



The cluster randomized crossover trial: The effects of attrition in the AB/BA design and how to account for it in sample size calculations

Mirjam Moerbeek

Abstract

Background/Aims: This article studies the effect of attrition in the cluster randomized crossover trial. The focus is on the two-treatment two-period AB/BA design where attrition occurs during the washout period. Attrition may occur at either the subject level or the cluster level. In the latter case, clusters drop out entirely and provide no measurements in the second period. Subject attrition can only occur in the cohort design, where each subject receives both treatments. Cluster attrition can also occur in the cross-sectional design, where different subjects are measured in the two time periods. Furthermore, this article explores two different strategies to account for potential levels of attrition: increasing sample size and replacing those subjects who drop out by others.

Methods: The statistical model that takes into account the nesting of subjects within clusters, and the nesting of repeated measurements within subjects is presented. The effect of attrition is evaluated on the basis of the efficiency of the treatment effect estimator. Matrix algebra is used to derive the relation between efficiency, the degree of attrition, cluster size and the intraclass correlations: the within-cluster within-period correlation, the within-cluster between-period correlation and (in the case of a cohort design) the within-subject correlation. The methodology is implemented in two Shiny Apps.

Results: Attrition in a cluster randomized crossover trial implies a loss of efficiency. Efficiency decreases with an increase of the attrition rate. The loss of efficiency due to attrition of subjects in a cohort design is largest for small number of subjects per cluster-period, but it may be repaired to a large degree by increasing the number of subjects per cluster-period or by replacing those subjects who drop out by others. Attrition of clusters results in a larger loss of efficiency, but this loss does not depend on the number of subjects per cluster-period. Repairing for this loss requires a large increase in the number of subjects per cluster-period. The methodology of this article is illustrated by an example on the effect of lavender scent on dental patients' anxiety.

Conclusion: This article provides the methodology of exploring the effect of attrition in cluster randomized crossover trials, and to repair for attrition. As such, it helps researchers plan their trial in an appropriate way and avoid underpowered trials. To use the methodology, prior estimates of the degree of attrition and intraclass correlation coefficients are needed. It is advocated that researchers clearly report the estimates of these quantities to help facilitate planning future trials.

Keywords

Cluster randomized trial, crossover, attrition, efficiency, statistical power

Introduction

A common research design in biomedicine is the cluster randomized trial.^{1–5} Clusters are randomized to treatment conditions, so all subjects within the same cluster receive the same treatment. Outcomes of subjects in the same cluster tend to be correlated; hence, the cluster

Department of Methodology and Statistics, Utrecht University, Utrecht, The Netherlands

Corresponding author:

Mirjam Moerbeek, Department of Methodology and Statistics, Utrecht University, PO Box 80140, 3508 TC Utrecht, The Netherlands.
Email: m.moerbeek@uu.nl

randomized trial is less efficient than the individual randomized trial. It is obvious that efficiency can be improved by increasing sample size. However, this is not always feasible and alternative strategies have been proposed, such as reducing residual variance by including covariates,⁶⁻⁹ balancing, matching and stratification,^{5,10} taking repeat measurements^{5,11-13} and introducing a crossover design.¹⁴⁻¹⁷

With the crossover design, clusters are randomized to a sequence of treatment conditions rather than just one treatment condition. As such, each cluster serves as its own control, which has an advantageous effect on efficiency. The simplest crossover design is the so-called AB/BA design in which one group of clusters receives treatment A in the first time period and treatment B in the second, and the other group of clusters receives both treatments in reverse order.

Two different variants may be distinguished depending on whether different or the same subjects are enrolled in the two time periods or, stated differently, on whether a cross-sectional or cohort design is used. In the case of a cross-sectional design, the cross-over is at the level of the cluster. This design is suitable when it is not possible to use the same subjects in both time periods. Consider a trial in which two programs to help adolescents to better support their peers with a mental health problem were compared in the school setting.¹⁸ Both programs were designed for year 10 students, and hence different cohorts were used in the 2 years. Another reason for using different cohorts is that interventions that promote to change participants' behavior can often not be undone at the subject level.

In the case of a cohort design, the crossover is at the level of the subject. An example is a study in which the New Nordic Diet was compared with standard lunches.^{19,20} Nine Danish municipal schools were randomized to treatment sequences; participants were children 8-11 years old. Apparently, the researchers were willing to believe the effect of the intervention of the first time period did not carry over into the second.

The cohort and cross-sectional design have been evaluated on the basis of their efficiency.¹⁷ As is obvious, the cohort design is more efficient since each subject receives both conditions and therefore serves as its own control. This is true when indeed two measurements are made on each subject. In longitudinal studies, which crossover studies essentially are, attrition is the rule rather than the exception. Attrition rates of 5%-10% are not uncommon in crossover studies and they can be as high as 25%.²¹ The rate of attrition and reasons for attrition depend on the type of treatments, subjects and clusters, the treatment sequence and the duration between the two treatment periods. In the dietary study, 3.4% of children who entered the second period did not complete the trial because they did not like the school meals, changed school or class, disliked the measurements or found them time consuming, were

lost to follow-up, or for other reasons. An attrition rate of 7% was found in a study on the effects of an exercise program on musculoskeletal symptoms and physical capacity among nursing staff in four geriatric wards.²² Attrition only occurred in the group where the exercise program was followed by the control and one of the reasons for attrition was not being interested in the control. Other reasons were pregnancy, sick leave, starting to study or a new job.

Attrition may also occur at the cluster level. In a cross-sectional design, a cluster may drop out when it is not able or willing to recruit subjects for the second time period. Other reasons for attrition are the trial is too time-consuming or too difficult to implement, the professional who delivers treatment is unavailable in the second time period, or the cluster is not interested to implement the treatment of the second time period.

This article extends previous work on cluster randomized crossover trials¹⁷ by taking attrition into account. The efficiency of the crossover design is related to the degree of attrition. The cross-sectional design without attrition is compared with the cohort design with subject attrition and it is studied at what degrees of attrition the first outperforms the latter. In addition, two strategies to repair for the loss of efficiency are explored: increasing sample size and replacing those subjects who drop out by others. It is assumed that missing data are missing (completely) at random. Statistical models for attrition that is informative are more complicated and outside the scope of this article.²³ Furthermore, a restriction to linear models for a quantitative outcome is made.

Methods

Review of the crossover design

The focus of this contribution is on the AB/BA design.²⁴⁻²⁶ Each cluster receives each of the two treatments A and B, but in separate time periods, leading to the formation of two cluster-periods. It is assumed that carryover is absent, for instance, by using a washout period of sufficient length. The model that relates the outcome variable to treatment condition should take into account different levels of nesting: clusters, subjects and, in the case of a cohort design, measurements within subjects. An appropriate model is the multilevel model.²⁷⁻³⁰

Cross-sectional design. The multilevel model for the cross-sectional design is given by

$$y_{hij} = \gamma_0 + \gamma_1 x_{1h} + \gamma_2 x_{2hj} + u_{hj} + e_{ij} \quad (1)$$

where y_{hij} is the quantitative outcome of subject $i = 1, \dots, m$ in cluster $j = 1, \dots, k$ in time period $h = 1, 2$. The total number of clusters, k , is assumed

even, with $k/2$ clusters per treatment sequence. The predictor x_{1h} denotes the time period and is equal to 0 for the first period and 1 for the second. The predictor x_{2hj} indicates the treatment received by cluster j in time period h and is equal to 0 for the control and 1 for the intervention. u_{hj} is the random effect of cluster j in time period h and e_{hij} is the random effect of subject i in cluster j in time period h . They have a mean of zero and variance $var(e_{ij}) = \sigma_I^2$ and $var(u_{1j}) = var(u_{2j}) = \sigma_{CP}^2 + \sigma_C^2$, and covariance $cov(u_{1j}, u_{2j}) = \sigma_{CP}^2$. Here, σ_C^2 is the variance in mean outcome between clusters, σ_{CP}^2 is the variance in mean outcome between cluster-periods and σ_I^2 is the residual variance.

The total variance is $var(y_{hij}) = \sigma_T^2 = \sigma_I^2 + \sigma_{CP}^2 + \sigma_C^2$. The within-cluster within-period correlation is the correlation between outcomes of two different subjects i and i' in the same cluster-period

$$cor(y_{1ij}, y_{1i'j}) = cor(y_{2ij}, y_{2i'j}) = \frac{\sigma_{CP}^2 + \sigma_C^2}{\sigma_I^2 + \sigma_{CP}^2 + \sigma_C^2} = \rho \tag{2}$$

The within-cluster between-period correlation is the correlation between outcomes of two subjects in the same cluster and different periods

$$cor(y_{1ij}, y_{2i'j}) = \frac{\sigma_{CP}^2}{\sigma_I^2 + \sigma_{CP}^2 + \sigma_C^2} = \eta \tag{3}$$

The weights γ_0 , γ_1 and γ_2 are the overall mean outcome in the first time period for the control condition, the period effect and treatment effect, respectively. The latter is of primary interest and it is estimated by taking the difference in the two period means across all clusters. The related variance is

$$var(\hat{\gamma}_2) = \frac{2\sigma_T^2(1 + (m - 1)\rho - m\eta)}{mk} \tag{4}$$

The derivation of this equation is given elsewhere.¹⁵ Equation (4) shows that the treatment effect estimate is less efficient when ρ increases and/or when η and/or m decreases.

Cohort design. With the cohort design, each subject is measured in both time periods. This implies an extra level, the repeated measures level, is needed

$$y_{hij} = \gamma_0 + \gamma_1 x_{1h} + \gamma_2 x_{2hj} + u_{hj} + e_{ij} + m_{hij} \tag{5}$$

The term m_{hij} is the random effect at the measurement level. The three random effects have a mean of zero and variances $var(m_{hij}) = \tilde{\sigma}_M^2$, $var(e_{ij}) = \tilde{\sigma}_I^2$ and $var(u_{1j}) = var(u_{2j}) = \tilde{\sigma}_{CP}^2 + \tilde{\sigma}_C^2$, and covariance $cov(u_{1j}, u_{2j}) = \tilde{\sigma}_{CP}^2$. Note that a tilde is used to distinguish these variances from those in the cross-sectional design.

The total variance is $var(y_{hij}) = \sigma_T^2 = \tilde{\sigma}_{CP}^2 + \tilde{\sigma}_C^2 + \tilde{\sigma}_I^2 + \tilde{\sigma}_M^2$. The within-cluster within-period correlation is

$$cor(y_{1ij}, y_{1i'j}) = cor(y_{2ij}, y_{2i'j}) = \frac{\tilde{\sigma}_C^2 + \tilde{\sigma}_{CP}^2}{\tilde{\sigma}_C^2 + \tilde{\sigma}_{CP}^2 + \tilde{\sigma}_I^2 + \tilde{\sigma}_M^2} = \rho \tag{6}$$

The within-cluster between-period correlation is

$$cor(y_{1ij}, y_{2i'j}) = \frac{\tilde{\sigma}_C^2}{\tilde{\sigma}_C^2 + \tilde{\sigma}_{CP}^2 + \tilde{\sigma}_I^2 + \tilde{\sigma}_M^2} = \eta \tag{7}$$

The within-cluster within-subject correlation is the correlation between the two measurements within a subject

$$cor(y_{1ij}, y_{2ij}) = \frac{\tilde{\sigma}_C^2 + \tilde{\sigma}_I^2}{\tilde{\sigma}_C^2 + \tilde{\sigma}_{CP}^2 + \tilde{\sigma}_I^2 + \tilde{\sigma}_M^2} = \xi \tag{8}$$

As for the cross-sectional design, the treatment effect is estimated by taking the difference in means over the clusters. This estimator has variance

$$var(\hat{\gamma}_2) = \frac{2\sigma_T^2(1 - \xi + (m - 1)(\rho - \eta))}{mk} \tag{9}$$

See the Online Supplement for the derivation. Equation (9) shows that the treatment effect is estimated with higher variance when ρ increases, and/or when η and/or ξ and/or m decreases.

The effects of attrition

Attrition is defined as the situation where clusters or subjects drop out during the washout period. This implies an outcome is measured on them in the first time period but not in the second. Simple closed-form expressions for the variance of the treatment effect estimator cannot be derived; hence, matrix algebra is used to explore the detrimental effects of attrition.

First, the multilevel model for the case without attrition is given and then it is shown how attrition is taken into account. The data are organized such that the first m observations are for time period 1 and the second m observations are for time period 2. The multilevel model for cluster j for the cohort design in matrix notation is

$$y_j = X_j \gamma + u_j + e_j + m_j \tag{10}$$

where y_j is the vector of length $2m$ with response variables; X_j is the $2m \times 3$ design matrix for the fixed part of the model; $\gamma = (\gamma_0, \gamma_1, \gamma_2)'$ is the vector of length 3 with regression coefficients; and u_j , e_j and m_j are vectors of length $2m$ with random terms at the cluster, subject and repeated measures levels. $V_j = cov(y_j) = cov(u_j + e_j + m_j)$ is the $2m \times 2m$ covariance matrix of the responses. The vector of regression coefficients is estimated as

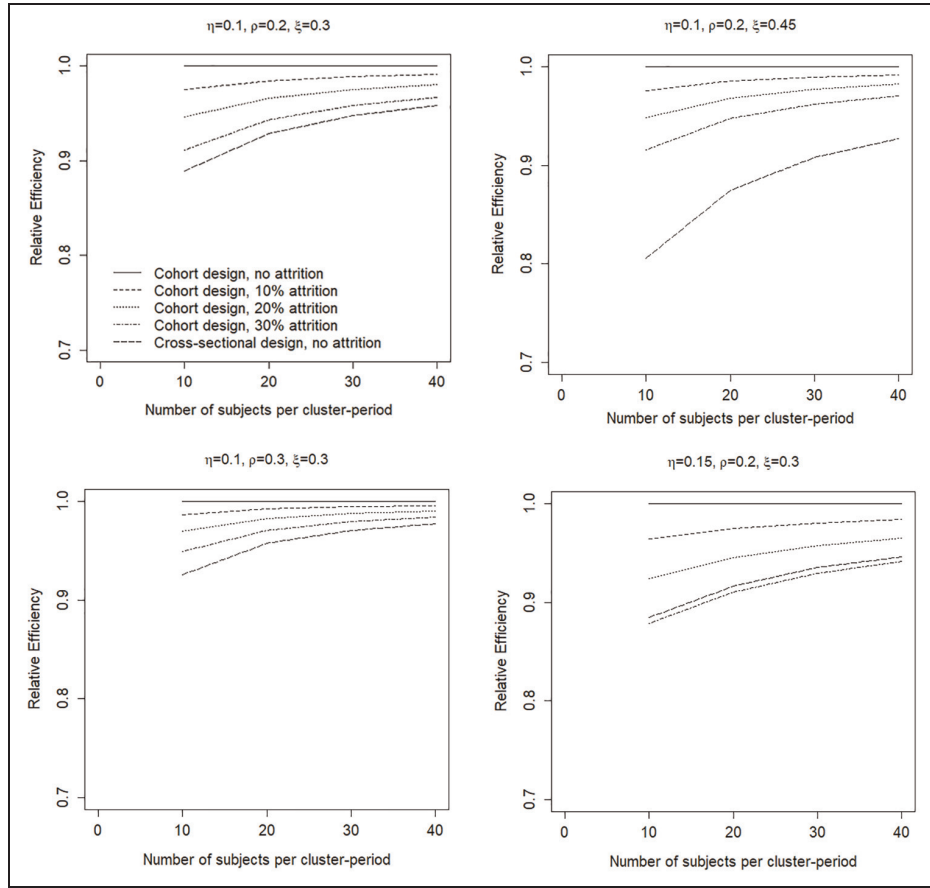


Figure 1. The effects of subject attrition in a cohort design as a function of the number of subjects per cluster-period and for four combinations of values of the correlation coefficients. η is the within-cluster between-period correlation, ρ is the within-cluster within-period correlation, and ξ is the within-cluster within-subject correlation. The legend in the upper left panel holds for all panels.

$$\hat{\gamma} = \left(\sum_{j=1}^k X_j' \hat{V}_j^{-1} X_j \right)^{-1} \sum_{j=1}^k X_j' \hat{V}_j^{-1} y_j \quad (11)$$

and its corresponding covariance matrix is estimated as

$$\hat{var}(\hat{\gamma}) = \left(\sum_{j=1}^k X_j' \hat{V}_j^{-1} X_j \right)^{-1} \quad (12)$$

The model for the cross-sectional design follows from equation (10) by removing the vector m_j .

Results

Attrition of subjects in a cohort design

Suppose that a proportion p_j of the m subjects in cluster j drops out. This implies removing the last $p \times m$ entries from vectors y_j, u_j, e_j and m_j ; the last $p \times m$ rows from matrix X_j ; and the last $p \times m$ rows and columns from matrix V_j . Note that attrition rates may vary across clusters and treatment sequences.

Figure 1 shows the effects of various realistic degrees of subject attrition in a cohort design (constant across

clusters), as a function of the number of subjects per cluster-period and for four combinations of the correlation coefficients η, ρ and ξ . The cohort design without attrition is used as a reference (i.e. its relative efficiency is 1) and the cross-sectional design without attrition is also included. As is obvious, the cohort design becomes less efficient when more subjects drop out, especially so when the number of subjects per cluster-period is small. However, the loss of efficiency is most often small and only slightly depends on the correlation coefficients. The largest loss in efficiency is observed in the lower right panel: the relative efficiency is 0.88 when 10 subjects per cluster are included but 30% of them drop out. In that case, the cohort design is slightly less efficient than the cross-sectional design without attrition.

Attrition of clusters in a cohort or cross-sectional design

The focus is on the scenario where some of the k clusters drop out. The attrition rates may vary across treatment sequences. For those clusters that drop out the last m entries from vectors y_j, u_j, e_j and m_j are removed,

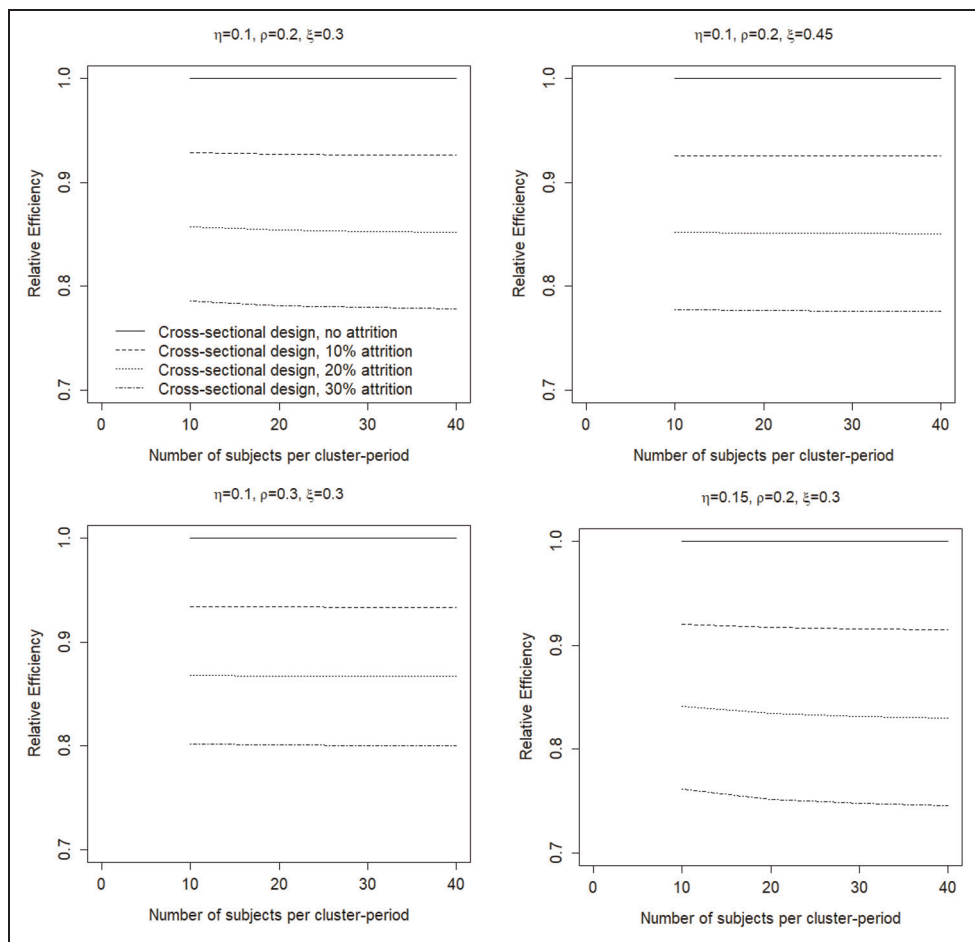


Figure 2. The effects of cluster attrition in a cohort design as a function of the number of subjects per cluster-period and for four combinations of values of the correlation coefficients. η is the within-cluster between-period correlation, ρ is the within-cluster within-period correlation, and ξ is the within-cluster within-subject correlation. The legend in the upper left panel holds for all panels.

and so are the last m rows from the matrix X_j and the last m rows and m columns from the matrix V_j .

Figure 2 shows the effects of various realistic degrees of cluster attrition in a cohort design (constant across treatment sequences), as a function of the number of subjects per cluster-period and for four combinations of correlation coefficients η , ρ and ξ . The cohort design without attrition serves as a reference. It is obvious the cohort design becomes less efficient when the attrition rate increases. The relative efficiency hardly depends on the number of subjects per cluster-period, or the combination of the correlation coefficients η , ρ and ξ . A comparison with Figure 1 shows attrition of clusters results in a larger loss of efficiency than attrition of subjects. The smallest relative efficiency is found in the lower right panel: it is equal to 0.75 when 30% of the clusters drop out and each cluster-period includes $m = 40$ subjects.

The results of cluster attrition in a cross-sectional design are very similar and a figure is not included. In general, the effects of attrition of clusters are slightly

less severe for a cross-sectional design than for a cohort design.

How to account for attrition in sample size calculations

The aim of a cluster randomized crossover trial is to detect the treatment effect γ_2 with sufficient probability. The statistical power $1 - \beta$ in a two-sided test follows from

$$z_{1-\beta} + z_{1-\alpha/2} = \frac{\gamma_2}{\sqrt{\text{var}(\gamma_2)}} \quad (13)$$

where $1 - \alpha$ is the type I error rate and $z_{1-\alpha/2}$ and $z_{1-\beta}$ are the $100(1 - \alpha/2)\%$ and $100(1 - \beta)\%$ standard normal deviates.

Power is related to sample sizes k and m through $\text{var}(\gamma_2)$. An a priori sample size calculation that ignores the possibility of attrition will result in an underpowered study in case attrition turns out to be present. The

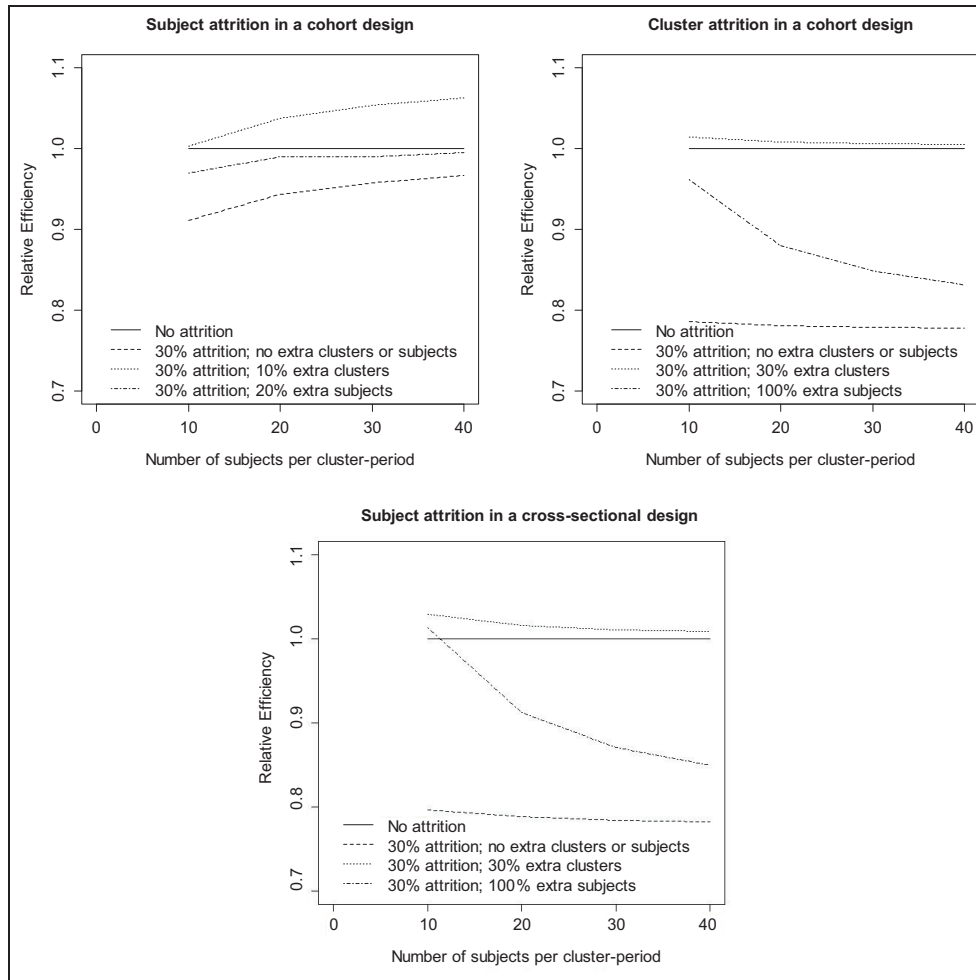


Figure 3. Repairing for anticipated attrition. For all graphs, the within-cluster between-period correlation $\eta = 0.1$, the within-cluster within-period correlation $\rho = 0.2$ and the within-cluster within-subject correlation, $\xi = 0.3$.

aim of this section is to explore strategies to adjust for potential attrition at the study design stage.

Anticipated attrition. Anticipated attrition is defined as attrition that a researcher anticipates before the start of the trial. The most obvious strategy to repair for such attrition is increasing the number of subjects and/or number of clusters from the start of the study onwards. This implies that in both time periods, higher sample sizes m and/or k are used than those that are based on an a priori power analysis for a study without attrition.

Figure 3 explores the effects of including extra subjects or clusters. The values of the correlation coefficients are $\eta = 0.1$, $\rho = 0.2$ and $\xi = 0.3$ and a constant attrition rate of 30% of either subjects or clusters is taken into account. The design without attrition is used as a reference. The top left panel shows results for attrition of subjects in a cohort design. Including 20% extra subjects results in an efficiency of almost 1. Including just 10% extra clusters results in a relative efficiency above 1 for all values m .

Repairing for attrition is a bigger challenge when clusters drop out, as is shown in the other panels of Figure 3. Increasing the number of clusters with 30% results in a relative efficiency slightly above 1. The effects of increasing the number of subjects are most evident for a small number of subjects per cluster-period ($m = 10$), but even then the number of subjects should be doubled to reach a relative efficiency near 1. For a large number of subjects per cluster-period ($m = 40$), the effect of increasing m is only small. A similar effect was found in cluster randomized trials without crossover: the effect of increasing cluster size is relatively large at small cluster sizes but levels off at higher cluster sizes, especially so when the intraclass correlation is large.³¹ Furthermore, it should be noted that the loss of efficiency as a result of cluster attrition is somewhat larger for a cohort design than for a cross-sectional design. A Shiny App is available at <https://utrecht-university.shinyapps.io/CRXO1/> to explore the effects of anticipated attrition and the effects of increasing sample size.

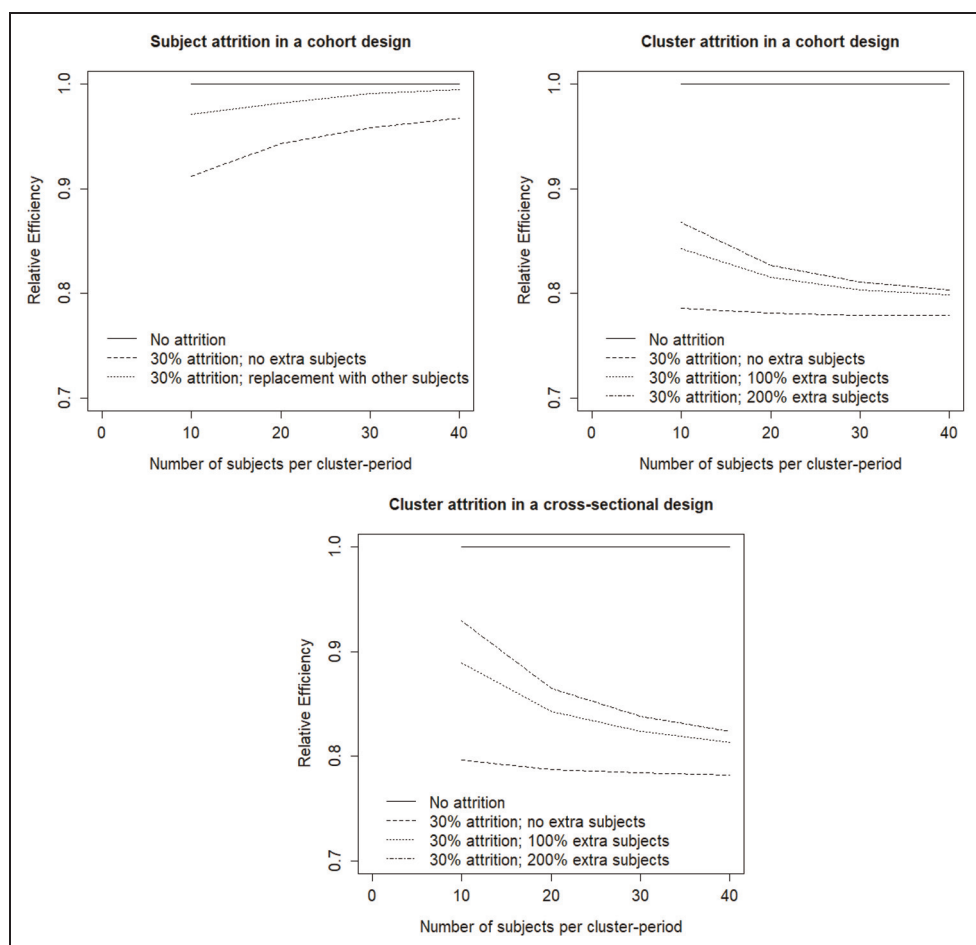


Figure 4. Repairing for unanticipated attrition. For all graphs the within-cluster between-period correlation $\eta = 0.1$, the within-cluster within-period correlation $\rho = 0.2$ and the within-cluster within-subject correlation $\xi = 0.3$.

Unanticipated attrition. Unanticipated attrition is defined as attrition that a researcher does not anticipate before the start of the trial. In other words, only during the washout period does it become clear that some subjects or clusters will not continue to the second time period. The only means to repair for the loss of efficiency is increasing the number of subjects in the second time period within those clusters that do not drop out.

Sampling more subjects in the second time period than originally planned may raise issues with respect to informed consent and also administrative and practical issues. For instance, the number of subjects that may be enrolled in the second time period may be limited because a limited amount of material and equipment to deliver the treatments was ordered before the start of the study, or because a limited number of professionals who are eligible to deliver the treatments were contracted. In the worst case, the number of subjects that can be enrolled in the second time period cannot be larger than the number that was originally planned.

Figure 4 shows the effects of enrolling extra subjects in the second time period. As in Figure 3, the design

without attrition is used as a reference. In all graphs, the values of the correlation coefficients are $\eta = 0.1$, $\rho = 0.2$ and $\xi = 0.3$ and 30% attrition is taken into account.

The panel at the top left shows the effect of replacing those subjects who drop out in a cohort design by others. The design thus becomes a mixture of a cohort and cross-sectional design: those subjects who do not drop out are measured in both time periods (i.e. they are the cohort part of the study) and all others in just one time period (i.e. they are the cross-sectional part). In other words, the study has an open cohort sampling structure. As is obvious, the efficiency of this strategy is between that of a study without attrition and that of a study with 30% attrition that does not replace subjects who drop out. This strategy is an effective means to increase efficiency, especially so for larger number of subjects per cluster-period.

Adjusting sample size for unanticipated attrition of clusters is more difficult than repairing for anticipated attrition. The top right and bottom panels show that even increasing the number of subjects in time period 2

Table 1. Summary of different designs for the lavender scent example.

Design	Sample size	Power
Individual randomized trial	$k = 1$ cluster, $n = 340$ subjects per cluster	0.31
	$k = 1$ cluster, $n = 1228$ subjects per cluster	0.80
Cluster randomized trial parallel groups	$k = 20$ clusters, $m = 157$ subjects per cluster	0.80
Cluster randomized trial Crossover, cross-sectional	$k = 20$ clusters, $m = 36$ subjects per cluster-period	0.80
Cluster randomized trial Crossover, cohort	$k = 20$ clusters, $m = 26$ subjects per cluster-period	0.81
Cluster randomized trial Crossover, cohort 25% attrition of subjects	No repair: $k = 20$ clusters, $m = 26$ subjects in cluster-period 1	0.75
	Increase m : $k = 20$ clusters, $m = 30$ subjects in cluster-period 1	0.80
	Replacement: $k = 20$ clusters, $m = 26$ subjects per cluster-period	0.79

by 200% does not result in a relative efficiency near 1. The effect of increasing the number of subjects in time period 2 becomes smaller when the number of subjects per cluster period increase. In addition, this strategy is somewhat less effective for the cohort design than for the cross-sectional design. For a Shiny App, the reader is referred to <https://utrecht-university.shinyapps.io/CRXO2/>.

Example: effects of lavender scent on dental patients' anxiety

This example is based on a study on the effect of lavender scent versus no odor on dental patients' anxiety.³² A total of 340 patients agreed to participate; these were from one practice and randomization was done at the patient level. One of the outcomes was the Modified Dental Anxiety Scale. The difference between the two conditions was insignificant ($F(1338) = 2.17, p > 0.05$). The estimates were as follows: mean = 9.84, standard deviation (SD) = 4.74 (lavender) and mean = 10.65, SD = 5.40 (control), and hence the effect size was small (Cohen's $d = 0.16$) and a total sample size of 1228 would have been needed to detect such an effect with 80% power in a two-sided test with $\alpha = 0.05$. The power for a study with 340 subjects is only 0.31.

A sample of 1228 can hardly be recruited within a reasonable amount of time within just one dental practice. It is therefore needed to involve multiple dental practices in a cluster randomized trial. Suppose $k = 20$ dental practices are willing to participate. The required cluster size m depends on the intraclass correlation coefficient. Let us use an a priori estimate of $\rho = 0.01$. Such a small value is common in large clusters such as dental practices where the interaction among patients is limited. Then $m = 157$ patients per cluster are needed. The total sample size is $mk = 3140$, which is almost 10 times as large as in the original study.

A smaller sample size is needed in a crossover design. Suppose an a priori estimate of the within-cluster between-period correlation is $\eta = 0.005$. Then $m = 36$ subjects per cluster-period are needed in a cross-sectional design, or $2m = 72$ per cluster, less than half the sample size as for a cluster randomized trial without crossover. The total number of subjects is $2mk = 1440$, and since each subject is measured only once, this is also the total number of measurements.

A cohort design further decreases the required number of subjects per cluster. Here, an a priori estimate $\xi = 0.3$ of the within-cluster within-subject correlation is used. The number of subjects per cluster is $m = 26$, resulting in a total number of subjects of $mk = 520$. As each subject is measured twice, the total number of measurements is $2mk = 1040$. The cohort design thus has a smaller number of subjects and measurements than the cross-sectional design.

The risk of the cohort design is attrition of subjects. Let us assume an attrition rate of 25% per cluster, irrespective of treatment sequence. The power for a design with 26 subjects per cluster slightly drops to 0.75. The loss of power may be repaired by including 30 subjects per cluster; all of them are measured in period 1 but only 75% of them are measured in period 2. Another strategy is replacing those subjects who drop out by other subjects. In that case, the power is 0.79. All designs are summarized in Table 1.

Conclusions and discussion

Attrition results in a loss of efficiency. In the case of subject attrition in a cohort design, the effect of attrition becomes smaller when the number of subjects per cluster-period increases. However, this sample size hardly influences the effect of attrition in the case clusters drop out. Cluster attrition results in a larger loss of efficiency than subject attrition. It should, therefore, be

avoided at all costs because it is difficult to repair, especially so when the number of subjects per cluster-period is large. It may be more cost-efficient to spend money on incentives to motivate those clusters who tend to drop out to remain in the trial than to spend money on recruiting extra subjects.

The effect of attrition of clusters is somewhat larger in a cohort design than in a cross-sectional design, but the cohort design may still be more efficient in the case attrition occurs. For each study at hand, these two designs should be compared for realistic a priori estimates of the degree of attrition and intraclass correlation coefficients. The final choice should not only be based on statistical but also on practical considerations. A cohort design is only recommended when the intervention can be undone at the level of the subject. Furthermore, the study should be planned such that subjects can indeed participate in both time periods.

The results of this study may help researchers plan a cluster randomized crossover study in the case attrition is anticipated. The efficiency of a design depends on the intraclass correlations and degree of attrition, and a plausible a priori estimate for each of them should be available. An overview of papers that lists estimates of intraclass correlation coefficients for cluster randomized trials without crossover is given in Table 11.1 of Moerbeek and Teerenstra.³³ A similar table should also become available for cluster randomized trials with crossover. Researchers conducting such a trial are therefore encouraged to clearly report the estimates of the intraclass correlation coefficients. In addition, they are encouraged to report attrition rates and reasons for attrition at the subject and cluster level so that this information can be used in planning future trials, see also item 16 in proposed reporting items for cluster randomized crossover trials.³⁴

An obvious extension of this work is to more complicated cluster randomized crossover trials that include more than two treatments and/or more than two time periods, where the attrition rate may change over time. It may also be interesting to study the optimal number of time periods in a trial of fixed duration, as a function of attrition and for different correlation structures. It has been shown that for the continuous-decay within-cluster correlation structure, greater numbers of crossovers result in greater study power,³⁵ and it should be verified if this is still the case of attrition is present.

Furthermore, it is also necessary to study the effect of attrition for the cluster randomized stepped wedge design where the number of time periods, and hence the number of measurements, is typically larger than two.^{36,37} This implies there are multiple points in time where attrition may occur; hence the results of this article do not easily translate to the stepped wedge design.

Other directions for future research are exploring the effects of attrition in the case where clusters and

subjects may drop out simultaneously, and for non-continuous outcome variables. Another avenue of future research would be the impact of non-random attrition, that is, when missing data are missing not at random.


Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship and/or publication of this article.

ORCID iD

Mirjam Moerbeek  <https://orcid.org/0000-0001-5537-1237>

Supplemental material

Supplemental material for this article is available online.

References

1. Campbell MJ and Walters SJ. *How to design, analyse and report cluster randomised trials in medicine and health related research*. Chichester: John Wiley & Sons, 2014.
2. Donner A and Klar N. *Cluster randomization trials*. London: Edward Arnold, 2000.
3. Eldridge S and Kerry S. *A practical guide to cluster randomised trials in health services research*. Chichester: John Wiley & Sons, 2012.
4. Hayes RJ and Moulton LH. *Cluster randomised trials*. Boca Raton, FL: CRC Press, 2009.
5. Murray D. *Design and analysis of group-randomized trials: a review of recent developments*. New York: Oxford University Press, 1998.
6. Bloom HS, Richburg-Hayes L and Black AR. Using covariates to improve precision for studies that randomize schools to evaluate educational interventions. *Educ Eval Policy Anal* 2007; 29(1): 30–59.
7. Konstantopoulos S. The impact of covariates on statistical power in cluster randomized designs: which level matters more. *Multivariate Behav Res* 2012; 47(3): 392–420.
8. Raudenbush SW. Statistical analysis and optimal design for cluster randomized studies. *Psychol Meth* 1997; 2(2): 173–185.
9. Moerbeek M. Power and money in cluster randomized trials: when is it worth measuring a covariate. *Stat Med* 2006; 25(15): 2607–2617.
10. De Hoop E, Teerenstra S, van Gaal BG, et al. The “best balance” allocation led to optimal balance in cluster-controlled trials. *J Clin Epidemiol* 2012; 65(2): 132–137.
11. Murray DM and Blstein JL. Methods to reduce the impact of intraclass correlation in group-randomized trials. *Eval Rev* 2003; 27(1): 79–103.
12. Murray DM, Hannan PJ, Wolfinger RD, et al. Analysis of data from group-randomized trials with repeat observations on the same groups. *Stat Med* 1998; 17(14): 1581–1600.

13. Murray DM, Lee Van Horn M, Hawkins JD, et al. Analysis strategies for a community trial to reduce adolescent ATOD use: a comparison of random coefficient and ANOVA/ANCOVA models. *Contemp Clin Trials* 2006; 27(2): 188–206.
14. Arnup SJ, McKenzie JE, Hemming K, et al. Understanding the cluster randomised crossover design: a graphical illustration of the components of variation and a sample size tutorial. *Trials* 2017; 18: 381.
15. Giraudeau B, Ravaud P and Donner A. Sample size calculation for cluster randomized cross-over trials. *Stat Med* 2008; 27(27): 5578–5585.
16. Harrison DA and Brady AR. Sample size and power calculations using the noncentral t-distribution. *Stata J* 2004; 4(2): 142–153.
17. Rietbergen C and Moerbeek M. The design of cluster randomized crossover trials. *J Educ Behav Stat* 2011; 36(4): 472–490.
18. Hart LM, Morgan AJ, Rossetto A, et al. Helping adolescents to better support their peers with a mental health problem: a cluster-randomised crossover trial of teen mental health first aid. *Aust N Z J Psychiatry* 2018; 52(7): 638–651.
19. Damsgaard CT, Dalskov SM, Petersen RA, et al. Design of the OPUS School Meal Study: a randomised controlled trial assessing the impact of serving school meals based on the New Nordic Diet. *Scand J Public Health* 2012; 40(8): 693–703.
20. Damsgaard CT, Dalskov SM, Laursen RP, et al. Provision of healthy school meals does not affect the metabolic syndrome score in 8–11-year-old children, but reduces cardiometabolic risk markers despite increasing waist circumference. *Br J Nutr* 2014; 112(11): 1826–1836.
21. Low JL, Lewis SM and Prescott P. Assessing robustness of crossover designs to subjects dropping out. *Stat Comput* 1999; 9(3): 219–227.
22. Skargren E and Oberg B. Effects of an exercise program on musculoskeletal symptoms and physical capacity among nursing staff. *Scand J Med Sci Sports* 1996; 6(2): 122–130.
23. Ho WK, Matthews JNS, Henderson R, et al. Dropouts in the AB/BA crossover design. *Stat Med* 2012; 31(16): 1675–1687.
24. Brown BW Jr. The crossover experiment for clinical trials. *Biometrics* 1980; 36(1): 69–79.
25. Jones B and Kenward MG. *Design and analysis of crossover trials*. Boca Raton, FL: Chapman & Hall; CRC Press, 2002.
26. Senn S. *Cross-over trials in clinical research*. Chichester: John Wiley & Sons, 2002.
27. Goldstein H. *Multilevel statistical models*. 4th ed. Chichester: John Wiley & Sons, 2011.
28. Hox JJ, Moerbeek M and Van de Schoot R. *Multilevel analysis: techniques and applications*. 3rd ed. Boca Raton, FL: Routledge, 2018.
29. Snijders TAB and Bosker RJ. *Multilevel analysis: an introduction to basic and advanced multilevel modeling*. London: SAGE, 2012.
30. Raudenbush SW and Bryk AS. *Hierarchical linear models: applications and data analysis methods*. Thousand Oaks, CA: SAGE, 2002.
31. Hemming K, Girling AJ, Sitch AJ, et al. Sample size calculations for cluster randomisation controlled trials with a fixed number of clusters. *BMC Med Res Methodol* 2011; 11: 102.
32. Kritsidima M, Newton T and Asimakopoulou K. The effects of lavender scent on dental patient anxiety levels: a cluster randomised-controlled trial. *Community Dent Oral Epidemiol* 2010; 38(1): 83–87.
33. Moerbeek M and Teerenstra T. *Power analysis of trials with multilevel data*. Boca Raton, FL: CRC Press, 2016.
34. Arnup SJ, Forbes AB, Kahan BC, et al. The quality of reporting in cluster randomised crossover trials: proposal for reporting items and an assessment of reporting quality. *Trials* 2016; 17: 575.
35. Grantham KL, Kasza J, Heritier S, et al. How many times should a cluster randomized crossover trial cross over. *Stat Med* 2019; 38(25): 5021–5033.
36. Hussey MA and Hughes JP. Design and analysis of stepped wedge cluster randomized trials. *Contemp Clin Trials* 2007; 28(2): 182–191.
37. Woertman W, de Hoop E, Moerbeek M, et al. Stepped wedge designs could reduce the required sample size in cluster randomized trials. *J Clin Epidemiol* 2013; 66(7): 752–758.