

Sample size adjustments for varying cluster sizes in cluster randomized trials with binary outcomes analyzed with second-order PQL mixed logistic regression

Citation for published version (APA):

Candel, M. J. J. M., & van Breukelen, G. J. P. (2010). Sample size adjustments for varying cluster sizes in cluster randomized trials with binary outcomes analyzed with second-order PQL mixed logistic regression. *Statistics in Medicine*, 29(14), 1488-1501. <https://doi.org/10.1002/sim.3857>

Document status and date:

Published: 30/06/2010

DOI:

[10.1002/sim.3857](https://doi.org/10.1002/sim.3857)

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.

Download date: 20 Mar. 2025

Sample size adjustments for varying cluster sizes in cluster randomized trials with binary outcomes analyzed with second-order PQL mixed logistic regression

Math J. J. M. Candel*[†] and Gerard J. P. Van Breukelen

Adjustments of sample size formulas are given for varying cluster sizes in cluster randomized trials with a binary outcome when testing the treatment effect with mixed effects logistic regression using second-order penalized quasi-likelihood estimation (PQL). Starting from first-order marginal quasi-likelihood (MQL) estimation of the treatment effect, the asymptotic relative efficiency of unequal versus equal cluster sizes is derived. A Monte Carlo simulation study shows this asymptotic relative efficiency to be rather accurate for realistic sample sizes, when employing second-order PQL. An approximate, simpler formula is presented to estimate the efficiency loss due to varying cluster sizes when planning a trial. In many cases sampling 14 per cent more clusters is sufficient to repair the efficiency loss due to varying cluster sizes. Since current closed-form formulas for sample size calculation are based on first-order MQL, planning a trial also requires a conversion factor to obtain the variance of the second-order PQL estimator. In a second Monte Carlo study, this conversion factor turned out to be 1.25 at most. Copyright © 2010 John Wiley & Sons, Ltd.

Keywords: cluster randomized trials; mixed effects logistic regression; quasi-likelihood estimation; sample size; varying cluster sizes

1. Introduction

Many 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 pupils nested within schools or the effect of medication given to patients nested within general practices. In cluster randomized trials, such clusters (e.g. schools or general practices) instead of subjects are randomly assigned to each of the treatment arms [1, 2]. The observations on outcome variables within the clusters may be correlated. If the outcome variable is binary, such as in smoking prevention studies, where a person has started smoking or not, or in clinical trials, where patients may or may not have been cured, mixed effects logistic models may be a suitable tool for capturing this correlation in analyzing the data.

Of primary interest is testing the treatment effect of, for instance, a prevention program or a new medication. Formulas for sample size calculation in case of mixed effects logistic regression analysis have been derived, yielding the numbers of clusters and numbers of subjects within a cluster such that power is maximized given a fixed budget for sampling [3, 4]. These studies, however, assume equal cluster sizes. Owing to variation in actual cluster sizes, but also due to nonresponse or dropout, unequal cluster sizes are rather common.

For a mixed effects linear regression of quantitative outcomes, the efficiency loss due to varying cluster sizes has been studied, both when estimating the fixed parameters [5, 6] and when estimating the variance components [6, 7]. Guidelines have been provided how to repair this efficiency loss to maintain the desired power levels. The present study will do the same when estimating and testing the fixed parameter representing the treatment effect in case of a binary outcome analyzed with mixed effects logistic regression. The issue is how much power is lost, and how to compensate for this loss.

Similar to Moerbeek *et al.* [3] and Moerbeek and Maas [4], treatment effect estimation through penalized quasi-likelihood (PQL) is assumed. PQL starts from a linearized version of the logistic model in maximizing the likelihood. The asymptotic loss of efficiency due to varying cluster sizes will be analytically examined for a more simple linearization known as first-order marginal

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

*Correspondence to: Math J. J. M. Candel, Department of Methodology and Statistics, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands.

[†]E-mail: math.candel@stat.unimaas.nl

quasi-likelihood (MQL). Through an extensive Monte Carlo simulation study, this expression for the loss of efficiency will be checked for its appropriateness in case of realistic sample sizes and in case second-order PQL is used, for both maximum likelihood (ML) and restricted maximum likelihood (REML) estimation.

Furthermore, the sample size formulas for mixed effects logistic models in case of equal cluster sizes are based on the first-order MQL-estimator of the treatment effect [3]. This means that there may be an effect of using second-order PQL instead of first-order MQL on the variance of the treatment effect estimator. Correction factors to convert the variance of the MQL-estimator into the variance of the PQL-estimator, as established by Moerbeek *et al.* [3], will also be studied here. Together with the information on the efficiency loss due to varying cluster sizes, this will show to what extent the sample sizes as planned, which are based on the first-order MQL-estimator for equal cluster sizes, should be corrected when using second-order PQL with varying cluster sizes.

The structure of the paper is as follows. In Section 2 we will explain how the mixed effects logistic model describes the relationship between the treatment and a binary outcome for cluster randomized trials, and we will explain quasi-likelihood estimation of the parameters of this model. Section 3 will introduce the relative efficiency of unequal versus equal cluster sizes for the treatment effect estimator and will provide an explicit expression for this relative efficiency in case of first-order MQL estimation. In addition a Taylor approximation of the asymptotic relative efficiency will be given, which is useful in taking care of varying cluster sizes when planning a cluster randomized trial. Section 4 will delineate the design and the results of a Monte Carlo simulation study that examines the relative efficiency of unequal versus equal cluster sizes when employing second-order PQL for realistic cluster sizes. This section will evaluate how accurate the asymptotic relative efficiency based on MQL as well as its Taylor approximation are, in describing the simulated relative efficiencies. Section 5 describes a second simulation study to establish the correction factor for converting the variance of the first-order MQL-estimator into the variance of the second-order PQL-estimator. In Section 6 it will be illustrated how the results can be used in planning a trial. The paper closes with some implications for the planning phase of cluster randomized trials as well as some issues for further research.

2. A mixed effects logistic regression model for cluster randomized trials

2.1. The mixed effects logistic regression model

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. In case of equal cluster sizes, we denote $n_j = n$ for $j = 1, \dots, K$. The total number of persons is $N = \sum_{j=1}^K n_j$. The dependent variable is a binary outcome, denoted as y_{ij} (0 or 1) for person i in cluster j . Let π_{ij} denote the probability that $y_{ij} = 1$. The outcome y_{ij} is assumed to be Bernoulli distributed and can be written as the sum of π_{ij} and a residual score ε_{ij} :

$$y_{ij} = \pi_{ij} + \varepsilon_{ij}, \quad (1)$$

with ε_{ij} having expectation 0 and variance $\pi_{ij}(1 - \pi_{ij})$.

Let X_j denote the treatment to which cluster j is allocated, which is coded +1 for clusters in the treatment condition and -1 for clusters in the control condition. The mixed effects logistic regression model defines $\pi_{ij} = f(\beta_{0j} + \beta_1 X_j)$, with f the logistic function, which can be rewritten in a linear form by taking the natural logarithm of $\pi_{ij}/(1 - \pi_{ij})$:

$$\ln\left(\frac{\pi_{ij}}{1 - \pi_{ij}}\right) = \beta_{0j} + \beta_1 X_j, \quad \text{with } \beta_{0j} = \beta_0 + u_{0j}. \quad (2)$$

With this coding scheme for X_j , β_0 represents the grand mean of $\ln(\pi_{ij}/(1 - \pi_{ij}))$, and β_1 is half the treatment effect in the metric of $\ln(\pi_{ij}/(1 - \pi_{ij}))$. Furthermore, u_{0j} represents a random cluster effect, and is assumed to be independently normally distributed with variance σ_0^2 and mean 0.

A relevant concept is the intraclass correlation, which is the correlation between two randomly drawn persons from the same cluster. The logistic regression model can be seen as arising from variations between subjects on an underlying continuous variable. When these variations are described by the logistic distribution, the mixed effects logistic model results (e.g. [8]). The within-cluster variance on this latent variable is $\pi^2/3$ [8, p. 223] and the intraclass correlation, ρ , in terms of the scores on this latent variable is equal to $\rho = \sigma_0^2 / (\sigma_0^2 + \pi^2/3)$.

2.2. Quasi-likelihood estimation of the parameters

The model for cluster randomization has three parameters that have to be estimated: two fixed regression weights, β_0 and β_1 , and one variance component, σ_0^2 . These parameters can be estimated through maximizing the marginal likelihood function. Since this likelihood function is intractable for the mixed effects logistic model (e.g. [4]), two approaches can be taken in estimating the model parameters: numerical integration and quasi-likelihood estimation. The first approach approximates the likelihood function itself, the second approximates the nonlinear response model as given in equation (2) by a Taylor series expansion. Numerical integration is implemented in, for instance, PROC NLMIXED of SAS [9] and MIXOR [10], quasi-likelihood estimation is implemented in GLIMMIX of SAS [9] and in MLwiN [11]. When the approximation by a Taylor series is only based on the fixed parameters (β_0 and β_1), the method is referred to as MQL. When it is also based on the random effects (u_{0j}), it is called PQL. Depending on the degree of extension of the Taylor approximation, one can distinguish between first- and second-order quasi-likelihood

estimation. Once the logistic model has been linearized through a Taylor series expansion, the model parameters can be estimated with common estimation procedures for the linear mixed effects model, such as 'Iterative Generalized Least Squares' (IGLS) and 'Restricted Iterative Generalized Least Squares' (RIGLS) [12]. These algorithms yield, respectively, ML and REML estimates for the parameters in the linearized logistic model.

Second-order PQL as implemented in MLwiN has been shown to yield accurate estimates of the intervention effect in the model in equation (2) [13]. The bias was rather small, and the mean-squared error smaller than for numerical integration as implemented in MIXOR and somewhat larger than, but close to, the mean-squared error for first-order PQL. Second-order PQL also turned out to perform better in statistical tests of the treatment effect than the other estimation methods. For SAS routines [14], it was shown that, although PQL was somewhat more biased than numerical integration, it clearly had smaller mean-squared error. Both studies [13, 14] furthermore concluded that the convergence rates were much higher for PQL than for numerical integration. In what follows we will therefore consider second-order PQL.

3. Asymptotic relative efficiency of unequal versus equal cluster sizes for first-order MQL

3.1. Closed-form expression of the relative efficiency

We will present an expression for the relative variance of the treatment effect estimator when comparing unequal to equal cluster sizes. Let $\text{var}(\hat{\beta}_1|\xi)$ denote the variance of the treatment effect estimator given a design ξ . Let ξ_{equal} denote a design with equal cluster sizes, and let ξ_{unequal} denote a design with unequal cluster sizes, but with the same number of clusters K and the same number of subjects N as ξ_{equal} . The relative efficiency of unequal versus equal cluster sizes for the estimator of the treatment effect, $\text{RE}(\hat{\beta}_1)$, is defined as

$$\text{RE}(\hat{\beta}_1) = \frac{\text{Var}(\hat{\beta}_1|\xi_{\text{equal}})}{\text{Var}(\hat{\beta}_1|\xi_{\text{unequal}})} \tag{3}$$

Deriving expressions for the variance of the treatment effect estimator in equation (3) in case of second-order PQL will result in an expression that will be too complicated to be of practical use. It, among others, will require the specification of the K values for the cluster-specific intercepts (cf. [4]). Instead we will consider the first-order MQL-estimator, and will examine how these results translate to the second-order PQL-estimator in Section 4.

Let \bar{n} denote the average cluster size of the K clusters. Let w_{jt} be defined as $w_{jt} = (\sigma_0^2 + \sigma_{\text{it}}^2/n_j)^{-1}$ and let w_{jc} be defined as $w_{jc} = (\sigma_0^2 + \sigma_{\text{ic}}^2/n_j)^{-1}$, in which

$$\sigma_{\text{it}}^2 = 2 + e^{-\beta_0 - \beta_1} + e^{\beta_0 + \beta_1} \quad \text{and} \quad \sigma_{\text{ic}}^2 = 2 + e^{-\beta_0 + \beta_1} + e^{\beta_0 - \beta_1} \tag{4}$$

are the within-cluster variances for the treatment and control condition after a particular transformation of y_{ij} (see Appendix A for details). For the case of equal cluster sizes, we have $n_j = \bar{n}$, for $j = 1, \dots, K$, and the weight w_{jt} is denoted as w_{et} and the weight w_{jc} is denoted as w_{ec} . For the first-order MQL estimation, the following expression can be derived for $\text{RE}(\hat{\beta}_1)$ (see Appendix A):

$$\text{RE}(\hat{\beta}_1) = \left(\frac{\sum_{j=1}^{K/2} w_{jt} \sum_{j=K/2+1}^K w_{jc}}{\sum_{j=1}^{K/2} w_{jt} + \sum_{j=K/2+1}^K w_{jc}} \right) \frac{2}{K} \left(\frac{w_{\text{et}} + w_{\text{ec}}}{w_{\text{et}} w_{\text{ec}}} \right) \tag{5}$$

The following properties can be shown to hold for $\text{RE}(\hat{\beta}_1)$:

1. The relative efficiencies do not depend on the number of clusters K , but through the weights w_{jt} , w_{jc} , w_{et} and w_{ec} , do depend on the intercept variance, the distribution of cluster sizes and the fixed regression weights.
2. When we multiply each n_j by a factor $c > 0$ and divide σ_0^2 by the same factor, the weights w_{jt} and w_{jc} are multiplied by the same factor c , which does not affect the relative efficiency in equation (5). As a consequence, the minimum of the relative efficiency will not depend on \bar{n} , but will only be achieved at another value of σ_0^2 and thus of ρ .
3. When $\sigma_0^2 \rightarrow 0$ (and thus $\rho \rightarrow 0$) or $\sigma_0^2 \rightarrow \infty$ (and thus $\rho \rightarrow 1$), we have $\text{RE} \rightarrow 1$. For $0 < \sigma_0^2 < \infty$ (and thus $0 < \rho < 1$), we can show that $\text{RE} < 1$ (making use of the Jensen inequality; see [15, p. 72]), implying that equal cluster sizes are optimal.

In the special case of homogeneous within-cluster variances, that is, $\sigma_{\text{it}}^2 = \sigma_{\text{ic}}^2$ in equation (5), we obtain the expression for the RE for the linear mixed model as derived by Van Breukelen *et al.* [5, p. 2593]. The RE as derived for the MQL-estimator in equation (5) is thus a generalization of the RE for the ML estimator obtained in a mixed effects linear regression in the case of heterogeneous within-cluster variances.

3.2. Taylor approximation of the relative efficiency

For planning the sample sizes of a study, it will be useful to have an approximation of the relative efficiency that does not need the specification of the exact distribution of cluster sizes. For such practical purposes, we will present a second-order Taylor approximation of the relative efficiency in equation (5). The Taylor approximation is based on a stochastic cluster size

distribution, whereas equation (5) is the relative efficiency for a particular realization from such a stochastic cluster size distribution. Consider n_j in equation (5) for $j=1, \dots, K$ as independent realizations of a random variable with expectation μ_n and standard deviation σ_n . Let $CV = \sigma_n / \mu_n$ be the coefficient of variation of cluster sizes. Furthermore, let $\lambda_t = (\mu_n / (\mu_n + \alpha_t))$ with $\alpha_t = \sigma_{\epsilon t}^2 / \sigma_0^2$, and $\lambda_c = (\mu_n / (\mu_n + \alpha_c))$ with $\alpha_c = \sigma_{\epsilon c}^2 / \sigma_0^2$. The following second-order Taylor approximation of the RE in equation (5) can then be derived (see Appendix B):

$$RE_T = \frac{\{1 - CV^2 \lambda_t (1 - \lambda_t)\} \{1 - CV^2 \lambda_c (1 - \lambda_c)\} (\lambda_t + \lambda_c)}{\lambda_t + \lambda_c - CV^2 \{\lambda_t^2 (1 - \lambda_t) + \lambda_c^2 (1 - \lambda_c)\}}. \quad (6)$$

According to this approximation the relative efficiency depends on the mean cluster size, the intercept variance, the fixed regression weights and the coefficient of variation of the cluster sizes, CV. When $\sigma_0^2 \rightarrow 0$ (and thus $\lambda_t \rightarrow 0$ and $\lambda_c \rightarrow 0$) or $\sigma_0^2 \rightarrow \infty$ (and thus $\lambda_t \rightarrow 1$ and $\lambda_c \rightarrow 1$), it can be shown that $RE_T \rightarrow 1$. Furthermore, if $CV \leq 1$ it holds that $RE_T \leq 1$ for $0 < \sigma_0^2 < \infty$. The expression in equation (6) can be reformulated as

$$RE_T^{-1} = \left(\frac{\lambda_t}{\lambda_t + \lambda_c} \right) \{1 - CV^2 \lambda_c (1 - \lambda_c)\}^{-1} + \left(\frac{\lambda_c}{\lambda_t + \lambda_c} \right) \{1 - CV^2 \lambda_t (1 - \lambda_t)\}^{-1}. \quad (7)$$

From equation (7) it is evident that as CV increases, RE_T^{-1} increases and thus RE_T decreases. From equation (7) it also can be derived that the minimum of RE_T is attained at $\lambda_t = \lambda_c = \frac{1}{2}$. For these values $\lambda_t(1 - \lambda_t)$ and $\lambda_c(1 - \lambda_c)$ are maximum, and thus $\{1 - CV^2 \lambda_t(1 - \lambda_t)\}^{-1}$ and $\{1 - CV^2 \lambda_c(1 - \lambda_c)\}^{-1}$ are at the same maximum. Therefore, RE_T is minimized at $\lambda_t = \lambda_c = \frac{1}{2}$, and the minimum is $(1 - CV^2 / 4)$.

Similar to the RE in equation (5), the Taylor approximation given in equations (6) and (7) is a generalization of the one derived by Van Breukelen *et al.* [5, p. 2594] for the case of a mixed effects linear regression, with $\sigma_{\epsilon t}^2$ and $\sigma_{\epsilon c}^2$ representing the heterogeneous variances in the treatment and control condition, respectively. When $\sigma_{\epsilon t}^2 = \sigma_{\epsilon c}^2$, equations (6) and (7) reduce to the Taylor approximation in Van Breukelen *et al.* [5, p. 2594].

In what follows we will examine through an extensive Monte Carlo simulation study, to what extent the asymptotic results on the RE (see equation (5)) as well as the Taylor approximation (see equation (6)), which both are derived for the first-order MQL estimation, hold for the second-order PQL and for numbers of clusters and cluster sizes that are representative of realistic samples (Section 4). After that, we will also study how the variance of the first-order MQL-estimator of the treatment effect needs to be corrected in order to obtain the variance of the second-order PQL-estimator of the treatment effect in case of equal cluster sizes (Section 5). This correction factor is necessary when planning sample sizes based on the variance of the MQL estimator (see [3]), but employing second-order PQL in the actual analysis of the data. This correction factor has also been examined by Moerbeek *et al.* [3] for smaller numbers of clusters and cluster sizes, and by Moerbeek and Maas [4] for a model involving two independent binary variables. Once the variance of the MQL estimator can be translated into the variance of the PQL estimator through the correction factor, we are able to adapt sample sizes further for the loss of efficiency due to varying cluster sizes. In what follows, we will first examine the efficiency loss due to varying cluster sizes (Section 4) and then will consider the correction factor for second-order PQL (Section 5).

4. Monte Carlo investigation of the asymptotic relative efficiency for realistic sample sizes and second-order PQL

4.1. Design

For all simulations we assumed 50–50 per cent allocation, that is, 50 per cent of the clusters are assigned to each of the two treatment conditions. To investigate the relative efficiency for realistic sample sizes and second-order PQL, the following factors were systematically varied: (1) the number of clusters and cluster sizes, (2) the frequency distribution of the cluster sizes within each of the treatment arms, (3) the size of the intraclass correlation, ρ , (4) the estimation method and (5) the values of the fixed parameters, β_0 and β_1 . Table I displays the choices made for these factors. The motivation is given in what follows.

4.1.1. Number of clusters and cluster sizes. We will consider a situation where the cluster size is small, and a situation where the number of clusters is small: $K=54$ with $\bar{n}=24$, and $K=24$ with $\bar{n}=80$. These sample sizes will yield sufficient power for realistic effects and realistic intraclass correlations as encountered in cluster randomized trials. For instance, for $\beta_0=1$ and $\beta_1=0.5$, such that the median probabilities in the treatment and control condition are 0.82 and 0.62, respectively, and assuming equal cluster sizes, the power exceeds 0.80 for $K=54$ and $\bar{n}=24$ in case $\rho \leq 0.28$ (i.e. $\sigma_0^2 \leq 1.28$), and exceeds 0.80 for $K=24$ and $\bar{n}=80$ in case $\rho \leq 0.16$ (i.e. $\sigma_0^2 \leq 0.63$) (cf. [3]). The number of clusters and cluster sizes were deliberately chosen to be modest, so as to study the applicability of the asymptotic results on the relative efficiency for realistic sample sizes.

4.1.2. Frequency distribution. Five different cluster size distributions were studied: (1) a unimodal, (2) a uniform, (3) a bimodal, (4) a positively skewed and (5) a negatively skewed distribution. Three different cluster sizes, g_a, g_b, g_c , with respective frequencies f_a, f_b and f_c were employed. A large range of cluster sizes, namely $R=36$ for $\bar{n}=24$ and $R=120$ for $\bar{n}=80$ (each giving $R/\bar{n}=1.5$),

Table I. Overview of the factors varied in the Monte Carlo simulation study to examine the efficiency loss due to varying cluster sizes.

Factor	Levels
Distribution of cluster sizes*	Uniform distribution, unimodal distribution, bimodal distribution, positively skewed distribution, negatively skewed distribution
Intraclass correlation	$\rho=0.02$ up to 0.30, with steps of 0.02
Estimation method	Second-order PQL: IGLS, RIGLS
Sample sizes (K, \bar{n})	(54, 24), (24, 80)
β_0	0, 2
β_1	0.25, 0.5, 1.0

*A more detailed description of the distributions can be found in Table II.

Table II. Distributions of cluster sizes examined in the Monte Carlo simulations for (a) $K=54$ and $\bar{n}=24$ and (b) $K=24$ and $\bar{n}=80$.

Distribution*	Cluster sizes			Cluster frequencies			CV
	g_a	g_b	g_c	f_a	f_b	f_c	
(a) [†]							
Uniform distribution	6	24	42	18	18	18	0.62
Unimodal distribution	6	24	42	10	34	10	0.46
Bimodal distribution	6	24	42	24	6	24	0.71
Positively skewed distribution	10	28	46	24	18	12	0.59
Negatively skewed distribution	4	16	28	4	10	40	0.31
(b) [‡]							
Uniform distribution	20	80	140	8	8	8	0.63
Unimodal distribution	20	80	140	4	16	4	0.44
Bimodal distribution	20	80	140	10	4	10	0.70
Positively skewed distribution	30	90	150	10	8	6	0.61
Negatively skewed distribution	10	70	130	6	8	10	0.61

* 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); CV = coefficient of variation.

[†]Cluster sizes and cluster frequencies are chosen such that the total number of clusters is equal to 54, the average cluster size is 24, the cluster sizes are even and the difference between the largest and the smallest cluster size is equal to 36 (except for the negatively skewed distribution, where the range is 24 to prevent $g_a < 0$).

[‡]Cluster sizes and cluster frequencies are chosen such that the total number of clusters is equal to 24, the average cluster size is 80, the cluster sizes are even and the difference between the largest and the smallest cluster size is equal to 120.

was considered. The exception was the negatively skewed distribution with $\bar{n}=24$, where—in order to prevent negative cluster sizes—the range was limited to $R/\bar{n}=1$. For symmetrical distributions with $\bar{n}=24$, this implies that the smallest $n=6$ and the largest $n=42$. For symmetrical distributions with $\bar{n}=80$, this implies that the smallest $n=20$ and the largest $n=140$.

A numerical study of the asymptotic relative efficiencies derived for MQL estimation (see equation (5)) showed that as the range of cluster sizes increases (for a fixed average cluster size), the relative efficiency becomes smaller. This also follows from the Taylor approximation of the asymptotic relative efficiency as given in equation (7). Since the accuracy of the asymptotic RE as well as the Taylor approximation is especially important in case of large efficiency loss, large ranges of cluster sizes were considered, that is, we have $R/\bar{n}=1.5$. Details of the cluster size distributions can be found in Table II.

4.1.3. Intraclass correlation. The intercept variance σ_0^2 took 15 different values (running from 0.067 to 1.41) such that the intraclass correlation ranged from 0.02 to 0.30 with steps of 0.02. This range encompasses values as included in similar studies for cluster randomized trials [3, 4, 16].

4.1.4. Estimation method. Two different likelihood-based estimation methods were considered: ML estimation and REML estimation. ML and REML estimation are comparable in terms of bias of the treatment effect estimator. REML estimation, however, is often preferred since it leads to less biased estimates of the variance components and thereby also to less biased estimates of the

Table III. Medians and ranges of the cluster-specific outcome probabilities for different values of the fixed regression weights and the random intercept variance.

β_0	β_1	σ_0^2 *	Condition			
			Treatment		Control	
			Median	(LOB, UPB) [†]	Median	(LOB, UPB) [†]
0	0.25	0.067	0.56	(0.43, 0.68)	0.44	(0.32, 0.57)
0	0.5	0.067	0.62	(0.50, 0.73)	0.38	(0.27, 0.50)
0	1	0.067	0.73	(0.62, 0.82)	0.27	(0.18, 0.38)
2	0.25	0.067	0.90	(0.85, 0.94)	0.85	(0.77, 0.91)
2	0.5	0.067	0.92	(0.88, 0.95)	0.82	(0.73, 0.88)
2	1	0.067	0.95	(0.92, 0.97)	0.73	(0.62, 0.82)
0	0.25	1.410	0.56	(0.11, 0.93)	0.44	(0.07, 0.89)
0	0.5	1.410	0.62	(0.13, 0.95)	0.38	(0.05, 0.87)
0	1	1.410	0.73	(0.20, 0.97)	0.27	(0.03, 0.80)
2	0.25	1.410	0.90	(0.47, 0.99)	0.85	(0.35, 0.98)
2	0.5	1.410	0.92	(0.53, 0.99)	0.82	(0.29, 0.98)
2	1	1.410	0.95	(0.65, 1.00)	0.73	(0.20, 0.97)

*Only extreme values for σ_0^2 are presented, corresponding to intraclass correlations of $\rho=0.02$ and $\rho=0.30$.

[†]LOB=lower bound of the range; UPB=upper bound of the range.

standard error of the fixed effects [17, 18]. ML and REML estimates were obtained with MLwiN through IGLS and RIGLS, respectively, applied to the second-order linearization of the logistic regression model (second-order PQL).

4.1.5. Values of fixed parameters. The relative variance of the MQL estimator of the treatment effect also depends on the fixed parameters (see equation (5)). Therefore different values for the regression weights, β_0 and β_1 , were chosen: $\beta_0=0, 2$, and $\beta_1=0.25, 0.5$ and 1 . For the model in equation (2), approximately 95 per cent of the cluster-specific probabilities are within the interval $(1/(1+\exp(-\beta_0-\beta_1+2\sigma_0)), 1/(1+\exp(-\beta_0-\beta_1-2\sigma_0)))$ for the treatment condition, and within the interval $(1/(1+\exp(-\beta_0+\beta_1+2\sigma_0)), 1/(1+\exp(-\beta_0+\beta_1-2\sigma_0)))$ for the control condition. Table III presents an overview of the intervals for the cluster-specific outcome probabilities being covered in our simulation study.

4.2. Simulation procedure

For each of the $5 \times 15 \times 2 \times 2 \times 3 = 900$ simulation conditions, 2000 data sets were generated, each representing the data for 54 clusters consisting on the average of 24 persons, or the data for 24 clusters consisting an average of 80 persons. Given the processing times of the simulations for each condition, 2000 replications yielded sufficient stability in the estimates of the relative efficiency and were still practically feasible. The simulations as well as the estimation of the model parameters were performed using version 2.02 of MLwiN [11]. Both IGLS and RIGLS were employed to obtain, respectively, the ML and REML estimates of the model parameters. In model estimation, the convergence criterion was set to 0.001 and the number of iterations was limited to 100. For each simulation condition, MLwiN converged in more than 99 per cent of the data sets. Non-convergence mainly occurred for the simulation conditions in which $K=54, \bar{n}=24, \beta_0=0$ and $\rho=0.02$. The non-converged runs were omitted in calculating the RE as defined in equation (3).

4.3. Results on the asymptotic relative efficiency

We display the distributions giving the most extreme results in terms of the RE: the unimodal distribution on the one hand (Figure 1) and the bimodal distribution on the other (Figure 2). As can be seen, for each distribution the asymptotic RE approximates the simulated RE rather well. The same is true for other values of the fixed parameters and for the other values of \bar{n} and K that were examined. An example is the uniform distribution of cluster sizes, with $\bar{n}=80, K=24$, and $\beta_0=0.0, \beta_1=0.5$, as displayed in Figure 3. Note that in this case the asymptotic relative efficiency has its minimum at $\rho=0.024$, and due to the choice of the values for ρ in Figure 3, the relative efficiency seems to be monotonically increasing.

In line with the asymptotic results, unequal cluster sizes are less efficient than equal cluster sizes when estimating the treatment effect. The minimum value for the RE depends on the type of distribution, with a unimodal distribution leading to the least efficiency loss (see Figure 1) and the bimodal distribution leading to the most efficiency loss (see Figure 2). The efficiency losses for a uniform, a positively skewed and a negatively skewed distribution are in between. Note that the different distributions also have a different coefficient of variation, as Table II shows.

According to equation (7), a larger coefficient of variation of the cluster size distribution is associated with a smaller minimum of the RE. Figure 4 displays the RE for a uniform distribution with cluster sizes $g_a=60, g_b=80, g_c=100$ and respective cluster frequencies $f_a=f_b=f_c=8$, giving a CV of 0.20. Comparing the uniform distributions in Figure 3 (CV=0.63) and Figure 4 (CV=0.20)

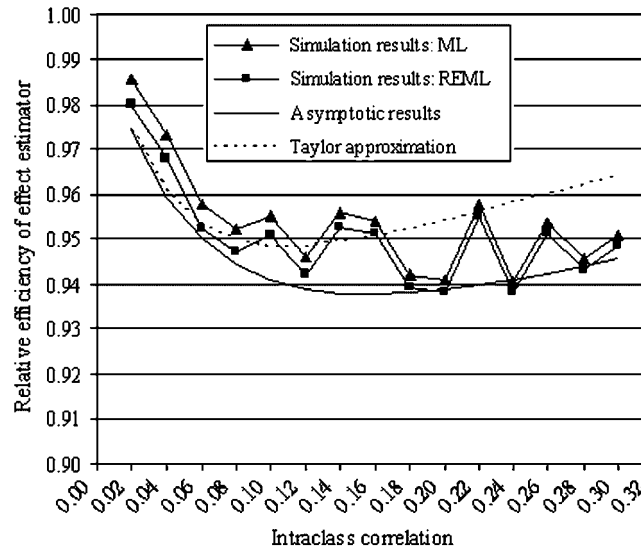


Figure 1. Relative efficiency of the treatment effect estimator for a unimodal distribution of cluster sizes, with $K=54$, $\bar{n}=24$, $CV=0.46$, $\beta_0=2$ and $\beta_1=0.25$. Displayed are also the asymptotic relative efficiency of the ML estimator and its second-order Taylor approximation.

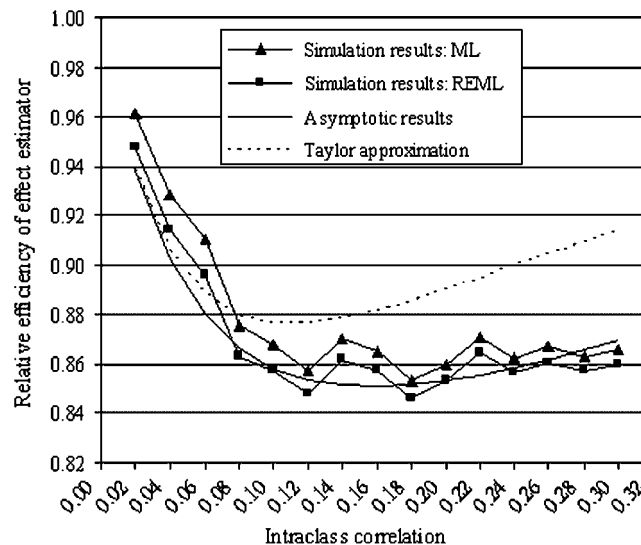


Figure 2. Relative efficiency of the treatment effect estimator for a bimodal distribution of cluster sizes, with $K=54$, $\bar{n}=24$, $CV=0.71$, $\beta_0=2$ and $\beta_1=0.25$. Displayed are also the asymptotic relative efficiency of the ML estimator and its second-order Taylor approximation.

illustrates that the minimum of the simulated RE increases with a decreasing coefficient of variation. However, for a fixed coefficient of variation, the minimum RE still may vary for different shapes of the cluster size distribution. For instance, in case $K=24$ and $\bar{n}=80$, the minima of both the asymptotic and simulated relative efficiency differ for a positively, a negatively skewed and a uniform distribution (not shown here), although they have similar coefficients of variation (see Table II). Hence, the maximum efficiency loss depends on the coefficient of variation of the cluster sizes, and also, but to a smaller extent, on the shape of the distribution.

The efficiency loss for ML estimators is somewhat less than the efficiency loss for REML estimators. Furthermore, the minimum value of the simulated RE does not appear to be strongly affected by the values selected for the fixed regression weights in the simulation study (not shown here).

4.4. Results on the Taylor approximation of the relative efficiency

Comparing the uniform distributions in Figures 3 and 4 showed that the minimum of the simulated RE increases with a decreasing coefficient of variation. This is in line with its Taylor approximation, RE_T . The variations in the minimum of the simulated RE due to variations of the fixed regression coefficients turned out to be rather small. This also is in line with numerical results for the minimum RE_T as a function of the fixed effects (not shown here).

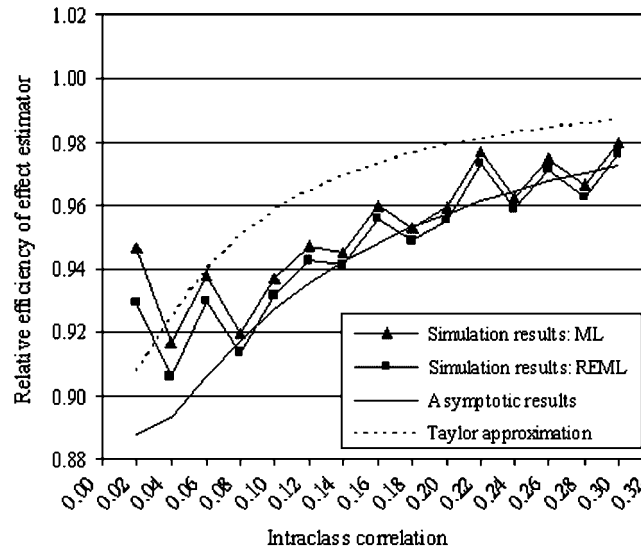


Figure 3. Relative efficiency of the treatment effect estimator for a uniform distribution of cluster sizes, with $K=24$, $\bar{n}=80$, $CV=0.63$, $\beta_0=0$ and $\beta_1=0.5$. Displayed are also the asymptotic relative efficiency of the ML estimator and its second-order Taylor approximation.

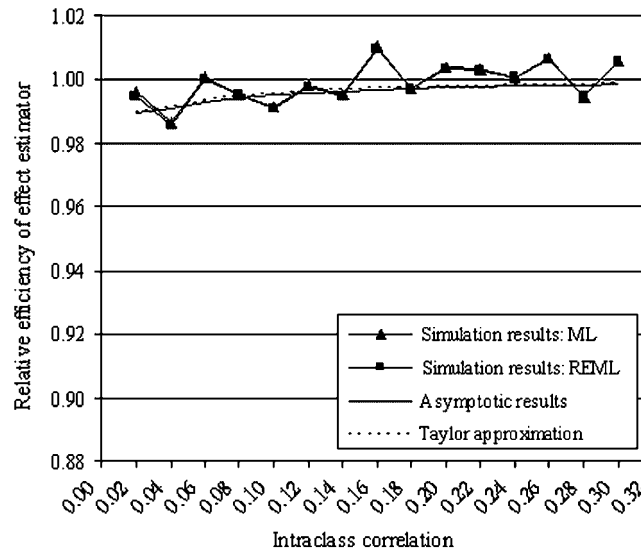


Figure 4. Relative efficiency of the treatment effect estimator for a uniform distribution of cluster sizes, with $K=24$, $\bar{n}=80$, $CV=0.20$, $\beta_0=0$ and $\beta_1=0.5$. Displayed are also the asymptotic relative efficiency of the ML estimator and its second-order Taylor approximation.

Figures 1–4 show that the approximation of the simulated RE by RE_T in most cases is acceptable. For the unimodal, uniform and (not shown) positively skewed distribution, RE_T has an approximation error less than 0.03. For the bimodal and (not shown) negatively skewed distribution, the approximation error can become almost 0.06, but the error at the minimum is smaller (less than 0.04). Plots for other values of the fixed regression weights and for other values of K and \bar{n} gave similar approximation errors. Extending the Taylor approximation with a third- and fourth-order term to also capture the skewness and kurtosis of the cluster size distribution (see Appendix B, equation (B2)) leads to approximation errors less than 0.03 for all distributions. These limits of approximation error also hold for the minima of the RE.

5. Monte Carlo investigation of the MQL-correction factor for equal cluster sizes

5.1. Design of the simulation study

To examine the correction factor for translating the variance of the first-order MQL estimator of the treatment effect for equal cluster sizes into the variance of the second-order PQL estimator of the effect, we calculated their ratio in an extensive Monte Carlo simulation study. Several combinations of the values of σ_0^2 , β_0 , β_1 , n and K were considered. The selected values are displayed in Table IV and, when combined, yield $15 \times 3 \times 3 \times 2 \times 2 = 540$ simulation conditions. Note that the design of the previous simulation

Table IV. Overview of the factors varied in the Monte Carlo simulation study to determine the conversion factors for the variance of the MQL estimator.

Parameter	Values
β_0	0, 1, 2
β_1	0.25, 0.5, 1.0
ρ	0.02–0.30 (with steps of 0.02)
n	24, 80
K	24, 54
Estimation method	Second-order PQL: IGLS, RIGLS

Table V. Conversion factors for calculating the variance of the second-order PQL estimator from the variance of the first-order MQL estimator, obtained through Monte Carlo simulations. The average and the maximum (between brackets) conversion factors are given.

Estimation method	Range of ρ	$K = 54$		$K = 24$	
		$n = 80$	$n = 24$	$n = 80$	$n = 24$
ML	0.02–0.06	1.01 (1.14)	1.00 (1.12)	1.01 (1.10)	1.06 (1.16)
	0.08–0.12	1.02 (1.14)	1.03 (1.18)	1.02 (1.11)	1.06 (1.16)
	0.14–0.18	1.02 (1.08)	1.03 (1.19)	1.04 (1.14)	1.07 (1.21)
	0.20–0.24	1.02 (1.16)	1.04 (1.15)	1.04 (1.18)	1.07 (1.17)
	0.26–0.30	1.03 (1.09)	1.05 (1.15)	1.05 (1.14)	1.08 (1.16)
REML	0.02–0.06	1.01 (1.14)	1.01 (1.12)	1.01 (1.10)	1.07 (1.18)
	0.08–0.12	1.02 (1.15)	1.04 (1.19)	1.03 (1.12)	1.07 (1.18)
	0.14–0.18	1.02 (1.09)	1.04 (1.20)	1.05 (1.15)	1.09 (1.25)
	0.20–0.24	1.03 (1.17)	1.05 (1.16)	1.05 (1.19)	1.10 (1.20)
	0.26–0.30	1.03 (1.10)	1.07 (1.17)	1.06 (1.16)	1.11 (1.19)

is extended in that one more value for β_0 and two more combinations of n and K are examined. For each simulation condition, 2000 data sets were generated and analyzed using version 2.02 of MLwiN [11]. For each simulation condition, non-convergence occurred in less than 1 per cent of the data sets, non-convergence mainly occurred for conditions in which $K=24$ and $n=24$ and for the conditions mentioned in Section 4.2 (the two simulation studies overlapped). Based on the analysis results, the ratio of the variance of the second-order PQL estimates of the intervention effect and the variance of the first-order MQL estimates of the intervention effect was calculated for each simulation condition.

5.2. Results on the correction factor

The results of the second simulation study are summarized in Table V. As can be seen, the average and maximum correction factors for REML are somewhat larger than the corresponding correction factors for ML. This implies that for REML, the number of clusters needs to be increased somewhat more compared with ML, not only in order to compensate for the efficiency loss, but also since the variance of the PQL estimator is somewhat larger than the variance of the MQL estimator. Hence, using REML requires the largest adjustments of the sample size.

For both ML and REML, the correction factors become larger for larger intraclass correlations, for smaller cluster sizes and for smaller numbers of clusters. These results are in line with those of Moerbeek *et al.* [3], and although they considered smaller values for K and n , the correction factors were of similar magnitude. Based on the results in Table V, a conservative correction factor appears to be 1.25, although 1.10 is more representative.

6. Application in planning a trial

When planning a cluster randomized trial to examine the treatment effect through second-order PQL, the results can be used as follows. The relative efficiency of unequal cluster sizes when using the second-order PQL estimator can be approximated by the relative efficiency for the first-order MQL estimator. A reasonable approximation of the relative efficiency for the MQL estimator in turn is the Taylor approximation given in equation (6). One can determine the expected or the minimum RE based on this approximation. To correct for the efficiency loss, one can simply divide the number of clusters in both treatment arms by the expected RE or by the minimum RE (cf. [5]). However, when using the variance of the first-order MQL estimator in calculating the

sample size, also an estimation has to be made of the correction factor to translate this into the variance of the second-order PQL estimator [3]. For this purpose, the correction factors shown in Table V can be used.

To illustrate how to adjust sample sizes, assume we are going to replicate a study among general practices in the county of New Hampshire (U.K.) [19]. In a cluster randomized trial, we want to examine whether an educational programme leads to a higher sensitivity among physicians to detect depression. Suppose we have a budget of C British pounds, the costs per patient are c_1 pounds and the costs for each practice are c_2 pounds. A commonly used cost function is (see e.g. [3])

$$C = nKc_1 + Kc_2. \quad (8)$$

Note that the costs may also simply be the total number of persons involved in a study, namely by choosing $c_1 = 1$ and $c_2 = 0$. The optimal design minimizes the variance of the effect estimator given a budget C . The number of clusters and the number of persons within a cluster minimizing this variance are given by [3, p. 21]

$$K = \frac{C}{(\delta/\sigma_0)\sqrt{c_1c_2} + c_2} \quad \text{and} \quad n = \frac{\delta}{\sigma_0} \sqrt{\frac{c_2}{c_1}}, \quad (9)$$

where $\delta^2 = (\sigma_{\epsilon t}^2 + \sigma_{\epsilon c}^2)/2$. For these optimal K and n , the variance of the MQL-estimator of the treatment effect can be expressed as

$$\text{Var}(\hat{\beta}_1) = \frac{(\sigma_0\sqrt{c_2} + \delta\sqrt{c_1})^2}{C}. \quad (10)$$

The budget C may be chosen such that a particular power is realized. Note that equation (10) in many cases will underestimate the variance of the PQL estimator; hence, the variance of the MQL estimator still has to be multiplied by the correction factor, implying that the budget should be increased to $C^* = C \times$ correction factor, to obtain the same variance for the PQL estimator. From equation (9) it is evident that the number of clusters K then also increases with this same factor, whereas the cluster size does not change.

Suppose $C = 152000$ pounds, and $c_1 = 60$ and $c_2 = 1200$ are the costs per patient and general practice, respectively. Assume an intraclass correlation of $\rho = 0.05$ (or $\sigma_0^2 = 0.17$) and a sensitivity to detect depression of 0.35 in the control group and of 0.45 in the intervention group (cf. [19]). This implies that $\beta_0 = -0.425$, $\beta_1 = 0.218$ and $\delta = 2.058$. According to equation (9) the optimal number of general practices and patients per general practice would then be $K = 59.86$ and $n = 22.32$, say 60 practices of 23 patients each. When employing REML, a suitable correction factor for $\rho = 0.05$ would be 1.12 (see Table V), implying that $K = 59.86 \times 1.12 = 67.04$, and the required budget becomes $C^* = C \times 1.12 = 170240$ pounds.

How to compensate for the efficiency loss due to varying cluster sizes? For this study the frequency distribution of the cluster sizes (i.e. sizes of the general practices) is positively skewed and the corresponding CV can be estimated to be 0.62 (see [20]). According to the Taylor approximation in equation (6), using $\sigma_0^2 = 0.17$ and $\mu_n = 23$ so that $\lambda_t = 0.49$ and $\lambda_c = 0.47$, the RE is 0.90. The efficiency can then be restored by taking $K = 67.04/0.90$, which (after rounding up and allowing for 50–50 per cent allocation of clusters) amounts to $K = 76$ clusters (with $n = 23$).

7. Conclusions and discussion

When testing the treatment effect in cluster randomized trials, care has to be taken of the clustering structure in each of the treatment arms when analyzing the data. For a binary outcome, mixed effects logistic regression is a common tool for capturing this clustering. In the case of equal cluster sizes, formulas for calculating the sample size have been derived by Moerbeek *et al.* [3]. In planning a trial one should, however, also consider the rather common situation of unequal cluster sizes and compensate for the expected loss of efficiency. The present study examined this loss by deriving expressions for the asymptotic relative efficiency of unequal versus equal cluster sizes in case the treatment effect is estimated through the first-order MQL. The extent to which these asymptotic results hold for realistic sample sizes and the second-order PQL, which is a better method of estimating the treatment effect, was studied through an extensive Monte Carlo simulation study. In addition a Taylor approximation was derived, yielding a simpler expression of the relative efficiency, which may be useful in correcting for the efficiency loss in planning a study. Finally, since sample size formulas for cluster randomized trials are based on the expression for the variance of the MQL-estimator [3], correction factors for translating the variance of the MQL estimator into the variance of the PQL estimator were determined through a second Monte Carlo simulation study.

The simulated REs for small sample sizes and second-order PQL estimation were rather adequately described by the asymptotic REs derived for the first-order MQL estimation. In most cases the relative efficiency exceeded 0.88, implying that the efficiency loss can be compensated by sampling $((1/0.88) - 1) \times 100$ per cent = 14 per cent more clusters. For a bimodal distribution and a negatively skewed distribution (having the largest coefficients of variation), the relative efficiency could be lower, and in case of a bimodal distribution could become as low as 0.83, which can be repaired by sampling $((1/0.83) - 1) \times 100$ per cent = 21 per cent more clusters. The difference between the estimation methods ML and REML was small. However, there was a consistent advantage of the ML estimator over the REML estimator. ML estimation thus appears somewhat more robust to varying cluster sizes than REML. On the other hand, REML estimation is often preferred since it leads to less biased estimates of the variance components and of the standard errors of the fixed effects [17, 18].

From the second-order Taylor approximation it becomes clear that the relative efficiency becomes smaller as the coefficient of variation of the cluster sizes increases. The Taylor approximation of the simulated relative efficiency was reasonable, the approximation error for most cases being less than 0.03. The exception is the bimodal and negatively skewed distribution where the approximation error could be up to 0.06. Improvements are possible by extending the Taylor approximation by a third- and fourth-order term, but its use then also requires specifying the skewness and kurtosis of the cluster size distribution.

The correction factors for translating the variance of the MQL estimator into variance of the PQL estimator were somewhat larger for REML than for ML. The correction factor could become as large as 1.21 for ML and as large as 1.25 for REML. Combined with a relative efficiency of 0.88, this would imply, as a conservative estimate, taking $(1.25 \times (1/0.88) - 1) \times 100$ per cent = 42 per cent more clusters than calculated on the basis of the variance of the MQL estimator for equal cluster sizes. In the extreme case of a bimodal distribution of cluster sizes, the number of clusters could increase by $(1.25 \times (1/0.83) - 1) \times 100$ per cent = 51 per cent.

Several cluster size distributions involving more than three cluster sizes as well as models involving other values for the parameters were examined. The results were in line with those obtained in the present Monte Carlo simulation study, indicating the generalizability of the results.

The results of the present study resemble those for a linear mixed model analysis of cluster randomized trials [5, 6]. In those studies, the efficiency losses for the treatment effect estimator for similar coefficients of variation were comparable, although slightly larger: the relative efficiency in most cases exceeded 0.85, except for a bimodal distribution where the relative efficiency could become as low as 0.82. Further, the losses were adequately described by the asymptotic relative efficiencies and Taylor approximations of these asymptotic efficiencies for the linear mixed model case just as for the mixed logistic case. Finally, ML appeared to be more robust against varying cluster sizes than REML in the mixed linear case as well.

Cluster randomized trials are a practical alternative for multicenter trials, which assign part of the persons within each cluster to one of the treatments. Multicenter trials may not be feasible due to the risk of treatment contamination or, as in the case of community-based interventions or interventions taking place in a classroom setting, for logistical reasons. On the other hand, multicenter trials are more efficient than cluster randomized trials for quantitative as well as binary outcomes [3, 21]. Multicenter trials also allow for estimating the between-cluster variability of the effectiveness of the intervention. Asymptotic and simulation results on the efficiency loss due to varying center sizes have already been obtained for the case of a quantitative outcome analyzed with mixed linear regression [5–7]. Therefore, in future research it would be useful to examine the efficiency loss for multicenter trials with a binary outcome.

Appendix A: Derivation of the RE for unequal cluster sizes in case of MQL estimation of the treatment effect

The mixed effects logistic regression model can be formulated as $\pi_{ij} = f(\beta_0 + u_{0j} + \beta_1 X_j)$, with f the logistic function. If the observed score for person i in cluster j is denoted as y_{ij} , which is either 0 or 1, we further have $y_{ij} = \pi_{ij} + \varepsilon_{ij}$, with ε_{ij} the residual score representing the difference between the expected and the observed score, having expectation 0 and variance $\pi_{ij}(1 - \pi_{ij})$. Assuming a first-order linear approximation of f , and after a suitable transformation of y_{ij} (see [4] for details), we obtain $y_{ij}^* = \beta_0 + u_{0j} + \beta_1 X_j + \varepsilon_{ij}^*$, where the transformed error term is $\varepsilon_{ij}^* = \varepsilon_{ij}(\pi_{ij}(1 - \pi_{ij}))^{-1}$. This model is very similar to the linear mixed model formulation for a continuous outcome [5], the difference being that the residual scores ε_{ij}^* are now defined in a specific way (see below), leading to within-cluster variances that are different between the treatment and the control arm.

For $j = 1, \dots, K/2$ (first half of the clusters receiving the experimental treatment), let

$$\mathbf{x}_j = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ \cdot & \cdot \\ 1 & 1 \end{bmatrix},$$

and for $j = K/2 + 1, \dots, K$ (second half of the clusters receiving the control treatment)

$$\mathbf{x}_j = \begin{bmatrix} 1 & -1 \\ 1 & -1 \\ \cdot & \cdot \\ 1 & -1 \end{bmatrix}.$$

Let the vector of observed scores on the transformed dependent variable in cluster j be denoted as

$$\mathbf{y}_j^* = \begin{bmatrix} y_{1j}^* \\ y_{2j}^* \\ \vdots \\ y_{n_{jj}}^* \end{bmatrix},$$

and let the vector of residual scores in cluster j be denoted by

$$\boldsymbol{\varepsilon}_j^* = \begin{bmatrix} \varepsilon_{1j}^* \\ \varepsilon_{2j}^* \\ \vdots \\ \varepsilon_{n_{jj}}^* \end{bmatrix},$$

of which each element, for $j=1, \dots, K/2$, has variance $\sigma_{\varepsilon_{jt}}^2 = \text{Var}(\varepsilon_{ij}(\pi_{ij}(1-\pi_{ij}))^{-1}) = \text{Var}(\varepsilon_{ij}) \times (\pi_{ij}(1-\pi_{ij}))^{-2} = (\pi_{ij}(1-\pi_{ij}))^{-1} = 2 + e^{-\beta_0 - \beta_1} + e^{\beta_0 + \beta_1}$, and for $j=K/2+1, \dots, K$, has variance $\sigma_{\varepsilon_{jc}}^2 = 2 + e^{-\beta_0 + \beta_1} + e^{\beta_0 - \beta_1}$. Note that $\sigma_{\varepsilon_{jt}}^2, \sigma_{\varepsilon_{jc}}^2 \geq 4$ for all β_0 and β_1 . The difference with the linear mixed model formulation for cluster randomized trials by Van Breukelen *et al.* [5] is that the residual variances depend on the fixed parameters and are heterogeneous unless $\beta_0=0$ or $\beta_1=0$.

In matrix notation the model for the scores of \mathbf{y}_j^* of cluster $j=1, \dots, K/2$ can be formulated as

$$\mathbf{y}_j^* = \mathbf{X}_j \boldsymbol{\beta} + \mathbf{1} u_{0j} + \boldsymbol{\varepsilon}_j^*, \tag{A1}$$

where $\boldsymbol{\beta}^T = (\beta_0, \beta_1)$ and $\mathbf{1}$ is a vector of n_j ones. Let \mathbf{I} denote the identity matrix and \mathbf{J} a square matrix with only ones. The variance-covariance matrix of the scores of \mathbf{y}_j^* for cluster $j=1, \dots, K$ can be derived as

$$\mathbf{V}_j = \sigma_{\varepsilon}^2 \mathbf{I} + \sigma_0^2 \mathbf{J}, \tag{A2}$$

where we define $\sigma_{\varepsilon}^2 = \sigma_{\varepsilon_{jt}}^2$ for $j=1, \dots, K/2$, and $\sigma_{\varepsilon}^2 = \sigma_{\varepsilon_{jc}}^2$ for $j=K/2+1, \dots, K$. The inverse of this matrix is obtained by applying a result of Searle *et al.* [22, p. 443]:

$$\mathbf{V}_j^{-1} = \frac{1}{\sigma_{\varepsilon}^2} \left(\mathbf{I} - \mathbf{J} \left(\frac{\sigma_0^2}{n_j \sigma_0^2 + \sigma_{\varepsilon}^2} \right) \right). \tag{A3}$$

Define $w_{jt} = n_j / (n_j \sigma_0^2 + \sigma_{\varepsilon_{jt}}^2)$ and $w_{jc} = n_j / (n_j \sigma_0^2 + \sigma_{\varepsilon_{jc}}^2)$. First-order MQL in combination with IGLS reduces to ML estimation for a first-order linear approximation of the logistic link function f . In this case the information matrix of the ML-estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ can be elaborated as

$$\begin{aligned} \text{Inf} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} &= \sum_{j=1}^{K/2} \mathbf{x}_j^T \mathbf{V}_j^{-1} \mathbf{x}_j + \sum_{j=K/2+1}^K \mathbf{x}_j^T \mathbf{V}_j^{-1} \mathbf{x}_j \\ &= \begin{bmatrix} \sum_{j=1}^{K/2} w_{jt} + \sum_{j=K/2+1}^K w_{jc} & \sum_{j=1}^{K/2} w_{jt} - \sum_{j=K/2+1}^K w_{jc} \\ \sum_{j=1}^{K/2} w_{jt} - \sum_{j=K/2+1}^K w_{jc} & \sum_{j=1}^{K/2} w_{jt} + \sum_{j=K/2+1}^K w_{jc} \end{bmatrix}. \end{aligned} \tag{A4}$$

The asymptotic variance-covariance matrix is obtained by taking the inverse, which can be written as

$$\text{VarCov} \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^{K/2} w_{jt} + \sum_{j=K/2+1}^K w_{jc} & -\sum_{j=1}^{K/2} w_{jt} + \sum_{j=K/2+1}^K w_{jc} \\ -\sum_{j=1}^{K/2} w_{jt} + \sum_{j=K/2+1}^K w_{jc} & \sum_{j=1}^{K/2} w_{jt} + \sum_{j=K/2+1}^K w_{jc} \end{bmatrix} \times \frac{1}{4 \sum_{j=1}^{K/2} w_{jt} \sum_{j=K/2+1}^K w_{jc}}. \tag{A5}$$

The variance of $\hat{\beta}_1$ can be read from this matrix as

$$\text{Var}(\hat{\beta}_1) = \frac{\sum_{j=1}^{K/2} w_{jt} + \sum_{j=K/2+1}^K w_{jc}}{4 \sum_{j=1}^{K/2} w_{jt} \sum_{j=K/2+1}^K w_{jc}}. \tag{A6}$$

Note that the preceding steps are the same when deriving the expression for the variance of the ML estimator of the treatment effect in case of a mixed effects linear regression. A similar expression for $\text{Var}(\hat{\beta}_1)$ is then obtained, with $\sigma_{\epsilon t}^2$ and $\sigma_{\epsilon c}^2$, however, representing the within-cluster variances within the treatment and control arm in a linear mixed effects model.

In case of equal cluster sizes (with \bar{n} the cluster size), let w_{jt} and w_{jc} be denoted as w_{et} and w_{ec} . The relative efficiency of unequal versus equal cluster sizes in terms of $\text{Var}(\hat{\beta}_1)$ can be seen to be equal to

$$\begin{aligned} \text{RE}(\hat{\beta}_1) &= \frac{\sum_{j=1}^{K/2} w_{jt} \sum_{j=K/2+1}^K w_{jc}}{\sum_{j=1}^{K/2} w_{jt} + \sum_{j=K/2+1}^K w_{jc}} \times \frac{(K/2)w_{et} + (K/2)w_{ec}}{(K/2)w_{et}(K/2)w_{ec}} \\ &= \left(\frac{\sum_{j=1}^{K/2} w_{jt} \sum_{j=K/2+1}^K w_{jc}}{\sum_{j=1}^{K/2} w_{jt} + \sum_{j=K/2+1}^K w_{jc}} \right) \frac{2}{K} \left(\frac{w_{et} + w_{ec}}{w_{et}w_{ec}} \right) \end{aligned} \tag{A7}$$

Employing the Jensen inequality (see e.g. [15, p. 72]), it can be shown that $\text{RE}(\hat{\beta}_1)$ in equation (A7) is smaller than 1. It can also be shown that whenever $\sigma_0^2 \rightarrow \infty$ ($\rho \rightarrow 1$) or $\sigma_0^2 \rightarrow 0$ ($\rho \rightarrow 0$), then $\text{RE}(\hat{\beta}_1) \rightarrow 1$.

Let \bar{w}_t be the average of w_{jt} for $j=1, \dots, K/2$ in the treatment arm, and let \bar{w}_c be the average of w_{jc} for $j=K/2+1, \dots, K$ in the control arm. Equation (A7) can then be rewritten as

$$\text{RE}(\hat{\beta}_1) = \left(\frac{\frac{1}{w_{et}} + \frac{1}{w_{ec}}}{\frac{1}{\bar{w}_t} + \frac{1}{\bar{w}_c}} \right). \tag{A8}$$

If $\sigma_{\epsilon t}^2$ and $\sigma_{\epsilon c}^2$ were the within-cluster variances in the treatment and control arm in a mixed effects linear regression, then equation (A8) would also be the RE for the ML estimator of the treatment effect for this model. Actually, if $\sigma_{\epsilon t}^2 = \sigma_{\epsilon c}^2$ equation (A8) reduces to the expression of the RE as derived by Van Breukelen *et al.* [5, p. 2593], which starts from homogeneous within-cluster variances.

Appendix B: Taylor series approximation of the RE for unequal cluster sizes in case of MQL estimation of the treatment effect

To derive a Taylor series approximation of the asymptotic RE for unequal cluster sizes when using the first-order MQL estimator (see equation (5)), we assume that the cluster sizes n_j ($j=1, \dots, K$) are realizations of a random variable U having mean μ_n and standard deviation σ_n . Furthermore, let $\alpha_t = \sigma_{\epsilon t}^2 / \sigma_0^2$ and $\alpha_c = \sigma_{\epsilon c}^2 / \sigma_0^2$. Equation (5) can be considered a moment estimator of

$$\left(\frac{E\left(\frac{U}{U+\alpha_t}\right) E\left(\frac{U}{U+\alpha_c}\right)}{\left(\frac{\mu_n}{\mu_n+\alpha_t}\right) \left(\frac{\mu_n}{\mu_n+\alpha_c}\right)} \right) \left(\frac{\left(\frac{\mu_n}{\mu_n+\alpha_t}\right) + \left(\frac{\mu_n}{\mu_n+\alpha_c}\right)}{E\left(\frac{U}{U+\alpha_t}\right) + E\left(\frac{U}{U+\alpha_c}\right)} \right). \tag{B1}$$

If we define $\lambda_t = (\mu_n / (\mu_n + \alpha_t))$, $\lambda_c = (\mu_n / (\mu_n + \alpha_c))$ and the coefficient of variation of the random variable U by $\text{CV} = \sigma_n / \mu_n$, then by derivations similar to Van Breukelen *et al.* [5, pp. 2601–2602], we obtain the following fourth-order Taylor approximations of the expectation parts of the expression in (B1):

$$\begin{aligned} E\left(\frac{U}{U+\alpha_t}\right) &\approx \lambda_t [1 - \text{CV}^2 \times \lambda_t \times (1 - \lambda_t) \times (1 - \text{skew} \times \text{CV} \times \lambda_t + (\text{kurt} + 3) \times \text{CV}^2 \times \lambda_t^2)] \quad \text{and} \\ E\left(\frac{U}{U+\alpha_c}\right) &\approx \lambda_c [1 - \text{CV}^2 \times \lambda_c \times (1 - \lambda_c) \times (1 - \text{skew} \times \text{CV} \times \lambda_c + (\text{kurt} + 3) \times \text{CV}^2 \times \lambda_c^2)], \end{aligned} \tag{B2}$$

where $\text{skew} = E(U - \mu_n)^3 / \sigma_n^3$ and $\text{kurt} = (E(U - \mu_n)^4 / \sigma_n^4) - 3$. When employing a second-order Taylor approximation, we obtain

$$E\left(\frac{U}{U+\alpha_t}\right) \approx \lambda_t (1 - \text{CV}^2 \times \lambda_t \times (1 - \lambda_t)) \quad \text{and} \quad E\left(\frac{U}{U+\alpha_c}\right) \approx \lambda_c (1 - \text{CV}^2 \times \lambda_c \times (1 - \lambda_c)) \tag{B3}$$

resulting, upon substitution into equation (B1), in the following Taylor approximation of the relative efficiency for unequal cluster sizes:

$$\text{RE}_T = \frac{\{1 - \text{CV}^2 \lambda_t (1 - \lambda_t)\} \{1 - \text{CV}^2 \lambda_c (1 - \lambda_c)\} (\lambda_t + \lambda_c)}{\lambda_t + \lambda_c - \text{CV}^2 \{ \lambda_t^2 (1 - \lambda_t) + \lambda_c^2 (1 - \lambda_c) \}}. \tag{B4}$$

If $\sigma_{\epsilon t}^2$ and $\sigma_{\epsilon c}^2$ were the within-cluster variances in the treatment and control arm, respectively, of a mixed effects linear regression, the approximation in equation (B4) would generalize the Taylor approximation of the RE for a mixed effects linear regression to

the case of heterogeneous within-cluster variances. If $\sigma_{\epsilon t}^2 = \sigma_{\epsilon c}^2$ (so that $\lambda_t = \lambda_c = \lambda$) equation (B4) reduces to $RE_T = 1 - CV^2 \lambda(1 - \lambda)$, which is the expression of the Taylor approximation of the RE as derived by Van Breukelen *et al.* [5, p. 2594] for the case of homogeneous within-cluster variances.

References

1. Donner A, Klar N. Cluster randomization trials in epidemiology: theory and application. *Journal of Statistical Inference and Planning* 1994; **42**:37–56.
2. Raudenbush SW. Statistical analysis and optimal design for cluster randomized trials. *Psychological Methods* 1997; **2**:173–185.
3. Moerbeek M, Van Breukelen GJP, Berger MPF. Optimal experimental designs for multilevel logistic models. *Journal of the Royal Statistical Society Series D—The Statistician* 2001; **50**:17–30.
4. Moerbeek M, Maas CJM. Optimal experimental designs for multilevel logistic models with two binary predictors. *Communications in Statistics: Theory and Methods* 2005; **34**:1151–1167.
5. Van Breukelen GJP, Candel MJJM, Berger MPF. Relative efficiency of unequal versus equal cluster sizes in cluster randomized and multicentre trials. *Statistics in Medicine* 2007; **26**:2589–2603.
6. Candel MJJM, Van Breukelen GJP, Kotova L, Berger MPF. Optimality of unequal cluster sizes in multilevel studies with small sample sizes. *Communications in Statistics: Simulation and Computation* 2008; **37**:222–239.
7. Van Breukelen GJP, Candel MJJM, Berger MPF. Relative efficiency of unequal cluster sizes for variance component estimation in cluster randomized and multicentre trials. *Statistical Methods in Medical Research* 2008; **17**:439–458.
8. Snijders TAB, Bosker RJ. *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. Sage: London, 1999.
9. Littell RC, Milliken GA, Stroup WW, Wolfinger RD. *SAS System for Mixed Models*. SAS institute, Inc.: Cary, NC, 1996.
10. Hedeker D, Gibbons RD. MIXOR: a computer program for mixed-effects ordinal regression analysis. *Computer Methods and Programs in Biomedicine* 1996; **49**:157–176.
11. Rasbash J, Steele F, Browne W, Prosser B. *A User's Guide to MLwin*, version 2. Centre for Multilevel Modelling, University of Bristol: U.K., 2005.
12. Goldstein H. Restricted unbiased iterative generalized least squares estimation. *Biometrika* 1989; **76**:622–623.
13. Moerbeek M, Van Breukelen GJP, Berger MPF. A comparison of estimation methods for multilevel logistic models. *Computational Statistics* 2003; **18**:19–37.
14. Callens M, Croux C. Performance of likelihood-based estimation methods for multilevel binary regression models. *Journal of Statistical Computation and Simulation* 2005; **75**:1003–1017.
15. Mood AM, Graybill FA, Boes DC. *Introduction to the Theory of Statistics* (3rd edn). McGraw-Hill: Tokyo, 1974.
16. Heo M, Leon AC. Performance of mixed effects logistic regression model for binary outcomes with unequal cluster size. *Journal of Biopharmaceutical Statistics* 2005; **15**:513–526.
17. Brown H, Prescott R. *Applied Mixed Models in Medicine*. Wiley: Chichester, 2006.
18. Snijders T, Bosker R. *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. Sage: London, 1999.
19. Thompson C, Kinmonth AL, Stevens L, Pevele RC, Stevens A, Ostler KJ, Pickering RM, Baker NG, Henson A, Preece J, Cooper D, Campbell MJ. Effects of a clinical-practice guideline and practice-based education on detection and outcome of depression in primary care: Hampshire Depression Project randomised controlled trial. *The Lancet* 2000; **355**:185–191.
20. Eldridge SM, Ashby D, Kerry S. Sample size for cluster randomized trials: effect of coefficient of variation of cluster size and analysis method. *International Journal of Epidemiology* 2006; **35**:1292–1300.
21. Moerbeek M, Van Breukelen GJP, Berger MPF. Design issues for experiments in multilevel populations. *Journal of Educational and Behavioral Statistics* 2000; **25**:271–284.
22. Searle SR, Casella G, McCulloch CE. *Variance Components*. Wiley: New York, 1992.