

RESEARCH ARTICLE

Efficient design of cluster randomized trials with treatment-dependent costs and treatment-dependent unknown variances

Gerard J.P. van Breukelen^{1,2}  | Math J.J.M. Candel¹ 

¹Department of Methodology and Statistics, CAPHRI Care and Public Health Research Institute, Maastricht University, PO Box 616, 6200 MD, The Netherlands

²Department of Methodology and Statistics, Graduate School of Psychology and Neuroscience, Maastricht University, PO Box 616, 6200 MD, The Netherlands

Correspondence

Gerard J.P. van Breukelen, Department of Methodology and Statistics, CAPHRI Care and Public Health Research Institute, Maastricht University, PO Box 616, 6200 MD, The Netherlands.

Email: gerard.

vbreukelen@maastrichtuniversity.nl

Cluster randomized trials evaluate the effect of a treatment on persons nested within clusters, where treatment is randomly assigned to clusters. Current equations for the optimal sample size at the cluster and person level assume that the outcome variances and/or the study costs are known and homogeneous between treatment arms. This paper presents efficient yet robust designs for cluster randomized trials with treatment-dependent costs and treatment-dependent *unknown* variances, and compares these with 2 practical designs. First, the maximin design (MMD) is derived, which maximizes the minimum efficiency (minimizes the maximum sampling variance) of the treatment effect estimator over a range of treatment-to-control variance ratios. The MMD is then compared with the optimal design for homogeneous variances and costs (balanced design), and with that for homogeneous variances and treatment-dependent costs (cost-considered design). The results show that the balanced design is the MMD if the treatment-to control cost ratio is the same at both design levels (cluster, person) and within the range for the treatment-to-control variance ratio. It still is highly efficient and better than the cost-considered design if the cost ratio is within the range for the squared variance ratio. Outside that range, the cost-considered design is better and highly efficient, but it is not the MMD. An example shows sample size calculation for the MMD, and the computer code (SPSS and R) is provided as supplementary material. The MMD is recommended for trial planning if the study costs are treatment-dependent and homogeneity of variances cannot be assumed.

KEYWORDS

cluster randomized trial, heterogeneous variance, maximin design, optimal design, sample size, study costs

1 | INTRODUCTION

Randomized experiments are commonly seen as the best method for evaluating the effect of some new treatment in public health and medicine. However, randomized assignment of individuals is neither always feasible nor always

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2018 The Authors. *Statistics in Medicine* published by John Wiley & Sons Ltd.

desirable. For instance, a healthy food program for restaurants in high schools can only be implemented at the school level, and so we can randomly assign schools, but not pupils, to treatment or control. Further, even if randomization of individuals is possible, this may be undesirable due to a risk of treatment contamination. For instance, it may be very difficult for a life style counselor in family practice to switch between 2 methods of counseling when switching between patients. Treatment contamination reduces the outcome difference between groups and the power to detect a difference. For these reasons, cluster randomized trials,^{1,2} also known as group randomized trials,³ are frequently run. In such trials organizational units (eg, schools or family practices), called clusters, are randomly assigned to treatment or control, and all individuals within the same cluster receive the treatment to which their cluster was assigned. Cluster randomized trials are thus a natural alternative to the randomized experiment when individuals are nested within organizations, such as pupils in schools or patients in health centres. Some examples of cluster randomized trials are a smoking prevention trial and a stress management trial, both run in primary schools with schools as unit of assignment,^{4,5} and the assessment of COPD burden trial, run in primary and secondary care.⁶

However, cluster randomization gives a lower power and precision than individual randomization because outcome variation between clusters within the same treatment group inflates the sampling variance of the treatment effect estimator by a factor known as the design effect. This design effect can easily be as large as 2 or 3, depending on the so-called intraclass correlation (*ICC*), which is the ratio of outcome variance between clusters to total outcome variance, and on the sample size per cluster.⁷ It is thus important to design a cluster randomized trial efficiently within constraints like feasibility and the budget for sampling, treating and measuring clusters and persons. In particular, the sample size at each design level (cluster, person) must be cost-effective, that is, the power to detect a treatment effect of interest must be maximized at minimum costs. Equations for the optimal sample size in a cluster randomized trial, taking into account outcome variance and costs at each level (cluster, person), have been published.^{8,9} These equations are based on 3 restrictive assumptions, however: (1) homogeneity of outcome variance and costs between treatment groups, at the cluster level and at the individual level, (2) a known *ICC*, and (3) an equal sample size per cluster.

The assumption of an equal sample size per cluster was relaxed by showing that the increased sampling variance of the treatment effect arising from cluster size variation, can be approximated by a simple function of the coefficient of variation (*CV*) of cluster size, and that this increase can often be compensated by sampling 10% to 20% more clusters.^{10,11}

The assumption of a known *ICC* was recently dropped by the derivation of maximin designs (MMDs) which are robust against misspecification of the *ICC*.¹² Two other methods to handle the unknown *ICC* are the Bayesian approach and group sequential trials.¹³⁻¹⁵ This leaves the assumption of homogeneous costs and variances to be relaxed, and there are good reasons for relaxing that assumption.

With respect to costs, the growing literature on cost-effectiveness cluster randomized trials¹⁶ testifies to the heterogeneity of costs between treatments. Some of the costs for running a trial may also differ between treatment arms, for instance if one of the treatments is so new that the staff of the participating clusters needs to be trained into the proper use of that treatment. With respect to outcome variance, substantial heterogeneity of variance, with variance ratios as high as 8 or even 12, was found in some studies in clinical psychology.¹⁷ For cluster randomized trials in medical and health care, no similar review is known to us. Going through all cluster randomized trials in a published review,¹⁸ and restricting ourselves to quantitative outcomes, we found only a handful of trials reporting the outcome variance or SD per arm, and these suggest at least some heterogeneity. For instance, a control-to-treatment variance ratio of 3 was reported for the outcome "confidence" in a trial on the effects of a training of general practitioners and nurses on patient-centered counseling and material use for Type 2 diabetes patients.¹⁹ Similarly, a variance ratio of 1.7 was found for systolic blood pressure in a comparison between computer based clinical decision support and usual care for management of hypertension in primary care.²⁰ Both variance ratios are significantly different from one according to the classical *F*-test (no adjustment for clustering could be made because both publications only report the total variance and not the *ICC*). Further, significant heterogeneity of variance, with a variance ratio of 1.35 for the outcome patient satisfaction, was reported in a trial comparing nurse practitioners with general practitioners.²¹

For binary outcomes and count data, heterogeneity of variance is even bound to occur if treatments differ in mean outcome. The present paper focusses on quantitative outcomes, but its general approach can also be applied to other outcomes. For example, all equations in the next section have an equivalent for binary outcomes.²²

Going back to the literature on optimal design and sample sizes, the homogeneity of variance assumption was relaxed for randomized experiments without clustering by several authors,²³⁻²⁵ and for experiments with clustering in 1 treatment arm by Moerbeek and Wong.²⁶ Heterogeneity of costs per cluster in a cluster randomized trial was considered by Liu,²⁷ assuming homogeneous variances and sample size per cluster. Allowing heterogeneous costs and heterogeneous variances, Candel and Van Breukelen²⁸ derived the optimal number of clusters per treatment arm, given a fixed

instead of optimal cluster size, which was allowed to differ between treatment arms, however. Lemme et al²⁹ derived the optimal number of clusters as well as the optimal cluster size per treatment arm, but application of their result requires the outcome variance per treatment arm per design level to be known in the design stage. Finally, Wu, Wong and Crespi³⁰ derived the optimal treatment to control allocation ratio for a cluster randomized trial with a binary outcome and heterogeneous costs and ICCs, given a fixed instead of optimal sample size per cluster, which was restricted to be the same for both treatment arms. Candel and Van Breukelen as well as Wu et al also discussed MMDs to handle dependence of the optimal design on unknown parameters.

This paper extends the literature in 3 ways. First, both the number of clusters and the cluster size per treatment arm of a cluster randomized trial are optimized as a function of treatment-dependent costs and treatment-dependent variances at each design level, cluster and individual. In this respect it extends^{27,28,30}. Secondly, the problem that the variances are unknown in the design stage is addressed by MMD, as in Candel and Van Breukelen²⁸ and Wu et al³⁰, which only requires the user to specify a maximum for the *ICC* and a range for the treatment-to-control outcome variance ratio. Within these ranges we derive robust and efficient designs. Third and last, we compare these designs not only with the commonly used balanced design, which is optimal under homogeneity of variances and costs, but also with the design that is optimal for heterogeneous costs but homogeneous variances. The latter design is a practical alternative to the balanced design because costs, unlike variances, can be known in the design stage.

The outline of this paper is as follows. First, the mixed model for analyzing a cluster randomized trial is specified and the optimal sample size per level is given as a function of the *ICC* and the cluster-to-person cost ratio, for the case of homogeneous variances and costs (Section 2) and for the heterogeneous case (Section 3). Because costs can be known in the design stage while variances cannot in general, we then derive the MMD (Section 4). This is the design with maximum efficiency (minimum sampling variance) of the treatment effect in the worst case scenario, that is, in the case where the treatment-to-control variance ratio is such that the efficiency is at its minimum (ie, that the sampling variance is at its maximum), hence the name MMD. We compare the minimum efficiency of this design with that of the balanced design, and with that of the optimal design for heterogeneous costs and homogeneous variances. The comparison is made as a function of the range for the treatment-to-control variance ratio, and of the treatment-to-control cost ratio. Throughout the paper, we assume an equal sample size per cluster within treatment arms, but not between arms. Cluster size variation within arms can be handled as in^{10,11,31} which often means increasing the number of clusters per treatment arm with 10% or so. Furthermore, it is assumed that the data will be analysed with a statistical model that takes into account the nesting of subjects in clusters and that allows for heterogeneity of variance at each design level. To help the reader keep track of all mathematical symbols in this paper, Appendix A lists all symbols, with their meaning and the section where they are first used.

2 | OPTIMAL DESIGN UNDER HOMOGENEOUS COSTS AND VARIANCES

Let us start with the statistical model for the treatment effect on a quantitative outcome Y in a cluster randomized trial:

$$Y_{ij} = \beta_{0j} + \beta_1 X_j + \varepsilon_{ij}, \quad \text{where } \beta_{0j} = \beta_0 + u_{0j}, \quad \text{giving:}$$

$$Y_{ij} = \beta_0 + \beta_1 X_j + u_{0j} + \varepsilon_{ij}. \quad (1)$$

Here, Y_{ij} is the outcome for subject i in cluster j , and X_j is the treatment assigned to cluster j ($1 = \text{treatment}$, $0 = \text{control}$). Parameter β_1 is the mean outcome difference between treated and control, that is, the treatment effect. The intercept β_{0j} is the sum of a fixed mean intercept β_0 plus a random cluster effect u_{0j} which is assumed to have a normal distribution with mean zero and variance $\sigma_{u_0}^2$. Finally, ε_{ij} is the sum of subject and measurement error effects, with mean zero and variance σ_ε^2 within each cluster. The outcomes of any 2 subjects i and i' in the same cluster j are correlated due to the shared cluster effect u_{0j} , and this intraclass or intracluster correlation (*ICC*) is:

$$\rho = \frac{\sigma_{u_0}^2}{\sigma_y^2} = \frac{\sigma_{u_0}^2}{\sigma_{u_0}^2 + \sigma_\varepsilon^2}, \quad (2)$$

where σ_y^2 is the total outcome variance within conditions.

Because the aim of a cluster randomized trial is to estimate the treatment effect β_1 as precisely as possible and to have a maximum power for testing this effect, a logical criterion for the design efficiency is the sampling variance of the treatment effect estimator, $Var(\hat{\beta}_1)$, which for model 1 and 50:50 treatment allocation is²²:

$$\text{Var}(\hat{\beta}_1) = [(n - 1)\rho + 1] \times \frac{4 \sigma_y^2}{n K}. \quad (3)$$

Here, n is the sample size per cluster, K is the total number of clusters sampled, and so nK is the total number of subjects sampled. The factor $[(n - 1)\rho + 1]$ is the design effect (DE) and indicates the sample size needed for a cluster randomized trial relative to that for a classical trial with individual randomization. If ρ is zero, there is no outcome variation between clusters, and cluster randomization is as efficient as individual randomization, as reflected by a DE of one. In practice, the ICC is usually between 0.01 and 0.10 or 0.25, depending on the field of application,^{18,32,33} and so the DE is larger than one and increases with the sample size per cluster, n . For large n , $\text{Var}(\hat{\beta}_1)$ approaches $\frac{4 \rho \sigma_y^2}{K}$, which depends on the variance and sample size at the cluster level only.

We can increase the power of the trial by increasing n or K , and the latter is more effective because increasing n also increases the DE , thus undoing part of the power gain. Unfortunately, increasing the sample size not only increases the study power, but also the total study cost (money, time and effort), and the cost per cluster may be larger than the cost per subject, raising the question of the optimal sample size per level (cluster, person). To answer this question, we need to combine the optimality criterion in Equation 3 with a cost function relating study costs to sample size. Following,^{8,9} we assume a fixed cost c per cluster and cost s per subject, where c and s are in the same currency (eg, US\$ or Euro). The total budget B needed for K clusters of n subjects each, ignoring all study costs that do not depend on the sample size, then is:

$$B = cK + snK = K(c + sn), \quad (4)$$

where $(c + sn)$ is the total cost per cluster including the cost per subject within the cluster. The optimal design now minimizes $\text{Var}(\hat{\beta}_1)$ in Equation 3 as a function of n and K , given the constraint in Equation 4. Assuming $\rho \in (0, 1)$, this gives^{8,9}:

$$n^* = \sqrt{\left(\frac{1-\rho}{\rho}\right)\left(\frac{c}{s}\right)}, \quad K^* = \frac{B}{(c + sn^*)}. \quad (5)$$

where the superscript $*$ indicates the optimal design. So, the optimal sample size per cluster, n^* , increases with the cluster-to-subject cost ratio c/s , and with the subject-to-cluster variance ratio $\left(\frac{1-\rho}{\rho}\right)$. The optimal number of clusters, K^* , then follows from the budget constraint in Equation 4. Note that the total budget B only affects the optimal number of clusters K^* , but not the optimal sample size per cluster n^* . So, if the budget is increased to gain power, then the extra budget must be spent on more clusters rather than on more subjects per cluster. Further, the optimal design in Equation 5 clearly depends on the ICC , which is usually unknown in the design stage. So, the design is locally optimal only, that is, optimal for 1 specific value of the unknown ICC . Combining Equations 3 and 5 gives $\text{Var}^*(\hat{\beta}_1)$ for the optimal design, which is the smallest possible $\text{Var}(\hat{\beta}_1)$, yielding the largest possible power and precision, given our budget, costs, and outcome variance and ICC :

$$\text{Var}^*(\hat{\beta}_1) = \frac{4g(\rho)\sigma_y^2}{B}, \quad \text{where: } g(\rho) = \left(\sqrt{\rho c} + \sqrt{(1-\rho)s}\right)^2. \quad (6)$$

So the smallest possible sampling variance of the treatment effect estimator increases with the costs c and s , in a way which depends on the ICC . Further, $g(\rho)$ increases with the ICC to a maximum $g(\rho) = c + s$, attained if $\rho = \frac{c}{c+s}$. This last case gives a sample size $n^* = 1$ per cluster by Equation 5, so that the design is no longer nested and $\text{Var}(\hat{\beta}_1)$ no longer depends on ρ , see Equation 3. So, Equations 5 and 6 only apply if $0 < \rho < \frac{c}{c+s}$, which is a reasonable assumption given that the ICC is typically between 0.01 and 0.25,^{18,32,33} and that $c > s$ will almost always hold. If $\text{Var}^*(\hat{\beta}_1)$ in Equation 6 is still too large for sufficient power and precision, the only solution is to increase the budget B , and spend the extra money on sampling more clusters, see Equation 5. A simple equation for computing the number of clusters needed for a given effect size d (defined as $d = \frac{\beta_1}{\sigma_y}$, analogous to Cohen's d for non-nested trials,³⁴) power, type I error rate α , cost ratio c/s , and ICC , is given in Van Breukelen and Candel.⁷

3 | OPTIMAL DESIGN UNDER HETEROGENEOUS COSTS AND VARIANCES

Equations 1 to 6 assume that the outcome variances and the costs are homogeneous between treatment arms. The assumption of homogeneous costs is unrealistic for expensive or time-consuming treatments and was relaxed by Liu.²⁷ The assumption of homogeneous variance is not realistic either, because a treatment which affects the outcome mean will also affect its variance, unless the treatment effect is the same for all clusters and all subjects. The optimal design results thus need to be generalized to heterogeneous costs and variances. In terms of Equation 1 this means that σ_{u0}^2 and σ_{ϵ}^2 and thus also the total outcome variance σ_y^2 and the ICC ρ can differ between treatment arms. In terms of Equation 4, the costs c and s can likewise differ between arms.

Now, the treatment effect estimator $\hat{\beta}_1$ is the difference between 2 independent sample means: that of the treated clusters, and that of the control clusters.²⁷ So $Var(\hat{\beta}_1)$ is under heterogeneity of variance equal to:

$$Var(\hat{\beta}_1) = [(n_t - 1)\rho_t + 1] \frac{\sigma_{Yt}^2}{n_t K_t} + [(n_c - 1)\rho_c + 1] \frac{\sigma_{Yc}^2}{n_c K_c}, \tag{7a}$$

for any design (n_t, K_t, n_c, K_c) . Here, ρ_t is the ICC, σ_{Yt}^2 is the total outcome variance, n_t is the sample size per cluster, and K_t is the number of clusters, in the treated arm, and likewise $\rho_c, \sigma_{Yc}^2, n_c$ and K_c in the control arm. The minimum of $Var(\hat{\beta}_1)$ is found in 2 steps: First, assume a given budget B_t for the treated arm and B_c for the control arm, and minimize $Var(\hat{\beta}_1)$ in Equation 7a subject to the same budget constraint as in Equation 4, but now per treatment arm. Secondly, given this minimum $Var(\hat{\beta}_1)$ as a function of the budgets B_t and B_c , let B denote the total budget for both treatment arms and let f denote the fraction of that budget which is spent on the treated arm, so $B_t = fB$ and $B_c = (1 - f)B$. Then, find that f which minimizes $Var(\hat{\beta}_1)$ as obtained in step 1. Below these 2 steps are elaborated. Step 1 gives as optimal design for the treated arm respectively the control arm:

$$n_t^* = \sqrt{\left(\frac{1-\rho_t}{\rho_t}\right)\left(\frac{c_t}{s_t}\right)}, \quad K_t = \frac{B_t}{(c_t + s_t n_t^*)}, \tag{7b}$$

$$n_c^* = \sqrt{\left(\frac{1-\rho_c}{\rho_c}\right)\left(\frac{c_c}{s_c}\right)}, \quad K_c = \frac{B_c}{(c_c + s_c n_c^*)}, \tag{7c}$$

Here, c_t is the cost per cluster, and s_t is the cost per subject, in the treated arm, and likewise c_c and s_c in the control arm. Note that the optimal sample size per treated cluster, n_t^* , does not depend on the budget for the treated arm, B_t , and n_c^* does not depend on B_c . Only the numbers of clusters per arm, K_t and K_c , depend on the budgets and are not optimal yet. Substituting Equations 7b and 7c into Equation 7a gives upon rewriting:

$$Var(\hat{\beta}_1) = \frac{g_t(\rho_t)\sigma_{Yt}^2}{B_t} + \frac{g_c(\rho_c)\sigma_{Yc}^2}{B_c}, \tag{7d}$$

for the design (n_t^*, K_t, n_c^*, K_c) as specified by Equations 7b and 7c, with $g_t(\rho_t) = (\sqrt{\rho_t c_t} + \sqrt{(1-\rho_t) s_t})^2 \in [\text{Min}(s_t, c_t), \text{Sum}(s_t, c_t)]$ and $g_c(\rho_c) = (\sqrt{\rho_c c_c} + \sqrt{(1-\rho_c) s_c})^2 \in [\text{Min}(s_c, c_c), \text{Sum}(s_c, c_c)]$.

Note that Equation 7a reduces to Equation 3 if $\sigma_{Yt}^2 = \sigma_{Yc}^2, \rho_t = \rho_c, n_t = n_c$ and $K_t = K_c = K/2$ hold, and Equation 7d reduces to Equation 6 if additionally $c_t = c_c$ and $s_t = s_c$ and $B_t = B_c$ hold.

In step 2 of the design optimization, substitute in Equation 7d $B_t = fB$ and $B_c = (1 - f)B$, where f is the fraction of the budget B that is spent on the treatment arm, and then minimize Equation 7d as a function of f , given B . This gives as optimal split of the budget between the treated and control arm:

$$\frac{f^*}{1-f^*} = \frac{\sigma_{Y_t} \sqrt{g_t(\rho_t)}}{\sigma_{Y_c} \sqrt{g_c(\rho_c)}} \quad (8)$$

The optimal number of treated clusters, K_t^* , then follows from Equation 7b by letting $B_t = f^* B$, and likewise the optimal number of control clusters, K_c^* follows from Equation 7c by letting $B_c = (1 - f^*)B$. Combining Equation 7d with Equation 8 results in the following minimum variance of the treatment effect estimator under heterogeneity:

$$\text{Var}^* \left(\hat{\beta}_1 \right) = \left(\sigma_{Y_t} \sqrt{g_t(\rho_t)} + \sigma_{Y_c} \sqrt{g_c(\rho_c)} \right)^2 / B, \quad (9)$$

which reduces to Equation 6 if the costs and the variances are homogeneous. In the latter case, the optimal design is balanced, that is, $K_t^* = K_c^*$ and $n_t^* = n_c^*$.

Another special case is that of homogeneous variances but heterogeneous costs, giving $\frac{f^*}{1-f^*} = \frac{\sqrt{g_t(\rho)}}{\sqrt{g_c(\rho)}}$. We refer to this as the cost-considered (cc) design. The optimal design of Liu²⁷ is a special case of the cc design in that Liu assumes $n_t = n_c$. This is optimal only if $\frac{c_t}{s_t} = \frac{c_c}{s_c}$, see Equations 7b and 7c. Given homogeneity of variances and $\frac{c_t}{s_t} = \frac{c_c}{s_c}$, implying $n_t^* = n_c^*$, the optimal budget allocation ratio is $\frac{f^*}{1-f^*} = \sqrt{\frac{c_t}{c_c}}$, and the optimal cluster allocation ratio is $\frac{K_t^*}{K_c^*} = \sqrt{\frac{c_c}{c_t}}$, which depends on the costs per cluster only. However, it can be shown that this cluster allocation ratio is also optimal if $\frac{c_t}{s_t} \neq \frac{c_c}{s_c}$, provided the variances are homogeneous and n_t, n_c are optimized with Equations 7b and 7c instead of letting $n_t = n_c$.

4 | SOLVING THE PROBLEM OF LOCAL OPTIMALITY: MAXIMIN DESIGN

As Equations 7b, 7c, and 8 show, the optimal budget split between the treated and control arms, and the optimal sample size per design level (cluster, subject), depend on the costs and variances per design level, per arm, making the optimal design locally optimal only. But while the costs can be known in the design stage, the variances cannot, and so the optimal design is vulnerable to misspecification of the variances. A solution is the MMD. Maximin design consists of 4 steps, and each is elaborated after this list:

1. Specify the parameter space, that is, the region containing all plausible values of the unknown parameter vector $(\sigma_{Y_t}^2, \rho_t, \sigma_{Y_c}^2, \rho_c)$ on which the optimal design depends;
2. Specify the design space, that is, the set of all candidate designs (n_t, K_t, n_c, K_c) ;
3. For each design in the design space find its worst case, here: the maximum of $\text{Var} \left(\hat{\beta}_1 \right)$, denoted as $\max \text{Var} \left(\hat{\beta}_1 \right)$, as a function of the unknown parameter vector $(\sigma_{Y_t}^2, \rho_t, \sigma_{Y_c}^2, \rho_c)$;
4. Finally, select the design with the smallest $\max \text{Var} \left(\hat{\beta}_1 \right)$, denoted as $\min \max \text{Var} \left(\hat{\beta}_1 \right)$, or equivalently, the design with the maximum minimum efficiency. This is the MMD, denoted as $(n_t^m, K_t^m, n_c^m, K_c^m)$, which is robust against misspecification of the unknown parameters in that it optimizes the worst case.

Concerning step 1, some constraints are needed on the parameter space, because Equation 7a implies that $\text{Var} \left(\hat{\beta}_1 \right)$ is an increasing function of the variance in each treatment arm for any given design (n_t, K_t, n_c, K_c) , and Equation 9 implies the same for $\text{Var}^* \left(\hat{\beta}_1 \right)$ of the optimal design. A finite maximum is therefore imposed on the (unknown) sum of the outcome variances, $\sigma_{Y_t}^2 + \sigma_{Y_c}^2 \leq V_{\max}$, to prevent an infinitely large $\text{Var} \left(\hat{\beta}_1 \right)$ as the worst case in step 3. Further, assume a bounded SD ratio $\frac{\sigma_{Y_t}}{\sigma_{Y_c}} \in \left[\frac{1}{u}, u \right]$, where u ($u \geq 1$) is based upon prior knowledge and determines the possible amount of heterogeneity of variance, with $u = 1$ giving homogeneous variance, and $u \rightarrow \infty$ allowing extremely heterogeneous variance. This section focusses on 2-sided intervals for the SD ratio, but Section 6 covers 1-sided intervals. Combining $\frac{\sigma_{Y_t}}{\sigma_{Y_c}} \in \left[\frac{1}{u}, u \right]$ with $\sigma_{Y_t}^2 + \sigma_{Y_c}^2 \leq V_{\max}$ gives $\sigma_{Y_t}^2 \in \left[\frac{V_{\max}}{1+u^2}, \frac{u^2 V_{\max}}{1+u^2} \right]$ and likewise for $\sigma_{Y_c}^2$. Finally, in step 1 we assume a common maximum possible value ρ_{\max} for the unknown ICC in each arm, but allow the actual ICC to differ

between arms. This ρ_{\max} may depend on the field of application, for instance 0.10 in primary care,^{18,32} or 0.25 in education,³³ and is assumed to satisfy $0 < \rho_{\max} < \min\left(\frac{c_t}{c_t + s_t}, \frac{c_c}{c_c + s_c}\right)$ for reasons given in Section 2 below Equation 6.

Concerning step 2, restrict the design space to all designs satisfying the following constraints: $n_t \geq 1, K_t \geq 1, n_c \geq 1, K_c \geq 1$ and $K_t(c_t + n_t s_t) + K_c(c_c + n_c s_c) = B$, where B is the total study budget for all costs that depend on the sample size. Without the budget constraint, step 4 would give an infinitely large sample size as the best design.

Concerning step 3, note that $Var(\hat{\beta}_1)$ is an increasing function of the ICC and of the variance in each arm, see Equations 7a and 7d and the text below Equation 6. Therefore, given any design (n_t, K_t, n_c, K_c) , $Var(\hat{\beta}_1)$ is maximized (worst case) by letting $\rho = \rho_{\max}$ and $\sigma_{Y_t}^2 + \sigma_{Y_c}^2 = V_{\max}$ in Equation 7a or 9;

Finally, concerning step 4, note that for any given parameter vector $(\sigma_{Y_t}^2, \rho_t, \sigma_{Y_c}^2, \rho_c)$, including those from step 3, $Var(\hat{\beta}_1)$ is minimized by taking the optimal design. So the MMD can be derived from Equation 7d instead of Equation 7a. To begin with, this gives as maximin sample size per cluster $n_t^m = \sqrt{\left(\frac{c_t}{s_t}\right)\left(\frac{1-\rho_{\max}}{\rho_{\max}}\right)}$ in the treated arm,

and likewise $n_c^m = \sqrt{\left(\frac{c_c}{s_c}\right)\left(\frac{1-\rho_{\max}}{\rho_{\max}}\right)}$ in the control arm, and as worst case the following sampling variance for the treatment effect:

$$\max Var(\hat{\beta}_1) = \frac{g_t(\rho_{\max})\sigma_{Y_t}^2}{fB} + \frac{g_c(\rho_{\max})(V_{\max}-\sigma_{Y_c}^2)}{(1-f)B} = \frac{g_c(\rho_{\max})}{B} \left(\frac{p^2\sigma_{Y_t}^2}{f} + \frac{V_{\max}-\sigma_{Y_t}^2}{1-f} \right) \tag{10}$$

where $p = \sqrt{\frac{g_t(\rho_{\max})}{g_c(\rho_{\max})}}$, which is a function of the costs c and s in each treatment arm, and of ρ_{\max} . As a special case,

$$\frac{c_t}{s_t} = \frac{c_c}{s_c} \text{ gives } p = \sqrt{\frac{c_t}{c_c}} = \sqrt{\frac{s_t}{s_c}}$$

It is important to note that, once the costs and ρ_{\max} and V_{\max} have been specified by the user, $\max Var(\hat{\beta}_1)$ in Equation 10 depends on 2 unknowns only: the total outcome variance in the treatment arm, $\sigma_{Y_t}^2$, and the fraction of the budget B spent on the treatment arm, f . So we can first maximize Equation 10 as a function of $\sigma_{Y_t}^2$ within its constraint

$\sigma_{Y_t}^2 \in \left[\frac{V_{\max}}{1+u^2}, \frac{u^2 V_{\max}}{1+u^2} \right]$, and then minimize that maximum as a function of the budget split. This gives the maximin budget ratio $\frac{f^m}{1-f^m}$, and the corresponding $\min \max Var(\hat{\beta}_1)$ shown in Table 1 (for detailed derivations, see Appendix B).

The MMD $(n_t^m, K_t^m, n_c^m, K_c^m)$ itself is not given in Table 1, but is obtained by filling in Equations 7b and 7c, using the budget per arm as given in Table 1, and the known costs for that arm, and the maximum ICC, $\rho = \rho_{\max}$.

As Table 1 shows, the maximin budget ratio depends on whether p is within the range $\left[\frac{1}{u}, u \right]$ for the SD ratio or not. If p is within this range, the budget ratio and $\min \max Var(\hat{\beta}_1)$ do not depend on that range. If $p > u$, then treatment is more expensive than control ($p > 1$) and costs are more heterogeneous than variances. Consequently, more

TABLE 1 Maximin budget split and maximum variance of the treatment effect for heterogeneous costs and variances given a fixed maximum total variance V_{\max} and a fixed total budget B , as a function of the relation between u (square root of maximum treated-to-control variance ratio) and p (square root of treated-to-control cost ratio)

Relation of p to u	Budget Split $\frac{f^m}{1-f^m}$	Maximum $Var(\hat{\beta}_1)$ for the Maximin Design	Equation nr for Referencing
$\frac{1}{u} \leq p \leq u$	p^2	$\frac{g_c(\rho_{\max}) V_{\max}}{B} \times (1+p^2)$	(11)
$p > u$	pu	$\frac{g_c(\rho_{\max}) V_{\max}}{B} \times \frac{(pu+1)^2}{(u^2+1)}$	(12)
$p < \frac{1}{u}$	$\frac{p}{u}$	$\frac{g_c(\rho_{\max}) V_{\max}}{B} \times \frac{(p+u)^2}{(u^2+1)}$	(13)

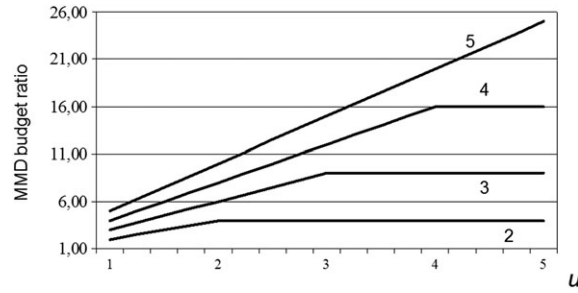


FIGURE 1 Maximin budget ratio treatment-to-control, based on the minimum efficiency ($\max\text{Var}(\hat{\beta}_1)$) criterion, as a function of the range $\left[\frac{1}{u}, u\right]$ for the treatment-to-control SD ratio, for $p = 2, 3, 4, 5$ ($p = \sqrt{\text{of treatment-to-control cost ratio if } c/s \text{ is homogeneous}}$)

budget is allocated to treatment than to control ($pu > 1$). Likewise, $p < \frac{1}{u}$ implies that treatment is cheaper than control ($p < 1$) and that costs are more heterogeneous than variances, resulting in a higher budget for control ($\frac{p}{u} < 1$). Figure 1 plots the maximin budget ratio $\frac{f^m}{1-f^m}$ against u from 1 to 5 (implying a range from 0.04 to 25 for the variance ratio), for p from 2 to 5 (implying a cost ratio from 4 to 25 in the case $\frac{c_t}{s_t} = \frac{c_c}{s_c}$). If $p < 1$ (treatment cheaper than control), then Figure 1 applies in that it then shows the control-to-treatment budget ratio.

5 | COMPARISON WITH THE BALANCED DESIGN AND THE COST-CONSIDERED DESIGN

This section compares the MMD with the balanced design (which is optimal for homogeneous costs and homogeneous variances) and the cc design (which is optimal for homogeneous variances but heterogeneous costs). It does so by considering the relative efficiencies of all designs, as defined by the ratios of their $\max\text{Var}(\hat{\beta}_1)$'s.

5.1 | Relative efficiency of balanced design versus MMD

Consider first the case $\frac{c_t}{s_t} = \frac{c_c}{s_c}$ (homogeneous cluster-to-subject cost ratio), so that $n_t = n_c$ holds for both designs (see Equations 7b and 7c with $\rho = \rho_{\max}$), and $p^2 = \frac{g_t(\rho_{\max})}{g_c(\rho_{\max})} = \frac{c_t}{c_c} = \frac{s_t}{s_c}$, the treatment-to-control cost ratio. Balanced allocation $K_t = K_c$ then implies $\frac{f^{bal}}{1-f^{bal}} = p^2$, that is, the budget allocation ratio equals the cost ratio. This is the MMD if $p \in \left[\frac{1}{u}, u\right]$, that is, if costs are more homogeneous than variances, see Equation 11 in Table 1.

If $p > u$, the relative efficiency (RE) of the balanced design versus the MMD is the ratio of $\max\text{Var}(\hat{\beta}_1)$ in Equation 12 to that in Equation 11 in Table 1. If $p < \frac{1}{u}$, the RE is the ratio of $\max\text{Var}(\hat{\beta}_1)$ in Equation 13 to that in

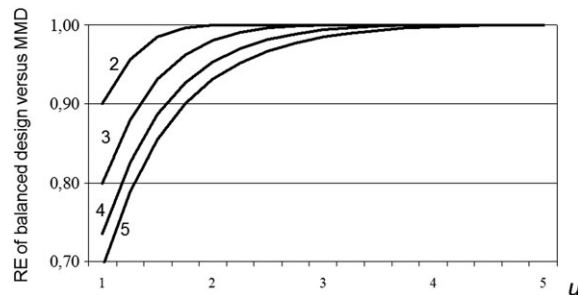


FIGURE 2 Relative efficiency of the balanced design versus the MMD as a function of the range $\left[\frac{1}{u}, u\right]$ for the unknown SD ratio, for various p (assuming homogeneity of c/s)

Equation 11. In both cases, the RE increases in u , and $RE > 0.95$ if $p \in \left[\frac{1}{u^2}, u^2 \right]$. Far outside these boundaries, the balanced design can be inefficient, with a minimum of $RE \rightarrow 0.50$ for $u \rightarrow 1$ and $p \rightarrow \infty$ (homogeneous variances, expensive treatment), and for $u \rightarrow 1$ and $p \rightarrow 0$ (homogeneous variances, expensive control). To show this, Figure 2 plots the RE of the balanced design versus MMD against u for various p .

For $\frac{c_t}{s_t} \neq \frac{c_c}{s_c}$, the RE of the balanced design is obtained as follows: First, compute the design with Equation 5, assuming $\rho = \rho_{\max}$ and $c = \frac{c_t + c_c}{2}$ and $s = \frac{s_t + s_c}{2}$. Then, compute $\max Var(\hat{\beta}_1)$ for this design using Equation 3, assuming $\rho = \rho_{\max}$ and letting $\sigma_{Yt}^2 = \frac{\sigma_{Yt}^2 + \sigma_{Yc}^2}{2}$.

5.2 | Relative efficiency of cost-considered design versus MMD

Because the balanced design is inefficient for very heterogeneous costs, and costs can be known in the design stage, an alternative to the balanced design is the cc design, which is optimal for heterogeneous costs ($p \neq 1$) and homogeneous variance ($u = 1$). Following Equation 8, the cc design has the budget allocation ratio $\frac{f^{cc}}{1-f^{cc}} = p$ (see the definition of p below Equation 10). This is the limiting case of the MMD for $u \rightarrow 1$ (homogeneous variances). If $\frac{c_t}{s_t} = \frac{c_c}{s_c}$, we get $p = \sqrt{\frac{c_t}{c_c}} = \sqrt{\frac{s_t}{s_c}}$. However, the RE of the cc design versus the MMD as discussed below is not restricted to this special case.

Using $\frac{f^{cc}}{1-f^{cc}} = p$, with $p = \sqrt{\frac{g_t(\rho_{\max})}{g_c(\rho_{\max})}}$, and using Equation 10, we get $\max Var(\hat{\beta}_1)$ for the cc design if the variance is actually heterogeneous:

$$\max Var(\hat{\beta}_1) = \frac{g_c(\rho_{\max})(1+p)}{B} \times (V_{\max} + (p-1)\sigma_{Yt}^2). \tag{14}$$

If $p = 1$, Equation 14 does not depend on σ_{Yt}^2 and gives the same result as Equation 11 in Table 1. But the design is balanced only if $\frac{c_t}{s_t} = \frac{c_c}{s_c}$ which, for $p = 1$, implies homogeneity of costs.

If $p > 1$ (treatment more expensive than control), then Equation 14 is maximized by maximizing σ_{Yt}^2 within the constraint $\sigma_{Yt}^2 \in \left[\frac{V_{\max}}{1+u^2}, \frac{u^2 V_{\max}}{1+u^2} \right]$ of Section 4, giving $\sigma_{Yt}^2 = \left(\frac{u^2 V_{\max}}{1+u^2} \right)$, and the resulting worst case, $\max Var(\hat{\beta}_1)$, is shown in Equation 15 in Table 2.

If $p < 1$ (treatment cheaper than control), then Equation 14 is maximized by minimizing σ_{Yt}^2 within the same constraint, giving $\sigma_{Yt}^2 = \left(\frac{V_{\max}}{1+u^2} \right)$ and as worst case Equation 16 in Table 2. Note that, under homogeneity of variances ($u = 1$), Table 1 and 2 give the same results.

TABLE 2 Cost-considered budget split and maximum variance of the treatment effect for heterogeneous costs and variances given a fixed maximum total variance V_{\max} and fixed total budget B , as a function of the range of p (square root of treated-to-control cost ratio)

Range of p ^(a)	Budget Split $\frac{f^{cc}}{1-f^{cc}}$	Maximum $Var(\hat{\beta}_1)$ for the Cost-Considered Design	Equation nr for Referencing
$p = 1$	p	$\frac{g_c(\rho_{\max})}{B} 2V_{\max}$	-
$p > 1$	p	$\frac{g_c(\rho_{\max})}{B} V_{\max} (1+p) \times \left(\frac{pu^2 + 1}{u^2 + 1} \right)$	(15)
$p < 1$	p	$\frac{g_c(\rho_{\max})}{B} V_{\max} (1+p) \times \left(\frac{p + u^2}{u^2 + 1} \right)$	(16)

^aThe cost-considered design assumes homogeneous variances and so its budget split does not depend on the SD ratio parameter u (square root of maximum variance ratio). But the maximum (worst case) $Var(\hat{\beta}_1)$ of the cost-considered design does depend on the actual heterogeneity of variance as expressed by u , except if $p = 1$.

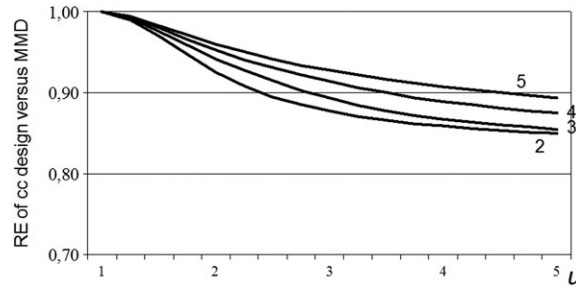


FIGURE 3 Relative efficiency of the cost-considered versus the MMD as a function of the range $\left[\frac{1}{u}, u\right]$ for the unknown SD ratio, for various p (without assuming homogeneity of c/s)

The relative efficiency (RE) of the cc design versus the MMD with respect to the optimality criterion $\max Var(\hat{\beta}_1)$ now follows from Equations 11 to 16, for instance, as the ratio of $\max Var(\hat{\beta}_1)$ in Equation 11 to that in Equation 15 if $1 < p < u$. Figure 3 plots the RE of the cc design versus the MMD as a function of u for various p . The results can be summarized thus:

If $p \in \left[\frac{1}{u}, u\right]$, the RE of the cc design is a decreasing function of u , with a minimum of 0.83 if $u \rightarrow \infty$ and $p \approx 2.41$ or $p \approx 0.41$, and a maximum of 1 if $u \rightarrow p$ and ($p \rightarrow 1$ or $p \rightarrow \infty$), or if $u \rightarrow \frac{1}{p}$ and ($p \rightarrow 1$ or $p \rightarrow 0$).

If $p \notin \left[\frac{1}{u}, u\right]$, the RE of the cc design again decreases in u , but now with a minimum of 0.89 if $u \rightarrow p$ and $p \approx 3.73$, or if $u \rightarrow \frac{1}{p}$ and $p \approx 0.27$, and a maximum of 1 if either $p \rightarrow \infty$ or $p \rightarrow 0$ or $u \rightarrow 1$. So the RE of the cc design compared with the MMD design, using as criterion $\max Var(\hat{\beta}_1)$, is always above 0.80.

5.3 | Relative efficiency of balanced versus cost-considered design

Finally, if $\frac{c_t}{s_t} = \frac{c_c}{s_c}$ (homogeneous cluster-to-person cost ratio), the RE of the balanced versus the cc design follows from Equations 11, 15, and 16 in Tables 1 and 2 and is plotted in Figure 4. If $p \in \left[\frac{1}{u^2}, u^2\right]$, the balanced design is more efficient. If $p \notin \left[\frac{1}{u^2}, u^2\right]$, the cc design is more efficient. At the interval boundaries, both designs are equally efficient.

Comparing all 3 designs, MMD, balanced and cc, we get for the case $\frac{c_t}{s_t} = \frac{c_c}{s_c}$:

For $p \in \left[\frac{1}{u}, u\right]$ balanced is optimal and cc is fairly efficient, with a minimum RE of 0.83;

For $p \in [u, u^2]$ and for $p \in \left[\frac{1}{u^2}, \frac{1}{u}\right]$, balanced is more efficient than the cc design, which in turn has a minimum RE of 0.89 once p is outside the interval $\left[\frac{1}{u}, u\right]$;

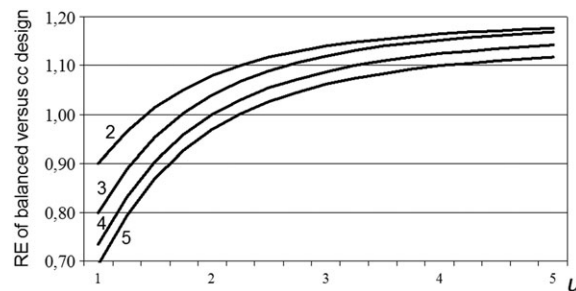


FIGURE 4 Relative efficiency of the balanced versus the cost-considered design as a function of the range $\left[\frac{1}{u}, u\right]$ for the unknown SD ratio, for various p (assuming homogeneity of c/s)

For p outside the interval $\left[\frac{1}{u^2}, u^2\right]$, the cc design is more efficient than the balanced design, and highly efficient due to its minimum RE of 0.89. The balanced design becomes inefficient if $u \rightarrow 1$ while p moves away from 1, for instance, $RE < 0.80$ if $u = 1$ and $p > 3$ or $p < 1/3$, meaning a homogeneous variance and a cost ratio above 9 or below 1/9.

5.4 | Relative efficiency of the MMD if the variances turn out to be homogeneous

Up to now, the variance was assumed to be heterogeneous and so the MMD was always at least as efficient as, and usually more efficient than, the balanced and the cc design, which both assume homogeneous variances. But how efficient is a MMD based on heterogeneous variance ($u > 1$ assumed) if the variance turns out to be homogeneous ($u = 1$ actually) in the analysis? Consider again the 3 cases in Equations 11 to 13.

If $\frac{1}{u} \leq p \leq u$ is assumed in the design stage, the MMD is balanced, at least for $\frac{c_t}{s_t} = \frac{c_c}{s_c}$ (Section 5.1). Its RE compared with the cc design if $u = 1$ (homogeneous variances) turns out to be the case, then follows from Section 5.3 and Figure 4 with $u = 1$;

If $p > u$ is assumed, the maximin budget ratio $\frac{f^m}{1-f^m} = pu$. Because $p > u > 1$, this budget split is in-between that for the cc design, $\frac{f^{cc}}{1-f^{cc}} = p$, and that for the balanced design, $\frac{f^{bal}}{1-f^{bal}} = p^2$. The MMD itself is obtained by filling in the budget and costs per treatment arm in Equations 7b and 7c, which depend on u only through the budget, and so the MMD is in-between the balanced design and the cc design. It then follows from Figure 4 with $u = 1$ that, if the variances are actually homogeneous, the MMD is always more efficient than the commonly used balanced design, and less efficient than the cc design. The differences between the 3 designs are small for p up to 2, and still acceptable for p up to 3, implying a treated-to-control cost ratio up to 4, respectively up to 9, in the case where $\frac{c_t}{s_t} = \frac{c_c}{s_c}$. Finally, the case $p < \frac{1}{u}$ becomes the case $p > u$ by switching the 2 treatment arms, thus giving the same RE results as the case $p > u$.

6 | MAXIMIN DESIGN FOR ONE-SIDED INTERVALS FOR THE VARIANCE RATIO

Sometimes, a 1-sided interval may perhaps be assumed for the SD ratio, for instance if we know from other studies that treatment always increases outcome variance (or always decreases it). This gives the following result (for details, see appendix B):

If $\frac{\sigma_{Y_t}}{\sigma_{Y_c}} \in [1, u]$ and $p < 1$, or if $\frac{\sigma_{Y_t}}{\sigma_{Y_c}} \in \left[\frac{1}{u}, 1\right]$ and $p > 1$, then the budget split for the MMD is $\frac{f^m}{1-f^m} = p$, which is the budget split for the cc design. The $\maxVar(\hat{\beta}_1)$ is then obtained from Table 1 or 2 by filling in $u = 1$ (see Appendix B, and note that Tables 1 and 2 give the same results if $u = 1$). In all other cases the MMD and $\maxVar(\hat{\beta}_1)$ are the same as for the 2-sided case in Table 1. So if the more expensive arm has the smaller variance,

TABLE 3 Maximin budget split $\frac{f}{1-f}$ for heterogeneous costs and variances as a function of the relation between u (maximum SD ratio) and p (square root of treated-to-control cost ratio), for 2- and 1-sided intervals for the SD ratio (note: Budget split p gives the cost-considered design; budget split p^2 gives the balanced design if $\frac{c_t}{s_t} = \frac{c_c}{s_c}$)

Relation of p to u				
Range for SD ratio	$p < \frac{1}{u}$	$\frac{1}{u} \leq p \leq 1$	$1 \leq p \leq u$	$u < p$
$\left[\frac{1}{u}, u\right]$	$\frac{p}{u}$	p^2	p^2	pu
$[1, u]$	p	p	p^2	pu
$\left[\frac{1}{u}, 1\right]$	$\frac{p}{u}$	p^2	p	p

then the MMD for 1-sided intervals for the SD ratio is the cc design. But if the more expensive arm has the larger variance, then the MMD for 1-sided intervals for the SD ratio is the same as for 2-sided intervals. The MMD for the various cases in this and the preceding sections are summarized in Table 3. It follows from this table that the maximin budget split is always between p (the cc budget split) and p^2 (the split for the balanced design if $\frac{c_t}{s_t} = \frac{c_c}{s_c}$).

7 | POWER AND SAMPLE SIZE CALCULATION

It has been shown how the MMD depends on 2 factors: $p = \sqrt{\frac{g_t(\rho_{\max})}{g_c(\rho_{\max})}}$, which is a function of the costs in each treatment arm and of the maximum ICC, and $\left[\frac{1}{u}, u\right]$, which is the range for the SD ratio $\frac{\sigma_{Yt}}{\sigma_{Yc}}$. It has also been shown how efficient the balanced and cc design are relative to the MMD with respect to their minimum efficiency (maximum sampling variance). For practical use, these results must be translated into a procedure for computing the sample size needed per treatment arm to have sufficient power to detect a treatment effect of given size. This procedure consists of the following 5 steps for the MMD (the cc design is obtained by choosing $\left[\frac{1}{u}, u\right] = [1, 1]$, and the balanced design is obtained by additionally letting $c_t = c_c, s_t = s_c$):

1. First, specify the costs c_t, s_t, c_c, s_c and choose plausible values for $\rho_{\max}, V_{\max}, \left[\frac{1}{u}, u\right]$ and the treatment effect β_1 . Instead of specifying V_{\max} and β_1 , one can also specify a smallest clinically relevant effect size $d = \frac{\beta_1}{\sqrt{V_{\max}/2}}$ (a generalization of Cohen's d to heterogeneous variances, see the end of Section 2). Because the maximin sample size per cluster depends on neither V_{\max} nor β_1 (see step 3 below), and the maximin number of clusters per treatment depends on them only through d (see steps 2 and 4 below), one can also, without loss of generality, specify d ;
2. Second, compute the sampling variance that gives the desired power level $(1 - \gamma)$, for a given type I error rate α and true treatment effect β_1 , with the following equation:

$$\text{Var}(\hat{\beta}_1) = \left(\frac{\beta_1}{z_{1-\alpha/2} + z_{1-\gamma}} \right)^2. \quad (17)$$

Here, z_p denotes the p -th percentile of the standard normal distribution. For 1-tailed testing, replace $z_{1-\alpha/2}$ with $z_{1-\alpha}$. Note that Equation 17 is also used for the classical RCT with individual randomization;

3. Third, compute the maximin sample size per treated cluster, n_t^m , with Equation 7b, using as costs c_t, s_t and as ICC $\rho = \rho_{\max}$, and the maximin sample size per control cluster, n_c^m , using as costs c_c, s_c and as ICC $\rho = \rho_{\max}$, and compute the functions $g_c(\rho_{\max})$ and $g_t(\rho_{\max})$ in Equation 7d and the ratio $p = \sqrt{\frac{g_t(\rho_{\max})}{g_c(\rho_{\max})}}$;
4. Fourth, compute the maximin budget ratio $\frac{f^m}{1-f^m}$ from p and $\left[\frac{1}{u}, u\right]$, the plausible range for the SD ratio, using Table 1. Then compute the total budget B by equating $\max \text{Var}(\hat{\beta}_1)$ as given in Table 1, with $\text{Var}(\hat{\beta}_1)$ as computed in step 2 with Equation 17, and compute the budget per treatment arm as $B_t^m = f^m B$ and $B_c^m = (1-f^m)B$;
5. Fifth and last, compute the number of clusters per treatment arm as $K_t^m = \frac{B_t^m}{(c_t + s_t n_t^m)}$, and the number of clusters per control arm as $K_c^m = \frac{B_c^m}{(c_c + s_c n_c^m)}$, following Equations 7b and 7c, and round both numbers upward to ensure sufficient power. The budget B will increase a bit by this.

This procedure is easily implemented in standard statistical software, and an implementation in SPSS and in R is available as supplementary material. Two additional remarks are needed, however. First, we do not round off the sample sizes per cluster, n_t^m and n_c^m , for 2 reasons: (1) to ensure that all further computations can use the equations in this paper without modification, and (2) because unplanned variation in sample size per cluster always occurs in practice, and so the average sample size per cluster will rarely be an integer anyway. Such unplanned variation leads to some

power loss which can be restored by increasing the number of clusters with a percentage that depends on the CV of cluster size, and an increase of the number of clusters with 10% respectively 20% is often a reasonable respectively very safe adjustment.¹⁰ This adjustment is not included into the presented syntax, because the user may prefer another percentage, depending on the expected CV. Of course the user can choose to round n_t^m and n_c^m upward after step 5 to guarantee sufficient power, and this will lead to some increase of the study budget that is needed.

Secondly, step 2 of the procedure above assumes that the test statistic for the treatment effect, $t = \hat{\beta}_1/SE$, has a standard normal distribution under the null hypothesis, and that $(\hat{\beta}_1 - \beta_1)/SE$ has a standard normal distribution under the alternative hypothesis, where $SE = \sqrt{Var(\hat{\beta}_1)}$. This ignores the fact that SE is a function of the unknown variances and must be estimated. The correct reference distribution is the Student t -distribution. This follows from the equivalence of mixed regression analysis of individual data following Equation 1 with the unpaired t -test of treatment versus control using clusters as units of analysis and cluster means as outcome.³⁵ Under heterogeneity of variance of the cluster means, the degrees of freedom for this t -test obey an expression due to Welch and Satterthwaite,^{36,37} which reaches a minimum $df = MIN(K_t - 1, K_c - 1)$ if the variance ratio goes to zero or infinity, and a maximum $df = SUM(K_t - 1, K_c - 1)$ if

TABLE 4 Maximin design $(n_t^m, K_t^m, n_c^m, K_c^m)$ and budget B^m needed to detect a treatment effect of medium size ($d = 0.50$) with a power of 90% using 2-tailed testing with $\alpha = 0.05$, as a function of the range for the SD ratio $\left[\frac{1}{u}, u\right]$, the maximum ICC ρ_{max} , and the study costs per cluster and per subject in the treated arm c_t, s_t , and control arm c_c, s_c

Input Parameters		Maximin Budget Split, Maximin Design, Total Budget (a)									
$\left[\frac{1}{u}, u\right]$	ρ_{max}	c_t, s_t	c_c, s_c	p	$\frac{f^m}{1-f^m}$	n_t^m	n_c^m	K_t^m	K_c^m	B^m	
[1,1]	0.10	200,10	200,10	1.00	1.00	13.42	13.42	14.04	14.04	11361.58	
		360,10	40,10	1.80	1.80	18.00	6.00	9.81	29.42	9680.00	
		200,18	200,2	1.46	1.46	10.00	30.00	13.45	13.45	10240.00	
		360,18	40,2	3.00	3.00	13.42	13.42	9.36	28.09	9289.76	
	0.20	200,10	200,10	1.00	1.00	8.94	8.94	24.33	24.33	15629.91	
		360,10	40,10	2.00	2.00	12.00	4.00	16.81	50.44	13360.00	
		200,18	200,2	1.33	1.33	6.67	20.00	23.54	23.54	14560.00	
		360,18	40,2	3.00	3.00	8.94	8.94	16.22	48.66	12851.26	
	[0.50,2]	0.10	200,10	200,10	1.00	1.00	13.42	13.42	14.04	14.04	11361.58
			360,10	40,10	1.80	3.24	18.00	6.00	12.61	21.01	10500.00
			200,18	200,2	1.46	2.14	10.00	30.00	15.97	10.93	10220.00
			360,18	40,2	3.00	6.00	13.42	13.42	13.11	19.66	11094.25
0.20		200,10	200,10	1.00	1.00	8.94	8.94	24.33	24.33	15629.91	
		360,10	40,10	2.00	4.00	12.00	4.00	22.42	33.62	14880.00	
		200,18	200,2	1.33	1.78	6.67	20.00	26.90	20.17	14800.00	
		360,18	40,2	3.00	6.00	8.94	8.94	22.71	34.06	15166.80	
[0.33,3]		0.10	200,10	200,10	1.00	1.00	13.42	13.42	14.04	14.04	11361.58
			360,10	40,10	1.80	3.24	18.00	6.00	12.61	21.01	10500.00
			200,18	200,2	1.46	2.14	10.00	30.00	15.97	10.93	10220.00
			360,18	40,2	3.00	9.00	13.42	13.42	14.04	14.04	11361.58
	0.20	200,10	200,10	1.00	1.00	8.94	8.94	24.33	24.33	15629.91	
		360,10	40,10	2.00	4.00	12.00	4.00	22.42	33.62	14880.00	
		200,18	200,2	1.33	1.78	6.67	20.00	26.90	20.17	14800.00	
		360,18	40,2	3.00	9.00	8.94	8.94	24.33	24.33	15629.91	

^aThe number of subjects per cluster, n_t^m and n_c^m , is not rounded for reasons discussed in the text. The number of clusters per arm, K_t^m and K_c^m , has to be rounded upward and increased with 2. The budget is based on the thus adjusted numbers of clusters.

the variance ratio goes to one. The df determine the percentiles $t_{1-\alpha/2}$ and $t_{1-\gamma}$, to be used in step 1 instead of $z_{1-\alpha/2}$ and $z_{1-\gamma}$ respectively, and thus also the budget and sample size needed, as these are proportional to $(t_{1-\alpha/2} + t_{1-\gamma})^2$. The case $df \rightarrow \infty$ gives the standard normal distribution, but the Student t -distribution is close to that for $df = 60$ or so. For smaller df , calculations in^{28,38} show that, at least if the number of clusters per arm is 8 or more, the power loss due to finite df is compensated by having 2 extra clusters per arm if $\alpha = 0.05$ two-tailed, or 4 extra clusters per arm if $\alpha = 0.01$ 2-tailed. With less than 8 clusters in a given arm, add 3 instead of 2 clusters to that arm if $\alpha = 0.05$, and 4 if $\alpha = 0.01$. This holds for a planned power of 80% or 90%, and under strong heterogeneity of variance. This adjustment is included into the computer code. Not included into the code for simplicity, but stated here for completeness, is the following exception: Under complete homogeneity of variance and costs the MMD becomes balanced, and if there are at least 8 clusters per arm we then need to add only 1 (if $\alpha = 0.05$) or 2 (if $\alpha = 0.01$) clusters per arm.³⁸ This only holds under complete homogeneity of variance, however, not if heterogeneity is allowed but the MMD happens to be balanced as discussed in Section 5.1. Having said this, let us now consider an example of MMD.

7.1 | Example

Strong heterogeneity of outcome variance has been found in randomized trials in clinical psychology, with variance ratios up to 12 or even 16,¹⁷ but results from cluster randomized trials in health care are scarce and suggest mild heterogeneity, such as a variance ratio of 3.¹⁹ Further, study costs per cluster and per subject are rarely reported. As an exception, Moerbeek et al reported a cluster-to-subject cost ratio $c/s = 26$ in a smoking prevention trial in primary education, but gave no information on the costs per arm.³⁹

Our example therefore makes the following assumptions in step 1 of the sample size calculation: Let $u = 1$ or 2 or 3 (homogeneous variances, respectively a treatment-to-control SD ratio between 0.50 and 2, or between 0.33 and 3). Further, the maximum ICC is $\rho_{\max} = 0.10$ or 0.20.^{18,32,33} Finally, the costs are homogeneous at each level (cluster, subject), or heterogeneous at the cluster level (with $\frac{c_t}{c_c} = 9$), or at the individual level (with $\frac{s_t}{s_c} = 9$), or at both levels (with $\frac{c_t}{c_c} = \frac{s_t}{s_c} = 9$), and the cluster-to-subject cost ratio $\frac{c_t}{s_t} = \frac{c_c}{s_c} = 20$ in case of homogeneity at both levels or of heterogeneity at both levels. These scenarios assume that treatment is at least as expensive as control. The reverse case gives the same designs apart from switching K_t^m and K_c^m and, except if $\frac{c_t}{s_t} = \frac{c_c}{s_c}$, also switching n_t^m and n_c^m (see Sections 3 and 4 for technical details).

Combining the input parameters above gives $3 \cdot 2 \cdot 4 = 24$ different scenarios for the MMD. For each scenario we let $\beta_1 = 5$ and $V_{\max} = 200$, so that the standardized treatment effect size $d = 0.50$,³⁴ where $d = \frac{\beta_1}{\sqrt{V_{\max}/2}}$ which, under homogeneity of variance reduces to $d = \frac{\beta_1}{\max s_y}$. Remember that β_1 and V_{\max} do not affect either the maximin sample sizes per cluster, n_t^m , n_c^m (see step 3 of the sample size calculation above), or the budget allocation ratio (see Table 1), or the cluster allocation ratio $\frac{K_t^m}{K_c^m}$ (steps 4 and 5 above). Instead, β_1 and V_{\max} only affect the total budget and the total number of clusters needed. This completes step 1 of the maximin sample size calculation. Next, in step 2, let $\alpha = 0.05$ two-tailed and let the desired power be 90%. This gives as maximum allowable sampling variance for the treatment effect, $Var(\hat{\beta}_1) = 2.3815$ by Equation 17. Following the steps 3-4-5 of the sample size calculation procedure above for each of the 24 scenarios then gives the maximin sample sizes in Table 4, with a separate row per scenario. To better see the relation between the input parameters and the resulting MMD, neither n_t^m , n_c^m nor K_t^m , K_c^m have been rounded off. As explained above, n_t^m , n_c^m need not be rounded off anyway in view of unplanned variation in sample size per cluster in practice, and K_t^m , K_c^m both need to be rounded upward and then increased by 2. The total budget in the last column of Table 4 is based on the thus adjusted K_t^m , K_c^m .

In the first 8 rows, the variances are homogeneous and so the MMD reduces to the cc design, as shown by the fact that the budget ratio $\frac{f^m}{1-f^m} = p$ (see Section 4 or appendix B for the definition of p). If the cluster-to-subject cost ratio is homogeneous, so $\frac{c_t}{s_t} = \frac{c_c}{s_c}$, then the sample size per cluster is homogeneous, that is, $n_t^m = n_c^m$ (rows 1, 4, 5 and 8). If the cluster cost is homogeneous, so $c_t = c_c$, then the number of clusters is homogeneous, that is, $K_t^m = K_c^m$ (rows 1, 3, 5 and 7). This only holds under homogeneity of variance, however (see the end of Section 3). Finally, if the costs are

completely homogeneous, so $c_t = c_c$ and $s_t = s_c$ both hold, then the MMD is balanced at each design level, that is, $n_t^m = n_c^m$ and $K_t^m = K_c^m$ (rows 1 and 5), in line with Tables 1 and 2 (noting that $p = 1$).

Rows 9 to 16 show the same cost scenarios but now for heterogeneity of variance with as range for the SD ratio $\frac{\sigma_{Yt}}{\sigma_{Yc}} \in \left[\frac{1}{2}, 2\right]$. Again, homogeneity of the cluster-to-subject cost ratio gives homogeneity of the sample size per cluster (rows 9, 12, 13, 16), and complete homogeneity of costs gives a balanced design at each level (rows 9 and 13), in line with the case $p = 1$ in Table 1. Also in line with Table 1, the budget allocation ratio is p^2 as long as $p \in \left[\frac{1}{u}, u\right]$, and is pu else (given that $p < 1$ does not occur in Table 4).

Finally, rows 17 to 24 show the same cost scenarios for heterogeneity of variance with as range for the SD ratio $\frac{\sigma_{Yt}}{\sigma_{Yc}} \in \left[\frac{1}{3}, 3\right]$. Homogeneity of the cluster-to-subject cost ratio now gives a balanced design at each level (rows 17, 20, 21, 24) because, in these rows, the budget allocation ratio p^2 equals the treatment-to-control cost ratio, which is 9 at each design level in rows 20 and 24, and 1 at each level in rows 17 and 21.

8 | DISCUSSION

Optimal sample sizes per level (cluster, subject) of a cluster randomized trial have been published, based on the assumptions that costs and variances at each level are (1) homogeneous between treatments, and (2) known in the design stage.^{8,9} Liu²⁷ extended this to the case of treatment-dependent costs, still assuming known and homogeneous variances. This assumption is problematic for 2 reasons. First, a treatment which affects the mean of an outcome variable can also be expected to affect its variance, making the assumption of *homogeneity* of variance realistic only under the null hypothesis. Secondly, trials are run to test hypotheses about the *unknown* mean outcome difference between treatments, making the assumption of *known* variances untenable. Unfortunately, optimal design strongly depends on the 4 variances involved, and misspecification of these variances can lead to an inefficient design.

Similar to Candel and Van Breukelen²⁸ and Wu et al,³⁰ but extending their results to optimal instead of fixed sample sizes per cluster, this paper addressed these problems by deriving a MMD which allows for heterogeneity of variances, and is robust against misspecification of the variances by maximizing the minimum efficiency over a range of plausible values for the unknown variances. The MMD was compared with the balanced design, which is optimal under homogeneity of costs and variances, and with the cc design, which is optimal under heterogeneity of costs and homogeneity of variances. The results can be summarized as follows: First, if the costs are less heterogeneous than the variances (more precisely, if $u^{-1} \leq p \leq u$), then the MMD allocates budget to each treatment arm proportionally to the costs as expressed by the function $g(\rho_{\max})$ for that arm, see Equations 8 and 11. If the cluster-to-person cost ratio c/s is furthermore the same in both arms, then the MMD is balanced. Secondly, if the costs are more heterogeneous than the variances, then the budget allocation to each arm increases with the costs and variance in that arm. Compared with the MMD, the balanced design is still highly efficient ($RE > 0.95$) if $u^{-2} \leq p \leq u^2$ (at least if c/s is homogeneous). The balanced design is inefficient outside that range, that is, if costs are very heterogeneous. Compared with the MMD, the cc design is never optimal (unless $u = 1$, ie, homogeneous variances), but always fairly to highly efficient, with a minimum RE of 0.83. Compared with the balanced design, the cc design is less efficient if $u^{-2} \leq p \leq u^2$, but more efficient else. Further, if it is known that $\sigma_{Yt}/\sigma_{Yc} \geq 1$ and $p < 1$, or that $\sigma_{Yt}/\sigma_{Yc} \leq 1$ and $p > 1$ (ie, treatment arm ordering by variances is opposite to ordering by costs), then the cc design is the MMD. In view of these results, the practical recommendation is to use the MMD (which is balanced in some cases). As the preceding section shows, sample size calculation is simple for the MMD, and the computer code for such calculations in SPSS respectively R is provided as supplementary material.

No study is without limitations, and here we mention a few of the present study. First of all, although the derivations for the maximin and cc designs and their comparison allowed the cluster-to-person cost ratio c/s to be treatment-dependent, the comparison of both with the balanced design was limited to the case of a homogeneous cost ratio. This was done to reduce the relative efficiencies (RE s) to functions of 2 parameters only, u and p , and thus to get a better understanding of how these RE s depend on the heterogeneity of variance, u , and the heterogeneity of costs, p . In Section 5, we explained how the RE of the balanced design can be obtained without assuming a homogeneous cost ratio.

Another limitation is that this study assumes a quantitative outcome. Wu et al derived the MMD for binary outcomes, but they assumed a fixed and homogeneous sample size per cluster.³⁰ Extending their work to optimal and

heterogeneous sample sizes per cluster is thus a potential topic for future work, as is the extension of MMD to multicentre/multisite trials, for which optimal designs under homogeneity of variance and costs are given by Raudenbusch and Liu and by Moerbeek et al,^{9,40} and under heterogeneity of variances and costs by Lemme et al.⁴¹

ORCID

Gerard J.P. van Breukelen  <http://orcid.org/0000-0003-0949-0272>

Math J.J.M. Candel  <http://orcid.org/0000-0002-2229-1131>

REFERENCES

1. Donner A, Klar N. *Design and Analysis of Cluster Randomization Trials in Health Research*. New York: Wiley; 2000.
2. Hayes RJ, Moulton LH. *Cluster Randomized Trials*. Boca Raton (FL): Chapman and Hall; 2009.
3. Murray DM. *Design and Analysis of Group-Randomized Trials*. New York: Oxford University Press; 1998.
4. Ausems M, Mesters I, Van Breukelen G, De Vries H. Short-term effects of a randomized computer-based out-of-school smoking prevention trial aimed at Dutch elementary schoolchildren. *Prev Med*. 2002;34(6):581-589. <https://doi.org/10.1006/pmed.2002.1021>
5. Kraag G, Van Breukelen GJP, Kok GJ, Hosman C. 'Learn young, learn fair', a stress management program for fifth and sixth graders: longitudinal results from an experimental study. *J Child Psychol Psychiatry*. 2009;50(9):1185-1195. <https://doi.org/10.1111/j.1469-7610.2009.02088.x>
6. Slok AHM, Kotz D, Van Breukelen GJP, et al. Effectiveness of the Assessment of Burden of COPD (ABC) tool on health-related quality of life in COPD patients: a cluster randomised controlled trial in primary and hospital care. *BMJ Open*. 2016;6(7):e011519. <https://doi.org/10.1136/bmjopen-2016-011519>
7. Van Breukelen GJP, Candel MJJM. Calculating sample sizes for cluster randomized trials: we can keep it simple and efficient! *J Clin Epidemiol*. 2012;65(11):1212-1218. <https://doi.org/10.1016/j.jclinepi.2012.06.002>
8. Raudenbush SW. Statistical analysis and optimal design for cluster randomized trials. *Psychol Methods*. 1997;2(2):173-185.
9. Moerbeek M, Van Breukelen GJP, Berger MPF. Design issues for experiments in multilevel populations. *J Educ Behav Stat*. 2000;25(3):271-284. <https://doi.org/10.3102/10769986025003271>
10. Van Breukelen GJP, Candel MJJM, Berger MPF. Relative efficiency of unequal versus equal cluster sizes in cluster randomized and multicentre trials. *Stat Med*. 2007;26(13):2589-2603. <https://doi.org/10.1002/sim.2740>
11. Candel MJJM, Van Breukelen GJP. Sample size adjustments for varying cluster sizes in cluster randomized trials with binary outcomes analyzed with second-order PQL mixed logistic regression. *Stat Med*. 2010;29(14):1488-1501. <https://doi.org/10.1002/sim.3857>
12. Van Breukelen GJP, Candel MJJM. Efficient design of cluster randomized and multicentre trials with unknown intraclass correlation. *Stat Methods Med Res*. 2015;24(5):540-556. <https://doi.org/10.1177/0962280211421344>
13. Rotondi MA, Donner A. Sample size estimation in cluster randomized educational trials: an empirical Bayes approach. *J Educ Behav Stat*. 2009;34(2):229-237. <https://doi.org/10.3102/1076998609332756>
14. Lake S, Kammann E, Klar N, Betensky R. Sample size re-estimation in cluster randomized trials. *Stat Med*. 2002;21(10):1337-1350. <https://doi.org/10.1002/sim.1121>
15. Van Schie S, Moerbeek M. Re-estimating sample size in cluster randomised trials with active recruitment within clusters. *Stat Med*. 2014;33(19):3253-3268. <https://doi.org/10.1002/sim.6172>
16. Gomes M, Grieve R, Nixon R, Edmunds W. Statistical methods for cost-effectiveness analyses that use data from cluster randomized trials a systematic review and check-list for critical appraisal. *Med Decis Making*. 2012;32(1):209-220.
17. Grissom RJ. Heterogeneity of variance in clinical data. *J Consult Clin Psychol*. 2000;68(1):155-165. <https://doi.org/10.1037//0022-006X.68.1.155>
18. Adams G, Gulliford MC, Ukoumunne OC, Eldridge S, Chinn S, Campbell MJ. Patterns of intracluster correlation from primary care research to inform study design and analysis. *J Clin Epidemiol*. 2004;57(8):785-794. <https://doi.org/10.1016/j.jclinepi.2003.12.013>
19. Woodcock AJ, Kinmonth AL, Campbell MJ, Griffin SJ, Spiegel NM. Diabetes care from diagnosis: effects of training in patient-centred care on beliefs, attitudes and behaviour of primary care professionals. *Patient Educ Couns*. 1999;37(1):65-79.
20. Montgomery AA, Fahey T, Peters TJ, MacIntosh C, Sharp DJ. Evaluation of computer based clinical decision support system and risk chart for management of hypertension in primary care: randomized controlled trial. *BMJ*. 2000;320(7236):686-690.
21. Roberts C, Roberts SA. The design and analysis of clinical trials with clustering effects due to treatment. *Clin Trials*. 2005;2(2):152-162.
22. Van Breukelen GJP. Optimal experimental design with nesting of persons in organizations. *J Psychol / Zeitschrift Fuer Psychologie*. 2013;221(3):145-159. <https://doi.org/10.1027/2151-2604/a000143>

23. Schouten HJA. Sample size formula with a continuous outcome for unequal group sizes and unequal variances. *Stat Med.* 1999;18(1):87-91. [https://doi.org/10.1002/\(SICI\)1097-0258\(19990115\)](https://doi.org/10.1002/(SICI)1097-0258(19990115))
24. Wong WK, Zhu W. Optimum treatment allocation rules under a variance heterogeneity model. *Stat Med.* 2008;27(22):4581-4595.
25. Guo JH, Luh WM. Efficient sample size allocation with cost constraints for heterogeneous-variance group comparison. *J Appl Stat.* 2013;40(12):2549-2563. <https://doi.org/10.1080/02664763.2013.819417>
26. Moerbeek M, Wong WK. Sample size formula for trials comparing group and individual treatments in a multilevel model. *Stat Med.* 2008;27(15):2850-2864. <https://doi.org/10.1002/sim.3115>
27. Liu X. Statistical power and optimum sample allocation ratio for treatment and control having unequal costs per unit of randomization. *J Educ Behav Stat.* 2003;28(3):231-248. <https://doi.org/10.3102/10769986028003231>
28. Candel MJJM, Van Breukelen GJP. Sample size calculation for treatment effects in randomized trials with fixed cluster sizes and heterogeneous intraclass correlations and variances. *Stat Methods Med Res.* 2015;24(5):557-573. <https://doi.org/10.1177/0962280214563100>
29. Lemme F, Van Breukelen GJP, Berger MPF. Efficient treatment allocation in 2x2 cluster randomized trials, when costs and variances are heterogeneous. *Stat Med.* 2016;35(24):4320-4334.
30. Wu S, Wong WK, Crespi CM. Maximin optimal designs for cluster randomized trials. *Biometrics.* 2017;73(3):916-926. <https://doi.org/10.1111/biom.12659>
31. Candel MJJM, Van Breukelen GJP. Repairing the efficiency loss due to varying cluster sizes in two-level two-armed randomized trials with heterogeneous clustering. *Stat Med.* 2016;35(12):2000-2015. <https://doi.org/10.1002/sim.6851>
32. Eldridge SM, Ashby D, Feder GS, Rudnicka AR, Ukoumunne OC. Lessons for cluster randomized trials in the twenty-first century: a systematic review of trials in primary care. *Clin Trials.* 2004;1(1):80-90. <https://doi.org/10.1191/1740774504cn006rr>
33. Hedges LV, Hedberg EC. Intraclass correlation values for planning group-randomized trials in education. *Educ Eval Policy Analysis.* 2007;29(1):60-87. <https://doi.org/10.3102/0162373707299706>
34. Cohen J. *Statistical Power Analysis for the Behavioral Sciences.* 2nd ed. Mahwah (NJ): Erlbaum; 1988.
35. Moerbeek M, Van Breukelen GJP, Berger MPF. A comparison between traditional methods and multilevel regression for the analysis of multicentre intervention studies. *J Clin Epidemiol.* 2003;56(4):341-350. [https://doi.org/10.1016/S0895-4356\(03\)00007-6](https://doi.org/10.1016/S0895-4356(03)00007-6)
36. Welch BL. The significance of the difference between two means when the population variances are unequal. *Biometrika.* 1938;29(3/4):350-362.
37. Satterthwaite FE. Synthesis of variance. *Psychometrika.* 1941;6(5):309-316.
38. Lemme F, Van Breukelen GJP, Candel MJJM, Berger MPF. The effect of heterogeneous variance on efficiency and power of cluster randomized trials with a balanced 2x2 factorial design. *Stat Methods Med Res.* 2015;24(5):574-593. <https://doi.org/10.1177/0962280215583683>
39. Moerbeek M, Van Breukelen GJP, Berger MPF, Ausems M. Optimal sample sizes in experimental designs with individuals nested within clusters. *Understanding Stat.* 2003;2(3):151-175.
40. Raudenbush SW, Liu X. Statistical power and optimal design for multisite trials. *Psychol Methods.* 2000;5(2):199-213. <https://doi.org/10.1037//1082-989X.5.2.199>
41. Lemme F, Van Breukelen GJP, Candel MJJM. Efficient treatment allocation in 2x2 multicentre trials, when costs and variances are heterogeneous. *Stat Med.* 2018;37(1):12-27. <https://doi.org/10.1002/sim.7499>

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: van Breukelen GJP, Candel MJJM. Efficient design of cluster randomized trials with treatment-dependent costs and treatment-dependent unknown variances. *Statistics in Medicine.* 2018;37:3027–3046. <https://doi.org/10.1002/sim.7824>

APPENDIX A

LIST OF SYMBOLS USED

Symbol	Interpretation	Introduced in section nr
β_1	The treatment effect of interest	2
σ_{u0}^2	Residual variance at the cluster level	2
σ_ε^2	Residual variance at the individual level	2
σ_Y^2	Total residual variance	2
ρ	Intraclass correlation	2
K	Total number of clusters sampled	2
n	Number of individuals sampled per cluster	2
Superscript * for any design factor	Optimal design value of that factor	2
c	Cost per cluster	2
s	Cost per subject	2
B	Budget for the study	2
$g(\rho)$	Shortcut for $(\sqrt{\rho c} + \sqrt{(1-\rho)s})^2$	2
Subscript t for any symbol	In the treated group	3
Subscript c for any symbol	In the control group	3
f	Fraction of the study budget spent on the treated arm	3
$f/(1-f)$	Budget allocation ratio	4
ρ_{\max}	Maximum plausible intraclass correlation	4
u and $1/u$	Maximum and minimum for the SD ratio $\frac{\sigma_{Yt}}{\sigma_{Yc}}$	4
V	Shortcut for $\sigma_{Yt}^2 + \sigma_{Yc}^2$	4
V_{\max}	Maximum plausible V	4
p	Shortcut for $\sqrt{\frac{g_t(\rho_{\max})}{g_c(\rho_{\max})}}$	4
Superscript m, bal, cc for $f/(1-f)$	Maximin, balanced, cost-considered	4,5

APPENDIX B

DERIVATION OF THE MAXIMIN DESIGN

This appendix derives the Maximin design from Equation 10 in Section 4:

$$\max Var(\hat{\beta}_1) = \frac{g_c(\rho_{\max})}{B} \left(\frac{p^2 \sigma_{Yt}^2}{f} + \frac{V_{\max} - \sigma_{Yt}^2}{1-f} \right), \text{ with } p = \sqrt{\frac{g_t(\rho_{\max})}{g_c(\rho_{\max})}}. \quad (\text{B1})$$

Equation B1 will first be maximized as a function of the unknown σ_{Yt}^2 within the joint constraints $\sigma_{Yt}^2 + \sigma_{Yc}^2 = V_{\max}$ and $\frac{\sigma_{Yt}}{\sigma_{Yc}} \in \left[\frac{1}{u}, u \right]$, implying $\sigma_{Yt}^2 \in \left[\frac{V_{\max}}{1+u^2}, \frac{u^2 V_{\max}}{1+u^2} \right]$.

That maximum will then be minimized as a function of the budget ratio $\frac{B_t}{B_c} = \frac{f}{1-f}$.

The derivative of (B1) with respect to σ_{Yt}^2 is 0, or <0 , or >0 if $\frac{f}{1-f} = p^2$, or $>p^2$, or $<p^2$, respectively. To find the MMD, consider each case in turn.

First, for $\frac{f}{1-f} = p^2$, the derivative of B1 with respect to σ_{Yt}^2 is zero, giving

$$\text{Var}(\hat{\beta}_1) = \frac{g_c(\rho_{\max})V_{\max}(1+p^2)}{B}, \quad (\text{B2})$$

which does not depend on the SD ratio $\frac{\sigma_{Yt}}{\sigma_{Yc}} \in \left[\frac{1}{u}, u\right]$ anymore.

Second, for $\frac{f}{1-f} > p^2$, the derivative of B1 is negative, so B1 is maximized by letting $\frac{\sigma_{Yt}}{\sigma_{Yc}} = \frac{1}{u}$, which by substitution into B1 gives $\max \text{Var}(\hat{\beta}_1) = \frac{g_c(\rho_{\max})V_{\max}}{B(u^2+1)} \left(\frac{p^2}{f} + \frac{u^2}{1-f}\right)$, which is minimized by $\frac{f}{1-f} = \frac{p}{u}$, giving

$$\max \text{Var}(\hat{\beta}_1) = \frac{g_c(\rho_{\max})V_{\max}}{B} \frac{(p+u)^2}{(u^2+1)}, \quad (\text{B3})$$

provided that $\frac{1}{u} > p$, so that $\frac{f}{1-f} = \frac{p}{u}$ is compatible with $\frac{f}{1-f} > p^2$.

Equation B3 is always smaller than B2, except if $\frac{1}{u} = p$, in which case they are equal.

Third and last, for $\frac{f}{1-f} < p^2$, the derivative of B1 is positive, so B1 is maximized by letting $\frac{\sigma_{Yt}}{\sigma_{Yc}} = u$, which by substitution into B1 gives

$$\max \text{Var}(\hat{\beta}_1) = \frac{g_c(\rho_{\max})V_{\max}}{B(u^2+1)} \left(\frac{p^2 u^2}{f} + \frac{1}{1-f}\right), \text{ which is minimized by } \frac{f}{1-f} = pu, \text{ giving}$$

$$\max \text{Var}(\hat{\beta}_1) = \frac{g_c(\rho_{\max})V_{\max}}{B} \frac{(pu+1)^2}{(u^2+1)}, \quad (\text{B4})$$

provided that $u < p$, so that $\frac{f}{1-f} = pu$ is compatible with $\frac{f}{1-f} < p^2$.

Equation B4 is always smaller than B2, except if $u = p$, in which case they are equal.

In short:

If $\frac{1}{u} \leq p \leq u$, the MMD is $\frac{f}{1-f} = p^2$ and $\max \text{Var}(\hat{\beta}_1)$ is given by B2.

If $\frac{1}{u} > p$ (implying $p < 1$), the MMD is $\frac{f}{1-f} = \frac{p}{u}$ and $\max \text{Var}(\hat{\beta}_1)$ obeys B3.

If $u < p$ (implying $p > 1$), the MMD is $\frac{f}{1-f} = pu$ and $\max \text{Var}(\hat{\beta}_1)$ obeys B4.

If the interval for the SD ratio is 1-sided, the following modifications apply

First, for $\frac{\sigma_{Yt}}{\sigma_{Yc}} \in [1, u]$ the only change is when $\frac{f}{1-f} > p^2$, so that the derivative of B1 is negative. Now B1 is maximized by letting $\frac{\sigma_{Yt}}{\sigma_{Yc}} = 1$, which by substitution into B1 gives

$$\max \text{Var}(\hat{\beta}_1) = \frac{g_c(\rho_{\max})V_{\max}}{2B} \left(\frac{p^2}{f} + \frac{1}{1-f}\right), \text{ which is minimized by } \frac{f}{1-f} = p, \text{ giving}$$

$$\max \text{Var}(\hat{\beta}_1) = \frac{g_c(\rho_{\max})V_{\max}}{B} \frac{(p+1)^2}{2}, \quad (\text{B5})$$

provided that $p < 1$, so that $\frac{f}{1-f} = p$ is compatible with $\frac{f}{1-f} > p^2$.

Note that Equation B5 can be obtained from Equation B3 by letting $u = 1$ in B3.

Secondly, for $\frac{\sigma_{Yl}}{\sigma_{Yc}} \in \left[\frac{1}{u}, 1 \right]$ the only change is when $\frac{f}{1-f} < p^2$, so that the derivative of B1 is positive. Now B1 is maximized by letting $\frac{\sigma_{Yl}}{\sigma_{Yc}} = 1$, which by substitution into B1 and minimization gives Equation B5 again, provided $p > 1$, so that $\frac{f}{1-f} = p$ is compatible with $\frac{f}{1-f} < p^2$. Note that Equation B5 can be obtained from Equation B4 by letting $u = 1$ in B4.