

Optimality of equal vs. unequal cluster sizes in multilevel intervention studies: A Monte Carlo study for small sample sizes

Citation for published version (APA):

Candel, M. J. J. M., van Breukelen, G. J. P., Kotova, L., & Berger, M. P. F. (2008). Optimality of equal vs. unequal cluster sizes in multilevel intervention studies: A Monte Carlo study for small sample sizes. *Communications in Statistics-Simulation and Computation*, 37(1), 222-239. <https://doi.org/10.1080/03610910701724052>

Document status and date:

Published: 01/01/2008

DOI:

[10.1080/03610910701724052](https://doi.org/10.1080/03610910701724052)

Document Version:

Publisher's PDF, also known as Version of record

Document license:

Taverne

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

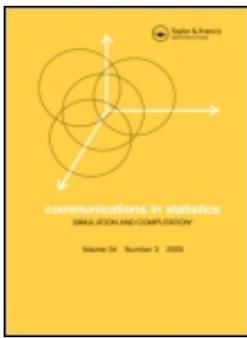
www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.



Optimality of Equal vs. Unequal Cluster Sizes in Multilevel Intervention Studies: A Monte Carlo Study for Small Sample Sizes

Math J. J. M. Candel , Gerard J. P. Van Breukelen , Larissa Kotova & Martijn P. F. Berger

To cite this article: Math J. J. M. Candel , Gerard J. P. Van Breukelen , Larissa Kotova & Martijn P. F. Berger (2007) Optimality of Equal vs. Unequal Cluster Sizes in Multilevel Intervention Studies: A Monte Carlo Study for Small Sample Sizes, Communications in Statistics - Simulation and Computation, 37:1, 222-239, DOI: [10.1080/03610910701724052](https://doi.org/10.1080/03610910701724052)

To link to this article: <https://doi.org/10.1080/03610910701724052>



Published online: 03 Jan 2008.



Submit your article to this journal [↗](#)



Article views: 179



View related articles [↗](#)



Citing articles: 2 View citing articles [↗](#)

Statistics in Health

Optimality of Equal vs. Unequal Cluster Sizes in Multilevel Intervention Studies: A Monte Carlo Study for Small Sample Sizes

MATH J. J. M. CANDEL¹, GERARD J. P. VAN
BREUKELEN¹, LARISSA KOTOVA²,
AND MARTIJN P. F. BERGER¹

¹Department of Methodology and Statistics, Maastricht University,
Maastricht, The Netherlands

²Rabobank Nederland, Utrecht, The Netherlands

Optimality of equal versus unequal cluster sizes in the context of multilevel intervention studies is examined. A Monte Carlo study is done to examine to what degree asymptotic results on the optimality hold for realistic sample sizes and for different estimation methods. The relative D-criterion, comparing equal versus unequal cluster sizes, almost always exceeded 85%, implying that loss of information due to unequal cluster sizes can be compensated for by increasing the number of clusters by 18%. The simulation results are in line with asymptotic results, showing that, for realistic sample sizes and various estimation methods, the asymptotic results can be used in planning multilevel intervention studies.

Keywords D -optimality; D_s -optimality; Mean squared error; Multilevel intervention studies; Relative efficiency; (Restricted) maximum likelihood; Unequal cluster sizes.

Mathematics Subject Classification Primary 62K05; Secondary 65C05.

1. Introduction

Many study designs evaluating the effect of an intervention are characterized by a nesting of subjects within clusters. Examples are the effect of a prevention program for citizens nested within cities or the effect of medication applied to patients nested within hospitals. Depending on whether such clusters are randomized to the intervention or whether persons within such clusters are randomized to the

Received August 24, 2006; Accepted May 18, 2007

Address correspondence to Math J. J. M. Candel, Department of Methodology and Statistics, Maastricht University, P.O. Box 616, Maastricht 6200, MD, The Netherlands; E-mail: Math.Candel@stat.unimaas.nl

intervention, we will call these designs cluster randomized trials and multicenter trials, respectively. The data obtained with such designs are often analyzed with multilevel analysis or mixed effects regression (Verbeke and Molenberghs, 2000).

In planning cluster randomized trials and multicenter studies, several design issues are important: (1) randomization at person or randomization at the cluster level, (2) the choice of the number of clusters, and (3) the choice of the number of persons within each cluster. Formulas for optimal numbers of persons and optimal numbers of clusters have been derived in several studies (e.g., Headrick and Zumbo, 2005; Liu, 2003; Moerbeek et al., 2001; Raudenbush, 1997; Raudenbush and Liu, 2000). These formulas assume the number of persons in each cluster are all equal. In practice, however, equal cluster sizes are rarely feasible. This may be due to variation in actual cluster size, but also due to non response or dropout.

The issue that we want to examine is whether, and to what extent, unequal cluster sizes are less efficient. There are asymptotic results for cluster randomized trials, showing that equal cluster sizes are asymptotically optimal for estimating fixed effects (Ankenman et al., 2003; Giovagnoli and Sebastiani, 1989). This implies that unequal cluster sizes lead to loss of power. Ankenman et al. (2003) also show that equal cluster sizes are asymptotically optimal for the variance components if the intraclass correlation is larger than 0.50. For multicenter studies, equal cluster sizes are also optimal asymptotically, when estimating fixed effects, but this result is obtained assuming that there is no correlation between random intercept and random slope (Van Breukelen et al., 2007).

In the present study, we will first complement these studies by providing asymptotic results for both cluster randomized and multicenter trials. Second, we will examine to what extent the results on the efficiency of unequal vs. equal cluster sizes hold in case the number of clusters and the number of persons within a cluster are of realistic sizes. This will be done using an extensive Monte Carlo simulation study.

The article is structured as follows. Section 2 presents the multilevel model or mixed effects regression model for both cluster and person randomization. In Sec. 3, several criteria for evaluating the performance of equal vs. unequal cluster sizes will be presented. Section 4 provides the explicit expressions for the asymptotic relative efficiencies for both cluster randomized trials and multicenter trials. Section 5 delineates the design of the simulation study that examines the relative efficiency of equal versus unequal cluster sizes in case of realistic sample sizes and will report upon the results of this simulation. The article closes with some implications for the planning phase of cluster randomized trials and multicenter trials. Issues for further research are also discussed.

2. The Multilevel Model for Intervention Studies

2.1. Model for Cluster Randomized Trials

Let there be K clusters, half of which are randomly allocated to one of two conditions: treatment or control. In cluster j ($j = 1, \dots, K$), there are n_j persons, with all persons receiving the treatment to which this cluster was allocated. The total number of persons is $N = \sum_{j=1}^K n_j$. The dependent variable is a quantitative

outcome, denoted as Y_{ij} , for person i in cluster j . The model for analyzing the data is as follows:

$$Y_{ij} = \beta_{0j} + \beta_1 X_j + \varepsilon_{ij}, \quad \text{with } \beta_{0j} = \beta_0 + u_{0j}, \quad (1)$$

where X_j denotes the treatment to which cluster j is allocated, and is coded as +1 for treated and -1 for control clusters. With this coding scheme, β_0 represents the grand mean of the outcome, and β_1 is half the treatment effect. The terms ε_{ij} and u_{0j} represent a random person and random cluster effect, which are assumed to be independently normally distributed with variances σ^2 and σ_0^2 , respectively. The mean of these random effects is zero. The model for cluster randomization actually has four parameters that are estimated: two fixed regression weights, β_0 and β_1 , and two variance components, σ_0^2 and σ^2 . A relevant concept is the intraclass correlation, which is the correlation between the outcome measures for two randomly drawn persons from the same cluster. This is denoted as ρ and is equal to $\rho = \sigma_0^2 / (\sigma_0^2 + \sigma^2)$.

2.2. Model for Multicenter Trials

In the case of multicenter trials, there is randomized allocation of persons within clusters. Some of the persons within a center get the treatment and others do not. In this case, the model is as follows:

$$Y_{ij} = \beta_{0j} + \beta_{1j} X_{ij} + \varepsilon_{ij}, \quad \text{with } \beta_{0j} = \beta_0 + u_{0j} \quad \text{and} \quad \beta_{1j} = \beta_1 + u_{1j}, \quad (2)$$

where X_{ij} again is coded as +1 for the treatment group and -1 for the control group. Note that X_{ij} also has a subscript for persons, indicating that this variable may vary from person to person within a cluster. New to this model is the random slope effect u_{1j} , which is assumed to be normally distributed with variance σ_1^2 and has a covariance σ_{01} with the random intercept u_{0j} . All random effects have a mean 0.

In the case of person randomization, the expression for the intraclass correlation is more complicated, and depends on whether we consider the correlation between the outcomes of two persons within the same cluster and the same treatment, or the correlation between the outcomes of two persons within the same cluster but with different treatments. We will give the expression for two persons getting the same treatment, assuming $\sigma_{01} = 0$: $\rho = (\sigma_0^2 + \sigma_1^2) / (\sigma_0^2 + \sigma_1^2 + \sigma^2)$. It will furthermore be useful to define the variance ratio η of the intercept variance and the slope variance: $\eta = \sigma_0^2 / \sigma_1^2$. Note that this model has six parameters: two fixed regression weights, β_0 and β_1 , and four (co)variance components, σ^2 , σ_0^2 , σ_1^2 , and σ_{01} .

The model parameters can be estimated through maximum likelihood (ML) or restricted maximum likelihood estimation (REML) (Verbeke and Molenberghs, 2000, pp. 42–43). One may obtain negative estimates for the variance components. If one wants to interpret these parameters as variances, negative estimates should be reset to zero (truncation), as variances cannot be negative. Truncating negative estimates to zero, however, leads to biasedness in the estimates. Therefore, others have argued for a marginal model interpretation of the multilevel model, in which case negative values for the individual variance components are allowed, as long as the variance-covariance matrix of the outcome variable is positive definite (Verbeke and Molenberghs, 2000, p. 53). Both truncation and non truncation of estimators of the variance components will be considered in this article.

3. Evaluation Criteria

In comparing equal vs. unequal cluster sizes, different criteria will be examined. First, we will consider the bias, variance, and mean squared error for the separate model parameters (Mood et al., 1974, pp. 291–294). We will compare equal and unequal cluster sizes on these three criteria, by taking ratios. This will be denoted as relative bias, relative variance, and relative mean squared error, respectively. For cluster randomization, we are primarily interested in the estimator of the intervention effect, β_1 , and the estimator of the random intercept variance, σ_0^2 . For person randomization, we are primarily interested in the estimator of the intervention effect, β_1 , and the estimator of the random slope variance, σ_1^2 , reflecting the inter-individual variation of this intervention effect.

Focusing on the variances of the estimators involved in the multilevel models is justified to the extent that the mean squared error consists of the variance of a parameter estimate (Mood et al., 1974, p. 293). Stated otherwise, focusing on the variances is justified in case the estimators have small biases. For all model parameters involved, a commonly used criterion is the determinant of the variance-covariance matrix of the parameter estimators. This criterion is known as the *D*-criterion.

In terms of this *D*-criterion, we can compare equal cluster sizes to unequal cluster sizes. Let θ_f denote the vector of the f fixed parameters and let θ_r denote the vector of r variance components. Then we define $\theta^T = [\theta_f^T, \theta_r^T]$ as the vector of all, $p = f + r$, parameters in the model. Note that for the model for cluster randomization $p = 4$, whereas for person randomization $p = 6$ or $p = 5$ (when $\sigma_{01} = 0$).

Let ξ denote the design of a study. The covariance matrix of the estimators $\hat{\theta}$, given a design ξ , is denoted as $\text{cov}(\hat{\theta} | \xi)$. Let $\text{Det}(\text{cov}(\hat{\theta} | \xi))$ denote the determinant of this variance-covariance matrix of the estimators $\hat{\theta}$. In what follows, let ξ^* denote a design with equal cluster sizes and let ξ denote a design with unequal cluster sizes, but with the same number of clusters, K , and the same total sample size, N , as ξ^* . The relative efficiency (RE) of design ξ compared to design ξ^* in terms of the *D*-criterion is:

$$\text{RE}(D) = \left(\frac{\text{Det}(\text{cov}(\hat{\theta} | \xi^*))}{\text{Det}(\text{cov}(\hat{\theta} | \xi))} \right)^{1/p} \quad (3)$$

In case we are interested in either the fixed parameters or the variance components of the model, separately, we can employ the D_s -criterion. This is similar to the *D*-criterion but then applied to a subset of the model parameters (Atkinson and Donev, 1996, p. 109). The RE in terms of the D_s -criterion can be defined for the fixed parameters, θ_f , and for the variance components, θ_r , as follows:

$$\begin{aligned} \text{RE}(D_s(\text{fixed})) &= \left(\frac{\text{Det}(\text{cov}(\hat{\theta}_f | \xi^*))}{\text{Det}(\text{cov}(\hat{\theta}_f | \xi))} \right)^{1/f} \quad \text{and} \\ \text{RE}(D_s(\text{random})) &= \left(\frac{\text{Det}(\text{cov}(\hat{\theta}_r | \xi^*))}{\text{Det}(\text{cov}(\hat{\theta}_r | \xi))} \right)^{1/r} \end{aligned} \quad (4)$$

The inverse of the RE gives the number of times that design ξ , with unequal cluster sizes, has to be replicated in order to give the same statistical information about the parameters as the design ξ^* , with equal cluster sizes.

The D -criterion and D_s -criterion are chosen for two reasons. First, a design that minimizes the D -criterion or D_s -criterion in fact minimizes the area of the simultaneous confidence interval of the parameters (Atkinson and Donev, 1996, pp. 106–107). Second, D -optimal designs are independent of the scale of the variables contained in the design matrices (Atkinson and Donev, 1996).

4. Asymptotic Relative Efficiencies

Since asymptotically there are no correlations between ML estimators of the fixed parameters and the variance components (see, e.g., McCulloch and Searle, 2001, p. 176), the relation between the relative efficiencies in terms of the D -criterion and the D_s -criterion for the fixed parameters and the variance components is as follows:

$$\text{RE}(D) = \sqrt{\text{RE}(D_s(\text{fixed})) \times \text{RE}(D_s(\text{random}))}. \quad (5)$$

4.1. Cluster Randomized Trials

For cluster randomized trials, we can derive asymptotic expressions for the relative efficiencies in terms of the D_s -criterion. Let \bar{n} denote the average cluster size of the K clusters and let w_j be defined as: $w_j = (\sigma_0^2 + \sigma^2/n_j)^{-1}$. For the case of equal cluster sizes, we have $n_j = \bar{n}$, and the weight w_j is denoted as w_e . In Van Breukelen et al. (2007) and Appendix A1, respectively, it is shown that the expressions for the relative efficiencies based on the D_s -criteria are as follows:

$$\text{RE}(D_s(\text{fixed})) = \left(\frac{\bar{n} + (1 - \rho)/\rho}{\bar{n}} \right) \times \frac{1}{K} \sum_{j=1}^K \left(\frac{n_j}{n_j + (1 - \rho)/\rho} \right), \quad \text{and} \quad (6)$$

$$\text{RE}(D_s(\text{random})) = \sqrt{\frac{N \sum_{j=1}^K w_j^2 - \left(\sum_{j=1}^K w_j \right)^2}{K(N - K)w_e^2}}. \quad (7)$$

Van Breukelen et al. (2007) and Appendix A1, respectively, also show that the asymptotic relative efficiencies for the individual parameter estimators are: $\text{RE}(\hat{\beta}_0) = \text{RE}(\hat{\beta}_1) = \text{RE}(D_s(\text{fixed}))$ and $\text{RE}(\hat{\sigma}_0^2) \approx (\text{RE}(D_s(\text{random})))^2$.

4.2. Multicenter Trials

For multicenter trials, the asymptotic results for the relative efficiency for the fixed parameter estimators are quite elaborate and therefore are only given in the Appendix (Appendix A2). These can be used in numerically evaluating the asymptotic relative efficiency of a design. No results were derived for the relative efficiency with respect to the variance components.

5. Simulation Study of the Relative Performance for Small Sample Sizes

We will examine, through an extensive Monte Carlo simulation study, to what extent the asymptotic results hold for numbers of clusters and cluster sizes that

are small. In addition to the relative variances of the parameter estimates, the relative biases and relative mean squared errors will be examined. In the case of multicenter trials, there are no asymptotic results for variance components, so that the Monte Carlo results provide some insights into the relative efficiency for these parameter estimators.

5.1. Design

In the Monte Carlo simulations, three factors were systematically varied when studying cluster randomized trials: (1) the size of the intraclass correlation, (2) the frequency distribution of the cluster sizes, and (3) the estimation method. In the case of person randomization, two additional factors were varied: (4) the value of the variance ratio $\eta = \sigma_0^2/\sigma_1^2$ and (5) the correlation between the random intercept and random slope (ρ_{01}). Table 1 summarizes the choices made for these factors, their motivation is given in what follows.

5.1.1. Intraclass Correlation. The variance components, σ_0^2 for cluster randomization, and $\sigma_0^2 + \sigma_1^2$ for person randomization, took values ranging from 1–30, with the error variance σ^2 simultaneously varying from 99–70, to keep the total variance constant (at 100). As a result, the intraclass correlation, ρ , ranged by steps of size 0.01 from 0.01–0.30, which encompasses realistic values as encountered in cross-sectional studies (Smeeth and Siu-Woon, 2002).

5.1.2. Frequency Distribution. To examine the effect of unequal cluster sizes, five different frequency distributions of cluster sizes were studied: (1) a uniform distribution; (2) a unimodal distribution; (3) a bimodal distribution; (4) a positively and (5) negatively skewed distribution. Three cluster sizes were assumed, with a large range of cluster sizes, namely 36 (28 in the case of a negatively skewed distribution, in order to prevent negative cluster sizes). Since $\bar{n} = 24$, for symmetrical

Table 1
Overview of the conditions of the Monte Carlo simulations

Factor	Levels
Intraclass correlation	$\rho = 0.01$ upto 0.30, with steps of 0.01
Distribution of cluster sizes ¹	Uniform distribution, unimodal distribution, bimodal distribution, positively skewed distribution, negatively skewed distribution
Estimation method	ML with truncation, ML without truncation REML with truncation, REML without truncation
<i>Only for person randomization</i>	
Intercept-slope correlation	0.0, 0.50, and -0.50
Variance ratio ²	1

Note 1. A more detailed description of the different distributions can be found in Table 2.

Note 2. Since a numerical investigation based on asymptotic results showed that a variance ratio of 1 yielded the minimum relative efficiency in terms of $D_s(\text{fixed})$, the Monte Carlo simulations for small samples were limited to a variance ratio of 1.

Table 2
Distributions of cluster sizes examined in the Monte Carlo simulations

Distribution ^{1,2}	Cluster sizes			Cluster frequencies		
	g_a	g_b	g_c	f_a	f_b	f_c
Uniform distribution	6	24	42	14	14	14
Unimodal distribution	6	24	42	8	26	8
Bimodal distribution	6	24	42	18	6	18
Positively skewed distribution	12	30	48	20	16	6
Negatively skewed distribution	4	18	32	6	12	24

Note 1: f_a = number of clusters of size g_a (small), f_b = number of clusters of size g_b (medium), f_c = number of clusters of size g_c (large)

Note 2: Cluster sizes and cluster frequencies are chosen such that the total number of clusters is equal to 42, the average cluster size is 24, the difference between the largest and smallest cluster size is equal to 36 (except for the negatively skewed distribution, where the range is 28 to prevent $g_a < 0$) and all cluster size frequencies and cluster sizes are even.

distributions this implies that the smallest $n = 6$ and the largest $n = 42$. Note that the cluster sizes were deliberately chosen to be small, to study the applicability of the asymptotic results.

A numerical study based on the asymptotic relative efficiencies showed that as the range of cluster sizes increases, the relative efficiency becomes smaller. Since we are primarily interested in the minimum value of relative performance, only large ranges were considered. An overview of the cluster sizes and their frequency of occurrence is given in Table 2.

5.1.3. Estimation Methods. Four different, commonly used, estimation methods were considered: Maximum Likelihood (ML) estimation and Restricted Maximum Likelihood (REML) estimation, combined with either truncation of negative estimates of the variance components to 0, or without such truncation.

5.1.4. Intercept-Slope Correlation. In the case of person randomization, it is possible to estimate also a random slope for the intervention effect. A numerical study based on the asymptotic results for the fixed effect parameters showed that the minimum relative efficiency is smallest for the case of no correlation. To also get an impression of the relative efficiency, in case there is a correlation between random intercept and slope, we also consider $\rho_{01} = 0.50$. This also covers the case of $\rho_{01} = -0.50$, since for a 50/50 allocation ratio, the relative efficiencies of the fixed parameter estimates are symmetrical in $\rho_{01} = 0$.

5.1.5. Variance Ratio. We assume the variance ratio $\eta = \sigma_0^2/\sigma_1^2 = 1$. Through a numerical study examining the asymptotical relative efficiency based on D_s (fixed), involving values of η from 0.02–1, with steps of 0.02, and from 1.0–100 with steps of 2, it was established for different types of distributions, for different intraclass correlations and for different intercept-slope correlations that $\eta = 1$ gives the lowest value of the relative efficiency. Since we are interested in the minimum value for the relative efficiency, we only consider $\eta = 1$ in this study.

For all simulations, we chose realistic numbers of clusters and persons within clusters: 42 clusters, with an average cluster size of 24. We assumed 50/50 allocation, that is, 50% of the clusters are assigned to one of two treatment conditions in case of cluster randomization, and 50% of the persons within a cluster are assigned to each of the two treatment conditions in case of person randomization. The fixed intercept, β_0 , was set at 50, whereas the fixed effect parameter, β_1 , was set at 5, without loss of generality, because the covariances of parameter estimators are asymptotically independent of the fixed effects (see Van Breukelen et al., 2007; Appendices A1 and A2). The total variance, $\sigma_0^2 + \sigma_1^2 + \sigma^2$, was, without loss of generality, set equal to 100.

5.2. Simulation Procedure

For each of the simulation conditions, 10,000 data sets were generated, each representing the data for 42 clusters consisting of, on average, 24 persons. The simulations as well as the estimation of the model parameters were performed using version 1.10.0007 of MLwiN (Rasbash et al., 2000). Both Iterative Generalized Least Squares and Restricted Iterative Generalized Least Squares were employed to obtain, respectively, the ML and REML estimates of the model parameters. Both estimation with and without truncation of the variance component estimates was carried out.

In model estimation, the convergence criterion was set to 0.001 and there was no limitation on the maximum number of iterations. In almost all cases, MLwiN reached convergence. Exceptions occurred for person randomization, in which case for intraclass correlations smaller than 0.03 and positively skewed cluster size distributions, a non positive definite variance-covariance matrix of the outcome variable resulted. No results were obtained for this (small) range of intraclass correlations.

5.3. Results for Cluster Randomization

5.3.1. Bias, Variance and Mean Squared Error. The squared biases are less than 5% of the mean squared errors, and in most cases, even less than 1%. Hence, the relative mean squared errors are very close to the relative variances. For cluster randomization, the most relevant parameters are the intervention effect and the random intercept variance. The relative mean squared errors for these parameters are shown in Figs. 1 and 2 for the unimodal and the bimodal distribution, which give the most extreme results. The differences in estimation methods are mostly smaller than 0.05. For intraclass correlations smaller than 0.03, the differences can become larger, but maximally 0.13. Since there are hardly any differences between ML and REML and between truncation and non truncation of the variance component estimates, we only show the results for REML and truncation.

Figure 1 shows that, in line with the asymptotic results on the variance of the effect estimator, unequal cluster sizes yield estimates with higher-mean squared errors than equal cluster sizes. In most cases, however, the relative mean squared error of the estimator of the intervention effect exceeds 0.85. Only for the bimodal distribution, the relative mean squared error can become lower than 0.85, but always exceeds 0.82. Figure 2 shows, in line with the asymptotic results, that for small intraclass correlations, the relative mean squared errors of the intercept variance

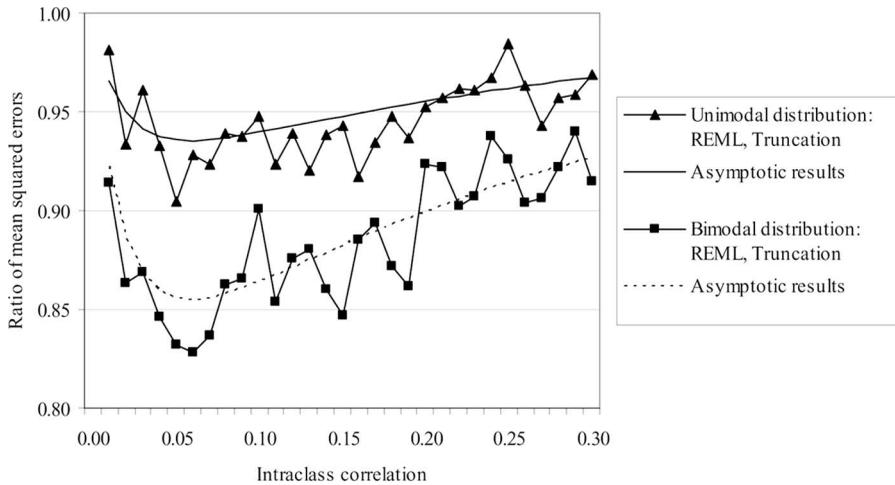


Figure 1. Relative mean squared errors of the fixed effect estimator for cluster randomized trials and REML estimation with truncation. Displayed is also the relative asymptotic variance of the ML estimator.

may be larger than 1, implying that unequal cluster sizes yield better estimates of the random intercept variance in these cases. Furthermore, the relative mean squared errors always exceed 0.80. The exception is again the bimodal distribution, in which case the relative mean squared error can become lower, but always exceeds 0.77.

For both the intervention effect and the variance of the random intercept, the relative mean squared errors can be modeled rather adequately by the relative variances as derived asymptotically. Examining the loss of mean squared error through these analytical results therefore seems appropriate.

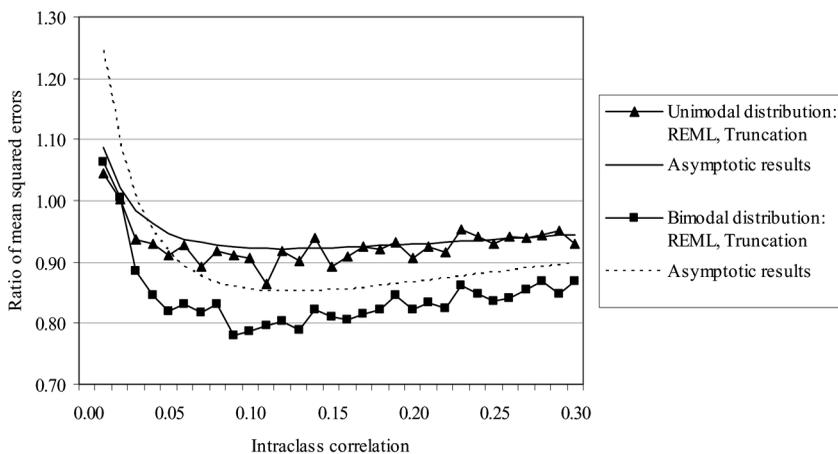


Figure 2. Relative mean squared errors of the intercept variance estimator for cluster randomized trials and REML estimation with truncation. Displayed is also the relative asymptotic variance of the ML estimator.

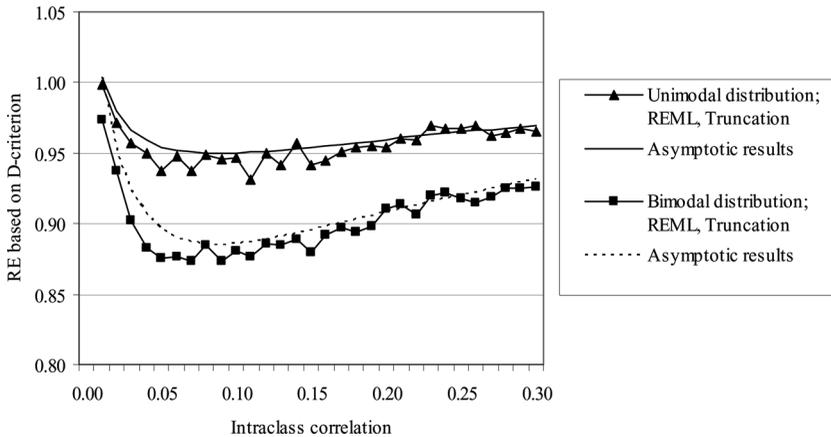


Figure 3. The relative D -criterion for cluster randomized trials and REML estimation with truncation. Displayed is also the relative D -criterion based on the asymptotic variances and covariances of the ML estimators of the model parameters.

5.3.2. D -Criterion and D_s -Criteria. Since the squared biases are small compared to the variances, we can also consider criteria based only on the variances of the estimators. The relative efficiencies in terms of the D -criterion are displayed in Fig. 3, both for the unimodal distribution and the bimodal distribution, which represent extreme situations. Since for both the D -criterion and the D_s -criterion there are hardly any differences between the results obtained with different estimation methods (always smaller than 0.05), only the results for REML in combination with truncation will be shown.

In Fig. 3, the relative efficiency in terms of the D -criterion is almost always lower than 1, in line with the asymptotic results. For the bimodal distribution, the relative efficiency can become as low as 0.85. For most distributions, however, similar to the results for the unimodal distribution, the relative efficiency exceeds 0.90.

Figures 4 and 5 show relative efficiencies in terms of the D_s -criterion for the fixed effects and the variance components, respectively. For the bimodal distribution, the relative efficiency is lowest but just below 0.85 for the fixed effects and above 0.85 for the variance components. Note that the relative efficiency for the variance components can become larger than 1 for values of the intraclass correlation close to 0. For very small intraclass correlations, unequal cluster sizes are more efficient when the estimation of variance components is at stake.

Finally, although the derived asymptotic relative efficiencies are somewhat higher than the simulated small sample relative efficiencies, the simulated relative efficiencies are rather adequately described by the asymptotic relative efficiencies.

5.4. Results for Person Randomization

5.4.1. Bias, Variance and Mean Squared Error. For the fixed effect estimators, the share of the squared bias in the mean squared error is less than 1%. Consequently, the mean squared errors are very close to the variances of the parameter estimators. The relative mean squared errors for the intervention effect are displayed in Fig. 6

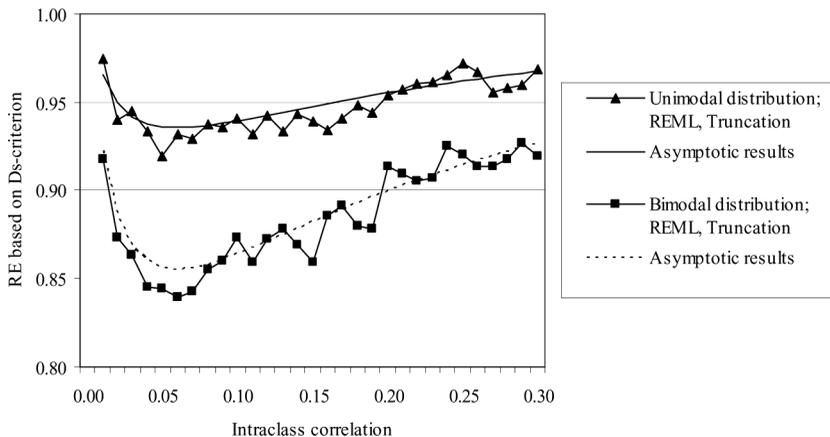


Figure 4. The relative D_s -criterion of the fixed effect parameters for cluster randomized trials and REML estimation with truncation. Displayed is also the relative D_s -criterion based on the asymptotic variances and covariance of the ML estimators of the fixed effect parameters.

for the extreme distributions (unimodal and bimodal), in case there is no correlation between random slope and random intercept. Since the results of the different estimation methods are very similar only the results for REML, with truncation of the variance component estimates, are shown.

In line with the asymptotic results on the variance of the intervention effect estimator, Fig. 6 shows that unequal cluster sizes yield estimates with higher mean squared errors than equal cluster sizes. For most cluster size distributions, the relative mean squared error of the effect estimator exceeds 0.85. Only for the bimodal distribution, the relative mean squared error can become somewhat lower than 0.85. A similar pattern of results was obtained in the case of the intercept-

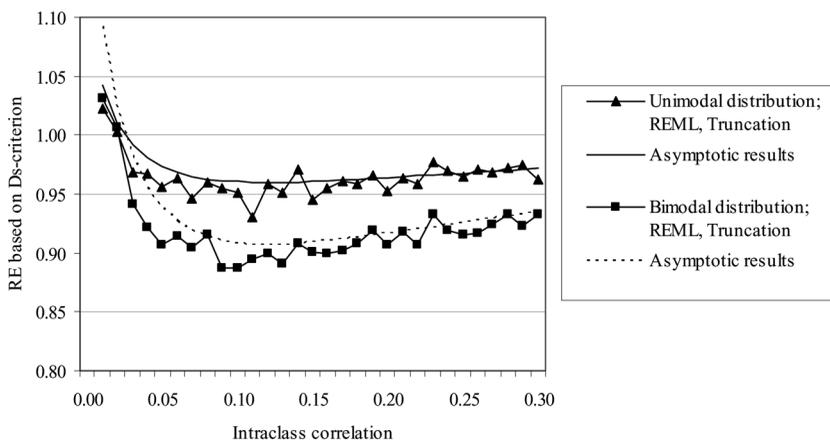


Figure 5. The relative D_s -criterion of the variance components for cluster randomized trials and REML estimation with truncation. Displayed is also the relative D_s -criterion based on the asymptotic variances and covariance of the ML estimators of the variance components.

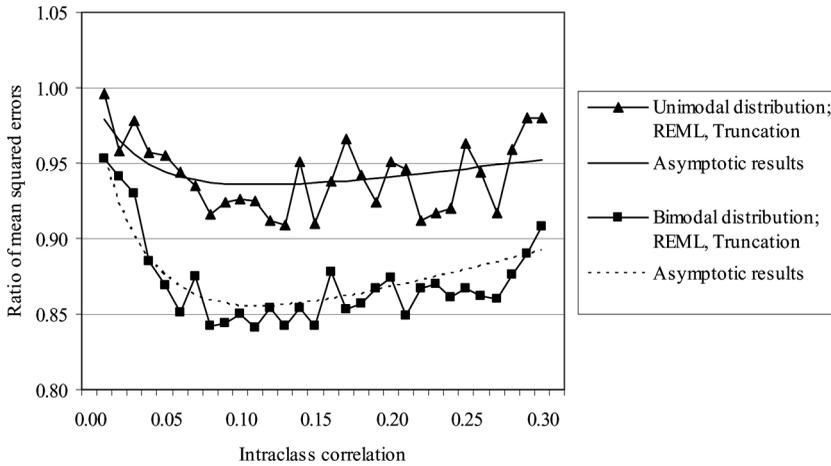


Figure 6. Relative mean squared errors of the fixed effect estimator for person randomized trials and REML estimation with truncation. Displayed is also the relative asymptotic variance of the ML estimator.

slope correlation $\rho_{01} = 0.50$, the minima of the relative mean squared errors being somewhat higher.

In Fig. 7, the relative mean squared error of the slope variance estimator can be larger than 1, implying that unequal cluster sizes can yield better estimates of the slope variance. For most distributions the relative mean squared errors exceed 0.85, the exception being the bimodal distribution, where the relative mean squared error always exceeds 0.77. Similar results are obtained in case of an intercept-slope correlation of $\rho_{01} = 0.50$.

For the random slope variance the squared biases can make up more than 10% of the mean squared errors. This occurred when the variance estimators were

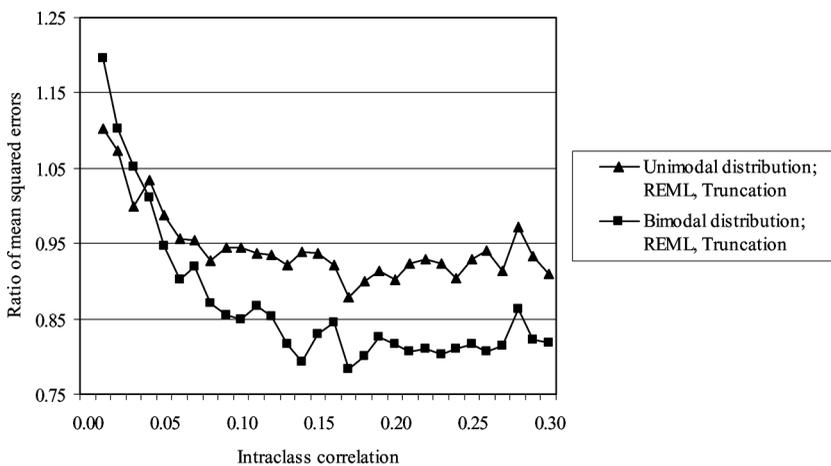


Figure 7. Relative mean squared errors of the slope variance estimator for person randomized trials and REML estimation with truncation.

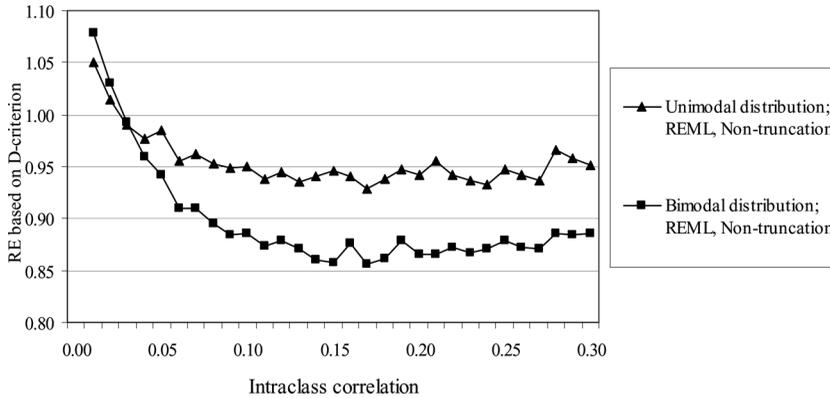


Figure 8. The relative D -criterion for person randomized trials and REML estimation without truncation.

truncated and the intra-class correlation was smaller than 0.03. For non truncation, the squared biases were less than 5% of the mean squared errors. In what follows, we will therefore present results for the D -criterion and the D_s -criterion for the variance component estimators only for non truncation. Since there were small differences between the different estimation methods for intra-class correlations higher than 0.02, these results are also representative of the other estimation methods.

5.4.2. D -Criterion and D_s -Criteria. The relative efficiencies in terms of the D -criterion are displayed in Fig. 8, for both the unimodal distribution and the bimodal distribution. The relative efficiency in terms of the D -criterion can become larger than 1, for small intra-class correlations. However, for larger intra-class correlations, unequal cluster sizes are clearly less efficient than equal cluster sizes, although the relative efficiency is never lower than 0.85. For most distributions, similar to the results for the unimodal distribution, the relative efficiency even exceeds 0.90. Similar results were obtained for $\rho_{01} = 0.50$.

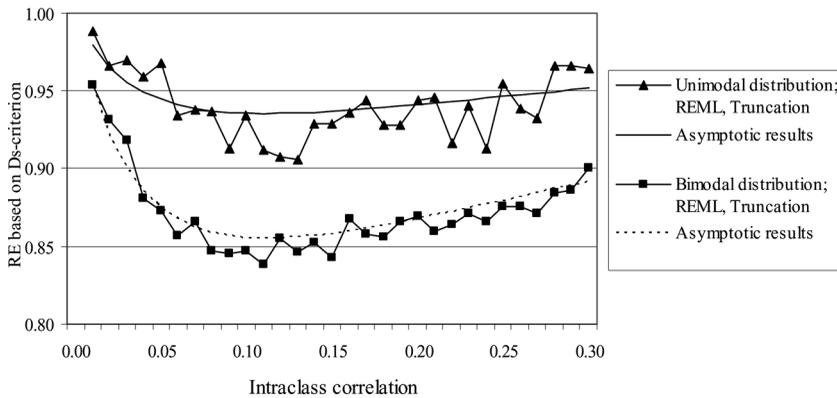


Figure 9. The relative D_s -criterion of the fixed effect parameters for person randomized trials and REML estimation with truncation. Displayed is also the relative D_s -criterion based on the asymptotic variances and covariance of the ML estimators of the fixed effect parameters.

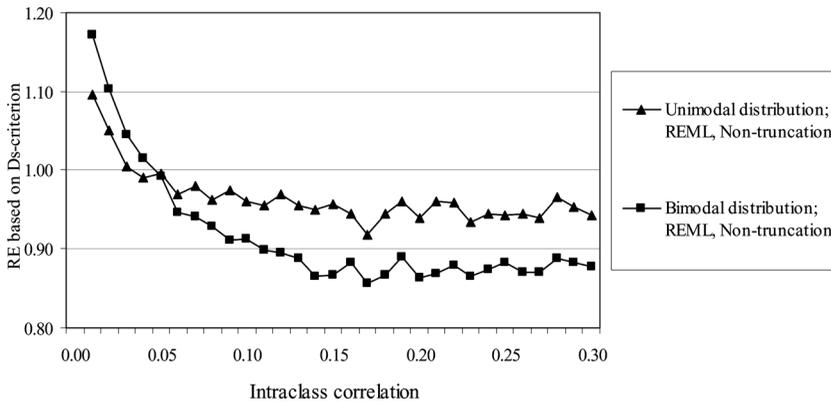


Figure 10. The relative D_s -criterion of the variance components for person randomized trials and REML estimation without truncation.

Figures 9 and 10 show relative efficiencies in terms of the D_s -criterion for the fixed effects and the variance components, respectively. For the bimodal distribution, the relative efficiency for the fixed effects is lowest but still exceeds 0.82. For the remaining distributions the relative efficiencies are always near to or above 0.90. The relative efficiency for the variance components always exceeds 0.85, and can become larger than 1 for values of the intraclass correlation close to 0. For very small intraclass correlations, also in the case of person randomization, unequal cluster sizes turn out to be more efficient when the estimation of variance components is at stake. The results again are similar in case of $\rho_{01} = 0.50$.

Finally, for the D_s -criterion in case of the fixed effect estimators, the derived asymptotic relative efficiencies are rather close to the simulated small sample relative efficiencies.

6. Conclusion and Discussion

The optimality of unequal vs. equal cluster sizes was studied for multilevel intervention studies with small samples. Asymptotic results on the relative variances of the intervention effect estimators were given, as well as on efficiencies in terms of the D_s -criterion for the fixed effects for both cluster and person randomization. For cluster randomization asymptotic results on the relative variance of variance component estimators were also given, as well as the relative efficiency in terms of the D -criterion and the D_s -criterion for the variance components. The degree to which these results hold for small samples was studied through a Monte Carlo simulation study. The differences between the estimation methods used turned out to be very small. More importantly, in most cases the small sample results were rather adequately described by the asymptotic results, implying that these results can be used in planning a multilevel intervention study.

The loss of efficiency in terms of the D -criterion and the D_s -criterion for variance components turned out to be maximally 15%, which means that this loss can be compensated by sampling 18% more clusters. The loss of efficiency in terms of the D_s -criterion for the fixed effects turned out to be maximally 18%, which means that this loss can be compensated by sampling 22% more clusters. The

relative mean squared errors for the intervention effect turned out to be at least 0.82 for both cluster and person randomization. The relative mean squared error for the random intercept variance estimator (as examined for the case of cluster randomization) and for the slope variance estimator (as examined for the case of person randomization) were at least 0.77.

These results were validated for other realistic cluster size distributions, involving more than three cluster sizes, involving other values for the model parameters and involving both cluster and person randomization. Results for these examples were in line with the results obtained in the present Monte Carlo simulation study.

Although asymptotic results were derived for a general allocation ratio of p , the Monte Carlo study only focused on a 50/50 allocation ratio. In case the costs of allocating a person to an experimental group and a control group are not the same, unequal allocation ratios may be more efficient (Liu, 2003). Further studying to which degree the asymptotic results hold for other allocation ratios would therefore be a useful extension of the present study.

Since in many intervention studies a binary outcome measure is used, a final useful extension of the present study would involve the logistic mixed effects model. Formulas for the asymptotic relative efficiencies could be derived, which also could be tested for their practical usefulness through a Monte Carlo simulation study, similar to the present study.

Appendix A1. Relative Efficiency in Terms of the D_s -Criterion of the Variance Components in the Case of Cluster Randomized Trials

The model is defined as in Eq. (1). Searle et al. (1992, p. 89), but also Ankenman et al. (2003, p. 393) give expressions for the asymptotic variance-covariance matrix of the ML-estimators of the variance components:

$$\text{cov} \begin{bmatrix} \hat{\sigma}_0^2 \\ \hat{\sigma}^2 \end{bmatrix} = \begin{bmatrix} \frac{N-K}{\sigma^4} + \sum_{j=1}^K \frac{w_j^2}{n_j^2} & -\sum_{j=1}^K \frac{w_j^2}{n_j} \\ -\sum_{j=1}^K \frac{w_j^2}{n_j} & \sum_{j=1}^K w_j^2 \end{bmatrix} \frac{2\sigma^4}{N \sum_{j=1}^K w_j^2 - \left(\sum_{j=1}^K w_j\right)^2}, \quad (\text{A.1})$$

with w_j being defined as: $w_j = 1/(\sigma_0^2 + \sigma^2/n_j)$. Note that the asymptotic variance-covariance matrix is independent of the allocation ratio used in cluster randomized trials. Searle et al. (1992, p. 90) remark that the determinant of the matrix in the multiplication in (A.1) is equal to $\frac{N \sum_{j=1}^K w_j^2 - (\sum_{j=1}^K w_j)^2}{\sigma^4}$, which implies that the D_s -criterion can be elaborated as:

$$D_s(\text{random}) = \frac{4\sigma^4}{N \sum_{j=1}^K w_j^2 - \left(\sum_{j=1}^K w_j\right)^2}. \quad (\text{A.2})$$

Let w_e be defined as: $w_e = 1/(\sigma_0^2 + \sigma^2/\bar{n})$. The relative efficiency can then be shown to be equal to:

$$\text{RE}(D_s(\text{random})) = \sqrt{\frac{\left(N \sum_{j=1}^K w_j^2 - \left(\sum_{j=1}^K w_j\right)^2\right)}{K(N-K)w_e^2}}. \quad (\text{A.3})$$

The relative efficiency for the random intercept variance can also be derived from Eq. (A.1). We obtain:

$$\text{RE}(\hat{\sigma}_0^2) = \left(\frac{\left(N \sum_{j=1}^K w_j^2 - \left(\sum_{j=1}^K w_j \right)^2 \right)}{K(N-K)w_e^2} \right) \times \left(\frac{(N-K) + K \left(\frac{\sigma^2}{\bar{n}\sigma_0^2 + \sigma^2} \right)^2}{(N-K) + \sum_{j=1}^K \left(\frac{\sigma^2}{n_j \sigma_0^2 + \sigma^2} \right)^2} \right). \quad (\text{A.4})$$

The right factor in the multiplication lies between $(N-K)/N$ and $N/(N-K)$. Note that whenever \bar{n} is large, these boundaries will be close to one. Hence:

$$\text{RE}(\hat{\sigma}_0^2) \approx \frac{\left(N \sum_{j=1}^K w_j^2 - \left(\sum_{j=1}^K w_j \right)^2 \right)}{K(N-K)w_e^2} = (\text{RE}(D_s(\text{random})))^2. \quad (\text{A.5})$$

Appendix A2. Relative Efficiency for the D_s -Criterion of Fixed Parameters in the Case of Person Randomized Trials

The model is defined as in Eq. (2). Let \mathbf{X}_j be the design matrix for cluster j , and let \mathbf{V}_j be the covariance matrix of the outcome variable for cluster j . Let p be the proportion of persons within each cluster allocated to the treated group, and let $(1-p)$ be the proportion of persons allocated to the control group. We then have:

$$\mathbf{X}_j = \begin{bmatrix} \mathbf{1}_{n_j p} & \mathbf{1}_{n_j p} \\ \mathbf{1}_{n_j(1-p)} & -\mathbf{1}_{n_j(1-p)} \end{bmatrix}, \quad \text{and asymptotically also} \quad (\text{A.6})$$

$$\mathbf{V}_j = \begin{bmatrix} \mathbf{J}_{n_j p \times n_j p} (\sigma_0^2 + \sigma_1^2 + 2\sigma_{01}) + \mathbf{I}_{n_j p \times n_j p} \sigma^2 & \mathbf{J}_{n_j p \times n_j(1-p)} (\sigma_0^2 - \sigma_1^2) \\ \mathbf{J}_{n_j(1-p) \times n_j p} (\sigma_0^2 - \sigma_1^2) & \mathbf{J}_{n_j(1-p) \times n_j(1-p)} (\sigma_0^2 + \sigma_1^2 - 2\sigma_{01}) + \mathbf{I}_{n_j(1-p) \times n_j(1-p)} \sigma^2 \end{bmatrix}. \quad (\text{A.7})$$

The asymptotic covariance matrix of the ML estimators of the fixed parameters can be expressed as (Verbeke and Molenberghs, 2000, p. 56):

$$\mathbf{V} = \text{cov} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = \left(\sum_{j=1}^K \mathbf{X}_j^T \mathbf{V}_j^{-1} \mathbf{X}_j \right)^{-1}. \quad (\text{A.8})$$

We introduce the following notations: $\alpha = (\sigma_0^2 + \sigma_1^2 + 2\sigma_{01})$, $\beta = (\sigma_0^2 - \sigma_1^2)$, $\gamma = \sigma_0^2 + \sigma_1^2 - 2\sigma_{01}$, and $\xi_j = \alpha - \frac{\beta^2(n_j(1-p))}{\sigma^2 + \gamma(n_j(1-p))}$. Making use of a result on inverses of partitioned matrices (Harville, 1997, p. 99, Theorem 8.5.11), the inverse \mathbf{V}_j^{-1} can be shown to be equal to:

$$\mathbf{V}_j^{-1} = \begin{bmatrix} v_{j11} & v_{j12} \\ v_{j12} & v_{j22} \end{bmatrix}, \quad \text{in which}$$

$$v_{j11} = \frac{1}{\sigma^2} \left(\mathbf{I}_{n_j p \times n_j p} - \frac{\xi_j}{\sigma^2 + \xi_j n_j p} \mathbf{J}_{n_j p \times n_j p} \right),$$

$$\begin{aligned}
 v_{j12} &= -\left(\frac{\beta}{(\sigma^2 + \xi_j n_j p)(\sigma^2 + \gamma n_j(1-p))}\right) \mathbf{J}_{n_j p \times n_j(1-p)}, \quad \text{and} \quad (\text{A.9}) \\
 v_{j22} &= \frac{1}{\sigma^2} \mathbf{I}_{n_j(1-p) \times n_j(1-p)} + \left(\frac{1}{(\sigma^2 + \gamma n_j(1-p))}\right) \\
 &\quad \times \left(\frac{\beta^2 n_j p}{(\sigma^2 + \gamma n_j(1-p))(\sigma^2 + \xi_j n_j p)} - \frac{\gamma}{\sigma^2}\right) \mathbf{J}_{n_j(1-p) \times n_j(1-p)}.
 \end{aligned}$$

Pre- and postmultiplication of this inverse matrix by the design matrix \mathbf{X}_j and summation across all K clusters yields the inverse of the covariance matrix of fixed effect parameter estimators:

$$\begin{aligned}
 \mathbf{V}^{-1} &= \begin{bmatrix} v_{11} & v_{12} \\ v_{12} & v_{22} \end{bmatrix}, \quad \text{in which} \\
 v_{11} &= \sum_{j=1}^K \frac{n_j p}{(\sigma^2 + \xi_j n_j p)} \left(\frac{\sigma^2 + (\gamma - \beta)n_j(1-p)}{\sigma^2 + \gamma n_j(1-p)}\right) \\
 &\quad + \sum_{j=1}^K \frac{n_j(1-p)}{(\sigma^2 + \gamma n_j(1-p))} \left(\left(\frac{\beta n_j p}{(\sigma^2 + \xi_j n_j p)}\right) \left(\frac{(\beta - \gamma)n_j(1-p) - \sigma^2}{\sigma^2 + \gamma n_j(1-p)}\right) + 1\right), \\
 v_{12} &= \sum_{j=1}^K \frac{n_j p}{(\sigma^2 + \xi_j n_j p)} \left(\frac{\sigma^2 + (\gamma + \beta)n_j(1-p)}{\sigma^2 + \gamma n_j(1-p)}\right) - \sum_{j=1}^K \frac{n_j(1-p)}{(\sigma^2 + \gamma n_j(1-p))} \quad (\text{A.10}) \\
 &\quad \times \left(\left(\frac{\beta n_j p}{(\sigma^2 + \xi_j n_j p)}\right) \left(\frac{\sigma^2 + (\beta + \gamma)n_j(1-p)}{\sigma^2 + \gamma n_j(1-p)}\right) + 1\right), \quad \text{and} \\
 v_{22} &= \sum_{j=1}^K \frac{n_j p}{(\sigma^2 + \xi_j n_j p)} \left(\frac{\sigma^2 + (\gamma + \beta)n_j(1-p)}{\sigma^2 + \gamma n_j(1-p)}\right) \\
 &\quad + \sum_{j=1}^K \frac{n_j(1-p)}{(\sigma^2 + \gamma n_j(1-p))} \left(\left(\frac{\beta n_j p}{(\sigma^2 + \xi_j n_j p)}\right) \left(\frac{\sigma^2 + (\beta + \gamma)n_j(1-p)}{\sigma^2 + \gamma n_j(1-p)}\right) + 1\right).
 \end{aligned}$$

The covariance matrix \mathbf{V} is then obtained as:

$$\mathbf{V} = \begin{bmatrix} v_{22} & -v_{12} \\ -v_{12} & v_{11} \end{bmatrix} \times \frac{1}{(v_{11}v_{22} - v_{12}^2)}. \quad (\text{A.11})$$

No simple expressions result for the determinant of this matrix. But the above expressions can be used in numerically evaluating the relative efficiency of designs in terms of the D_s -criterion for the fixed effect parameters.

In our simulation study, we studied the situation where the intercept and slope variances are equal, and the allocation proportion $p = 0.5$. If we additionally assume there is no intercept-slope correlation, $\rho_{01} = 0$, one can derive from (A.11) that:

$$\text{RE}(\hat{\beta}_0) = \text{RE}(\hat{\beta}_1) = \text{RE}(D_s(\text{fixed})), \quad (\text{A.12})$$

similar to the relative efficiencies for cluster randomization. In this case also a simple expression for these relative efficiencies arises, similar to Eq. (6):

$$\text{RE}(D_s(\text{fixed})) = \left(\frac{\bar{n} + 2(1 - \rho)/\rho}{\bar{n}} \right) \times \frac{1}{K} \sum_{j=1}^K \left(\frac{n_j}{n_j + 2(1 - \rho)/\rho} \right), \quad (\text{A.13})$$

with the intraclass correlation, ρ , however, as defined for multicenter trials.

References

- Ankenman, B. E., Aviles, A. I., Pinheiro, J. C. (2003). Optimal designs for mixed-effects models with two random nested factors. *Statistica Sinica* 13:385–401.
- Atkinson, A. C., Donev, A. N. (1996). *Optimum Experimental Design*. Oxford: Clarendon Press.
- Giovagnoli, A., Sebastiani, P. (1989). Experimental designs for mean and variance estimation in variance components models. *Computational Statistics & Data Analysis* 8:21–28.
- Harville, D. A. (1997). *Matrix Algebra from a Statistician's Perspective*. New York: Springer.
- Headrick, T. C., Zumbo, B. D. (2005). On optimizing multi-level designs: Power under budget constraints. *Australian New Zealand Journal of Statistics* 47:219–299.
- Liu, X. (2003). Statistical power and optimum sample allocation ratio for treatment and control having unequal costs per unit of randomization. *Journal of Educational and Behavioral Statistics* 28:231–248.
- McCulloch, C. E., Searle, S. R. (2001). *Generalized, Linear, and Mixed Models*. New York: Wiley.
- Moerbeek, M., Van Breukelen, G. J. P., Berger, M. P. F. (2001). Optimal experimental designs for multilevel models with covariates. *Communications in Statistics A* 30:2683–2697.
- Mood, A. M., Graybill, F. A., Boes, D. C. (1974). *Introduction to the Theory of Statistics*. 3rd ed. Tokyo: McGraw-Hill.
- Rasbash, J., Browne, W., Goldstein, H., Yang, M., Plewis, I., Healy, M., Woodhouse, G., Draper, D., Langford, I., Lewis, T. (2000). *A User's Guide to MLwiN*. 2nd ed. London: Centre for Multilevel Modelling, Institute of Education, University of London.
- Raudenbush, S. W. (1997). Statistical analysis and optimal design for cluster randomized trials. *Psychological Methods* 2:173–185.
- Raudenbush, S. W., Liu, X. (2000). Statistical power and optimal design for multisite randomized trials. *Psychological Methods* 5:199–213.
- Searle, S., Casella, G., McCulloch, C. (1992). *Variance Components*. New York: Wiley.
- Smeeth, L., 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* 23:409–421.
- Van Breukelen, G. J. P., Candel, M. J. J. M., Berger, M. P. F. (2007). Relative efficiency of unequal versus equal cluster sizes in cluster randomized and multicentre trials. *Statistics in Medicine* 26:2589–2603.
- Verbeke, G., Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. New York: Springer.