

Relative efficiency of unequal cluster sizes for variance component estimation in cluster randomized and multicentre trials

Citation for published version (APA):

van Breukelen, G. J. P., Candel, M. J. J. M., & Berger, M. P. F. (2008). Relative efficiency of unequal cluster sizes for variance component estimation in cluster randomized and multicentre trials. *Statistical Methods in Medical Research*, 17(4), 439-458. <https://doi.org/10.1177/0962280206079018>

Document status and date:

Published: 01/01/2008

DOI:

[10.1177/0962280206079018](https://doi.org/10.1177/0962280206079018)

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.

Relative efficiency of unequal cluster sizes for variance component estimation in cluster randomized and multicentre trials

Gerard JP van Breukelen, Math JJM Candell, Martijn PF Berger Department of Methodology and Statistics, Maastricht University, Maastricht, The Netherlands

Cluster randomized and multicentre trials evaluate the effect of a treatment on persons nested within clusters, for instance patients within clinics or pupils within schools. Although equal sample sizes per cluster are generally optimal for parameter estimation, they are rarely feasible. This paper addresses the relative efficiency (RE) of unequal versus equal cluster sizes for estimating variance components in cluster randomized trials and in multicentre trials with person randomization within centres, assuming a quantitative outcome. Starting from maximum likelihood estimation, the RE is investigated numerically for a range of cluster size distributions. An approximate formula is presented for computing the RE as a function of the mean and variance of cluster sizes and the intraclass correlation. The accuracy of this approximation is checked and found to be good. It is concluded that the loss of efficiency for variance component estimation due to variation of cluster sizes rarely exceeds 20% and can be compensated by sampling 25% more clusters.

1 Introduction

The effect of a medical or behavioural treatment is often evaluated by a cluster randomized trial. This means that organizational units such as schools or medical centres, called clusters, are randomly allocated to treatment conditions, and all persons sampled from such clusters receive the same treatment assigned to that cluster. Examples are the study of a smoking prevention programme for pupils in schools¹ and community interventions against infectious diseases.² Data analysis has to take the nesting of persons within clusters into account, by including into the statistical model both a random person effect and a random cluster effect, giving a multilevel or mixed-effect model. For quantitative outcomes, this model is known as the one-way classification³ and as the random intercept model.

Instead of randomizing clusters, one may randomize persons within each cluster, which is statistically more efficient,⁴ but not always feasible. An example of this design are multicentre trials in the medical field, which are carried out if the required number of patients cannot be included by any single medical centre. The centres in a multicentre

Address for correspondence: Gerard JP van Breukelen, Department of Methodology and Statistics, Maastricht University, PO Box 616, 6200 MD Maastricht, The Netherlands. E-mail: gerard.vbreukelen@stat.unimaas.nl

trial can be treated as fixed or as random, depending on several considerations.⁵ This paper treats them as random and refers to them as clusters. The statistical model has an additional random effect to allow cluster by treatment interaction, and is known as the two-way crossed classification mixed model³ and as the random coefficient model.

Equal cluster sizes, that is, equal numbers of persons sampled per cluster, have been shown to be optimal for estimating all fixed parameters with maximum precision,⁶ and at least for intraclass correlations above 0.50 also for variance component estimation.⁷ Furthermore, equal cluster sizes simplify data analysis.⁸ However, they are rarely feasible in practice. Among others, variation of actual cluster size and non-response and dropout of persons generate unequal cluster sizes in the trial. Ignoring the issue of bias due to selective dropout, the question then is how much power is lost by cluster size variation and how to compensate for this loss. This is the topic of the present paper.

Some results on the relative efficiency (RE) of unequal cluster sizes for estimating the fixed parameters of the statistical model have been obtained for cluster randomized trials^{6,9,10} and for multicentre trials.^{6,11} In randomized trials, the primary interest is usually in fixed parameters, but the standard errors of fixed parameter estimators are functions of the variance components, that is, of the variances of the random effects. Also, variance component estimation can be an aim in itself, for instance in quality control studies of health care where the amount of health outcome variance between medical centres is of interest.^{12,13} Likewise, estimating cluster by treatment interaction in multicentre trials is needed to assess the generalizability of an average treatment effect found to individual clinical centres. Therefore, addressing the RE of unequal versus equal cluster sizes for variance components is a useful addendum to the study of the RE for fixed parameters. Khuri and Lee studied the effect of unequal cluster sizes on estimation and testing of the outcome variance between clusters in the one-way classification model.^{14,15} They did so by fitting polynomial functions of cluster size variation and variance components to the numerically computed RE for several amounts of cluster size variation and several values of the variance components.

This paper expands the results of Khuri and Lee as follows. Assuming maximum likelihood (ML) estimation, we will show 1) that equal cluster sizes are optimal for variance component estimation, and 2) that the RE of unequal versus equal cluster sizes for variance component estimation is usually high. A Taylor approximation to this RE will be presented as a simple function of the intraclass correlation, mean cluster size and the coefficient of variation of cluster size. This adds to the work of Khuri and Lee in that our approximation is a simple function which does not require estimation of polynomial weights, and applies not only to cluster randomized trials, but also to multicentre trials with cluster by treatment interaction. The structure of this paper is as follows. Section 2 presents the model and estimation procedure for cluster randomized trials, and Section 3 introduces the design criteria that we use and defines RE in the context of nested designs. Formulae for the RE of unequal cluster sizes with respect to variance component estimation are given in Section 4. Section 5 applies these formulae to several cluster size distributions, and evaluates the accuracy of the Taylor approximation. Finally, multicentre trials with person randomization within clusters are treated in Section 6. The paper ends with some conclusions about the practical use of our results in planning trials, and some remarks about unfinished business.

2 Cluster randomized trials

2.1 Model

Suppose that a new treatment is evaluated by a cluster randomized trial, where K clusters (e.g., general practices or schools) are randomly allocated to treatment or control, and control refers to no treatment or to a standard treatment. Within each cluster j ($j = 1, \dots, K$) a total of n_j patients is included and all patients are given the treatment to which their cluster was allocated. Let the effectiveness of the treatment be expressed as its effect on the average of a quantitative outcome Y . The model for data analysis is:

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

where Y_{ij} is the outcome for person i in cluster j , and X_j is the treatment to which cluster j is allocated and will be coded +1 for treated and -1 for control clusters, although other codings are possible. Given this coding scheme, β_0 is the grand mean of the outcome, and β_1 is half the mean outcome difference between treatment and control. Finally, u_{0j} and e_{ij} are a random cluster effect and a person-level residual, assumed to be independently and normally distributed with variances σ_0^2 and σ_e^2 , respectively. The outcomes Y_{ij} and $Y_{i'j}$ of two persons i and i' in an arbitrary cluster j are correlated, and this intraclass correlation is equal to

$$\rho = \frac{\sigma_0^2}{\sigma_0^2 + \sigma_e^2} \quad (2)$$

where the numerator is the covariance of Y_{ij} and $Y_{i'j}$ and the denominator is the variance.

2.2 Covariance matrix of variance component estimators

Ignoring the individual random effects u_{0j} , model (1) has four parameters of interest: two fixed regression parameters and two variance components, summarized by the parameter vector $\Theta = (\beta_0, \beta_1, \sigma_e^2, \sigma_0^2)'$ which can be estimated with the ML method. Asymptotically, the ML estimators of the fixed parameters $(\beta_0, \beta_1)'$ are orthogonal to the variance component estimators, of which the asymptotic covariance matrix is^{3,7}:

$$\text{VarCov} \begin{pmatrix} \hat{\sigma}_e^2 \\ \hat{\sigma}_0^2 \end{pmatrix} = \frac{2}{D} \times \begin{bmatrix} \sum_{j=1}^K w_j^2 & -\sum_{j=1}^K \frac{w_j^2}{n_j} \\ -\sum_{j=1}^K \frac{w_j^2}{n_j} & \frac{N-K}{\sigma_e^4} + \sum_{j=1}^K \frac{w_j^2}{n_j^2} \end{bmatrix} \quad (3)$$

where D is the determinant of the matrix $[\]$ at the right of (3), and

$$w_j = \left(\sigma_0^2 + \frac{\sigma_e^2}{n_j} \right)^{-1} \quad (4)$$

is the weight given to cluster j in estimating the fixed parameters, which increases with the cluster size at a rate that depends on the variance components, with $N = \sum_{j=1}^K n_j$ being the total number of persons in all K clusters. D can be rewritten as (Ref. [3], p. 90):

$$D = \sigma_e^{-4} \times \left\{ N \sum_{j=1}^K w_j^2 - \left(\sum_{j=1}^K w_j \right)^2 \right\} \quad (5)$$

which result will be needed in the sequel. Note that the determinant of matrix (3) equals $4/D$.

3 Design criteria

If model (1) is correct, the ML estimators of all four parameters of (1) are asymptotically unbiased and their covariance matrix can be used as a criterion for evaluating a design ξ . A well-known criterion from optimal design theory is the D -criterion,^{16,17} which has the advantage that it does not depend on the scale or coding of the predictors, here the treatment indicator X_j . A design is D -optimal if it minimizes the determinant $\text{Det}(\text{VarCov}(\hat{\Theta}))$ of the covariance matrix of estimators. Since $\hat{\Theta} = (\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma}_e^2, \hat{\sigma}_0^2)'$ and the fixed estimators are asymptotically orthogonal to the variance component estimators, $\text{Det}(\text{VarCov}(\hat{\Theta}))$ is the product of the determinants of two submatrices, that for the fixed part, and that in Equation (3) for the variance component part.

This paper uses the following optimality criteria:

1. the determinant of the covariance matrix in (3),
2. $\text{Var}(\hat{\sigma}_0^2)$,
3. $\text{Var}(\hat{\sigma}_e^2)$, and
4. $\text{Var}(\bar{n}\hat{\sigma}_0^2 + \hat{\sigma}_e^2)$, where \bar{n} denotes the mean cluster size.

The first criterion concerns the determinant of a submatrix of $\text{VarCov}(\hat{\Theta})$, and is therefore an example of what is known as the D_S -criterion in optimal design.¹⁶ It has a natural interpretation, since the square root of the determinant of (3) is proportional to the area of the confidence ellipsoid for both variance components. The last criterion is an example of the c -criterion. A c -optimal design minimizes the variance of a linear combination $c'\hat{\Theta}$, where c is a vector of weights. In this case, $c' = (0, 0, 1, \bar{n})$. The present c -criterion is of importance because the variances of the ML estimators of $(\beta_0, \beta_1)'$ are roughly proportional to $(\bar{n}\sigma_0^2 + \sigma_e^2)$ in case of unequal cluster sizes, and exactly proportional for equal cluster sizes. The second and third criterion are both special cases of the c -criterion as well as of the D_S -criterion.

Using the D -criterion, and denoting the optimal design as ξ^* , the RE of a design ξ is defined as:

$$RE = \left(\frac{\det(\text{VarCov}(\hat{\Theta}_s)|\xi^*)}{\det(\text{VarCov}(\hat{\Theta}_s)|\xi)} \right)^{1/p} \tag{6}$$

where Θ_s denotes the subvector of parameters of interest, and p is its dimensionality. Applied to the variance components, this means $\Theta_s = (\sigma_e^2, \sigma_0^2)'$ and $p = 2$. Likewise, $p = 1$ for the RE with respect to a single parameter, such as σ_0^2 . The inverse of the RE gives the number of times design ξ has to be replicated in order to give the same statistical information about Θ_s as the optimal design ξ^* .

The next section gives closed forms for the RE of unequal to equal cluster sizes based on each of the criteria above, and a simple Taylor approximation to the RE. In the sequel, ξ and ξ^* will be the design with unequal and equal cluster sizes, respectively, because equal cluster sizes are optimal in general.

4 Relative efficiency of unequal versus equal cluster sizes

4.1 RE: formal expression and some properties

The RE for variance component estimation, based on the D -criterion, follows from Equations (3) and (4), and is equal to:

$$RE = \left(\frac{N \sum_{j=1}^K w_j^2 - (\sum_{j=1}^K w_j)^2}{(N - K)Kw_e^2} \right)^{1/2} \tag{7}$$

Where w_e is w_j in Equation (4) with $n_j = \bar{n}$. Note that $w_e \neq \bar{w}$ (where \bar{w} is the mean of the K weights), except if $\rho = 0$. Equation (7) can be rewritten into

$$RE = \left(\frac{N SS(w) + (N - K)K\bar{w}^2}{(N - K)Kw_e^2} \right)^{1/2} \approx \left(\frac{s_w^2 + \bar{w}^2}{w_e^2} \right)^{1/2} \text{ for large } \bar{n} \tag{8}$$

where $SS(w)$ and \bar{w} are the sum of squares and the mean of the K weights w_j , respectively, and $s_w^2 = SS(w)/K$. The approximation in (8) assumes that \bar{n} is large enough to warrant that $N/(N - K) \approx 1$ (remember that $N = K\bar{n}$). Equation (8) shows that the RE has the following three properties, at least if \bar{n} is large enough to use the approximation:

1. The RE is independent of the number of clusters K .
2. The RE is also independent of \bar{n} in the sense that a larger (smaller) value for \bar{n} gives the same RE at a suitably chosen smaller (larger) value of ρ .
3. For $\rho \rightarrow 0$, the RE is larger than 1, and for $\rho \rightarrow 1$ the RE approaches 1. The RE will be seen to be less than 1 for other values of ρ .

To see that the RE is independent of K , note that the approximation in (8) depends on the mean and variance of w_j and on w_e only. To see that the RE is independent of \bar{n} , note that $w_j = \sigma_0^{-2} n_j / (n_j + \alpha)$ where $\alpha = (1 - \rho) / \rho$. Multiplying n_j and α by a factor > 0 does not affect w_j , \bar{w} , or w_e in (8) apart from changing σ_0^{-2} in all w values. Finally, to see that $\text{RE} > 1$ as $\rho \rightarrow 0$, note that if $\rho \rightarrow 0$, then $\bar{w} \rightarrow w_e$ in (8) and so the numerator of (8) is larger than the denominator. Although $s_w^2 \rightarrow 0$ if $\rho \rightarrow 0$, the net result is $\text{RE} > 1$ as we will see later. This counterintuitive result will be discussed in more detail in Section 5.

The RE per variance component follows from (3) and (5). It can be shown that if \bar{n} is large, the covariance matrix in (3) is almost diagonal and the following approximations hold, with RE denoting the RE in Equations (7) and (8):

$$\text{RE}(\hat{\sigma}_0^2) \approx \text{RE}^2, \quad (9)$$

$$\text{RE}(\hat{\sigma}_e^2) \approx 1,$$

and

$$\text{RE}(\bar{n}\hat{\sigma}_0^2 + \hat{\sigma}_e^2) \approx \text{RE}^2$$

For practical applications, it is important to know how the RE in Equation (7) behaves as a function of cluster size variation. The REs will therefore be studied in two ways. First, a Taylor approximation of Equation (7) will be derived, which expresses the RE as a function of the intraclass correlation ρ and the mean and standard deviation of cluster sizes only (this section). Second, Equation (7) will be investigated numerically for various cluster size distributions in order to evaluate its behaviour and the accuracy of the Taylor approximation (next section). Since the REs in Equation (9) are simple functions of the RE in (7), they will not be addressed in detail anymore.

4.2 Taylor approximation of the RE

Let the n_j values in Equation (7) be i.i.d. with expectation μ_n and standard deviation σ_n . The following Taylor approximation of (7) can then be obtained (see Appendix A):

$$\text{RE}_T = (1 + \text{CV}^2(1 - \lambda)(1 - 3\lambda))^{1/2} \quad (10)$$

where $\text{CV} = \sigma_n / \mu_n$, the coefficient of variation of cluster size, and $\lambda = \mu_n / (\mu_n + \alpha)$, which increases from $\lambda \approx 0$ for $\rho \approx 0$ to $\lambda = 1$ for $\rho = 1$ (since $\alpha = (1 - \rho) / \rho$). The RE_T has as extremes:

$$\begin{aligned} \max \text{RE}_T &\approx (1 + \text{CV}^2)^{1/2} \text{ if } \lambda \approx 0 \iff \rho \approx 0 \\ \min \text{RE}_T &= \left(1 - \frac{\text{CV}^2}{3}\right)^{1/2} \text{ if } \lambda = \frac{2}{3} \iff \rho = \frac{2}{\mu_n + 2} \end{aligned}$$

and $\text{RE}_T = 1$ if $\lambda = 1/3$ or $\lambda = 1$, that is, if $\rho = 1/(2\mu_n + 1)$ or $\rho = 1$.

The maximum is in line with the result from Equation (8) that the $\text{RE} > 1$ when $\rho \rightarrow 0$. In fact, Equation (10) implies $\text{RE} > 1$ for all $\rho < 1/(2\mu_n + 1)$. The minimum

RE is a decreasing function of the CV of cluster size only, and this minimum is attained at a ρ -value almost proportional to the inverse mean cluster size. For instance, a CV of 0.60 leads to a minimum RE of 0.94, and as Equation (9) shows, a minimum RE of 0.88 for estimating the cluster variance σ_0^2 . Of course, the accuracy of Taylor approximation (10) must be tested for a relevant range of cluster size distributions. This will be done in the next section, where we also discuss the strange result of an RE > 1 for intraclass correlations close to zero.

5 Numerical evaluation of the relative efficiency and its Taylor approximation

5.1 Study conditions

The RE of Equation (7) was numerically investigated, and the accuracy of its Taylor approximation (10) was evaluated, under the following conditions:

1. Five different cluster size distributions: a uniform, unimodal, bimodal, positively skewed, and negatively skewed distribution.
2. Three different cluster sizes, $g_a < g_b < g_c$, with frequencies f_a, f_b, f_c .
3. A cluster size range $R (= g_c - g_a)$ increasing from 0 to almost $2\bar{n}$, where \bar{n} is again the mean cluster size.

Details of these conditions are listed in Table 1, showing that a large variation of cluster size was covered. As a benchmark, note that if cluster sizes are normally distributed, the CV is 0.50 at most, and if only the extreme cluster sizes of 2 and $2(\bar{n} - 1)$ occur, each with 50% probability, then the CV is almost 1.0. The number of clusters was fixed at $K = 120$ to have integer values for all cluster size frequencies, and the mean cluster size \bar{n} was fixed at 20. As shown in Section 4, the RE does not depend on K , and choosing a larger (smaller) \bar{n} gives the same RE at a smaller (larger) value of ρ .

For each of the five distributions, the RE was computed with Equations (7) and (10) and plotted against ρ from 0 to 0.25, for different amounts of cluster size variation. Larger values for ρ seem unrealistic for cluster randomized trials,^{2,18,19} and give an RE close to 1. Results for the positively and negatively skewed distributions were combined

Table 1 Cluster size distributions used for computing the relative efficiency

Distribution	Frequencies	Smallest size	Range R	CV
Uniform	$f_a = f_b = f_c = 40$	$g_a = 20 - R/2$	$R = 4, \dots, 36$	0.08–0.73
Unimodal	$f_a = f_c = 30, f_b = 60$	$g_a = 20 - R/2$	$R = 4, \dots, 36$	0.07–0.64
Bimodal	$f_a = f_c = 48, f_b = 24$	$g_a = 20 - R/2$	$R = 4, \dots, 36$	0.09–0.80
+Skewed	$f_a = 60, f_b = 40, f_c = 20$	$g_a = 20 - R/3$	$R = 12, \dots, 36$	0.22–0.67
–Skewed	$f_a = 20, f_b = 40, f_c = 60$	$g_a = 20 - 2R/3$	$R = 12, \dots, 24$	0.22–0.45

Note: 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), $K = 120$ clusters, $\bar{n} = 20$ = average cluster size, $\bar{n} = g_b$ for symmetric distributions, R = range of cluster sizes = $g_c - g_a$, restricted by $g_a > 0$.

into one plot, as the negatively skewed distribution allows only a limited CV to prevent $g_a < 0$. Finally, RE values above 1 were left out for reasons given at the end of this section.

5.2 Results for the RE

Figure 1 plots the RE for the uniform, unimodal, bimodal and skewed cluster size distributions, for the following ranges of cluster size R : $1.0\bar{n}$, $1.2\bar{n}$, $1.5\bar{n}$ and $1.8\bar{n}$. For all distributions and all cluster sizes ranges, the RE first drops to a minimum of about 0.90 at ρ between 0.10 and 0.20, and then increases slowly (returning to 1 for $\rho = 1$). Comparing curves within each plot shows that the minimum RE decreases, as cluster size variation increases. This is in line with the Taylor approximation (10). Comparisons between the four plots show that the minimum RE is highest for the positively skewed distribution, followed by the unimodal, uniform and bimodal distribution, respectively. The negatively skewed distribution is limited to $R = 1.2\bar{n}$ to prevent negative cluster sizes, giving an RE comparable to that of the bimodal distribution. As shown by Equation (9), the squared RE gives an approximation for the $RE(\hat{\sigma}_0^2)$ for cluster variance estimation and for the $RE(\bar{n}\hat{\sigma}_0^2 + \hat{\sigma}_e^2)$.

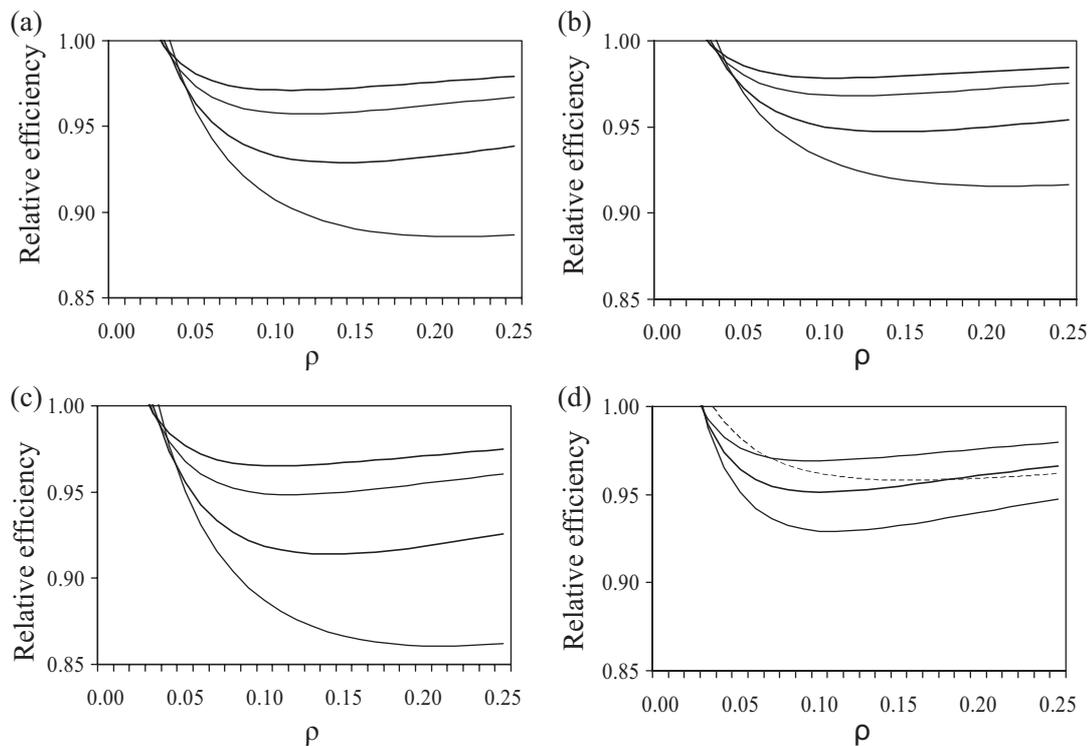


Figure 1 Relative efficiency of unequal versus equal cluster sizes plotted against the intraclass correlation ρ for different distributions of cluster size: (a) uniform, (b) unimodal, (c) bimodal, (d) skewed. Curves for range/mean of the cluster size = 1.0, 1.2, 1.5, 1.8, reading from top to bottom in the right half of each plot. In plot (d) solid curves show range/mean = 1.2, 1.5, 1.8 for positive skew, dashed curve shows range/mean = 1.2 for negative skew. For details, see Table 1.

5.3 Accuracy of the Taylor approximation of the RE

How accurate is the Taylor approximation in (10) of the RE for the cluster size distributions in Table 1? Figure 2(a) shows the RE for the uniform distribution and its approximation (10) for cluster sizes 10–20–30 ($range = \bar{n}$, $CV = 0.41$), and Figure 2(b) does the same for cluster sizes 5–20–35 ($range = 1.5\bar{n}$, $CV = 0.61$). In both plots the Taylor approximation is quite good, with an error that is always less than 0.01 in Figure 2(a) and less than 0.03 in Figure 2(b). Plots for the other distributions in Table 1 gave similar results, and for the positively skewed distribution the approximation error was always less than 0.01.

Plots for the extreme cluster size range of $1.8\bar{n}$ gave approximation errors up to 0.06 for the symmetric distributions, but note that in this case the two extreme cluster sizes of 2 and $2\bar{n} - 2$ each included 25–40% of all clusters. In practice, such extreme cluster sizes will not be encountered very often. Therefore, we also examined the RE for a uniform distribution with ten cluster sizes, such that the extreme cluster sizes of 2 and $2\bar{n} - 2$ each included 10% of all clusters. This gave an approximation error of 0.02 at worst. We finally plotted the RE for the distribution of general practice size in the UK,¹⁰ which is positively skewed with $CV = 0.63$ and range $R > 2\bar{n}$. This gave an error of 0.01 only.

In addition to the RE based on the D -criterion, we also approximated the $RE(\hat{\sigma}_0^2)$, by taking the square of Equation (10) in view of Equation (9). This gave an error

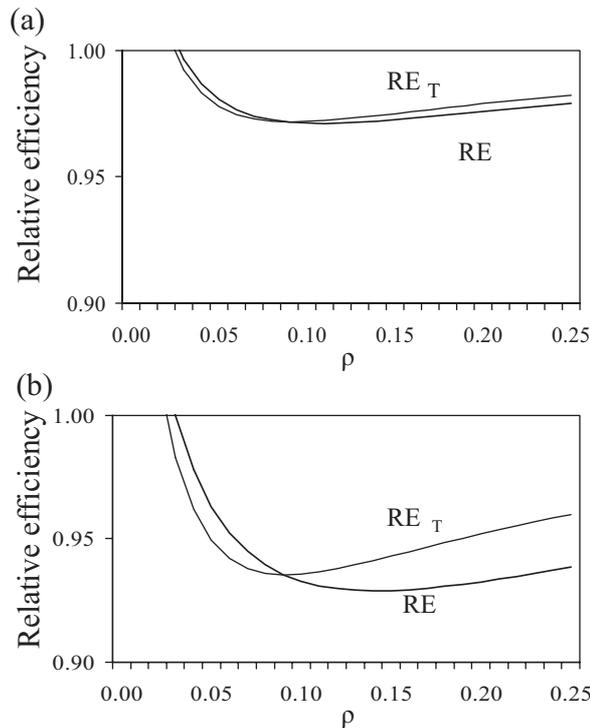


Figure 2 Relative efficiency (RE, thick line) of uniformly distributed cluster sizes 10–20–30 (a) and 5–20–35 (b), and the Taylor series approximation (RE_T , thin line).

almost twice as large as for the RE based on the D -criterion, suggesting that the Taylor approximation is inaccurate if the extreme cluster sizes 2 and $(2\bar{n} - 2)$ each occur with 25% probability.

5.4 Relative efficiencies greater than 1?

According to Equations (8) and (10), the RE of unequal cluster sizes for variance component estimation is greater than 1 if ρ is close to zero. This counterintuitive result follows from Equation (3) for the asymptotic covariance matrix of ML estimators as found in statistical literature.^{3,7} At first glance, this result might be attributed to bias in variance component estimation, introduced by truncation of negative estimates.³ Negative estimates occur more often as ρ approaches zero, and possibly more often for unequal cluster sizes. There are two reasons for rejecting this explanation, however. First, Equation (3) is based on asymptotics, reducing the probability of negative estimates to zero except if ρ is exactly 0. Second, the finding of an RE > 1 for ρ close to zero was confirmed by simulations with realistic sample sizes, for various estimation methods (ML and REML, without and with truncation) and for two efficiency criteria (variance or mean squared error of the variance component estimators). At present, we have no explanation for the RE > 1 finding. However, the RE for fixed parameter estimation is ≤ 1 for all ρ ,⁶ and the RE for variance components is ≤ 1 for almost all ρ , and so equal cluster sizes are optimal. To prevent misuse of the RE > 1 finding in trial planning, we left RE values larger than 1 out of the figures. However, these RE values can be inferred from the Taylor approximation (10) which gave an error less than 0.02 for all RE > 1 for all distributions except for the extreme range of $1.8\bar{n}$.

6 Multicentre trials

6.1 Model

In addition to cluster randomized trials, multicentre trials, with randomization of persons within clusters, are often done. This section extends our RE results to the case of multicentre trials with treatment by cluster interaction, assuming a 50:50 randomization per cluster. We have the following model for the effect of treatment X on a quantitative outcome Y for person i in cluster j :

$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + e_{ij} \quad (11)$$

where

$$\beta_{0j} = \beta_0 + u_{0j} \text{ and } \beta_{1j} = \beta_1 + u_{1j}$$

and X is coded +1 for treated and -1 for control, so that β_0 is the grand mean of Y , and β_1 is half the mean outcome difference between treatment and control. As before, u_{0j} and e_{ij} are a random cluster and person effect, independently and normally distributed with respective variances σ_0^2 and σ_e^2 . The random slope effect u_{1j} reflects treatment by cluster interaction, and is assumed to be normally distributed with variance σ_1^2 and

covariance σ_{01} with u_{0j} , and zero covariance with e_{ij} . The outcome variance is:

$$\text{Var}(Y_{ij}|x_{ij}) = \sigma_0^2 + \sigma_1^2 x_{ij}^2 + 2\sigma_{01}x_{ij} + \sigma_e^2 \tag{12}$$

showing that $\sigma_{01} \neq 0$ represents heterogeneity of outcome variance between treatments. Here, it will be assumed that $\sigma_{01} = 0$, which is in line with current literature.^{3-4,7-15} We will come back to this restriction in Section 7.

6.2 Covariance matrix of variance component estimators

Ignoring the individual random effects, model (11) has five parameters: two fixed regression parameters and three variance components (assuming $\sigma_{01} = 0$), which can be estimated with ML. The asymptotic covariance matrix $\text{VarCov}(\hat{\sigma}_e^2, \hat{\sigma}_0^2, \hat{\sigma}_1^2)'$ of the ML variance components estimators is derived in Appendix B, and is more complicated than, but looks like, that in Equation (3) for cluster randomized trials. Its determinant is:

$$\frac{8\sigma_e^4}{(\sum_{j=1}^K w_j^2 \sum_{j=1}^K v_j^2)N} \leq \text{Det} \left(\text{VarCov} \begin{pmatrix} \hat{\sigma}_e^2 \\ \hat{\sigma}_0^2 \\ \hat{\sigma}_1^2 \end{pmatrix} \right) \leq \frac{8\sigma_e^4}{(\sum_{j=1}^K w_j^2 \sum_{j=1}^K v_j^2)(N - 2K)} \tag{13}$$

where

$$w_j = \left(\sigma_0^2 + \frac{\sigma_e^2}{n_j} \right)^{-1} \quad \text{and} \quad v_j = \left(\sigma_1^2 + \frac{\sigma_e^2}{n_j} \right)^{-1} \tag{14}$$

are the weights given to cluster j in estimating β_0 and β_1 , respectively, and are (almost) the same as in Equation (4) for cluster randomization, increasing with the cluster size at a rate that depends on the variance components. The determinant in (13) moves from its lower bound to its upper bound as the cluster size variation decreases. If the mean cluster size \bar{n} is large enough to warrant that $2K/N \approx 0$ (where $N = K\bar{n}$), the two bounds in (13) coincide.

6.3 Relative efficiency of unequal versus equal cluster sizes

Combining Equations (6) and (13), we obtain the following RE of unequal versus equal cluster sizes for variance component estimation in multicentre trials:

$$\text{RE} \approx \left(\frac{[\text{SS}(w) + K(\bar{w})^2] \times [\text{SS}(v) + K(\bar{v})^2]}{K^2 w_e^2 v_e^2} \right)^{1/3} \tag{15}$$

which closely resembles the RE for cluster randomized trials in Equation (8), especially if the mean cluster size \bar{n} is large enough to let $(N - K)/N \approx 1$ in (8). In that case, the RE in (15) is equal to the geometric mean of the RE in (8) with the intraclass correlation defined as $\rho_0 = \sigma_0^2/(\sigma_0^2 + \sigma_e^2)$ and the RE in (8) with intraclass correlation $\rho_1 = \sigma_1^2/(\sigma_1^2 + \sigma_e^2)$, raised to the power 4/3. So the RE for multicentre trials is a very

similar function of the intraclass correlation and CV of cluster size as the RE for cluster randomized trials, but due to the power transformation it has a lower minimum than (8) if $\rho_0 = \rho_1$.

Treating the cluster size n_j as a random variable and then following the same derivations as in Appendix A for cluster randomized trials, gives the following Taylor approximation of the RE for multicentre trials:

$$RE_T = [(1 + CV^2(1 - \lambda_0)(1 - 3\lambda_0)) \times (1 + CV^2(1 - \lambda_1)(1 - 3\lambda_1))]^{1/3} \quad (16)$$

where $CV = \sigma_n/\mu_n$ is again the coefficient of variation of cluster size, and

$$\lambda_0 = \left(\frac{\mu_n}{\mu_n + \alpha_0} \right) \text{ with } \alpha_0 = \frac{1 - \rho_0}{\rho_0}, \text{ and}$$

$$\lambda_1 = \left(\frac{\mu_n}{\mu_n + \alpha_1} \right), \text{ with } \alpha_1 = \frac{1 - \rho_1}{\rho_1}$$

which resembles the Taylor approximation for cluster randomized trials in (10), being a geometric mean of two RE_T 's following (10), raised to the power 4/3. Its extremes are:

$$\max RE_T \approx (1 + CV^2)^{2/3} \text{ if } \rho_0 \approx 0 \text{ and } \rho_1 \approx 0$$

$$\min RE_T = \left(1 - \frac{CV^2}{3} \right)^{2/3} \text{ if } \rho_0 = \rho_1 = \frac{2}{\mu_n + 2}$$

which are equal to the extremes of the RE_T in (10), raised to the power 4/3.

The RE per variance component follows from equations for $\text{VarCov}(\hat{\sigma}_e^2, \hat{\sigma}_0^2, \hat{\sigma}_1^2)'$ in Appendix B. For large \bar{n} , $\text{VarCov}(\hat{\sigma}_e^2, \hat{\sigma}_0^2, \hat{\sigma}_1^2)'$ becomes diagonal, similarly to the covariance matrix (3) for cluster randomized trials, and the following approximations hold:

$$RE(\hat{\sigma}_0^2) \approx \left(\frac{SS(w) + K(\bar{w})^2}{Kw_e^2} \right) \approx (1 + CV^2(1 - \lambda_0)(1 - 3\lambda_0)) \quad (17)$$

$$RE(\hat{\sigma}_1^2) \approx \left(\frac{SS(v) + K(\bar{v})^2}{Kv_e^2} \right) \approx (1 + CV^2(1 - \lambda_1)(1 - 3\lambda_1))$$

$$RE(\hat{\sigma}_e^2) \approx 1$$

Note that, for \bar{n} large enough to let $(N - K)/K \approx 1$ in Equation (8), the $RE(\hat{\sigma}_0^2)$ in Equation (17) is equal to that for cluster randomized trials in Equations (8) and (9). Further, the $RE(\hat{\sigma}_1^2)$ in Equation (17) is of the same form, apart from replacing the cluster weights w_j with v_j . Observe that the RE in (15) is the geometric mean of the three REs in (17).

7 Discussion

Due to variation of actual centre (e.g., clinic or school) size and due to non-response and dropout, cluster randomized and multicentre trials show variation of cluster size. Since equal cluster sizes are optimal for parameter estimation, a good question is how much efficiency is lost by variation of cluster size and how to compensate for this loss. Focussing on variance component estimation in cluster randomized trials, this paper shows that the RE of unequal versus equal cluster sizes is at worst about 90% according to the D -criterion, and at worst about 80% for the $\text{Var}(\hat{\sigma}_0^2)$ criterion (see Equation 9). So the loss of efficiency is at most about 10 or 20%, depending on the criterion used, and can be compensated by sampling 11 or 25% more clusters than computed with sample size formulae for variance components under the assumption of equal cluster sizes.²⁰ Instead of increasing the number of clusters K , one might increase the average cluster size \bar{n} . But as Equations (4) and (14) show, this only reduces sampling variance at the person level, not at the cluster level. So this is an option only if increasing the number of clusters is impossible.

Additionally, we showed that the RE of unequal versus equal cluster sizes for variance component estimation can be approximated by a simple function of the mean and CV of cluster size and the intraclass correlation. This is again useful for adjusting sample sizes as computed under the assumption of equal cluster sizes. Interestingly, Khuri and Lee studied the RE of unequal versus equal cluster sizes for estimating cluster variance σ_0^2 , by fitting polynomial functions of the mean cluster size, intraclass correlation, and a measure of cluster size variation called ϕ to the RE.^{14,15} Their ϕ is equal to $1/(1 + \text{CV}^2)$. Our Taylor approximation adds to their work in the sense that it neither requires choosing a certain polynomial function nor estimating any polynomial coefficients.

Finally, we extended our results to multicentre trials with person randomization and cluster by treatment interaction, showing that the equations and Taylor approximation for the RE in cluster randomized trials by and large also hold for multicentre trials.

An unexpected result was the finding that the RE for variance components is larger than 1 if the intraclass correlation ρ is close to zero. Although a statistical artefact could be ruled out (see Section 5 for details), we have no explanation for the $\text{RE} > 1$ finding. Given both this lack of understanding and the fact that the RE for fixed parameter estimation is ≤ 1 for all ρ ,⁶ it is best to treat equal cluster sizes as optimal in planning trials.

Our approach was limited by the assumptions of 50:50 treatment allocation and homogeneity of variance components. These assumptions may be relaxed as follows for cluster randomized trials: Unequal allocation means that the number of treated clusters differs from the number of control clusters, but Equation (3) and thereby also the RE equations (7)–(10) hold for any allocation ratio. Heterogeneity of variance components implies that the variance components must be estimated per treatment and the RE of Equation (7) then holds per treatment arm. For multicentre trials, things are more complicated and new mathematical derivations beyond literature^{3,7} and this paper will be needed, since unequal allocation means that in each cluster the number of persons treated differs from the number untreated, and heterogeneity of variance introduces a covariance parameter σ_{01} (Equation 12).

Finally, to prevent misunderstanding about the generality of our results, we emphasize that the RE as based on D -criterion does not depend on the coding scheme for the predictors.^{16,17} In particular, using (0,1) coding for the treatment indicator X leads to the same RE for models (1) and (11) as the $(-1, +1)$ coding used in this paper. This invariance to the coding of X holds under the assumption of homogeneity of outcome variance [which for model (11) implies $\sigma_{01} = 0$ in case of $(-1, +1)$ coding and $\sigma_{01} = -\sigma_1^2/2$ in case of (0,1) coding], and also under heterogeneity of variance (which lets σ_{01} free in both codings).

Appendix A

Taylor series approximation (10) of the RE in Equation (7).

Step 1. Preliminaries

Let the n_j values be independent realizations of a random variable cluster size U with expectation μ_n and standard deviation σ_n . Treat all w_j in Equation (7) as realizations of the function $w(U)$ of U , where $w(U)$ satisfies Equation (4) and equals $w(U) = \sigma_0^{-2}U/(U + \alpha)$. Assume sufficiently large \bar{n} to use the approximation in (8). Finally, let $K \rightarrow \infty$ to obtain:

$$\text{RE} \approx \left(\frac{\sigma_w^2 + \mu_w^2}{w_e^2} \right)^{1/2} \quad (\text{A1})$$

where μ_w and σ_w^2 are the mean and variance of $w(U)$, and $w_e = w(\mu_n)$, and satisfy:

$$\begin{aligned} \mu_w &= \left(\frac{1}{\sigma_0^2} \right) E \left(\frac{U}{U + \alpha} \right), & \sigma_w^2 &= \left(\frac{1}{\sigma_0^4} \right) \text{Var} \left(\frac{U}{U + \alpha} \right) \\ w_e &= \left(\frac{1}{\sigma_0^2} \right) \left(\frac{\mu_n}{\mu_n + \alpha} \right), & \text{with } \alpha &= \frac{1 - \rho}{\rho}. \end{aligned}$$

Step 2. Approximation of μ_w

Defining $d = (U - \mu_n)$, the last term in μ_w can be rewritten as:

$$E \left(\frac{U}{U + \alpha} \right) = E \left(\frac{\mu_n + d}{\mu_n + \alpha + d} \right) = E \left\{ \left(\frac{\mu_n + d}{\mu_n + \alpha} \right) \left(1 + \left(\frac{d}{\mu_n + \alpha} \right) \right)^{-1} \right\}$$

of which the term below is a Taylor series [Ref. [21], p. 533, equation (34)]:

$$\left(1 + \left(\frac{d}{\mu_n + \alpha}\right)\right)^{-1} = \sum_{m=0}^{\infty} \left(\frac{-d}{\mu_n + \alpha}\right)^m$$

if $-(\mu_n + \alpha) < d < (\mu_n + \alpha)$ to ensure convergence. Since $d = U - \mu_n$ and $\alpha \geq 0$, this convergence condition will be satisfied, except for a small $P(U > 2\mu_n + \alpha)$ for strongly positively skewed distributions combined with a large ρ (= small α). So we have

$$E\left(\frac{U}{U + \alpha}\right) = E\left\{\left(\frac{\mu_n + d}{\mu_n + \alpha}\right) \sum_{m=0}^{\infty} \left(\frac{-d}{\mu_n + \alpha}\right)^m\right\} \tag{A2}$$

Step 3

Ignoring all terms d^m with $m > 2$, and rearranging terms, will give:

$$E\left(\frac{U}{U + \alpha}\right) \approx \lambda \times \left\{1 - CV^2 \times \lambda(1 - \lambda)\right\} \tag{A3}$$

where

$$\lambda = \left(\frac{\mu_n}{\mu_n + \alpha}\right) \in (0, 1]$$

and $CV = \sigma_n/\mu_n$ is the coefficient of variation of U .

Step 4. Approximation of σ_w^2

Applying the same Taylor series to the term $\text{Var}(U/(U + \alpha))$ in σ_w^2 , and ignoring terms d^m with $m > 1$ before taking the variance of $U/(U + \alpha)$, gives

$$\text{Var}\left(\frac{U}{U + \alpha}\right) \approx \lambda^2(1 - \lambda)^2 CV^2 \tag{A4}$$

Step 5. Approximation of the RE

Inserting (A3) and (A4) in Equation (A1), and rewriting gives:

$$RE_T = (1 + CV^2(1 - \lambda)(1 - 3\lambda) + CV^4(\lambda(1 - \lambda))^2)^{1/2} \tag{A5}$$

which, by ignoring the term of order CV^4 as CV is generally less than 1, gives Equation (10). Including the fourth order term into (10) hardly improved its accuracy in the study in Section 5.

Appendix B

$\text{VarCov}(\hat{\sigma}_e^2, \hat{\sigma}_0^2, \hat{\sigma}_1^2)'$ and its determinant for multicentre trials

Step 1. Covariance matrix for equal cluster sizes

In case of equal cluster sizes, the ML estimator $(\hat{\sigma}_e^2, \hat{\sigma}_0^2, \hat{\sigma}_1^2)'$ for model (11) satisfies:

$$\text{VarCov} \begin{pmatrix} \hat{\sigma}_e^2 \\ \hat{\sigma}_0^2 \\ \hat{\sigma}_1^2 \end{pmatrix} = \frac{2\sigma_e^4}{n^2(N-2K)} \times \begin{bmatrix} n^2 & -n & -n \\ -n & W & 1 \\ -n & 1 & V \end{bmatrix} \quad (\text{A6})$$

where

$$W = (n-2) \left(\frac{n}{w_e \sigma_e^2} \right)^2 + 1 \text{ and } V = (n-2) \left(\frac{n}{v_e \sigma_e^2} \right)^2 + 1$$

and n is the cluster size, and w_e, v_e are as defined in Equation (14), for equal cluster sizes.

Equation (A6) is obtained by starting with the following model (Ref. [3], p. 123):

$$\begin{aligned} Y_{ij} &= \beta_0 + \beta_1 + u_{0j} + u_{1tj} + e_{ij} \text{ for treated clusters} \\ Y_{ij} &= \beta_0 - \beta_1 + u_{0j} + u_{1cj} + e_{ij} \text{ for control clusters} \end{aligned}$$

where u_{1tj} and u_{1cj} are i.i.d. random variables representing treatment by cluster interaction. Model (11) differs from this model in that it adds the constraint $u_{1tj} + u_{1cj} = 0$. The two models are reparametrizations of each other, as the variance components (denoted σ^2) of model (11) can be obtained from the variance components (denoted τ^2) of the model above by the following equations (Ref. [3], pp. 123–6):

$$\sigma_1^2 = \frac{\tau_1^2}{2}, \quad \sigma_0^2 = \tau_0^2 + \frac{\tau_1^2}{2}, \quad \sigma_e^2 = \tau_e^2$$

Using these equations, and using the covariance matrix of $(\hat{\tau}_e^2, \hat{\tau}_0^2, \hat{\tau}_1^2)'$ for equal cluster sizes (Ref. [3], p. 156), Equation (A6) can be obtained.

Step 2. Information matrix for equal and unequal cluster sizes

Computing the inverse of (A6) gives the information matrix of $(\hat{\sigma}_e^2, \hat{\sigma}_0^2, \hat{\sigma}_1^2)'$ for equal cluster sizes. Filling in $K = 1$ and $N = n_j$ gives the information matrix for cluster j .

The information matrix for K unequal cluster sizes is the sum of these K matrices, giving

$$\text{Info} \begin{pmatrix} \hat{\sigma}_e^2 \\ \hat{\sigma}_0^2 \\ \hat{\sigma}_1^2 \end{pmatrix} = \frac{1}{2} \times \begin{bmatrix} \left(\frac{N-2K}{\sigma_e^4}\right) + \sum_{j=1}^K \frac{w_j^2 + v_j^2}{n_j^2} & \sum_{j=1}^K \frac{w_j^2}{n_j} & \sum_{j=1}^K \frac{v_j^2}{n_j} \\ \sum_{j=1}^K \frac{w_j^2}{n_j} & \sum_{j=1}^K w_j^2 & 0 \\ \sum_{j=1}^K \frac{v_j^2}{n_j} & 0 & \sum_{j=1}^K v_j^2 \end{bmatrix} \quad (\text{A7})$$

where w_j and v_j are defined by Equation (14). To prevent complicated expressions in the next step, the elements of this information matrix will be identified by the following notation:

$$\text{Info} \begin{pmatrix} \hat{\sigma}_e^2 \\ \hat{\sigma}_0^2 \\ \hat{\sigma}_1^2 \end{pmatrix} = \frac{1}{2} \times \begin{bmatrix} a & b & c \\ b & d & 0 \\ c & 0 & e \end{bmatrix}$$

Step 3. Covariance matrix for unequal cluster sizes

Finally, invert the information matrix (A7) to get the following covariance matrix of the ML estimator $(\hat{\sigma}_e^2, \hat{\sigma}_0^2, \hat{\sigma}_1^2)'$ for the case of unequal cluster sizes:

$$\text{VarCov} \begin{pmatrix} \hat{\sigma}_e^2 \\ \hat{\sigma}_0^2 \\ \hat{\sigma}_1^2 \end{pmatrix} = \frac{2}{(ade - b^2e - c^2d)} \times \begin{bmatrix} de & -be & -cd \\ -be & (ae - c^2) & bc \\ -cd & bc & (ad - b^2) \end{bmatrix} \quad (\text{A8})$$

where a, b, c, d, e are defined by (A7). The determinant of (A8) is:

$$\text{Det} \left(\text{VarCov} \begin{pmatrix} \hat{\sigma}_e^2 \\ \hat{\sigma}_0^2 \\ \hat{\sigma}_1^2 \end{pmatrix} \right) = \frac{8}{(ade - b^2e - c^2d)} \quad (\text{A9})$$

Step 4. Determinant of the covariance matrix

By substituting a, b, c, d, e and then rewriting (A9), we obtain:

$$\text{Det} \left(\text{VarCov} \begin{pmatrix} \hat{\sigma}_e^2 \\ \hat{\sigma}_0^2 \\ \hat{\sigma}_1^2 \end{pmatrix} \right) = \frac{8}{\left(D_w \sum_{j=1}^K v_j^2 + D_v \sum_{j=1}^K w_j^2 - \frac{N}{\sigma_e^4} \sum_{j=1}^K w_j^2 \sum_{j=1}^K v_j^2 \right)} \quad (\text{A10})$$

where

$$D_w = \left(\frac{N-K}{\sigma_e^4} + \sum_{j=1}^K \frac{w_j^2}{n_j^2} \right) \left(\sum_{j=1}^K w_j^2 \right) - \left(\sum_{j=1}^K \frac{w_j^2}{n_j} \right)^2$$

which is equal to the D in Equation (3), and D_v is defined analogously by replacing all w_j with v_j . Rewriting D_w and D_v with Equation (5) and rearranging terms in (A10) gives:

$$\text{Det} \left(\text{VarCov} \begin{pmatrix} \hat{\sigma}_e^2 \\ \hat{\sigma}_0^2 \\ \hat{\sigma}_1^2 \end{pmatrix} \right) = \frac{8\sigma_e^4}{\left(\sum_{j=1}^K w_j^2 \sum_{j=1}^K v_j^2 \right) (N - K\delta_w - K\delta_v)} \quad (\text{A11})$$

where

$$\delta_w = \frac{\left(\sum_{j=1}^K w_j \right)^2}{K \sum_{j=1}^K w_j^2} = \frac{K\bar{w}^2}{\sum_{j=1}^K w_j^2} \in (0, 1]$$

with $\delta_w = 1$ for equal cluster sizes and $\delta_w < 1$ else, and δ_v is defined likewise in terms of the v_j values.

Appendix C. Relative efficiency of unequal cluster sizes in multicentre trials

Using the results in Appendix B, the RE will be derived in terms of the D -criterion (derivations of the RE per variance component are available upon request).

RE based on the D -criterion

Using the RE definition in Equation (6), and using Equation (A11), we get

$$\text{RE} \geq \left(\frac{\sum_{j=1}^K w_j^2 \sum_{j=1}^K v_j^2}{K^2 w_e^2 v_e^2} \right)^{1/3} \quad (\text{A12})$$

with equality if $\bar{n} \rightarrow \infty$ such that $2K/N \approx 0$, making (A11) independent of δ_w and δ_v . Equation (A12) can be rewritten as

$$\text{RE} \geq \left(\frac{[\text{SS}(w) + K(\bar{w})^2] \times [\text{SS}(v) + K(\bar{v})^2]}{K^2 w_e^2 v_e^2} \right)^{1/3} \quad (\text{A13})$$

Treating the cluster size n_j as a random variable, and treating the cluster weight w_j as a random variable with expectation μ_w and variance σ_w^2 , and assuming K large enough to let $K/(K-1) \approx 1$, Equation (A13) can be rewritten as:

$$\text{RE} \geq \left(\left[\frac{\sigma_w^2 + \mu_w^2}{w_e^2} \right] \times \left[\frac{\sigma_v^2 + \mu_v^2}{v_e^2} \right] \right)^{1/3} \quad (\text{A14})$$

which allows the Taylor approximation in Equation (16) (for details, see Appendix A).

References

- 1 Ausems M, Mesters I, Van Breukelen G, De Vries H. Short-term effects of a randomised computer-based out-of-school smoking prevention trial aimed at Dutch elementary schoolchildren. *Preventive Medicine* 2002; **34**: 581–9.
- 2 Hayes RJ, Alexander NDE, Bennett S, Cousens SN. Design and analysis issues in cluster-randomized trials of interventions against infectious diseases. *Statistical Methods in Medical Research* 2000; **9**: 95–116.
- 3 Searle S, Casella G, McCulloch C. *Variance components*. Wiley, 1992.
- 4 Moerbeek M, Van Breukelen GJP, Berger MPF. Design issues for experiments in multilevel populations. *Journal of Educational & Behavioral Statistics* 2000; **25**: 271–84.
- 5 Senn S. Some controversies in planning and analyzing multicentre trials. *Statistics in Medicine* 1998; **17**: 1753–65.
- 6 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**: 2589–603.
- 7 Ankenman BE, Aviles AI, Pinheiro JC. Optimal designs for mixed-effects models with two random nested factors. *Statistica Sinica* 2003; **13**: 385–401.
- 8 Moerbeek M, Van Breukelen GJP, Berger MPF. A comparison between traditional methods and multilevel regression for the analysis of multi-center intervention studies. *Journal of Clinical Epidemiology* 2003; **56**: 341–50.
- 9 Manatunga AK, Hudgens MG, Chen S. Sample size estimation in cluster randomized studies with varying cluster size. *Biometrical Journal* 2001; **43**: 75–86.
- 10 Kerry SM, Bland JM. Unequal cluster sizes for trials in English and Welsh general practices: implications for sample size calculations. *Statistics in Medicine* 2001; **20**: 377–90.
- 11 Fedorov V, Jones B. The design of multicentre trials. *Statistical Methods in Medical Research* 2005; **14**: 205–48.
- 12 Van Berkestijn LGM, Kastein MR, Lodder A, DeMelker RA, Bartelink ML. How do we compare with our colleagues? Quality of general practitioner performance in consultations for non-acute abdominal complaints. *International Journal for Quality in Health Care* 1999; **11**: 475–86.
- 13 Campbell SM, Hann M, Hacker J, Burns C, Oliver D, Thapar A, Mead N, Gelb Safran D, Roland MO. Identifying predictors of high quality care in English general practice: observational study. *British Medical Journal* 2001; **323**: 784–7.

- 14 Khuri AI. A method for determining the effect of imbalance. *Journal of Statistical Planning and Inference* 1996; 55: 115–29.
- 15 Lee J. The effect of design imbalance on the power of the F-test of a variance component in the oneway random model. *Biometrical Journal* 2003; 45: 238–48.
- 16 Atkinson AC, Donev AN. *Optimum experimental design*. Clarendon Press, 1996.
- 17 Silvey, S. *Optimal design; an introduction to the theory of parameter estimation*. Chapman and Hall, 1980.
- 18 Campbell MJ. Cluster randomized trials in general (family) practice research. *Statistical Methods in Medical Research* 2000; 9: 81–94.
- 19 Smeeth L, Siu-Woon E. 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* 2002; 23: 409–21.
- 20 Moerbeek M, Wong WK (2002). Multiple-objective optimal designs for the hierarchical linear model. *Journal of Official Statistics* 2002; 18: 291–303.
- 21 Mood AM, Graybill FA, Boes DC. *Introduction to the theory of statistics*, third edition. McGraw-Hill, 1974.