

Sample size calculation in cost-effectiveness cluster randomized trials: optimal and maximin approaches

Citation for published version (APA):

Manju, A., Candel, M. J. J. M., & Berger, M. P. F. (2014). Sample size calculation in cost-effectiveness cluster randomized trials: optimal and maximin approaches. *Statistics in Medicine*, 33(15), 2538-2553. <https://doi.org/10.1002/sim.6112>

Document status and date:

Published: 10/07/2014

DOI:

[10.1002/sim.6112](https://doi.org/10.1002/sim.6112)

Document Version:

Publisher's PDF, also known as Version of record

Document license:

Taverne

Please check the document version of this publication:

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

[Link to publication](#)

General rights

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

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

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

www.umlib.nl/taverne-license

Take down policy

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

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Download date: 25 Apr. 2024

Sample size calculation in cost-effectiveness cluster randomized trials: optimal and maximin approaches

Md. Abu Manju,^{*†} Math J. J. M. Candel and Martijn P. F. Berger

In this paper, the optimal sample sizes at the cluster and person levels for each of two treatment arms are obtained for cluster randomized trials where the cost-effectiveness of treatments on a continuous scale is studied. The optimal sample sizes maximize the efficiency or power for a given budget or minimize the budget for a given efficiency or power. Optimal sample sizes require information on the intra-cluster correlations (ICCs) for effects and costs, the correlations between costs and effects at individual and cluster levels, the ratio of the variance of effects translated into costs to the variance of the costs (the variance ratio), sampling and measuring costs, and the budget. When planning, a study information on the model parameters usually is not available. To overcome this local optimality problem, the current paper also presents maximin sample sizes. The maximin sample sizes turn out to be rather robust against misspecifying the correlation between costs and effects at the cluster and individual levels but may lose much efficiency when misspecifying the variance ratio. The robustness of the maximin sample sizes against misspecifying the ICCs depends on the variance ratio. The maximin sample sizes are robust under misspecification of the ICC for costs for realistic values of the variance ratio greater than one but not robust under misspecification of the ICC for effects. Finally, we show how to calculate optimal or maximin sample sizes that yield sufficient power for a test on the cost-effectiveness of an intervention. Copyright © 2014 John Wiley & Sons, Ltd.

Keywords: cluster randomized trials; cost-effectiveness analysis; maximin design; optimal design; sample size calculation

1. Introduction

Decision makers have need of cost-effectiveness analyses (CEAs) to evaluate new healthcare technologies or interventions. These interventions are often evaluated with well-designed randomized controlled trials, where one randomly assigns individuals such as patients to treatments [1]. Randomized controlled trials with randomization carried out at the cluster level, that is, all the persons within the same cluster such as a hospital or general practitioner receive the same treatment, are called cluster randomized trials (CRTs). For example, a CRT in CEAs evaluated alternative interventions for preventing postnatal depression where general practices were randomly assigned to the interventions [2]. This study aimed at estimating incremental costs (UK pounds (£)) and effectiveness in terms of quality-adjusted life-years (QALYs) of interventions to prevent depression for postnatal women. In multicentre trials, on the other hand, individuals within clusters, as opposed to whole clusters, are randomly assigned to the treatments. Multicentre trials are more efficient [3] but, because of treatment contamination or logistic reasons, not always possible in practice. In these cases, CRTs are a suitable or, in many cases, even the only alternative.

A fundamental issue in CRTs is that the individual costs and effects data may vary noticeably between clusters. So CRTs on cost-effectiveness of interventions may result in a hierarchical or multilevel structure, that is, the responses of the individuals treated in the same centre or cluster are correlated. Failure

Department of Methodology and Statistics, Maastricht University, PO Box 616, 6200 MD, Maastricht, The Netherlands

*Correspondence to: Md. Abu Manju, Department of Methodology and Statistics, Maastricht University, PO Box 616, 6200 MD, Maastricht, The Netherlands.

†E-mail: abu.manju@maastrichtuniversity.nl

to acknowledge this between-cluster variability in CEAs may give misleading conclusions in an economic evaluation. The linear mixed model is a suitable framework to analyze the cost-effectiveness data gathered in CRTs [4].

An important part of the design of CRTs is the calculation of sample size (number of clusters and persons in each cluster), which gives high statistical power to show the cost-effectiveness of a treatment. The optimal sample size maximizes the efficiency of the treatment effect estimator and thus maximizes power of the Wald test on the treatment effect, given the constraints on sample sizes and costs for selecting samples. Formulas for computing optimal sample sizes for CRTs under a given budget or power are available in the literature, but these optimal designs are only based on the effectiveness of an intervention [3, 5, 6].

Nowadays researchers are not only interested in effectiveness but also try to include costs and effectiveness simultaneously in trials. In the literature on economic evaluation, there are sample size formulas that account for costs and effects at the same time [7–9]. The required sample sizes depend on the variance and correlation of costs and effects. However, the formulas for sample size calculation given by [7–9] are developed for single-level studies and thus ignore between-cluster variability, which is an important characteristic of cost-effectiveness CRTs. Ignoring the between-cluster variability in sample size calculation will lead to too small sample sizes, and thus to underpowered studies [3].

For the design of cost-effectiveness CRTs, the question is how many clusters and how many persons in each cluster are needed to prove the cost-effectiveness of an intervention. This paper proposes an optimal design for cost-effectiveness CRTs. Such an optimal design will be derived taking net monetary benefit [10] as an outcome variable, which accounts for both effectiveness and costs of two treatments.

The optimal design literature for CRTs on efficacy shows that optimal designs depend on the model parameters and requires knowledge of these parameters that may be unknown at the beginning of the study. This is called the local optimality problem [11, 12]. Although parameters are unknown, some information may be available from similar previous studies, yielding plausible ranges for those parameters.

Three approaches are available to deal with the local optimality problem, namely the sequential design approach, the Bayesian approach, and the maximin approach. The sequential approach updates the design based on updated parameter estimates and often requires a relatively long time to complete [13]. The Bayesian approach evaluates the expected efficiency of the design over the prior distributions of the model parameters and yields a design that performs well on average [14]. For the Bayesian approach, knowledge of the prior distributions of relevant parameters is needed. Finally, the maximin approach maximizes the minimum efficiency of the treatment effect estimator [11]. The maximin design thus has the best possible worst-case performance, in this sense differing from the Bayesian and the sequential approaches. From a practical point of view, this strategy is attractive in that the researcher only needs to specify plausible ranges for the unknown parameters. To accommodate the local optimality problem, the current paper proposes a maximin design for cost-effectiveness CRTs.

The paper is structured as follows. First, we present the model for analyzing individual level data on net monetary benefit that are gathered in CRTs. Second, we derive optimal sample sizes under a budget constraint for cost-effectiveness CRTs, minimizing the variance of the estimator of the average incremental net monetary benefit for comparing two interventions. Third, to accommodate uncertainty with respect to parameter values, maximin sample size formulas will be derived. We show through a numerical evaluation how robust these maximin sample sizes are against misspecification of parameter boundaries. Next, a formula is given to calculate the required sample sizes for a given power of the test on the cost-effectiveness of a treatment. We then apply our strategies for sample size calculation to a case study and discuss the findings. The final section summarizes the results, points out limitations of the present method, and gives suggestions for further research.

2. Model specification

In CRTs, the group or cluster is randomized to a treatment, not the individual. The key implication of CRTs for economic data is that cost-effectiveness of a particular treatment may vary markedly among the clusters (e.g., [10]). In addition, there may be a correlation in costs and effects not only among the individuals within the same cluster but also between clusters. In what follows, we will present a method for CEAs, accommodating both features of such data. In the treatment and control conditions, we have k_t and k_c clusters, respectively. In cluster j ($j = 1, 2, \dots, k_t$) of the treatment condition, there are m_j

persons, and in cluster j ($j = k_t + 1, k_t + 2, \dots, k_t + k_c$) of the control condition, there are n_j persons. Thus, the number of persons in the treatment group is $\sum_{j=1}^{k_t} m_j$, and the number of persons in the

control group is $\sum_{j=k_t+1}^{k_t+k_c} n_j$. Let C_{ij} and E_{ij} be the costs and effects of person i in cluster j , where $i = 1, 2, \dots, m_j$ for a person in the treatment arm, and $i = 1, 2, \dots, n_j$ for a person in the control arm.

A suitable technique to estimate incremental costs and effects of the treatment versus the control condition, allowing for clustering, is the following bivariate linear mixed model [4]:

$$E_{ij} = \alpha_0 + \alpha_1 t_j + u_{0j} + \delta_{ij}, \text{ and } C_{ij} = \gamma_0 + \gamma_1 t_j + v_{0j} + \varepsilon_{ij}, \quad (1)$$

where

$$\begin{pmatrix} u_{0j} \\ v_{0j} \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_u^2 & \rho_{uv}\sigma_u\sigma_v \\ & \sigma_v^2 \end{pmatrix} \right) \text{ and } \begin{pmatrix} \delta_{ij} \\ \varepsilon_{ij} \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_\delta^2 & \rho_{\delta\varepsilon}\sigma_\delta\sigma_\varepsilon \\ & \sigma_\varepsilon^2 \end{pmatrix} \right).$$

The intervention is represented by a variable t_j , which takes the value 0 if the j^{th} cluster is in the control condition and 1 if the j^{th} cluster is in the treatment condition. Incremental effects and costs are represented by slope coefficients α_1 and γ_1 . The random components u_{0j} and v_{0j} indicate the grand mean deviance of effects and costs for the j^{th} cluster, respectively, and are assumed to be bivariate normal with zero mean and variances σ_u^2 and σ_v^2 . The u_{0j} 's for different j ($j = 1, 2, \dots, k_t + k_c$) are independent from each other as well as the v_{0j} 's for different j ($j = 1, 2, \dots, k_t + k_c$), and u_{0j} is independent from v_{0k} for $j \neq k$. The models in Equation (1) take care of the hierarchical or nested structure by including random components for clusters. These models are known as multilevel or hierarchical models. Here we have a two-level model, with, for instance, patients being nested within hospitals. The correlation ρ_{uv} ($-1 \leq \rho_{uv} \leq 1$) represents the cost–effect correlation between random effects u_{0j} and v_{0j} at the cluster level. The error terms δ_{ij} and ε_{ij} also have a bivariate normal distribution with zero mean and variances σ_δ^2 and σ_ε^2 , respectively. The terms δ_{ij} for all i ($i = 1, 2, \dots, m_j$ or $i = 1, 2, \dots, n_j$) and j ($j = 1, 2, \dots, k_t + k_c$) are independent from each other. This also holds for the terms ε_{ij} for all i ($i = 1, 2, \dots, m_j$ or $i = 1, 2, \dots, n_j$) and j ($j = 1, 2, \dots, k_t + k_c$). The cost–effect correlation between random effects ε_{ij} and δ_{ij} at the individual level is $\rho_{\delta\varepsilon}$ ($-1 \leq \rho_{\delta\varepsilon} \leq 1$). The other model assumptions are that ε_{ij} and $\delta_{i'j'}$ are independent for $i \neq i'$ or $j \neq j'$, u_{0j} is independent from ε_{ij} and δ_{ij} , and v_{0j} is independent from ε_{ij} and δ_{ij} . The outcomes of individuals within the same cluster have a positive correlation, measured by the intra-cluster correlation (ICC) for effects $0 \leq \rho_e = \sigma_u^2 / (\sigma_u^2 + \sigma_\delta^2) \leq 1$ and costs $0 \leq \rho_c = \sigma_v^2 / (\sigma_v^2 + \sigma_\varepsilon^2) \leq 1$. Actually, the ICC describes the proportion of total variation due to clustering. In what follows, the variances of C_{ij} and E_{ij} are denoted as $V(E_{ij}) = \sigma_u^2 + \sigma_\delta^2$ and $V(C_{ij}) = \sigma_v^2 + \sigma_\varepsilon^2$.

In Equation (1), the costs and effects are expressed on their original scales, but the net-benefit framework gives us the opportunity to express both costs and effects on the same scale, either as net monetary benefit (NMB) or as net health benefit by using an additional term λ denoted as the ceiling ratio [15]. The ceiling ratio λ ($0 \leq \lambda < \infty$) translates efficacy into monetary units and is interpreted as the maximum value (on a monetary scale) associated with an extra unit of effect (e.g., health gain) [15]. Let NMB_{ij} represent the net monetary benefit for the i^{th} individual in the j^{th} cluster. This is defined as

$$NMB_{ij} = \lambda E_{ij} - C_{ij}. \quad (2)$$

Alternatively, $E_{ij} - \frac{C_{ij}}{\lambda}$ is the net health benefit for the i^{th} individual in the j^{th} cluster. As net health benefit can easily be translated into net monetary benefit and vice versa, we will only consider net monetary benefit in the following. Starting from the model in Equation (1), Equation (2) can be elaborated as

$$NMB_{ij} = \lambda (\alpha_0 + \alpha_1 t_j + u_{0j} + \delta_{ij}) - (\gamma_0 + \gamma_1 t_j + v_{0j} + \varepsilon_{ij}), \text{ or } = \beta_0 + \beta_1 t_j + w_{0j} + \omega_{ij}. \quad (3)$$

Here, the intercept $\beta_0 (= \lambda\alpha_0 - \gamma_0)$ is the average NMB in the control condition. Similarly, the average NMB for the treatment condition is $\beta_0 + \beta_1$. So $\beta_1 (= \lambda\alpha_1 - \gamma_1)$ stands for the average incremental NMB (INMB), and an intervention is cost-effective and thus should replace the control treatment, if the average INMB > 0 . For example, if a decision maker is willing to pay £30,000 for a year of life, the intervention

will be cost-effective if average $INMB > 0$ for $\lambda = \text{£}30,000$. For $\lambda = 0$ (or $\lambda \rightarrow \infty$), Equation (3) reduces to the second part of Equation (1), that is, the regression model of costs on treatment (or the first part of Equation (1), that is, the regression model of effects on treatment). It follows from Equation (1) that the random components $w_{0j} (= \lambda u_{0j} - v_{0j})$ and $\omega_{ij} (= \lambda \delta_{ij} - \varepsilon_{ij})$ are normally distributed with means of zeros and variances $V(w_{0j})$ and $V(\omega_{ij})$, respectively, where $V(w_{0j}) = \lambda^2 \sigma_u^2 + \sigma_v^2 - 2\lambda \rho_{uv} \sigma_u \sigma_v$ and $V(\omega_{ij}) = \lambda^2 \sigma_\delta^2 + \sigma_\varepsilon^2 - 2\lambda \rho_{\delta\varepsilon} \sigma_\delta \sigma_\varepsilon$. The w_{0j} 's are independent for all j , and ω_{ij} 's are independent for all i and j . The components w_{0j} and ω_{ij} are independent of each other.

3. Optimal sample sizes

In order to derive optimal sample sizes, we assume that the cluster sizes in the treatment and control conditions are all m and n , respectively, that is, $m_j = m$ for $j = 1, 2, \dots, k_t$ and $n_j = n$ for $j = k_t + 1, k_t + 2, \dots, k_t + k_c$. There are several methods to estimate the parameters of the linear mixed effects model in Equation (3), and maximum likelihood (ML) is the most common one. Verbeke and Molenberghs [16] gave a general formula for the ML estimator of fixed effects. From this expression, we can derive the ML estimator of the average $INMB$ (β_1) in Equation (3) as

$$\hat{\beta}_1 = \frac{1}{k_t m} \sum_{j=1}^{k_t} \sum_{i=1}^m NMB_{ij} - \frac{1}{k_c n} \sum_{j=k_t+1}^{k_t+k_c} \sum_{i=1}^n NMB_{ij}. \tag{4}$$

The asymptotic variance of the ML estimator is [16, pp. 46–49]

$$V(\hat{\beta}_1) = \left(\frac{1}{k_t} + \frac{1}{k_c}\right) V(w_{0j}) + \left(\frac{1}{k_t m} + \frac{1}{k_c n}\right) V(\omega_{ij}),$$

which can be written as (Appendix A)

$$V(\hat{\beta}_1) = \left\{ \left(\frac{1}{k_t} + \frac{1}{k_c}\right) A + \left(\frac{1}{k_t m} + \frac{1}{k_c n}\right) B \right\} V(C_{ij}), \tag{5}$$

where $A = \varphi \rho_e + \rho_c - 2\rho_{uv} \sqrt{\rho_e \rho_c \varphi}$ and $B = \varphi(1 - \rho_e) + (1 - \rho_c) - 2\rho_{\delta\varepsilon} \sqrt{(1 - \rho_e)(1 - \rho_c)\varphi}$. The parameter $\varphi = V(\lambda E_{ij}) / V(C_{ij})$ is the ratio of the total variance for effects translated into costs (λE_{ij}) to the total variance for costs (C_{ij}) ($0 \leq \varphi < \infty$). Equation (5) shows that the asymptotic variance of $\hat{\beta}_1$ depends on the number of clusters, the cluster sizes in both arms and the ICC for effects, ρ_e , and costs, ρ_c , the correlation at cluster level, ρ_{uv} , and individual level, $\rho_{\delta\varepsilon}$, and the variance ratio φ .

To derive the optimal sample sizes, the variance of the treatment effect estimator $\hat{\beta}_1$ is minimized under certain constraints with respect to sample sizes and costs for selecting samples. For the Wald test of $\hat{\beta}_1$ in Equation (4), this implies that the optimal design yields maximum power at lowest costs [11, 12]. Tokola *et al.* [17] proposed the following budget constraint:

$$C = k_t c_t + k_c c_c + k_t m s_t + k_c n s_c. \tag{6}$$

Here $c_t > 0$ and $c_c > 0$ are the costs for selecting a cluster in the treatment and control conditions, respectively. Furthermore, $s_t > 0$ and $s_c > 0$ are the costs for selecting a patient in the treatment and control clusters, respectively. It is very likely that the sampling and measuring costs for a cluster are larger than the costs for sampling and measuring an individual [3, 5, 17], so it is assumed that $c_t \geq s_t$ and $c_c \geq s_c$. Finally, C represents the total research budget.

Minimizing Equation (5) subject to the cost function in Equation (6), employing the method of Lagrange multipliers, yields the following optimal cluster sizes (for details, see Appendix A):

$$m_{opt} = \sqrt{\frac{c_t B}{s_t A}} \text{ and } n_{opt} = \sqrt{\frac{c_c B}{s_c A}}. \tag{7}$$

Equation (7) shows that the optimal cluster sizes m_{opt} and n_{opt} depend on the ICC for effects, ρ_e , and costs, ρ_c , the correlation at individual level, $\rho_{\delta\varepsilon}$, and cluster level, ρ_{uv} , the variance ratio, φ , but not on the total budget C . Also, the optimal cluster sizes m_{opt} and n_{opt} in the treatment and control conditions, respectively, decrease with increasing subject-specific costs in the treatment, s_t , and subject-specific

costs in the control condition, s_c , respectively. The optimal cluster sizes in both arms increase as the cluster-specific costs in the same arm increase.

Substituting the cluster sizes into Equation (6) gives the optimal number of clusters

$$k_{t_{opt}} = \frac{C\sqrt{A}}{\sqrt{c_t}\{(\sqrt{c_c A} + \sqrt{s_c B}) + (\sqrt{c_t A} + \sqrt{s_t B})\}} \text{ and } k_{c_{opt}} = \frac{k_{t_{opt}}\sqrt{c_t}}{\sqrt{c_c}}. \quad (8)$$

Equation (8) shows that the optimal number of clusters $k_{t_{opt}}$ and $k_{c_{opt}}$ depend on the ICC for effects, ρ_e , and costs, ρ_c , the correlation at individual, $\rho_{\delta\varepsilon}$, and cluster level, ρ_{uv} , the variance ratio, φ , and total budget C . The ratio of the optimal number of clusters in the treatment condition, $k_{t_{opt}}$, versus the number of clusters in the control condition, $k_{c_{opt}}$, is the square root of the ratio of cluster-specific costs in the control condition, c_c , versus the cluster-specific costs in the treatment condition, c_t .

Substituting the optimal values in Equations (7) and (8) into the variance formula in Equation (5) yields the optimal variance for the vector of parameters $\Pi = (\rho_e, \rho_c, \rho_{uv}, \rho_{\delta\varepsilon}, \varphi)$:

$$V(\hat{\beta}_1 | \Pi, opt) = \frac{\{(\sqrt{c_t} + \sqrt{c_c})\sqrt{A} + (\sqrt{s_t} + \sqrt{s_c})\sqrt{B}\}^2 V(C_{ij})}{C}. \quad (9)$$

Letting $\varphi \rightarrow \infty$ and setting the cluster-specific costs equal for both treatment arms (i.e., $c_t = c_c$) and the subject-specific costs (i.e., $s_t = s_c$), Equation (9) reduces to the optimal variance formula in Moerbeek *et al.* [3] for the linear mixed model for effects only.

The optimal design derivation does not provide any solutions for m, n, k_t , and k_c in case $A = 0$ or $B = 0$. Note that if $\rho_e = \rho_c = 0$, then $A = 0$, and if $\rho_e = \rho_c = 1$, then $B = 0$. So restricting the parameters ($\rho_e, \rho_c, \rho_{uv}, \rho_{\delta\varepsilon}$, and φ) such that $A > 0$ and $B > 0$ gives the optimal sample sizes by Equations (7) and (8).

4. Maximin sample sizes

To compute optimal sample sizes, a priori knowledge about the parameters $\rho_e, \rho_c, \rho_{uv}, \rho_{\delta\varepsilon}$, and φ is needed. Although some systematic reviews suggest plausible ranges for these parameters [18], the parameters are unknown before data analysis. An incorrect specification of the parameters may lead to an inefficient design. This is called the local optimality problem [11, 12]: the design is optimal only for a specific value of the parameter but not within the whole range of values. The approach of this paper to solve the local optimality problem is the maximin strategy, which guarantees a certain desired level of power for all parameter values in the range.

The maximin strategy is based on the efficiency $Eff = \{V(\hat{\beta}_1)\}^{-1}$ of a design [11] and consists of the following two steps:

- (1) Minimize the efficiency (Eff) of estimating the treatment effect estimator across the plausible ranges for the model parameters ($\rho_e, \rho_c, \rho_{uv}, \rho_{\delta\varepsilon}$, and φ). In this step, parameters are chosen to obtain the worst case in terms of efficiency.
- (2) Choose that sample size (m, n, k_t , and k_c), which maximizes this minimum efficiency given a fixed budget. The sample size that results in the second step optimizes the worst case and is called the maximin design (MMD).

As the efficiency is the inverse of the asymptotic variance in Equation (5), finding the minimum of efficiency, amongst others, depends on the range for ρ_{uv} and $\rho_{\delta\varepsilon}$. Based on a review by Gomes *et al.* [18], both ρ_{uv} and $\rho_{\delta\varepsilon}$ have negative lower boundaries. Let the realistic ranges for ρ_{uv} and $\rho_{\delta\varepsilon}$ thus be $\rho_{uv} \in [r_{1l}, r_{1u}]$, where $-1 \leq r_{1l} < 0$ and $r_{1l} \leq r_{1u} \leq 1$, and $\rho_{\delta\varepsilon} \in [r_{2l}, r_{2u}]$, where $-1 \leq r_{2l} < 0$ and $r_{2l} \leq r_{2u} \leq 1$, respectively. As shown in Appendix B, the efficiency is minimum if we start from the minimum of the two lower boundaries of ρ_{uv} and $\rho_{\delta\varepsilon}$, denoted as $r(-1 \leq r < 0) = \min\{r_{1l}, r_{2l}\}$. The spaces for other relevant parameters are the intervals $\rho_e \in [a, b]$, where $0 < a \leq b < 1$, $\rho_c \in [c, d]$, where $0 < c \leq d < 1$ and $\varphi \in [e, f]$, where $0 \leq e \leq f < \infty$. Note that 0 and 1 were excluded from the boundaries of ρ_e and ρ_c , because we could not find a minimum for the efficiency in these cases. It can be shown that the ICC for effects, ρ_e , and costs, ρ_c , minimize the efficiency when both are equal to their

maximum, that is, $\rho = \max\{b, d\}$. Finally, the efficiency is minimum at the upper bound of the variance ratio φ , that is, f . As shown in Appendix B, the minimum efficiency becomes

$$Eff^{\min} = \left[\left\{ \left(\frac{1}{k_t} + \frac{1}{k_c} \right) \rho + \left(\frac{1}{k_t m} + \frac{1}{k_c n} \right) (1 - \rho) \right\} (f + 1 - 2r\sqrt{f}) V(C_{ij}) \right]^{-1}, \quad (10)$$

where $(f + 1 - 2r\sqrt{f}) V(C_{ij})$ is the variance of the dependent variable NMB_{ij} in Equation (3) under the maximin values of the parameters.

The maximin sample sizes, which maximize the minimum efficiency in Equation (10), given the budget constraint in Equation (6), can be derived as (Appendix B)

$$m_{MMD} = \sqrt{\frac{c_t(1-\rho)}{s_t\rho}} \text{ and } n_{MMD} = \sqrt{\frac{c_c(1-\rho)}{s_c\rho}},$$

$$k_{tMMD} = \frac{C\sqrt{\rho}}{\sqrt{c_t}\left\{(\sqrt{c_c} + \sqrt{c_t})\sqrt{\rho} + (\sqrt{s_c} + \sqrt{s_t})\sqrt{(1-\rho)}\right\}} \text{ and } k_{cMMD} = \frac{k_{tMMD}\sqrt{c_t}}{\sqrt{c_c}}, \quad (11)$$

which implies that the cluster sizes m and n , under the maximin strategy, increase with the cost ratios c_t/s_t and c_c/s_c , respectively, and decrease with ρ . On the other hand, the number of clusters increases with ρ , the total budget C , and decreases with c_t/s_t and c_c/s_c .

The maximum variance of the treatment effect estimator for the given ranges of parameter values, when employing the maximin sample sizes, called the maximin variance (MMV), is given by (Equation (10))

$$MMV = V(\hat{\beta}_1 | \Pi^*, MMD) = \frac{\left\{ (\sqrt{c_t} + \sqrt{c_c})\sqrt{\rho} + (\sqrt{s_t} + \sqrt{s_c})\sqrt{(1-\rho)} \right\}^2 (f + 1 - 2r\sqrt{f}) V(C_{ij})}{C}, \quad (12)$$

where $\Pi^* = (\rho_e = \rho, \rho_c = \rho, \rho_{uv} = r, \rho_{\delta\varepsilon} = r, \varphi = f)$ is the vector of maximin boundaries. The MMV increases as specific sampling and measuring costs increase and as ρ increases, provided $\rho \leq 0.5$ and $c_t \geq s_t$ and $c_c \geq s_c$, but decreases as the total budget C increases. Equations (11) and (12) show that the maximin sample sizes and MMV depend on the choices of boundaries of the intervals for the parameters, so a different specification of these boundaries of the intervals will give different maximin sample sizes and corresponding MMV .

5. Robustness of maximin sample sizes

In this study, the maximin sample sizes are derived based on three choices for relevant model parameters. These choices are the minimum of the two lower boundaries of ρ_{uv} and $\rho_{\delta\varepsilon}$, denoted as r ($-1 \leq r < 0$), the maximum of the upper bounds of ρ_e and ρ_c , denoted as ρ , and the upper bound of φ , denoted as f . These choices minimize the efficiency or maximize $V(\hat{\beta}_1)$ in Equation (5). The question arises if the parameter values are not covered by the boundaries as assumed in the maximin strategy. In that case, the value of $V(\hat{\beta}_1)$ may become larger than the MMV . So, we have to check how robust the maximin sample sizes are against misspecification of boundaries of the parameter intervals. Here, misspecification means that the parameter values for ρ_e, ρ_c , and φ are larger than their assumed upper boundaries in the maximin strategy, and the parameter values for ρ_{uv} and $\rho_{\delta\varepsilon}$ are lower than their assumed lower boundaries.

Let $V(\hat{\beta}_1 | \Pi_1, MMD)$ be the variance of the treatment effect estimator in the maximin design ($m_{MMD}, n_{MMD}, k_{tMMD}$, and k_{cMMD}) for parameters outside the maximin boundaries, that is, $\Pi_1 = (\rho_e, \rho_c, \rho_{uv}, \rho_{\delta\varepsilon}, \varphi)$, with $\rho_e \geq \rho, \rho_c \geq \rho, \rho_{uv} \leq r, \rho_{\delta\varepsilon} \leq r$, and $\varphi \geq f$. Robustness is the extent to which results remain the same if the assumptions that are made to obtain those results are violated [19]. In our paper, the relative efficiency loss is used to quantify the degree of robustness of a design. The robustness

of the maximin design (*MMD*) against incorrect specification of the parameters $\rho_e, \rho_c, \rho_{uv}, \rho_{\delta\epsilon}$, and φ is expressed in terms of the following relative efficiency loss:

$$\begin{aligned}
 REL &= \frac{\{V(\hat{\beta}_1|\Pi^*, MMD)\}^{-1} - \{V(\hat{\beta}_1|\Pi_1, MMD)\}^{-1}}{\{V(\hat{\beta}_1|\Pi^*, MMD)\}^{-1}} \\
 &= 1 - \frac{V(\hat{\beta}_1|\Pi^*, MMD)}{V(\hat{\beta}_1|\Pi_1, MMD)} \\
 &= 1 - \frac{\{\rho + W(1 - \rho)\}(f + 1 - 2r\sqrt{f})}{(\varphi\rho_e + \rho_c - 2\rho_{uv}\sqrt{\rho_e\rho_c\varphi}) + W\{\varphi(1 - \rho_e) + (1 - \rho_c) - 2\rho_{\delta\epsilon}\sqrt{(1 - \rho_e)(1 - \rho_c)\varphi}\}},
 \end{aligned} \tag{13}$$

where $W = \frac{(\sqrt{s_t} + \sqrt{s_c})\sqrt{\rho}}{(\sqrt{c_t} + \sqrt{c_c})\sqrt{1 - \rho}}$.

When $\Pi_1 = \Pi^*$, that is, $\rho_{uv} = \rho_{\delta\epsilon} = r, \rho_e = \rho_c = \rho$ and $\varphi = f$, there is no misspecification, and the relative efficiency loss (*REL*) in Equation (13) is equal to zero. If $\rho_{uv} < r$ or $\rho_{\delta\epsilon} < r$, the *REL* becomes greater than zero. In the same way, the *REL* will be greater than zero if $\rho_e > \rho, \rho_c > \rho$ or $\varphi > f$. A *REL* greater than zero means that the variance $V(\hat{\beta}_1|\Pi_1, MMD)$ is larger than the variance $V(\hat{\beta}_1|\Pi^*, MMD)$, and we lose efficiency because of misspecification of parameters in the maximin design. The value for *REL* expresses the relative amount of efficiency loss due to misspecification of parameters in the *MMD*. Also Equation (13) shows that the *REL* depends on W , which is a function of the subject-specific costs and cluster-specific costs and ρ but is independent of the total budget, C .

Table I displays the parameter boundaries, which are used to investigate the *REL* numerically. The *MMD* and its *MMV* depend on the lower bounds for ρ_{uv} and $\rho_{\delta\epsilon}$ and the upper bounds for ρ_e, ρ_c and φ , so we consider only these bounds. In the robustness study, we consider the values of the parameters that are 1.33 times and twice as large as the values used in the *MMD*. Furthermore, Table I shows two different sets of maximin boundaries: one is from a review (based on 62 empirical studies) by Gomes *et al.* [18] except for φ , which is based on the six other studies [4, 10, 20–23], and another one in which the parameter spaces are assumed to be smaller. In some studies, researchers may find the Gomes *et al.* [18] intervals too broad. For instance, some other systematic reviews indicate that the ICC for effects rarely exceeds 0.025 in medical and health science studies [24]. As another illustration, a smaller value of the ICC for effects ($\rho_e = 0.04$) is reported by Morrell *et al.* [2] in a healthcare intervention study to be discussed in Section 7. This motivates considering a second set of smaller intervals and corresponding boundaries. So Table I contains large and small maximin boundaries and the corresponding boundary values as used in the robustness study to examine the effect of misspecification. As an example, the ICC for effects ρ_e has upper boundaries 0.3 and 0.1, respectively, which are used in the maximin strategy for two different *MMDs*. Then, in the robustness study, we consider $\rho_e = 0.4$ and $\rho_e = 0.13$ as 25% misspecifications of these boundaries, respectively, to study how much efficiency the *MMD* loses.

Table I. Boundary values used in the numerical robustness study of the maximin design (*MMD*).

Parameters	Assumed boundary values in the <i>MMD</i>		Boundary values in the robustness study			
			Large boundary		Small boundary	
	Large boundary	Small boundary	25% misspecification	50% misspecification	25% misspecification	50% misspecification
ρ_e (UB)	0.3	0.1	0.4	0.6	0.13	0.2
ρ_c (UB)	0.3	0.1	0.4	0.6	0.13	0.2
φ (UB)	50	20	66.67	100	26.67	40
ρ_{uv} (LB)	−0.5	−0.3	−0.67	−1.0	−0.4	−0.6
$\rho_{\delta\epsilon}$ (LB)	−0.5	−0.3	−0.67	−1.0	−0.4	−0.6

Note: The large maximin boundary values are based on a review by Gomes *et al.* [18] except for φ , which is based on the six other studies.

UB, upper bound; LB, lower bound.

Let $REL(\rho_e)$ and $REL(\rho_e^{25\%})$ be the relative efficiency losses in the *MMD* when the ICC for effects, ρ_e , is not misspecified (that is, $\rho_e = \rho$) and is misspecified by 25% (that is, $\rho_e = \rho(1-0.25)^{-1}$), respectively. In both cases, the other parameters $\Theta = (\rho_c, \rho_{uv}, \rho_{\delta\epsilon}, \varphi)$, with $\rho_c \geq \rho$, $\rho_{uv} \leq r$, $\rho_{\delta\epsilon} \leq r$, and $\varphi \geq f$, may or may not be misspecified, by either 25% or 50%. Then we define $\Delta_{\rho_e^{25\%}} = REL(\rho_e^{25\%}) - REL(\rho_e)$, which can be shown to be equal to

$$\Delta_{\rho_e^{25\%}} = \frac{\{V(\hat{\beta}_1 | \rho_e = \rho, \Theta, MMD)\}^{-1} - \{V(\hat{\beta}_1 | \rho_e^{25\%}, \Theta, MMD)\}^{-1}}{\{V(\hat{\beta}_1 | \Pi^*, MMD)\}^{-1}} \tag{14}$$

Equation (14) represents the amount of *REL* that is only due to misspecification of the ICC for effects, ρ_e , by 25%. This contribution to *REL* can be examined with other parameters $\Theta = (\rho_c, \rho_{uv}, \rho_{\delta\epsilon}, \varphi)$ possibly also being misspecified. Further, let $\bar{\Delta}_{\rho_e^{25\%}}$ be the efficiency loss of misspecifying the ICC for effects ρ_e by 25%, averaged across all possible choices of the other parameters $\Theta = (\rho_c, \rho_{uv}, \rho_{\delta\epsilon}, \varphi)$ that are misspecified maximally by 50%. The notation $\Delta_{\rho_e^{25\%}}$ can be generalized to all the parameters by the term $\Delta_{\theta p}$ and the corresponding average by $\bar{\Delta}_{\theta p}$, where p is either 25% or 50%, and θ stands for any one of $\rho_e, \rho_c, \rho_{uv}, \rho_{\delta\epsilon}$, and φ .

Table II displays the minimum, average ($\bar{\Delta}_{\theta p}$), and maximum efficiency loss for the misspecifications of each parameter in the case the maximin design is based on the large maximin boundaries in Table I. These results are shown for different values of W between 0 and 0.65. Notice that W in Equation (13)

Table II. Minimum (Min), average (Ave), and maximum (Max) efficiency losses for 25% and 50% misspecifications of each parameter for different values of W , when the other parameters are misspecified 50% at most.

W	Contribution to <i>REL</i> ($\Delta_{\theta p}$)	ICC for effects (ρ_e)		ICC for costs (ρ_c)		Correlation between costs and effects at cluster level (ρ_{uv})		Correlation between costs and effects at individual level ($\rho_{\delta\epsilon}$)		Variance ratio (φ)	
		25%	50%	25%	50%	25%	50%	25%	50%	25%	50%
0.00	Min	0.1001	0.2058	0.0031	0.0083	0.0059	0.0166	0.0000	0.0000	0.1001	0.2058
	Ave	0.1615	0.3307	0.0132	0.0345	0.0211	0.0579	0.0000	0.0000	0.1616	0.3307
	Max	0.2342	0.4768	0.0332	0.0849	0.0486	0.1301	0.0000	0.0000	0.2342	0.4768
0.10	Min	0.0811	0.1779	0.0029	0.0076	0.0062	0.0177	0.0004	0.0014	0.1157	0.2379
	Ave	0.1277	0.2800	0.0108	0.0279	0.0191	0.0528	0.0026	0.0078	0.1721	0.3521
	Max	0.1823	0.3989	0.0265	0.0680	0.0413	0.1122	0.0077	0.0225	0.2342	0.4768
0.20	Min	0.0656	0.1517	0.0026	0.0066	0.0064	0.0183	0.0010	0.0030	0.1295	0.2663
	Ave	0.1013	0.2347	0.0089	0.0226	0.0175	0.0489	0.0048	0.0141	0.1808	0.3698
	Max	0.1427	0.3309	0.0216	0.0550	0.0359	0.0985	0.0130	0.0373	0.2342	0.4768
0.30	Min	0.0526	0.1273	0.0019	0.0036	0.0065	0.0186	0.0016	0.0047	0.1418	0.2915
	Ave	0.0800	0.1941	0.0073	0.0182	0.0163	0.0456	0.0067	0.0194	0.1881	0.3847
	Max	0.1115	0.2711	0.0177	0.0449	0.0317	0.0878	0.0167	0.0477	0.2342	0.4768
0.40	Min	0.0407	0.1028	0.0008	0.0004	0.0066	0.0188	0.0022	0.0064	0.1527	0.3141
	Ave	0.0626	0.1576	0.0060	0.0145	0.0152	0.0429	0.0083	0.0240	0.1944	0.3975
	Max	0.0876	0.2200	0.0147	0.0367	0.0284	0.0791	0.0195	0.0555	0.2342	0.4768
0.50	Min	0.0306	0.0801	0.0000	-0.0042	0.0065	0.0188	0.0028	0.0082	0.1626	0.3344
	Ave	0.0481	0.1247	0.0049	0.0112	0.0144	0.0406	0.0098	0.0280	0.1998	0.4086
	Max	0.0681	0.1764	0.0132	0.0334	0.0257	0.0720	0.0218	0.0615	0.2342	0.4768
0.60	Min	0.0217	0.0583	-0.0015	-0.0095	0.0059	0.0171	0.0034	0.0099	0.1715	0.3527
	Ave	0.0358	0.0949	0.0039	0.0083	0.0136	0.0386	0.0110	0.0315	0.2046	0.4183
	Max	0.0535	0.1388	0.0127	0.0318	0.0246	0.0685	0.0235	0.0663	0.2342	0.4768
0.65	Min	0.0178	0.0482	-0.0022	-0.0117	0.0056	0.0163	0.0037	0.0108	0.1757	0.3613
	Ave	0.0303	0.0810	0.0035	0.0070	0.0133	0.0376	0.0116	0.0331	0.2068	0.4227
	Max	0.0469	0.1228	0.0124	0.0309	0.0246	0.0687	0.0243	0.0683	0.2342	0.4768

Note: In the maximin design, the ICC for effects and costs is set to $\rho = 0.3$, the cluster and individual level correlations are set to $r = -0.5$ based on Gomes *et al.* [18], and the variance ratio is set to $f = 50$ (from six other studies). W is a function of cluster-specific and subject-specific costs and ρ as defined in Equation (13).

cannot be larger than 0.65, as $c_t \geq s_t, c_c \geq s_c$, and $\rho = 0.3$. Table II shows that, when the ICC for effects, ρ_e , is misspecified by 25% and 50%, the maximum contribution to the efficiency loss can become 23.4% and 47.7%, respectively, occurring at $W = 0$. The case $W = 0$ means that the cost at cluster level is infinitely large compared with the subject-specific cost, therefore being a rather exceptional case. The contribution to the efficiency loss due to misspecification of ρ_e decreases with W . Similarly, the ICC for costs, ρ_c , contributes maximally to REL at $W = 0$: its contribution is 3.3% and 8.5%, for 25% and 50% misspecifications of ρ_c , respectively.

Table II shows that the contribution to the efficiency loss is much larger for ρ_e than for ρ_c . The efficiency loss of misspecifying ρ_e and ρ_c depends on the upper bound of φ . It can be shown that misspecification of ρ_e and ρ_c contributes equally to REL if $\varphi = 1$. If $\varphi > 1$, then ρ_e contributes to REL more than ρ_c , but if $\varphi < 1$, then ρ_c contributes more than ρ_e . The explanation is that if $\varphi > 1$, then the variance of effects translated into a monetary scale (λE_{ij}) is larger than the total variance for costs (C_{ij}). In that case, changing ρ_e has a larger impact on variance of the treatment effect estimator $V(\hat{\beta}_1)$ than changing ρ_c (Equation (5)). This in turn implies that misspecifying ρ_e will lead to stronger increase of the $V(\hat{\beta}_1)$ than misspecifying ρ_c .

For $W \geq 0.50$, the minimum efficiency loss due to misspecification of ρ_c can become negative (Table II), which means that we obtain some efficiency gain due to misspecification. This occurs in a few cases when other parameters are also misspecified. It should be noted that $W \geq 0.50$ results in maximin cluster sizes less than or equal to two in at least one of the treatment conditions. This follows from the definition of W (in terms of costs) and the results in Equation (11). So the efficiency gain

Table III. Minimum (Min), average (Ave), and maximum (Max) efficiency losses for 25% and 50% misspecifications of each parameter for different values of W , when the other parameters are misspecified 50% at most.

W	Contribution to REL ($\Delta_{\theta\rho}$)	ICC for effects (ρ_e)		ICC for costs (ρ_c)		Correlation between costs and effects at cluster level (ρ_{uv})		Correlation between costs and effects at individual level ($\rho_{\delta\varepsilon}$)		Variance ratio (φ)	
		25%	50%	25%	50%	25%	50%	25%	50%	25%	50%
		0.00	Min	0.0915	0.2054	0.0033	0.0098	0.0055	0.0160	0.0000	0.0000
	Ave	0.1466	0.3275	0.0142	0.0415	0.0195	0.0548	0.0000	0.0000	0.1603	0.3299
	Max	0.2120	0.4713	0.0344	0.0978	0.0429	0.1177	0.0000	0.0000	0.2303	0.4713
0.05	Min	0.0680	0.1674	0.0029	0.0088	0.0053	0.0156	0.0013	0.0038	0.1202	0.2482
	Ave	0.1057	0.2603	0.0106	0.0314	0.0154	0.0440	0.0049	0.0144	0.1739	0.3575
	Max	0.1499	0.3685	0.0248	0.0717	0.0318	0.0893	0.0115	0.0339	0.2303	0.4713
0.10	Min	0.0528	0.1381	0.0025	0.0076	0.0050	0.0146	0.0025	0.0075	0.1350	0.2788
	Ave	0.0807	0.2117	0.0084	0.0249	0.0128	0.0370	0.0081	0.0238	0.1828	0.3757
	Max	0.1131	0.296	0.0190	0.0556	0.0252	0.0719	0.0175	0.0509	0.2303	0.4713
0.15	Min	0.0423	0.1153	0.0022	0.0065	0.0046	0.0136	0.0036	0.0108	0.1461	0.3017
	Ave	0.0639	0.1751	0.0068	0.0203	0.0110	0.0321	0.0105	0.0306	0.1891	0.3886
	Max	0.0886	0.2434	0.0152	0.0446	0.0209	0.0601	0.0212	0.0611	0.2303	0.4713
0.20	Min	0.0345	0.0972	0.0018	0.0054	0.0043	0.0126	0.0046	0.0136	0.1547	0.3195
	Ave	0.0517	0.1465	0.0057	0.0168	0.0097	0.0284	0.0123	0.0357	0.1938	0.3983
	Max	0.0713	0.2025	0.0125	0.0368	0.0178	0.0516	0.0237	0.0678	0.2303	0.4713
0.25	Min	0.0282	0.0824	0.0015	0.0043	0.0039	0.0117	0.0055	0.0160	0.1616	0.3338
	Ave	0.0426	0.1236	0.0048	0.0141	0.0087	0.0254	0.0138	0.0398	0.1975	0.4058
	Max	0.0586	0.1700	0.0105	0.0308	0.0156	0.0454	0.0254	0.0727	0.2303	0.4713
0.30	Min	0.0231	0.0691	0.0011	0.0030	0.0035	0.0105	0.0062	0.0182	0.1673	0.3455
	Ave	0.0354	0.1049	0.0041	0.0119	0.0078	0.0230	0.0149	0.0431	0.2004	0.4119
	Max	0.0496	0.1445	0.0090	0.0265	0.0141	0.0413	0.0268	0.0763	0.2303	0.4713
0.33	Min	0.0206	0.0622	0.0009	0.0024	0.0033	0.0097	0.0066	0.0193	0.1702	0.3515
	Ave	0.0319	0.0951	0.0037	0.0108	0.0074	0.0218	0.0156	0.0447	0.2019	0.4149
	Max	0.0451	0.1326	0.0084	0.0246	0.0137	0.0398	0.0274	0.0781	0.2303	0.4713

Note: In the maximin design, the ICC for effects and costs are set to $\rho = 0.1$, the cluster and individual level correlations are set to $r = -0.3$, and the variance ratio is set to $f = 20$. W is a function of cluster-specific and subject-specific costs and ρ as defined in Equation (13).

only occurs when at least one of the treatment conditions has clusters with one or two individuals in each cluster. In practice, because of logistic reasons, CRTs often will have a larger number of persons in each cluster.

For the correlation at cluster level, ρ_{uv} , the maximum efficiency losses occur at $W = 0$, which are 4.9% and 13%, respectively. Misspecifying $\rho_{\delta\epsilon}$ leads to an even smaller efficiency loss when compared with the efficiency loss for misspecifying ρ_{uv} . The increases of REL due to misspecification of $\rho_{\delta\epsilon}$ become maximum at $W = 0.65$, which are 2.4% and 6.8%. The value 0.65 for W is its maximum for $\rho = 0.3$ and occurs when spending the same costs for selecting a cluster and a subject in each of the treatment arms. The effect of misspecification of the variance ratio φ is largest compared with the effect of misspecifying the other parameters. The largest contribution to REL due to misspecification of φ is about 47.7%.

In Table III, we examined the robustness of the MMD when the smaller boundaries of the parameters in Table I are taken as the maximin boundaries. The maximum value of W for $\rho = 0.1$ becomes 0.33. The results in Table III make clear that the efficiency losses of the MMD due to misspecification of parameters are similar to those in Table II. In addition, all possible combinations of small and large parameter boundaries additionally yield six different sets of maximin boundaries, which were also investigated. The results were analogous to those shown in Tables II and III and for that reason were omitted.

The numerical evaluations in Tables II and III show that the MMD is rather robust against misspecification of the parameters ρ_{uv} , $\rho_{\delta\epsilon}$, and ρ_c , whereas misspecifying the parameters ρ_e and φ result in relatively large efficiency losses. The robustness of the MMD against misspecification of the model parameters can be summarized as follows:

- (a) The contribution to REL due to misspecification of ρ_{uv} or $\rho_{\delta\epsilon}$ never exceeds 13%, and misspecification of the individual level correlation $\rho_{\delta\epsilon}$ leads to less efficiency loss than ρ_{uv} . So the MMD is rather robust against misspecification of the correlations at the cluster and individual levels.
- (b) We assumed an upper boundary for φ based on six available studies. For this upper bound and a smaller one (Table I), the MMD turned out to be rather sensitive to misspecification of this parameter.
- (c) For the upper boundaries of φ considered in this study, the MMD is robust under misspecification of the ICC for costs ρ_c , whereas the MMD is sensitive to the misspecification of the ICC for effects ρ_e . The maximum contribution to REL is about 47.7%.

6. Power of the test on cost-effectiveness of a treatment

The average $INMB$ (β_1) represents the cost-effectiveness of an intervention. Suppose we want to test the null hypothesis $H_0 : \beta_1 = 0$ against the alternative hypothesis $H_A : \beta_1 > 0$, that is, the intervention is cost-effective. To test these hypotheses, use can be made of the ML estimator $\hat{\beta}_1$ (Equation (4)), which is asymptotically normally distributed with variance $V(\hat{\beta}_1)$. The ratio $\hat{\beta}_1 / \sqrt{V(\hat{\beta}_1)}$ can be approximated by a standard normal distribution if the number of clusters is large enough [25]. For a one-sided test with alternative hypothesis $H_A : \beta_1 > 0$, with power $(1 - \gamma)$ at β_1 and with a type I error rate α , we have [25, p. 142]

$$(z_{1-\alpha} + z_{1-\gamma})^2 = \frac{\beta_1^2}{V(\hat{\beta}_1)}, \tag{15}$$

where $z_{1-\alpha}$ and $z_{1-\gamma}$ are the $(1 - \alpha)$ and $(1 - \gamma)$ percentiles of the standard normal cumulative distribution function. For a two-sided alternative $H_A : \beta_1 \neq 0$, the percentile value $z_{1-\alpha}$ in Equation (15) is replaced by $z_{1-\alpha/2}$. Cohen [26] introduced the effect size $ES = |\beta_1| / \sqrt{V(NMB_{ij})}$, where $V(NMB_{ij})$ is the variance of NMB_{ij} in Equation (3) and also proposed a classification into small (0.2), medium (0.5), and large (0.8) effect sizes. Equation (15) can be expressed in terms of the effect size (ES) in the following way:

$$(z_{1-\alpha} + z_{1-\gamma})^2 = \frac{(ES)^2(A + B)}{\left(\frac{1}{k_t} + \frac{1}{k_c}\right)A + \left(\frac{1}{k_{tm}} + \frac{1}{k_{cn}}\right)B}, \tag{16}$$

where A and B are defined as in Equation (5).

7. Application in planning a cost-effectiveness CRT

The optimal design as well as the *MMD* for the nested *NMB* model will be applied to a case study known as the PONDER study [2]. This study involves the assessment of two alternative interventions compared with usual care for preventing postnatal depression. The clusters (here: general practices) were randomly allocated to either a control group (usual care) or a treatment group (one of the two interventions termed as the Person Centered Approach and the Cognitive-Behavioral Approach). To establish the cost-effectiveness of the treatment group compared with the control group, data were obtained from 87 general practices ($k_t = 58$ treatment clusters and $k_c = 29$ control clusters) with, on the average, $m = n = 6$ women in each general practice. The effects and costs were expressed in terms of QALYs and UK pounds (£), respectively. A high ICC was found for costs, $\rho_c = 0.17$, but not for effects $\rho_e = 0.04$. There was a negative individual level correlation ($\rho_{\delta\epsilon} = -0.06$) between costs and effects, whereas the correlation between costs and effects at the cluster level was positive ($\rho_{uv} = 0.04$). For a ceiling ratio of $\lambda = 20,000$, the variance ratio φ turned out to be 4.32.

The design of the PONDER study is $k_t = 58, k_c = 29, m = n = 6$, which we call the actual design. In the sequel, the optimal design of the PONDER study is examined, based on the parameter estimates as found in the PONDER study by Morrell *et al.* [2], $\Pi_{PONDER} = (\rho_e = 0.04, \rho_c = 0.17, \rho_{uv} = 0.04, \rho_{\delta\epsilon} = -0.06, \varphi = 4.32)$. If we know the actual parameter values, we can calculate the optimal design. As this may not be the case in practice, a solution is to formulate a *MMD* based on the large maximin boundaries in Table I, $\Pi_{LARGE} = (\rho_e = 0.3, \rho_c = 0.3, \rho_{uv} = -0.5, \rho_{\delta\epsilon} = -0.5, \varphi = 50)$. Table IV shows four relative efficiencies of the aforementioned three designs. As the exact cost ratios for this study were not

Table IV. Relative efficiencies of the actual design of the PONDER study [2] (*ACT*), the maximin design (*MMD*), and the optimal design (*opt*), for the parameters in the PONDER study, Π_{PONDER} , and the parameters in Π_{LARGE} , specified for different cost ratios.

Cost ratio for treatment condition (c_t/s_t)	Cost ratio for control condition (c_c/s_c)	Subject specific cost ratio (s_t/s_c)	$\frac{V(\hat{\beta}_1 \Pi_{PONDER,opt})}{V(\hat{\beta}_1 \Pi_{PONDER,ACT})}$	$\frac{V(\hat{\beta}_1 \Pi_{PONDER,opt})}{V(\hat{\beta}_1 \Pi_{PONDER,MMD})}$	$\frac{V(\hat{\beta}_1 \Pi_{LARGE,MMD})}{V(\hat{\beta}_1 \Pi_{LARGE,ACT})}$	$\frac{V(\hat{\beta}_1 \Pi_{LARGE,MMD})}{V(\hat{\beta}_1 \Pi_{LARGE,opt})}$
2	2	1	0.89	0.84	0.69	0.80
2	10	1	0.90	0.81	0.87	0.81
2	50	1	0.69	0.80	0.89	0.84
10	2	1	0.72	0.81	0.70	0.81
10	10	1	0.78	0.80	0.88	0.82
10	50	1	0.69	0.80	0.96	0.85
50	2	1	0.41	0.80	0.53	0.84
50	10	1	0.47	0.80	0.66	0.85
50	50	1	0.53	0.81	0.84	0.87
2	2	4	0.67	0.84	0.52	0.80
2	10	4	0.74	0.82	0.67	0.81
2	50	4	0.73	0.80	0.81	0.82
10	2	4	0.52	0.81	0.54	0.81
10	10	4	0.58	0.80	0.66	0.82
10	50	4	0.64	0.80	0.84	0.84
50	2	4	0.30	0.80	0.43	0.85
50	10	4	0.34	0.81	0.50	0.86
50	50	4	0.40	0.81	0.63	0.87
2	2	9	0.56	0.84	0.44	0.80
2	10	9	0.62	0.82	0.55	0.80
2	50	9	0.68	0.81	0.73	0.82
10	2	9	0.45	0.81	0.47	0.81
10	10	9	0.49	0.80	0.56	0.82
10	50	9	0.56	0.80	0.72	0.84
50	2	9	0.27	0.81	0.39	0.86
50	10	9	0.29	0.81	0.44	0.86
50	50	9	0.34	0.81	0.53	0.87

Note: Maximin design (*MMD*) is always based on the parameter vector Π_{LARGE} in Table I, and the optimal design (*opt*) is based on the parameter estimates as found in the PONDER study, Π_{PONDER} .

known, we evaluated the efficiencies for several cost ratios, comprising the cost ratios of other studies. Raudenbush [5] used 2, 10, and 50 as the values for cluster to subject-specific cost ratios ($c_t/s_t, c_c/s_c$), and, similar to Liu [27], the subject-specific cost ratios of treatment to control arm (s_t/s_c) were equal to 4 and 9. For the first two relative efficiencies in Table IV, it is assumed that Π_{PONDER} is the set of parameter values; therefore, the *MMD* and actual design become suboptimal compared with the optimal design of the PONDER study. The first relative efficiency shows that the optimal design can become much more efficient than the actual design (up to three times as efficient) especially for large cost ratios. The second relative efficiency shows that the *MMD* loses at most 20% efficiency compared with the optimal design. The last two columns of Table IV assume Π_{LARGE} is the set of parameter values. In that case, the *MMD* becomes optimal compared with the other two designs. The efficiency of the actual design to the *MMD* is shown in the third relative efficiency column. As can be seen, the *MMD* can also be much more efficient than the actual design of the PONDER study. The last column shows that the design, which is optimal for Π_{PONDER} , is at most 20% less efficient than the *MMD* if the large maximin boundaries in Table I, Π_{LARGE} , are assumed to be the parameter values.

Next we consider how the optimal design, the *MMD*, and the actual design of the PONDER study relate to power for varying effect size (*ES*). Figure 1 displays the power against effect size for the optimal design (Equations (7) and (8)), the *MMD* (Equation (11)), and the actual design of the PONDER study for $\alpha = 0.05$ in a one-sided test. We compared these three designs, assuming Π_{PONDER} is the parameter vector, for four different sets of cost ratios. Notice that the *MMD* is based on the large ICC value $\rho = 0.3$ for both costs and effects, which results in smaller power for the *MMD* than the optimal design. The power of the optimal design and the *MMD* are larger than the actual design of the PONDER study. This is in line with the results of Table IV. The difference in power between the optimal design and actual design of the PONDER study increases as the cost ratios increase and the maximum difference

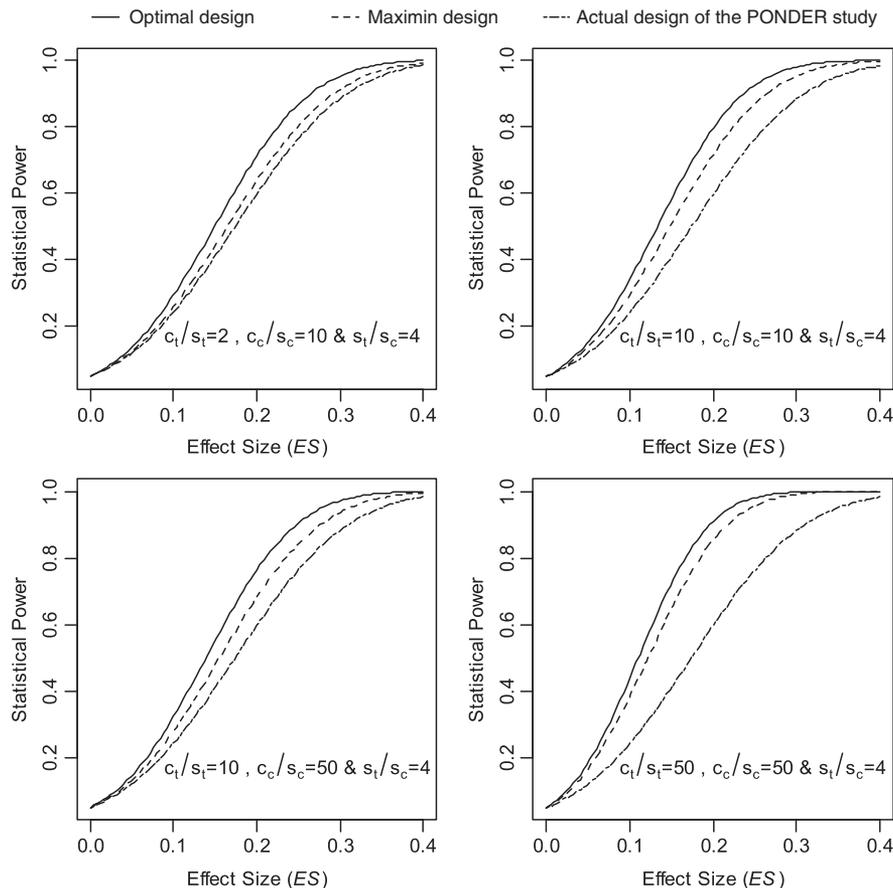


Figure 1. Statistical power for the optimal design, the *MMD*, and the actual design of the PONDER study for the parameter vector $\Pi_{PONDER} = (\rho_e = 0.04, \rho_c = 0.17, \rho_{uv} = 0.04, \rho_{\delta\varepsilon} = -0.06, \varphi = 4.32)$ as a function of the effect size, $ES = |\beta_1| / \sqrt{V(NMB_{ij})}$, under four different sets of cost ratios.

ranges between 0.08 and 0.33 for the effect sizes around 0.20. On the contrary, the power difference between the optimal design and the *MMD* remains approximately the same, even for larger cost ratios.

8. Discussion and conclusions

This paper makes two important contributions to sample size calculation in cost-effectiveness CRTs. First, our study takes into account the hierarchical characteristic of cost-effectiveness data, which were ignored by approaches available in the literature on sample size calculation for cost-effectiveness studies. Second, we propose formulas for sample size calculation that maximize the efficiency of the treatment effect estimator given a budget constraint. The hierarchical net-benefit regression was taken as a framework for the CEA of interventions. More specifically, we started from the model in Equation (3) that relates treatment condition to a quantitative outcome known as net monetary benefit. The model furthermore assumes that there are random cluster effects, thus taking into account the dependency between subjects within the same cluster. In this model, commonly the average *INMB* is of main interest, and an optimal design was derived to estimate the average *INMB* as efficiently as possible given a fixed research budget.

The optimal sample sizes require information on the ICCs for effects and costs, the correlations between costs and effects at the individual and cluster levels, the ratio of the variance of effects translated into costs to the variance of the costs (the variance ratio), sampling and measuring costs, and the budget. When planning a study, information on the model parameters usually is not available, and a maximin design (*MMD*) is proposed to handle this local optimality problem. A robustness study was carried out on the effects of misspecifying parameters in the maximin design. Although the numerical evaluation shows the *MMD* to be robust against misspecification of the ICC for costs and the correlation between costs and effects at the cluster and individual levels, the *MMD* may lose much efficiency because of misspecification of the ICC for effects and the variance ratio. Note that the upper boundaries for the variance ratio used in the numerical evaluation are greater than one, which implies more efficiency loss when the ICC for effects is misspecified. On the other hand, if the upper bound for the variance ratio is less than one, the *MMD* will become more sensitive against the misspecification of the ICC for costs. Further, as the ceiling ratio λ determines whether the variance ratio is greater than, or less than, or equal to one, λ also determines whether the *MMD* is more sensitive against misspecification of the ICC for effects or costs. The empirical evidence on the variance ratio is rather scarce, and its misspecification in the *MMD* showed large efficiency losses, so future studies should provide specific results on this parameter. Finally, a practical expression is given in Equation (16), which shows the relation between the sample sizes (either optimal or maximin), effect size, and the power of a test on the cost-effectiveness of two interventions.

The current study has some limitations. First, the distribution of costs as part of the outcome variable is assumed to be normal, but in practice, cost data may be positively skewed. However, some studies support the robustness against violating the normality assumption in hierarchical modeling [28, 29]. Note that the estimator $\hat{\beta}_1$ in Equation (4) is the difference between the averages of outcomes in the treatment arms. By the central limit theorem, these averages will be approximately normally distributed, even for skewed cost data, if there is a large number of clusters. Second, we assumed only equal numbers of individuals in each cluster in the treatment and control conditions, but in practice, because of non-response and dropout, cluster sizes may become unequal. One solution could be to use the correction methods as proposed by Van Breukelen *et al.* [30] and Candel *et al.* [31] to adjust the sample sizes as computed under the assumption of equal cluster sizes. Third, this study did not include any covariates in the models in Equations (1) and (3). However, for CRTs, the sample size formulas for a model without covariates also hold for a model with covariates [32]. Of course, adding covariates at either the cluster level or the individual level may decrease the variance components, which results in more precise estimates of the treatment's cost-effectiveness and increases statistical power. Hence if we would include covariates in the analysis, but would ignore this in the sample size calculation, our formulas would lead to the sample sizes that are too large, therefore being on the safe side in terms of efficiency. Obtaining cost-efficient sample sizes would of course require estimates of the variance components corrected for covariates. Fourth, the cluster level variance and the cluster level correlation, together with the individual level variance and individual level correlation, are assumed to be equal between the treatment arms. As some reviews [4, 33] suggest that the treatments may have unequal variances and correlations, an interesting issue for future research would be how the optimal and maximin design approach could account for this between-arm heterogeneity of model parameters.

Appendix A. Derivation of the optimal number of clusters and cluster sizes

The proof involves the following steps: First, we express the variance formula of the ML estimator of the treatment effect ($\hat{\beta}_1$) in terms of the model parameters $\rho_e, \rho_c, \rho_{uv}, \rho_{\delta\varepsilon}$, and φ . Second, we have to define a budget constraint and then obtain the number of clusters and cluster sizes, which minimize the variance of $\hat{\beta}_1$ subject to the specified budget constraint.

Step 1: The asymptotic variance of the ML estimator is [16, pp. 46–49]

$$V(\hat{\beta}_1) = \left(\frac{1}{k_t} + \frac{1}{k_c}\right) V(w_{0j}) + \left(\frac{1}{k_tm} + \frac{1}{k_cn}\right) V(\omega_{ij}), \tag{A.1}$$

in which there are variances $V(w_{0j}) = V(\lambda u_{0j} - v_{0j}) = \lambda^2 \sigma_u^2 + \sigma_v^2 - 2\lambda \rho_{uv} \sigma_u \sigma_v$ and $V(\omega_{ij}) = V(\lambda \delta_{ij} - \varepsilon_{ij}) = \lambda^2 \sigma_\delta^2 + \sigma_\varepsilon^2 - 2\lambda \rho_{\delta\varepsilon} \sigma_\delta \sigma_\varepsilon$. Now, $V(w_{0j}) = \lambda^2 \sigma_u^2 + \sigma_v^2 - 2\lambda \rho_{uv} \sigma_u \sigma_v = \lambda^2 \rho_e V(E_{ij}) + \rho_c V(C_{ij}) - 2\lambda \rho_{uv} \sqrt{\rho_e \rho_c} \sqrt{V(E_{ij})V(C_{ij})} = \left[\rho_e \left\{\frac{\lambda^2 V(E_{ij})}{V(C_{ij})}\right\} + \rho_c - 2\rho_{uv} \left\{\frac{\lambda \sqrt{V(E_{ij})}}{\sqrt{V(C_{ij})}}\right\} \sqrt{\rho_e \rho_c}\right] V(C_{ij}) = (\varphi \rho_e + \rho_c - 2\rho_{uv} \sqrt{\rho_e \rho_c \varphi}) V(C_{ij})$. Similarly, $V(\omega_{ij}) = \left\{\varphi(1 - \rho_e) + (1 - \rho_c) - 2\rho_{\delta\varepsilon} \sqrt{(1 - \rho_e)(1 - \rho_c)\varphi}\right\} V(C_{ij})$. Equation (A.1) then becomes

$$\begin{aligned} V(\hat{\beta}_1) &= \left[\left(\frac{1}{k_t} + \frac{1}{k_c}\right) (\varphi \rho_e + \rho_c - 2\rho_{uv} \sqrt{\rho_e \rho_c \varphi}) \right. \\ &\quad \left. + \left(\frac{1}{k_tm} + \frac{1}{k_cn}\right) \left\{\varphi(1 - \rho_e) + (1 - \rho_c) - 2\rho_{\delta\varepsilon} \sqrt{(1 - \rho_e)(1 - \rho_c)\varphi}\right\}\right] V(C_{ij}) \\ &= \left\{\left(\frac{1}{k_t} + \frac{1}{k_c}\right) A + \left(\frac{1}{k_tm} + \frac{1}{k_cn}\right) B\right\} V(C_{ij}), \end{aligned} \tag{A.2}$$

where $A = \varphi \rho_e + \rho_c - 2\rho_{uv} \sqrt{\rho_e \rho_c \varphi}$ and $B = \varphi(1 - \rho_e) + (1 - \rho_c) - 2\rho_{\delta\varepsilon} \sqrt{(1 - \rho_e)(1 - \rho_c)\varphi}$.

Step 2: To obtain the optimal sample sizes, we have to minimize $V(\hat{\beta}_1)$ in Equation (A.2), under a budget constraint $C = k_t c_t + k_c c_c + k_t m s_t + k_c n s_c$, where $c_t > 0, c_c > 0, s_t > 0$, and $s_c > 0$, and m, n, k_t , and k_c are positive integers, which give the following objective function:

$$L = V(\hat{\beta}_1) + \Lambda(k_t c_t + k_c c_c + k_t m s_t + k_c n s_c - C), \tag{A.3}$$

where Λ is the Lagrange multiplier. Setting the partial derivatives of Equation (A.3) with respect to m, n, k_t, k_c , and Λ equal to zero gives the optimal number of clusters and sample size per cluster in each of the treatment arms, as displayed in Equations (7) and (8) of the main text. Finally, inserting Equations (7) and (8) into Equation (A.2) yields the optimal variance as shown in Equation (9) of the main text.

Appendix B: Derivation of the maximin sample sizes and its MMV

First, we have to find out which values of the parameters in the assumed ranges minimize the efficiency of the treatment effect estimator. The efficiency of the treatment effect estimator can be derived from Equation (A.2) as

$$\begin{aligned} Eff &= \left\{V(\hat{\beta}_1)\right\}^{-1} = \left[P(\varphi \rho_e + \rho_c - 2\rho_{uv} \sqrt{\rho_e \rho_c \varphi}) \right. \\ &\quad \left. + Q\left\{\varphi(1 - \rho_e) + (1 - \rho_c) - 2\rho_{\delta\varepsilon} \sqrt{(1 - \rho_e)(1 - \rho_c)\varphi}\right\}\right]^{-1} \left\{V(C_{ij})\right\}^{-1}, \end{aligned} \tag{B.1}$$

where $P = \left(\frac{1}{k_t} + \frac{1}{k_c}\right)$, $Q = \left(\frac{1}{k_tm} + \frac{1}{k_cn}\right)$, and $(P - Q) \geq 0$ for all m, n, k_t , and k_c . The partial derivatives of Equation (B.1) with respect to ρ_{uv} and $\rho_{\delta\varepsilon}$ become $\frac{\partial Eff}{\partial \rho_{uv}} = 2Eff^2 V(C_{ij}) P \sqrt{\rho_e \rho_c \varphi} \geq 0$ and $\frac{\partial Eff}{\partial \rho_{\delta\varepsilon}} = 2Eff^2 V(C_{ij}) Q \sqrt{(1 - \rho_e)(1 - \rho_c)\varphi} \geq 0$, respectively. These partial derivatives show that

the efficiency always goes up as ρ_{uv} and $\rho_{\delta\epsilon}$ increase. So the minimum efficiency occurs at the lower boundaries of ρ_{uv} and $\rho_{\delta\epsilon}$. A systematic review of 62 empirical studies by Gomes *et al.* [18] shows that both ρ_{uv} and $\rho_{\delta\epsilon}$ have negative lower boundaries, say $r_{1l} (-1 \leq r_{1l} < 0)$ and $r_{2l} (-1 \leq r_{2l} < 0)$ (l stands for lower boundary), respectively. It is to be noted that the minimum efficiency with respect ρ_e and ρ_c also depends on the value of ρ_{uv} and $\rho_{\delta\epsilon}$. The Gomes *et al.* [18] review suggests $r_{1l} = r_{2l}$, but it may be that $r_{1l} \neq r_{2l}$. However, it turns out that in order to find values for ρ_e and ρ_c , which minimize the efficiency, it is necessary to assume that $\rho_{uv} = \rho_{\delta\epsilon} = \min\{r_{1l}, r_{2l}\} = r (-1 \leq r < 0)$. This minimizes the efficiency more than needed if $r_{1l} \neq r_{2l}$ but at least guarantees a certain power level also in this case. The minimum efficiency is then given by

$$Eff^* = \left[P(\varphi\rho_e + \rho_c - 2r\sqrt{\rho_e\rho_c\varphi}) + Q \left\{ \varphi(1 - \rho_e) + (1 - \rho_c) - 2r\sqrt{(1 - \rho_e)(1 - \rho_c)\varphi} \right\} \right]^{-1} \{V(C_{ij})\}^{-1}. \quad (B.2)$$

Taking the partial derivative of (B.2) with respect to φ gives $\frac{\partial Eff^*}{\partial \varphi} = -Eff^{*2} V(C_{ij}) \left[P \left(\rho_e - r\sqrt{\frac{\rho_e\rho_c}{\varphi}} \right) + Q \left\{ (1 - \rho_e) - r\sqrt{\frac{(1-\rho_e)(1-\rho_c)}{\varphi}} \right\} \right] \leq 0$, implying that Eff^* decreases with increasing φ . So Eff^* becomes minimum at the upper bound of φ , which is say $f (0 \leq f < \infty)$. As $P \geq Q$, $\frac{\partial Eff^*}{\partial \rho_e} = -Eff^{*2} V(C_{ij}) \left\{ (P - Q)\varphi - r\sqrt{\varphi} \left(P\sqrt{\frac{\rho_c}{\rho_e}} - Q\sqrt{\frac{1-\rho_c}{1-\rho_e}} \right) \right\} \leq 0$, if $\left(\sqrt{\frac{\rho_c}{\rho_e}} - \sqrt{\frac{1-\rho_c}{1-\rho_e}} \right) \geq 0$, which is equivalent to $(\rho_c - \rho_e\rho_c) \geq (\rho_e - \rho_e\rho_c)$ and then $\frac{\partial Eff^*}{\partial \rho_e} \leq 0$ whenever $\rho_c \geq \rho_e$. Similarly, $\frac{\partial Eff^*}{\partial \rho_c} \leq 0$, if $(\rho_e - \rho_e\rho_c) \geq (\rho_c - \rho_e\rho_c)$ or $\rho_e \geq \rho_c$. So the efficiency will be minimum for both ρ_e and ρ_c if $\rho_e = \rho_c = \rho = \max\{b, d\}$, where $\rho_e \in [a, b]$ with $0 < a \leq b < 1$ and $\rho_c \in [c, d]$ with $0 < c \leq d < 1$ are the realistic ranges for the parameters ρ_e and ρ_c , respectively. The minimum efficiency with respect to model parameters is thus obtained by substituting $\rho_{uv} = \rho_{\delta\epsilon} = r$, $\rho_e = \rho_c = \rho$ and $\varphi = f$ into Equation (B.1). This yields Equation (10) in the main text. Maximizing Equation (10) under the budget constraint in (6) yields the maximin sample sizes in Equation (11). The corresponding *MMV* is obtained by inserting Equation (11) into (10), giving Equation (12) of the main text.

References

1. Drummond M, Sculpher M, Torrance GW, O'Brien BJ, Stoddart GL. *Methods for the Economic Evaluation of Health Care Programmes*. Oxford University Press: Oxford, 2005.
2. Morrell CJ, Slade P, Warner R, Paley G, Dixon S, Walters SJ, *et al.* Clinical effectiveness of health visitor training in psychologically informed approaches for depression in postnatal women: pragmatic cluster randomised trial in primary care. *British Medical Journal* 2009; **338**:276–280.
3. Moerbeek M, Van Breukelen GJP, Berger MPF. Design issues for experiments in multilevel population. *Journal of Educational and Behavioral Statistics* 2000; **25**:271–284.
4. Grieve R, Nixon R, Thompson SG. Bayesian hierarchical models for cost-effectiveness analyses that use data from cluster randomized. *Medical Decision Making* 2010; **30**:163–175.
5. Raudenbush SW. Statistical analysis and optimal design for cluster randomized trials. *Psychological Methods* 1997; **2**:173–185.
6. Raudenbush ST, Liu X. Statistical power and optimal design for multisite randomized trials. *Psychological Methods* 2000; **5**:199–213.
7. Briggs A, Gray AM. Power and sample size calculations for stochastic cost-effectiveness analysis. *Medical Decision Making* 1998; **18**:81–92.
8. Laska EM, Meisner M, Siegel C. Power and sample size in cost-effectiveness analysis. *Medical Decision Making* 1999; **19**:339–343.
9. Zethraeus N, Johannesson M, Jönsson B, Löthgren M, Tambour M. Advantages of using the net-benefit approach for analysing uncertainty in economic evaluation studies. *Pharmacoeconomics* 2003; **21**:39–48.
10. Manca A, Rice N, Sculpher MJ, Briggs AH. Assessing generalisability by location in trial-based cost-effectiveness analysis: the use of multilevel models. *Health Economics* 2005; **14**:471–485.
11. Berger MPF, Wong WK. *An Introduction to Optimal Designs for Social and Bio-Medical Research*. Wiley: Chichester, 2009.
12. Atkinson AC, Donev AN, Tobias R D. *Optimum Experimental Designs, with SAS*. Oxford University Press: Oxford, 2007.
13. Wu CFJ. Efficient sequential designs with binary data. *Journal of the American Statistical Association* 1985; **80**:974–984.
14. Spiegelhalter DJ, Abrams KR, Myles JP. *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. Wiley: Chichester, 2007.
15. Stinnett A, Mullahy J. Net health benefits: a new framework for the analysis of uncertainty in cost-effectiveness analysis. *Medical Decision Making* 1998; **18**:S68–S80.
16. Verbeke G, Molenberghs G. *Linear Mixed Models for Longitudinal Data*. Springer: New York, 2000.

17. Tokola K, Larocque D, Nevalainen J, Oja H. Power, sample size and sampling costs for clustered data. *Statistics and Probability Letters* 2011; **81**:852–862.
18. Gomes M, Edmond SWN, Grieve R, Nixon R, Carpenter R, Thompson SG. Developing appropriate methods for cost-effectiveness analysis of cluster randomized trials. *Medical Decision Making* 2012; **32**:350–361.
19. Hampel FR. A general qualitative definition of robustness. *The Annals of Mathematical Statistics* 1971; **6**:1887–1896.
20. Grieve R, Nixon R, Thompson SG, Cairns J. Multilevel models for estimating incremental net benefits in multinational studies. *Health Economics* 2007; **16**:815–826.
21. Hoch JS, Briggs AH, Willan A. Something old, something new, something borrowed, something BLUE: a framework for the marriage of health econometrics and cost-effectiveness analysis. *Health Economics* 2002; **11**:415–430.
22. Briggs AH. A Bayesian approach to stochastic cost-effectiveness analysis. *International Journal of Technology Assessment in Health Care* 2001; **17**:69–82.
23. Willan AR, Lin DY. Incremental net benefit in randomized clinical trials. *Statistics in Medicine* 2001; **20**:1563–1574.
24. Murray DM, Varnell SP, Blitstein JL. Design and analysis of group-randomized trials: a review of recent methodological developments. *American Journal of Public Health* 2004; **94**:423–432.
25. Snijders TAB, Bosker RJ. *Multilevel Analysis*. Sage Publications: London, 1999.
26. Cohen J. A power primer. *Psychological Bulletin* 1992; **112**:155–159.
27. Liu X. Statistical power and optimum sample allocation ratio for treatment and control having unequal costs per unit of randomization. *Journal of Educational and Behavioral Statistics* 2003; **28**:231–248.
28. Cheong YK, Fotiu RP, Raudenbush SW. Efficiency and robustness of alternative estimators for two- and three-level models: the case of NAEP. *Journal of Educational and Behavioral Statistics* 2001; **26**:411–429.
29. Maas CJM, Hox JJ. The influence of violations of assumptions on multilevel parameter estimates and their standard errors. *Computational Statistics and Data Analysis* 2004; **46**:427–440.
30. Candel MJJM, Van Breukelen GJP, Kotova L, Berger MPF. Optimality of unequal cluster sizes in multilevel studies with small sample sizes. *Communications in Statistics: Simulation and Computation* 2008; **37**:222–239.
31. Van Breukelen GJP, Candel MJJM, Berger MPF. Relative efficiency of unequal versus equal cluster sizes in cluster randomized and multicentre trials. *Statistics in Medicine* 2007; **26**:2589–2603.
32. Moerbeek M, Van Breukelen GJP, Berger MPF. Optimal experimental designs for multilevel models with covariates. *Communications in Statistics—Theory and Methods* 2001; **30**:2683–2697.
33. Roberts C, Roberts SA. Design and analysis of clinical trials with clustering effects due to treatment. *Clinical Trials* 2005; **2**:152–162.