this paper it is assumed that there is no *period by treatment interaction* or *carry-over* effect. Hence, the model for the observations is:

$$Y_{iklm} = \mu + \pi_k + \phi X_{ik} + u_{ikl} + e_{iklm}$$

where μ is the mean for untreated subjects in the first time period and π_k is the fixed period effect, with $\pi_1 = 0$ and $\pi_2 = \pi$. The fixed treatment effect is denoted by ϕ, with X_{ik} indicating the treatment received by subject i in period k. Here, $X_{11} = X_{22} = 0$ for the observations in the control condition and $X_{12} = X_{21} = 1$ for observations in the experimental condition. As the subjects are not clustered, but two observations per subject

are available, there are random effects at subject and measurement level, denoted by u_{ikl} and e_{iklm} respectively. Both are independent and normally distributed with means zero and variances as displayed in Table 1. Again, the variance components add up to the total error variance σ^2. The proportion of the total error variance accounted for by the between-subject variation, or intra-subject correlation ρ_1, is calculated as $(\sigma_u^2)/(\sigma_u^2 + \sigma_e^2)$. It is the correlation between pairs of measurements in Period 1 and 2, taken on randomly selected subjects (Brown, 1980).

Brown (1980) stressed the efficiency of the simple crossover design relative to a completely randomized trial, in case the assumption of no *carry-over* is likely to be valid. The ratio of the error variance of a crossover trial and a simple randomized trial with an equal number of subjects included in both trials, is (Brown, 1980):

$$\frac{var(\hat{\phi}_{CO})}{var(\hat{\phi}_{SR})} = \tfrac{1}{2}(1 - \rho_1)$$

The right hand graph of Figure 1 shows a plot of this ratio against the intra-subject correlation coefficient ρ_1. Since post-test measurements in a crossover design are obtained from each subject in both periods, the crossover design basically is twice as efficient as a simple randomized trial if $\rho_1 = 0$. This efficiency increases as the intra-subject correlation increases.

3 The Cluster Randomized Crossover Design

By adopting the advantageous characteristics of the crossover design within the cluster randomized trial, one is able to deal with the inefficiency of the latter. Two possible fusions of the designs and their efficiency relative to the cluster randomized trial are discussed in this section.

3.1 The Cluster Randomized Cluster Crossover Design

Let us first consider the cluster randomized crossover design with crossover at cluster level (CRCC). In such a trial different subjects from each cluster are included in the separate periods of the trial. As a situation is assumed where the number of measurement occasions equals the number of periods in the trial, each subject is measured only once. A mixed-effect model is assumed, with fixed effects for treatment and period and random effects at cluster and subject level:

$$Y_{ijkl} = \mu + \pi_k + \phi X_{ik} + v_{ij} + u_{ijkl}$$

where again μ equals the mean for untreated subjects in the first time period. π_k and ϕ denote the fixed effects of period and treatment respectively. X_{ik} indicates the treatment received by subject i in period k, where $X_{11} = X_{22} = 0$ for subjects who received the control treatment and $X_{12} = X_{21} = 1$ for subjects who received the experimental treatment. The random effect at cluster level is denoted by v_{ij} and the lowest level random error u_{ijkl} is at subject level. These random terms are independent and normally distributed with means zero and variances as displayed in Table 2. The variance components add up to the total error variance σ^2.

—Table 2—

Table 2 presents formulae to calculate the variance of the treatment effect estimator under both cluster randomized crossover designs. The first formula is used to calculate $var(\hat{\phi})$ under the CRCC design for the assumed situation where the common correlation shared by two subjects within the same cluster and the same period equals the correlation between subjects within the same cluster but in different periods (denoted by η). For the

perhaps more realistic situation where $\eta \neq \rho_2$ Giraudeau et al. (2008) derived the variance formula that is presented in the second row of Table 2.

The amount of efficiency gained by implementing a crossover at cluster level within a cluster randomized trial relative to the basic cluster randomized trial, is again examined. The ratio of the variance of the treatment effect estimator for the CRCC and normal cluster randomized trial equals:

$$\frac{var(\hat{\phi}_{CRCC})}{var(\hat{\phi}_{CR})} = \frac{1-\rho_2}{1+(n_1-1)\rho_2}$$

for the situation where $\eta = \rho_2$. When $\eta \neq \rho_2$ the ratio is given by:

$$\frac{var(\hat{\phi}_{CRCC})}{var(\hat{\phi}_{CR})} = \frac{1+(\frac{1}{2}n_1-1)\rho_2-\frac{1}{2}n_1\eta}{1+(n_1-1)\rho_2}$$

Figure 2 shows the efficiency of the CRCC relative to the cluster randomized trial for $n_1 = 5$ and $n_1 = 50$, plotted against the intra-cluster correlation for different values of η. The solid line in the two plots represent the relationship between the efficiency and the intra-cluster correlation ρ_2, when $\eta = \rho_2$. At $\rho_2 = 0$ the CRCC trial is as efficient as a normal randomized trial, as we have an equal number of observations in both trials. As soon as the intra-cluster correlation increases the efficiency of the CRCC design takes a favorable turn. This effect is tempered by a smaller correlation between two subjects from the same cluster but from different periods. On the other hand, when we compare the two plots, it shows that the same effect is amplified by an increasing cluster size.

—Figure 2—

3.2 The Cluster Randomized Individual Crossover Design

The second type of cluster randomized crossover design is the trial with crossover at subject level (CRIC). Here, each cluster is randomly assigned to one of two arms of the trial, such that subjects within clusters in the first arm receive treatment A in the first period of the trial and treatment B in the second, and subjects within clusters included in the second arm receive treatment B in the first period and treatment A in the second period. A mixed-effects model is assumed, with fixed effects for treatment and period and random effects at cluster, subject and measurement level:

$$Y_{ijklm} = \mu + \pi_k + \phi X_{ik} + v_{ij} + u_{ijkl} + e_{ijklm}$$

where μ is the mean for untreated subjects in the first time period and π_k equals a fixed period effect. The fixed treatment effect is denoted by ϕ and X_{ik} indicates the treatment received by the subject, where $X_{11} = X_{22} = 0$ for the control condition and $X_{12} = X_{21} = 1$ for the experimental condition. The random effects at cluster, subject and measurement level are denoted by v_{ij}, u_{ijkl} and e_{ijklm} respectively. These random terms are independent and normally distributed with means zero and variances as displayed in Table 2. The variance components add up to the total error variance σ^2. The intra-cluster correlation ρ_2 is now calculated as $(\sigma_v^2)/(\sigma_v^2 + \sigma_u^2 + \sigma_e^2)$, whereas the intra-subject correlation ρ_1 is calculated as $(\sigma_u^2)/(\sigma_v^2 + \sigma_u^2 + \sigma_e^2)$.

Comparing the variances of the treatment effect estimator of the crossover and CRIC trial in Table 1 and 2, reveals that for the calculation of $var(\hat{\phi}$ only the lowest level error variance remains. This demonstrates that irrespective of the inefficient consequences of having clustered data, crossing over at subject level within a cluster randomized trial leads to a design that is statistically as efficient as a simple crossover trial.

The efficiency of the CRIC design relative to the cluster randomized trial is given by:

$$\frac{var(\hat{\phi}_{CRIC})}{var(\hat{\phi}_{CR})} = \frac{1}{2}\frac{1-\rho_1-\rho_2}{1+(n_1-1)\rho_2}.$$

This ratio is plotted against the intra-cluster correlation in Figure 3 for $n_1 = 5$ and $n_1 = 50$, at different values of intra-subject correlation ρ_1. The solid lines in the two plots show for $\rho_1 = 0$, that when $\rho_2 = 0$ the CRIC trial is twice as efficient as a cluster randomized trial. This can be explained by the fact that where subjects are measured only once in a cluster randomized trial, they are measured twice in a CRIC trial. The descending lines also indicate that the relative efficiency of the CRIC trial increases as the intra-cluster correlation increases. Moreover, this effect is stronger for larger cluster sizes. This relationship between the relative efficiency of the CRIC and the intra-cluster correlation also shows from the other lines in the plots. Moreover, the different origins of those lines indicate that the higher the intra-subject correlation the sooner the maximum efficiency is reached.

—Figure 3—

4 Optimal Allocation of Units

For the design of any experimental trial, an issue of great importance is the calculation of the required sample size in order to achieve a certain power level to find a presumed treatment effect. For a simple randomized trial this would come down to calculating the required number of subjects per intervention condition. For cluster randomized trials the sample sizes at both levels of the multilevel data structure can be estimated given either a fixed number of clusters or a fixed cluster size. In practice, however, often the interest of the researcher lies not only in finding the required sample size at each level to achieve a certain power level, but in finding the statistically as well as economically optimal allocation of both clusters and subjects.

The object of this section (that builds upon the paper of Moerbeek et al. (2000)) is twofold. The first issue is to provide formulae to calculate the optimal allocation of units for which $var(\hat{\phi})$ is minimized, under the condition that the budget for sampling and measuring units is fixed. Furthermore, a researcher might be interested in the minimal budget required to reach a specific power level, given the optimal allocation. Hence, the second issue is to describe how this budget can be calculated.

Starting with the first issue. The fixed budget is denoted by:

$$C = c_{mp} n_1 n_2 + c_c n_2 \tag{1}$$

where c_c are the costs to sample a cluster. $c_{mp} = (c_m M + c_p)$ denotes the costs to sample and measure a subject, with c_p representing the costs to sample a person within a cluster and c_m equals the costs to measure a person once. M indicates the number of measurements per subject, that is restricted to $M = 1$ for the CRCC trial and $M = 2$ for the CRIC.

To find the optimal n_1, n_2 is expressed in terms of the costs and n_1:

$$n_2 = \frac{C}{c_c + c_{mp} n_1} \tag{2}$$

This expression is then substituted into the formula for $var(\hat{\phi})$, which is then solved for n_1. For the cluster randomized trial with $var(\hat{\phi}_{CR})$ as displayed in Table 1 Moerbeek et al. (2000) found:

$$n_1 = \sqrt{\frac{1 - \rho_2}{\rho_2} \frac{c_c}{c_{mp}}}$$

for the optimal allocation of subjects. The optimal allocation of clusters given the fixed budget comes down to:

$$n_2 = \frac{C}{\sqrt{\frac{1-\rho_2}{\rho_2} c_{mp} c_c} + c_c}$$

They showed that by using a higher total budget C a lower $var(\hat{\phi}_{CR})$ can be obtained. This would result in sampling more units only at cluster level since the sample size at subject levels is independent of C (Moerbeek et al., 2000).

In a crossover trial the interest of the researcher lies in answering the question whether the scores on the outcome variable differ for the different treatment conditions. Therefore, the second issue to address in this section is the budget required for a test to have a specified power to find this difference in mean scores. In order to estimate the required budget, estimates of the expected size of the treatment effect and error variance are needed. Furthermore, a significance level α and desired power level $1-\beta$ have to be chosen. For a two-sided test, the well known formula to calculate the required $var(\hat{\phi})$ is given by:

$$var(\hat{\phi}) = \left(\frac{|\delta|}{z_{1-(\alpha/2)} + z_{1-\beta}} \right)^2$$

With use of the formulae as presented in Table 2 and 3 the required number of clusters and cluster size can be estimated. Substituting these values into Formula 1 gives the required budget.

4.1 Optimal Allocation in the Cluster Randomized Crossover Design

The procedure explained above was followed to find the optimal allocation of units for both the CRCC and CRIC design. In the discussion of the efficiency of the CRCC design, Figure 2 showed the efficiency of the CRCC plotted for different values of the inter-period correlation η. Two possibilities for the value of η can be distinguished, that is either $\eta = \rho$ or $\eta \neq \rho$. Formulae for calculating the optimal allocation of units under both situations were derived separately. When $\eta = \rho$ the formula for $var(\hat{\phi}_{CRCC})$ is as displayed in the first row Table 2. As the costs to sample a cluster are assumed to exceed the costs to sample and measure a subject, it immediately shows that optimal allocation comes down to sampling as many

subjects in as few clusters as possible, this way $var(\hat{\phi}_{CRCC})$ could become infinitely small. However, Snijders and Bosker (1999) advise to sample at least 10 clusters in order to obtain a representative sample. The formula to calculate the optimal cluster size is given in Table 3. The last column shows the formula to calculate $var(\hat{\phi}_{CRCC})$ under optimal allocation given the budget constraint.

When $\eta \neq \rho_2$ the formula for $var(\hat{\phi}_{CRCC})$ is as displayed in the second row Table 2. Substituting Formula 2 and solving for n_1, results in the formula for the optimal cluster size as given in Table 3. Here it is shown that the larger the costs to sample and measure a subject, and the larger ρ_2 and the smaller η, the smaller the number of subjects one should sample. To find the formula to calculate the optimal number of clusters the same procedure is followed. In general, the larger the total budget, the more clusters can be sampled. The larger the costs to sample a cluster and the smaller the costs to sample and measure a subject, the smaller the number of clusters to be sampled. The larger ρ_2 and the smaller η, the larger the number of clusters to be sampled.

For the CRIC design with the formula for the $var(\hat{\phi}_{CRIC})$ as displayed in Table 2 the cluster size n_1 should be as large as possible with a representative n_2. As shown from Table 3 the formulae for the optimal cluster size equals the one for the CRCC design when $\eta = \rho$. The formula to calculate $var(\hat{\phi}_{CRIC})$ under optimal allocation given the budget constraints is given in the last column of Table 3.

—Table 3—

4.2 Example

In the introduction section the simple crossover design was illustrated by a study where the effect of having breakfast on mood and school performance was tested within students. Presuming a multilevel data structure where students are nested within classes, translation

of this example to the cluster randomized crossover trial is easy to imagine. Cluster-wise allocation of subjects to intervention sequences would yield distributing breakfast to complete clusters, what might simplify the logistics. Crossover could either be at subject or at cluster level.

One of the results reported by Widenhorn-Müller et al. (2008) was a negative effect of having breakfast on negative affect in male students, with a treatment difference $\delta = -0.88$ and a standard deviance $\sigma = 3.67$. This result was found to be non-significant with reported $F(1, 102) = 3.08$ and $p = 0.085$. The number of students per measurement occasion was 52. Generally, a larger sample size would result in a smaller standard error of the treatment effect and a possible significant treatment effect. Therefore, the design of each trial demands a priori calculations of the optimal allocation of units to obtain a certain power level given a certain budget. In this section the formulae for optimal CRCC and CRIC designs will be applied to the breakfast example.

With $\sigma = 3.67$ the residual error variance at measurement level would equal $(3.67^2)/2 = 6.734$. Assuming an intra-subject correlation $\rho_1 = 0.60$ and an intra-cluster correlation $\rho_2 = 0.10$, would give an estimated decomposition of the total variance as displayed in Table 4. With use of these values the optimal allocation of units is calculated for both cluster randomized crossover designs under a certain budget. As the budget for the breakfast trial is not reported, let us assume the total budget equals $C = 400$, with $c_m = 0.5$, $c_p = 1$ and $c_c = 3$. A comparison is made between the results obtained by calculating the optimal allocation of students and classes under the CRCC and CRIC design.

—Table 4—

Using the CRCC design in this scenario would mean that each class is randomly assigned to a sequence of having and not having breakfast. A random half of the students in each class is included in the first period of the trial and the other half in the second period.

This way the effect of having breakfast on mood and performance is tested within classes. Section 3.1 showed that his aspect makes this design more efficient than a cluster randomized trial without crossover. An additional advantage is that only one half of the students will suffer from breakfast-deprivation. Using the formulae for optimal cluster size and number of clusters as displayed in Table 3, for the assumed situation $\eta = 0.07$ under the given budget constraints, gives $n_1 = 10.95$ and $n_2 = 20.59$. Rounding these to integers, as we are dealing with students and classes, gives $n_1 = 11$ and $n_2 = 20$. The total costs are calculated to be $C = 1.5 \cdot 11 \cdot 20 + 3 \cdot 20 = 390$ and $var(\hat{\phi}_{CRCC}) = 0.435$.

In a CRIC design complete classes either receive or do not receive breakfast on the morning of the first testing day. Students in all classes are tested on mood and school performance. On the morning of the second testing day, the other classes receive breakfast. Again measurements are obtained on each student. The effect of having breakfast is tested within students within classes. Section 3.2 showed how this design is even more efficient than the CRCC. Optimal allocation in a CRIC comes down to sampling as many students per class as possible in a representative number of classes. Sampling 10 classes would make $c_c n_2 = 3 \cdot 10 = 30$. So, the budget to sample and measure students equals $400 - 30 = 370$. The costs to sample and measure a student are $c_{pm} = 0.5 \cdot 2 + 1 = 2$. The total number of students to be sampled equals $370/2 = 185$. So, about 18 students per class can be sampled, which seems to be a realistic value. The variance of the treatment effect estimator would then equal $var(\hat{\phi}_{CRIC}) = 0.0748$, which is considerably smaller than $var(\hat{\phi}_{CRCC})$. To achieve a power level of $1 - \beta = .80$ with $\alpha = 0.05$ for a two-sided test $var(\hat{\phi}) = 0.0986$ is required. An even smaller value is achieved with the CRIC design. The optimal $var(\hat{\phi}_{CRCC})$ is much higher, making this design clearly underpowered. Using the formulae to calculate the optimal allocation under the cluster randomized design as given by Moerbeek et al. (2000) would give $var(\hat{\phi}_{CR}) = 0.656$. Which confirms the findings that the use of crossover within the cluster randomized design increases the statistical efficiency.

5 Conclusions and Discussion

Hierarchically structured data often force the social or medical researcher to randomize complete clusters of subjects to intervention conditions. In doing so, statistical inefficiency is induced by between-cluster variation. Adoption of a crossover design within a cluster randomized trial enables the researcher to deal with this inefficiency. Two procedures are possible. In a cluster randomized cluster crossover trial different subjects are included in each period of the trial. The treatment effect is calculated within clusters and between-cluster variation is eliminated from the variance of the treatment effect estimator. This results in reduced standard errors and hence enhances power. In a cluster randomized individual crossover trial, where the same subjects are included in each period, the treatment effect is calculated within subjects within clusters. The elimination of between-cluster as well as between-subject variation from the variance of the treatment effect estimator reduces the standard error to an even greater extent, which makes this design extremely efficient as compared to the cluster randomized design without crossover.

Despite its highly efficient nature, the CRIC design cannot always be adopted for the same reasons the simple crossover design cannot always be used. In principal the crossover design requires an underlying condition that does not really change over time. Furthermore, withdrawal of the treatment should reverse the effect of that treatment, so that the effect of a treatment given in a certain period does not last into a subsequent period. As explained before, a washout period can be implemented, but this is not always sufficient. For example, consider a trial where a researcher wants to examine the effect of two parenting-style intervention programs on children's problem behavior. One of the treatments given to half of the parents in the first period is so successful that the parents do not want to give up their new learned approach. It is likely that the effect of the first treatment lasts into the second period, making it impossible to compare the two treatments within these parents. The breakfast-study, however, is a nice example where the crossover design can be implemented at cluster as well as subject level.

Another shared disadvantage of the simple crossover and CRIC design is the issue of drop-out. As a crossover design consists of more than one study period, the duration of the study is usually longer than that of a simple randomized trial. The longer the duration of the study, the higher the drop-out rates. Logically, the treatment effect can not be calculated within a subject if that subject is absent in a subsequent period. Fortunately, both drawbacks of the CRIC design are irrelevant for the CRCC design, as different subjects are included within both periods.

A common approach to deal with the inefficiency induced by between-cluster variation is the use of covariates. Well selected covariates can explain part of the error variance, which will result in smaller standard errors and likewise higher statistical power. For both types of cluster randomized crossover trials the lowest level error variance is used to calculate the variance of the treatment effect estimator. Likewise, relevant covariates are those that explain variance at the lowest of the data structure. In the CRCC design, where the individual subjects make up the lowest level, included covariates should explain variance at subject level. For the CRIC design, where the lowest level refers to the measurements, relevant covariates are those that explain variance at measurement level.

For cluster randomized trials the allocation of clusters to intervention conditions is by definition a random procedure. When only few clusters are available, randomization could induce bias as clusters with different prognostic characteristics are not distributed evenly over the different intervention conditions. Therefore, it is much advised to balance the clusters over the conditions according to relevant covariates. In a CRCC or CRIC clusters are assigned to a sequence of interventions, so that each cluster receives each intervention in random order. As emphasized throughout this paper, the treatment effect is calculated within clusters or even within subjects within clusters. Next to gaining efficiency, this also protects from bias. This is especially relevant for the CRIC design where each subject receives each treatment. For the CRCC design, however, bias might be induced when subjects included in different periods differ significantly on relevant background variables.

To calculate the optimal allocation of clusters and subjects Section 4 provided formulae that assume known values for the error variance at each level of clustering or known intra-cluster or intra-subject correlation coefficients. In order to support the researcher in carefully planning an experimental trial, realistic estimates of the error variance or estimates of the intra-cluster and intra-subject correlation are required. Concerning cluster randomized trials without crossover Murray and Short (1997), Murray et al. (2001), Campbell et al. (2001), Murray et al. (2002), Smeeth and Siu-Woon (2002) and Adams et al. (2004) provide estimates of the intra-cluster correlation for different areas of research. Unfortunately, these estimates are not available for cluster randomized crossover trials. Accurate recording and documentation of intra-cluster and intra-subject correlation estimates obtained in current research, will provide more reliable calculations of optimal allocation of clusters and subject for future research.

References

Adams, G., Gulliford, M. C., Ukoumunne, O. C., Eldridge, S., Chinn, S., and Campbell, M. J. (2004). Patterns of intra-cluster correlation from primary care research to inform study design and analysis. *Journal of Clinical Epidemiology*, 57(8):785 – 794.

Brown, B. W. (1980). The crossover experiment for clinical trials. *Biometrics*, 36(1):69–79.

Campbell, M. K., Mollison, J., and Grimshaw, J. M. (2001). Cluster trials in implementation research: estimation of intracluster correlation coefficients and sample size. *Statistics in Medicine*, 20(3):391–399.

Donner, A. and Klar, N. (2000). *Design and Analysis of Cluster Randomization Trials in Health Research*. Londen, UK: Arnold.

Giraudeau, B., Ravaud, P., and Donner, A. (2008). Sample size calculation for cluster randomized cross-over trials. *Statistics in Medicine*, 27:5578–5585.

Goldstein, H. (2003). *Multilevel Statistical Models*. London: Edward Arnold.

Harrison, D. A. and Brady, A. R. (2004). Sample size and power calculations using the noncentral t-distribution. *The Stata Journal*, 4(2):142–153.

Hox, J. (2002). *Multilevel Analysis: Techniques and Applications*. Mahwah, NJ: Lawrence Erlbam Associates, Inc.

Jones, B. and Kenward, M. G. (2003). *Design and Analysis of Cross-over Trials*. CRC Press.

Moerbeek, M. (2005). Randomization of clusters versus randomization of persons within clusters: Which is preferable? *American Statistician*, 59(2):173 – 179.

Moerbeek, M. (2006). Power and money in cluster randomized trials: when is it worth measuring a covariate? *Statistics in Medicine*, 25(15):2607–2617.

Moerbeek, M., van Breukelen, G., and Berger, M. (2000). Design issues for experiments in multilevel populations. *Journal of Educational and Behavioral Statistics*, 25(3):271–284.

Murray, D. M. (1998). *Design and analysis of group-randomized trials*. New York, USA: Oxford U.P.

Murray, D. M., Alfano, C. M., Zbikowski, S. M., Padgett, L. S., Robinson, L. A., and Klesges, R. (2002). Intraclass correlation among measures related to cigarette use by adolescents: Estimates from an urban and largely african american cohort. *Addictive Behaviors*, 27(4):509 – 527.

Murray, D. M., Phillips, G. A., Birnbaum, A. S., and Lytle, L. A. (2001). Intraclass correlation for measures from a middle school nutrition intervention study: Estimates, correlates, and applications. *Health Education & Behavior*, 28(6):666–679.

Murray, D. M. and Short, B. J. (1997). Intraclass correlation among measures related to tobacco use by adolescents: Estimates, correlates, and applications in intervention studies. *Addictive Behaviors*, 22(1):1 – 12.

Raudenbush, S. (1997). Statistical analysis and optimal design for cluster randomized trials. *Psychological Methods*, 2:173–185.

Raudenbush, S. W. and Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods (Advanced Quantitative Techniques in the Social Sciences)*. Sage Publications, Inc.

Senn, S. (2002). *Cross-over Trials in Clinical Research*. Chichester, West Sussex, UK: Wiley.

Smeeth, L. and Siu-Woon, E. (2002). Intraclass correlation coefficients for cluster randomized trials in primary care: data from the mrc trial of the assessment and management of older people in the community. *Controlled Clinical Trials*, 23(4):409 – 421.

Snijders, T. A. and Bosker, R. J. (1999). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. London etc.: Sage Publishers.

Widenhorn-Müller, K., Hille, K., Klenk, J., and Weiland, U. (2008). Influence of having breakfast on cognitive performance and mood in 13- to 20-year-old high school students: Results of a crossover trial. *Pediatrics*, 122(2):279–284.

Table 1: Variance Components at Different Levels of the Multilevel Data Structure and Variances of the Treatment Effect Estimator.

	Variance Components			
Design	Cluster	Person	Measurement	$\text{Var}(\hat{\phi})$
SR		σ_u^{2*}		$\frac{4\sigma^2}{n}$
CR	σ_v^2	σ_u^{2*}		$\frac{4\sigma^2}{n}(1+(n_1-1)\rho_2)$
CO		σ_u^2	σ_e^{2*}	$\frac{2\sigma^2}{n}(1-\rho_1)$

Note: $*$ = lowest level error variance

Table 2: Variance Components at Different Levels of the Multilevel Data Structure and Variances of the Treatment Effect Estimator.

	Variance Components			
Design	Cluster	Person	Measurement	$\text{Var}(\hat{\phi})$
CRCC $\eta = \rho_2$	σ_v^2	σ_u^{2*}		$\frac{4\sigma^2}{n}(1-\rho_2)$
CRCC $\eta \neq \rho_2$	σ_v^2	σ_u^{2*}		$\frac{4\sigma^2}{n}\left[1+(\frac{1}{2}n_1-1)\rho_2-\frac{1}{2}n_1\eta\right]$
CRIC	σ_v^2	σ_u^2	σ_e^{2*}	$\frac{2\sigma^2}{n}(1-\rho_1-\rho_2)$

Note: $*$ = lowest level error variance

Table 3: Optimal Allocation of Units and Var($\hat{\phi}$) under Optimal Allocation.

Design	n_1	n_2	var($\hat{\phi}$)
CRCC $\eta = \rho_2$	$\frac{C-10c_c}{10c_{mp}}$	10	$\frac{4\sigma^2 c_{mp}(1-\rho_2)}{C-10c_c}$
CRCC $\eta \neq \rho_2$	$\sqrt{\frac{1-\rho_2}{\frac{1}{2}\rho_2-\frac{1}{2}\eta}\frac{c_c}{c_{mp}}}$	$\frac{C}{\sqrt{\frac{1-\rho_2}{\frac{1}{2}\rho_2-\frac{1}{2}\eta}c_{mp}c_c}+c_c}$	$\frac{\left(\sqrt{(1-\rho_2)c_{mp}}+\sqrt{\frac{1}{2}\rho_2-\frac{1}{2}\eta}\right)^2 4\sigma^2}{C}$
CRIC	$\frac{C-10c_c}{10c_{mp}}$	10	$\frac{2\sigma^2 c_{mp}(1-\rho_1-\rho_2)}{C-10c_c}$

Table 4: Variance Decomposition Breakfast-Example for CRCC and CRIC

	CRCC	%	CRIC	%
σ_v^2	2.245	10	2.245	10
σ_u^2	20.202	90	13.468	60
σ_e^2			6.734	30
σ^2	22.447	100	22.447	100

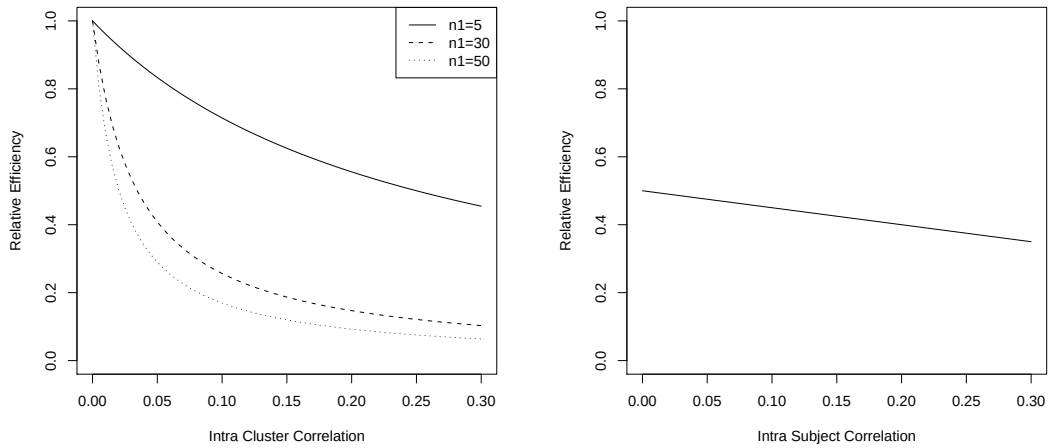

Figure 1: Efficiency of the SR design relative to the CR design as a function of ρ_2 for varying cluster sizes, on the left hand side. In the right plot, the efficiency of the CO design relative to the SR design as a function of ρ_1.

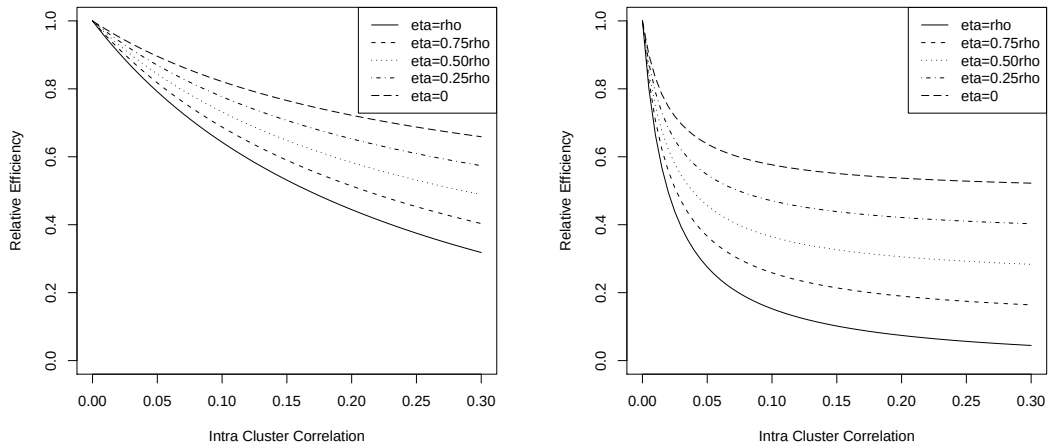

Figure 2: Efficiency of the CRCC design relative to the CR design as a function of ρ_2, for different values of η. Plotted in the left figure for $n_1 = 5$ and in the right for $n_1 = 50$.

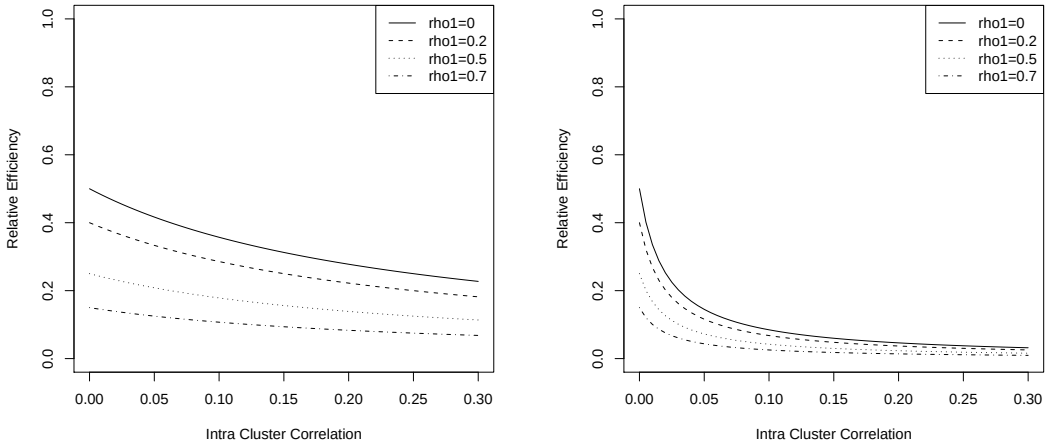

Figure 3: Efficiency of the CRIC design relative to the CR design as a function of ρ_2, for different values of ρ_1. Plotted in the left figure for $n_1 = 5$ and in the right for $n_1 = 50$.