



# SamP2CeT: an interactive computer program for sample size and power calculation for two-level cost-effectiveness trials

Md Abu Manju<sup>1</sup> · Math J. J. M. Candel<sup>1</sup> · Gerard J. P. van Breukelen<sup>1</sup>

Received: 15 August 2017 / Accepted: 3 August 2018 / Published online: 17 August 2018  
© The Author(s) 2018

## Abstract

The cost-effectiveness of interventions (e.g. new medical therapies or health care technologies) is often evaluated in randomized clinical trials, where individuals are nested within clusters, for instance patients within general practices. In such two-level cost-effectiveness trials, one can randomly assign treatments to individuals within clusters (multicentre trial) or to entire clusters (cluster randomized trial). Such trials need careful planning to evaluate the cost-effectiveness of interventions within the available research resources. The optimal number of clusters and the optimal number of subjects per cluster for both types of cost-effectiveness trials can be determined by using optimal design theory. However, the construction of the optimal design requires information on model parameters, which may be unknown at the planning stage of a trial. To overcome this problem, a maximin strategy is employed. We have developed a computer program **SamP2CeT** in R to perform these sample size calculations. **SamP2CeT** provides a graphical user interface which enables the researchers to optimize the numbers of clusters and subjects per cluster in their cost-effectiveness trial as a function of study costs and outcome variances. In case of insufficient knowledge on model parameters, **SamP2CeT** also provides safe numbers of clusters and subjects per cluster, based on a maximin strategy. **SamP2CeT** can be used to calculate the smallest budget needed for a user-specified power level, the largest power attainable with a user-specified budget, and also has the facility to calculate the power for a user-specified design. Recent methodological developments on sample size and power calculation for two-level cost-effectiveness trials have been implemented in **SamP2CeT**. This program is user-friendly, as illustrated for two published cost-effectiveness trials.

**Keywords** Cluster randomized trials · Cost-effectiveness analysis · Maximin design · Multicentre trials · Optimal design · Power · Sample size calculation

---

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s00180-018-0829-4>) contains supplementary material, which is available to authorized users.

---

✉ Md Abu Manju  
abu.manju@maastrichtuniversity.nl

Extended author information available on the last page of the article

## 1 Introduction

There is an increasing demand from decision makers for information not just on the effectiveness of new interventions, but on their cost-effectiveness (value for money). Cost-effectiveness studies of new medical therapies, treatments or health care technologies are therefore increasingly carried out and such studies frequently have a nested design, that is, costs and effectiveness data are collected at the individual person level, but these persons are nested within larger units called clusters, for instance, patients are nested within general practices. In such two-level nested designs, treatments are randomly assigned either to whole clusters (cluster randomized trial (CRT)) or to individuals within clusters (multicentre trial). Examples of cost-effectiveness CRTs are the SPHERE study (Ng et al. 2016) and the PONDER study (Morrell et al. 2009). Examples of multicentre cost-effectiveness trials are the EVALUATE trial (Sculpher et al. 2004; Manca et al. 2005) and the Comparison of European Stroke Costs and Survival study (Grieve et al. 2001, 2007).

The planning of such trials itself involves cost-effectiveness considerations, as the trial must have sufficient power and precision for testing and estimating the cost-effectiveness of the new treatment within the constraints of a budget for sampling, treating and measuring clusters and individuals. In particular, the question is how many clusters and how many individuals per cluster are needed to prove the cost-effectiveness of treatments. This problem can be addressed by optimal design theory. The optimal design, henceforth called the optimal sample size, is the design which maximizes the efficiency of the treatment effect estimator, and thus results in maximum power for testing the treatment effect, given a constraint on the total cost for sampling and measuring clusters and individuals. Alternatively, and equivalently, the optimal design is the design which minimizes the total costs for sampling and measuring given a constraint on the statistical power or efficiency. This paper presents software for computing the optimal design for cost-effectiveness CRTs and multicentre cost-effectiveness trials. The optimal design usually depends on some model parameters which are unknown in the design stage and is thus optimal only for a specific set of parameter values. This is known as the local optimality problem. One solution to this problem is the so-called maximin design (*MMD*). This design maximizes the minimum efficiency and can be considered a design which is efficient for a plausible range of values for the unknown model parameters (Berger and Wong 2009; Atkinson et al. 2007).

The statistical theory for constructing both optimal design and *MMD* for cost-effectiveness CRTs and multicentre cost-effectiveness trials is presented by Manju et al. (2014), Manju et al. (2015) This paper presents its software implementation in a computer program with a graphical user-interface (GUI). This computer program, called **Samp2CeT** (**S**ample size and **P**ower calculation for **2**-level **C**ost-effectiveness **T**rials), provides a user-friendly environment for researchers to design cost-effectiveness trials (CRT and multicentre trial) with regard to the optimal number of clusters and the optimal number of subjects per cluster. In case of limited information on model parameters, the **Samp2CeT** also provides maximin (i.e., worst case) numbers of clusters and subjects per cluster. The software presented here can either be used to minimize the budget for a desired level of power, or to maximize the power

for a fixed budget. In addition to fixing a budget or power, **SamP2CeT** also evaluates the power if the design is fixed. Note that other software programs have been developed for sample size calculation for cluster randomized and multicentre trials, such as OD (Raudenbush and Liu 2000; Spybrook et al. 2011), PINT (Snijders and Bosker 1993; Bosker et al. 2003) and ML-DE (Cools et al. 2008). These programs only take into account the effectiveness of interventions, whereas **SamP2CeT** also considers the costs of interventions. Furthermore, these software programs do not allow for optimizing the power for a given budget or minimizing the costs for a desired power level.

The paper is organized as follows. In Sect. 2, the design problem for cost-effectiveness trials is briefly discussed. The methodologies for CRTs and multicentre trials are outlined respectively in Sects. 3 and 4. The model, the cost function, the optimal design and the *MMD* for both CRTs and multicentre trials are explained. Section 5 describes the computer program and its input values. An illustration of the computer program with practical examples for both a CRT and a multicentre trial is provided in Sect. 6. Section 7 addresses the benefits of using optimal design and *MMD*. Finally, a discussion and conclusion are presented in Sect. 8.

## 2 Design problem for cost-effectiveness trials

A fundamental issue in both types of cost-effectiveness trials (CRTs and multicentre trials) is that individuals are nested within clusters, and that there will usually be variability in costs and effects not only between persons within clusters, but also between clusters, and so there is sampling variance at both design levels: person and cluster. Statistical methods that accommodate between cluster variability of outcomes in the analysis of cost-effectiveness CRTs and multicentre cost-effectiveness trials are relatively well established in the literature (Grieve et al. 2010; Gomes et al. 2012; Manju et al. 2014; Grieve et al. 2007; Manju et al. 2015). These methods address the variability in both costs and effects at both design levels, clusters and persons, and they also take into account the correlation between costs and effects at each level. The methods start from a bivariate mixed-effects model where costs and effects are assumed to follow a bivariate normal distribution<sup>1</sup>. This bivariate mixed-effects model is then used to estimate average incremental costs and effects from treatment as compared with the control condition. These average incremental costs and effects in turn are combined into a single outcome called the average incremental net-monetary benefit (*INMB*), which is a widely used measure of a treatment's cost-effectiveness (Stinnett and Mullahy 1998). The new treatment or intervention is said to be cost-effective if the average *INMB* exceeds 0<sup>2</sup>. In this case the intervention is considered for adoption

---

<sup>1</sup> Costs usually have a skewed distribution, but Gomes et al. (2012) and Manju (2016) show that statistical analysis of cost-effectiveness trials with the bivariate normal model is quite robust against skewed costs.

<sup>2</sup> If there is a health gain, this is equivalent to the Incremental Cost-Effectiveness Ratio (ICER) being smaller than the threshold willingness-to-pay parameter for a unit increase in health gain, and if there is a health loss, equivalent to the ICER being larger than the threshold willingness to pay parameter (Willan and Briggs 2006).

in health care practice. Technical details of the average *INMB* procedure are given in the next two sections.

When constructing an optimal design for cost-effectiveness trials, we want to estimate the average *INMB* as precisely, and to test it with as much power, as possible. Therefore, the optimal design minimizes the sampling variance (i.e. the squared standard error) of the estimator of the average *INMB*, subject to a certain cost function. This cost function relates the required research budget to the sampling costs per person and per cluster. The result of the design optimization are sample size formulae for the numbers of individuals and clusters that maximize the study power and precision given a fixed research budget, or that minimize the study costs given a desired level of precision and power for testing the average *INMB*.

These optimal design or sample size formulae depend on some model parameters which are not known in the planning stage, and so the optimal design is said to be locally optimal, (Berger and Wong 2009; Atkinson et al. 2007) that is, it typically only is optimal for specific values of model parameters, and not for other values. The maximin approach solves this local optimality problem by the following two-step procedure: (i) First, find that set of model parameter values within their plausible range of values, which minimizes the design efficiency, as expressed by the inverse of the sampling variance of the estimator of average *INMB*, and (ii) then choose that design which maximizes this minimum efficiency, thereby also maximizing the precision and power in the worst case scenario, given a fixed budget. The resulting design called the *MMD*, actually is the optimal design for the worst case.

### 3 Cluster randomized trials (CRTs)

#### 3.1 Model

Suppose that the cost-effectiveness of a new treatment is evaluated in a CRT, where  $k_t$  and  $k_c$  clusters (e.g., general practices or hospitals) are randomly assigned to the new treatment and control treatment respectively. Within each treatment cluster  $j$  ( $j = 1, 2, \dots, k_t$ ) there are  $m$  individuals (e.g., patients), and within each control cluster  $j$  ( $j = k_t + 1, k_t + 2, \dots, k_t + k_c$ ) there are  $n$  individuals. All individuals receive the treatment to which their cluster is allocated. To describe the modeling framework for the cost-effectiveness analysis of CRTs, the following bivariate linear mixed model is considered (Grieve et al. 2010; Gomes et al. 2012; Manju et al. 2014):

$$E_{ij} = \beta_0^e + \beta_1^e x_j + u_{0j}^e + \epsilon_{ij}^e, \text{ and } C_{ij} = \beta_0^c + \beta_1^c x_j + u_{0j}^c + \epsilon_{ij}^c,$$

$$\begin{bmatrix} u_{0j}^e \\ u_{0j}^c \end{bmatrix} \sim N \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{e0}^2 & \rho_{u0} \sigma_{e0} \sigma_{c0} \\ & \sigma_{c0}^2 \end{bmatrix} \right) \text{ and } \begin{bmatrix} \epsilon_{ij}^e \\ \epsilon_{ij}^c \end{bmatrix} \sim N \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{\epsilon^e}^2 & \rho_{\epsilon} \sigma_{\epsilon^e} \sigma_{\epsilon^c} \\ & \sigma_{\epsilon^c}^2 \end{bmatrix} \right), \tag{1}$$

where  $E_{ij}$  and  $C_{ij}$  are the effects and costs for individual  $i$  in cluster  $j$  respectively. The variable  $x_j$  represents the treatment assigned to cluster  $j$  and will be coded 0 and 1 for the control and treatment clusters respectively, although other coding schemes

are also possible and lead to the same sample size formulas for the optimal design and *MMD*. The coefficients  $\beta_0^e$  and  $\beta_0^c$  can be regarded as the average effects and costs respectively for the group of individuals receiving the control treatment. The random intercepts  $u_{0j}^e$  and  $u_{0j}^c$  reflect the deviations from the averages  $\beta_0^e$  and  $\beta_0^c$  for the  $j^{th}$  cluster respectively and are assumed to follow a bivariate normal distribution, with variances  $\sigma_{e0}^2$  and  $\sigma_{c0}^2$  and correlation  $\rho_{u0}$  ( $-1 \leq \rho_{u0} \leq 1$ ). The slope coefficients  $\beta_1^e$  and  $\beta_1^c$  correspond to the average incremental effects and costs from treatment respectively. Finally,  $\epsilon_{ij}^e$  and  $\epsilon_{ij}^c$  are individual level residuals for effects and costs respectively, assumed to follow a bivariate normal distribution, with variances  $\sigma_{\epsilon e}^2$  and  $\sigma_{\epsilon c}^2$  and correlation  $\rho_\epsilon$  ( $-1 \leq \rho_\epsilon \leq 1$ ).

In Eq. (1), the costs and effects are expressed on their original scales, but the net-monetary benefit (*NMB*) framework, where both the effects and costs of interventions are scaled onto the same monetary scale on the basis of a threshold value  $\lambda$ , offers a convenient way of modeling the data from cost-effectiveness CRTs (Stinnett and Mullahy 1998). Here *NMB* can be calculated for each individual  $i$  in cluster  $j$  as:  $NMB_{ij} = \lambda E_{ij} - C_{ij}$ , where the value  $\lambda$  ( $0 \leq \lambda < \infty$ ) is the threshold willingness to pay for a unit of health gain. Costs and effects are thus combined into the average incremental net-monetary benefit, *INMB* ( $\beta_1$ ), a measure of relative value for money of intervention as compared with control, which is defined as  $\beta_1 = \lambda\beta_1^e - \beta_1^c$ . In a CRT, the outcomes of two individuals in an arbitrary cluster are correlated, and these intraclass correlations (ICCs) for effects and costs are equal to  $0 \leq \rho_e = \frac{\sigma_{e0}^2}{\sigma_{e0}^2 + \sigma_{\epsilon e}^2} \leq 1$  and  $0 \leq \rho_c = \frac{\sigma_{c0}^2}{\sigma_{c0}^2 + \sigma_{\epsilon c}^2} \leq 1$  respectively. We furthermore define  $\varphi = \frac{\lambda^2(\sigma_{e0}^2 + \sigma_{\epsilon e}^2)}{\sigma_{c0}^2 + \sigma_{\epsilon c}^2}$ , and this  $\varphi$  ( $0 \leq \varphi < \infty$ ), which we call the variance ratio, can be interpreted as the ratio of total variance for effects translated into costs ( $\lambda E_{ij}$ ) to the total variance for costs ( $C_{ij}$ ). The ICCs and the variance ratio will be seen to play an important role in the design optimization.

### 3.2 Sample size calculation

The aim of optimal design theory is to find the design which minimizes some function of the covariance matrix of one or more parameter estimators of interest, and this function is called the optimality criterion. Here, the optimality criterion is simply the sampling variance of  $\hat{\beta}_1$ , the maximum likelihood (ML) estimator of the average *INMB*,  $\beta_1$ , which measures the treatment’s cost-effectiveness. The following asymptotic variance of the ML estimator of  $\beta_1$  has been derived from the general expression for the variance-covariance matrix of the fixed-effects estimators for a linear mixed model as proposed by Verbeke and Molenberghs (2000) and Manju et al. (2014):

$$\text{var}(\hat{\beta}_1) = \left\{ \left( \frac{1}{k_t} + \frac{1}{k_c} \right) A + \left( \frac{1}{k_{tm}} + \frac{1}{k_{cn}} \right) B \right\} v_c, \tag{2}$$

where  $A = \varphi\rho_e + \rho_c - 2\rho_{u0}\sqrt{\rho_e\rho_c\varphi}$ ,  $B = \varphi(1 - \rho_e) + (1 - \rho_c) - 2\rho_\epsilon\sqrt{(1 - \rho_e)(1 - \rho_c)\varphi}$  and  $v_c = \text{var}(C_{ij}) = \sigma_{c0}^2 + \sigma_{\epsilon c}^2$ . Note that  $A$  is the contribution

of random cluster effects, and  $B$  is the contribution of random person and measurement error effects, to the sampling variance of  $\hat{\beta}_1$ .

Designs are usually restricted by a budget. Therefore, it is important to take the costs for measuring and sampling clusters and individuals into account when planning a CRT. Let  $C$  represent the total budget for sampling, treating and measuring clusters and persons and let  $c_t > 0$  and  $c_c > 0$  be the cost per cluster in the treatment and control arm, respectively. Furthermore, let  $s_t > 0$  and  $s_c > 0$  be the cost per individual in the treatment and control cluster, respectively. The required budget (or total costs) is then given by Tokola et al. (2011) and Manju et al. (2014):

$$C = k_t c_t + k_c c_c + k_t m s_t + k_c n s_c. \quad (3)$$

Finding an optimal design means finding the  $m$ ,  $n$ ,  $k_t$  and  $k_c$  which minimize  $\text{var}(\hat{\beta}_1)$  in Eq. (2), and thus maximize the statistical power, given the budget constraint in Eq. (3). The optimal design formulas have been derived by Manju et al. (2014) and are here provided in Appendix A.

Computing the optimal design requires knowledge about the model parameters  $\rho_e$ ,  $\rho_c$ ,  $\rho_{u0}$ ,  $\rho_e$  and  $\varphi$ . As described in the previous section, since specifying these parameter values may not be possible at the planning stage of a trial, a *MMD* can be determined as an efficient design over a plausible range of values for these unknown model parameters. The expressions for the *MMDs* for planning a cost-effectiveness CRT, as derived by Manju et al. (2014), are presented in Appendix A. These *MMDs* depend on the upper bounds for plausible ranges of  $\rho_e$  and  $\rho_c$ .

The optimal design and the *MMD* as proposed by Manju et al. (2014) have been derived for a fixed budget. Therefore, these designs give the smallest variance of the treatment effect estimator and thus the largest precision and power, given the budget and given either the model parameter values (optimal design) or plausible ranges of these model parameters (maximin design). In designing a cost-effectiveness study, however, one will often aim at achieving a certain power for the treatment effect test instead of fixing the budget beforehand. This can be achieved by the same optimal and maximin design equations as for a fixed budget, however, by noting that the optimal respectively maximin cluster sizes  $m$  and  $n$  do not depend on the budget, whereas the optimal respectively maximin numbers of clusters do (see Appendix A for details). So if a given budget does not give sufficient power even after design optimization, then the budget must be increased and this will lead to an increase of the number of clusters  $k_t$  and  $k_c$ , not of the cluster sizes  $m$  and  $n$ . The required budget can thus be calculated either by repeated increases of the budget followed by computation of the optimal cluster sizes and power until the desired power is attained, or by rewriting the optimal sample size equations for a given budget into those for a given power. The latter approach underlies the present computer program **Samp2CeT**. To compute the power and required budget the user will also need to specify the effect size ( $ES$ ) of interest. The program uses as measure of  $ES$  an adaptation of the classical Cohen's  $d$  (Cohen 1992), to the present nested design. For details, see Appendix A.

A third and last way of using **Samp2CeT**, is to calculate the power for a fixed design, that is, for user-specified values of  $m$ ,  $n$ ,  $k_t$  and  $k_c$ . This is relevant if a researcher has several designs that are practically feasible, and wants to evaluate these in terms of

their power. In this case, the same procedure as above for computing the power of the optimal design for a given budget is followed, except that user-specified values of  $m$ ,  $n$ ,  $k_t$  and  $k_c$  are used instead of the optimal  $m$ ,  $n$ ,  $k_t$  and  $k_c$ . Furthermore, if the user does not have precise information on the model parameters that affect the power, the user can evaluate the worst case (lowest) power. In this case, the power of the user specified design in a worst case scenario results, comparable to the power of the *MMD* for a given budget.

## 4 Multicentre trials

### 4.1 Model

Next to CRTs, also multicentre trials, with randomization of individuals within clusters or centres, are often employed in cost-effectiveness studies. This section describes the modeling approach to analyze data from multicentre cost-effectiveness trials involving a treatment-by-centre interaction and assuming a  $m:n$  randomization in each centre  $j$  ( $j = 1, 2, \dots, k$ ). We have the following bivariate linear mixed model for the effect of treatment  $x$  on the quantitative outcomes, effects ( $E_{ij}$ ) and costs ( $C_{ij}$ ), for individual  $i$  in centre  $j$  (Grieve et al. 2007; Manju et al. 2015):

$$E_{ij} = \beta_{0j}^e + \beta_{1j}^e x_{ij} + u_{1j}^e x_{ij} + \epsilon_{ij}^e, \text{ and } C_{ij} = \beta_{0j}^c + \beta_{1j}^c x_{ij} + u_{1j}^c x_{ij} + \epsilon_{ij}^c, \\ \begin{bmatrix} u_{1j}^e \\ u_{1j}^c \end{bmatrix} \sim N \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{e1}^2 & \rho_{u1} \sigma_{e1} \sigma_{c1} \\ \rho_{u1} \sigma_{e1} \sigma_{c1} & \sigma_{c1}^2 \end{bmatrix} \right), \text{ and } \begin{bmatrix} \epsilon_{ij}^e \\ \epsilon_{ij}^c \end{bmatrix} \sim N \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{\epsilon e}^2 & \rho_{\epsilon} \sigma_{\epsilon e} \sigma_{\epsilon c} \\ \rho_{\epsilon} \sigma_{\epsilon e} \sigma_{\epsilon c} & \sigma_{\epsilon c}^2 \end{bmatrix} \right), \tag{4}$$

where  $x_{ij}$  is coded as  $b$  for treated and  $a$  control individuals ( $a$  and  $b$  are real numbers and  $a \neq b$ ), so that the average incremental effects and costs from treatment are  $(b - a)\beta_1^e$  and  $(b - a)\beta_1^c$  respectively. In case of a multicentre trial, the *NMB* $_{ij}$  and the average *INMB* are also defined as  $NMB_{ij} = \lambda E_{ij} - C_{ij}$  and  $(b - a)\beta_1 = (b - a)(\lambda\beta_1^e - \beta_1^c)$  respectively. The intercepts  $\beta_{0j}^e$  and  $\beta_{0j}^c$  for centre  $j$ , the main effect of centres for effects and costs respectively, may be either fixed or random. If the randomization is  $m:n$  in each centre  $j$  ( $j = 1, 2, \dots, k$ ), it makes no difference for the optimal design and *MMD* whether the components  $\beta_{0j}^e$  and  $\beta_{0j}^c$  are treated as random or fixed (Senn 2014). The random components  $u_{1j}^e$  and  $u_{1j}^c$ , represent the treatment-by-centre interactions for costs and effects respectively and are assumed to follow a bivariate normal distribution with the variances  $\sigma_{c1}^2$  and  $\sigma_{e1}^2$  and correlation  $\rho_{u1}$  ( $-1 \leq \rho_{u1} \leq 1$ ). We make the same distributional assumptions regarding the random individual level ( $\epsilon_{ij}^c, \epsilon_{ij}^e$ ) costs and effects in Eq. (4) as in Eq. (1). In deriving the optimal design and *MMD* for multicentre cost-effectiveness trials, Manju et al. (2015) used the following re-parameterization of the model parameters in Eq. (4):

$$0 \leq \theta_e = \frac{(b-a)^2 \sigma_{e1}^2}{(b-a)^2 \sigma_{e1}^2 + \sigma_{\epsilon e}^2} \leq 1, \quad 0 \leq \theta_c = \frac{(b-a)^2 \sigma_{c1}^2}{(b-a)^2 \sigma_{c1}^2 + \sigma_{\epsilon c}^2} \leq 1 \text{ and } 0 \leq \phi = \frac{\lambda^2 \{(b-a)^2 \sigma_{e1}^2 + \sigma_{\epsilon e}^2\}}{\{(b-a)^2 \sigma_{c1}^2 + \sigma_{\epsilon c}^2\}} < \infty. \text{ Here } \theta_e \text{ and } \theta_c \text{ play the same role as the ICCs in the CRT}$$



model as discussed in the previous part on cluster randomized trials, and  $\phi$  is similar to the ratio of the variance for effects translated into costs (i.e.,  $\lambda E_{ij}$ ) to the variance for costs, except that the intercept variances  $\sigma_{e0}^2$  and  $\sigma_{c0}^2$  are replaced with the slope variances  $\sigma_{e1}^2$  and  $\sigma_{c1}^2$  multiplied with a scaling factor  $(b-a)^2$  which depends on the coding of the treatment indicator  $x_{ij}$  in Eq. (4). Note that the quantities  $(b-a)^2\sigma_{e1}^2$  and  $(b-a)^2\sigma_{c1}^2$  do not depend on the treatment coding. For instance, changing from 0/1 coding to  $-1/+1$  coding makes  $(b-a)^2$  four times larger, but  $\sigma_{e1}^2$  four times smaller, since  $\beta_1^e + u_{1j}^e$  is the effects difference between treatment and control in centre  $j$  in case of 0/1 treatment coding, but only half that difference in case of  $-1/+1$  coding. Because of the similarity with the ICCs and the variance ratio for cluster randomized trials (CRTs), in the remainder of this paper we will call  $\theta_e$  and  $\theta_c$  the quasi-ICCs for effects and costs respectively, and denote  $\phi$  as the quasi-variance ratio.

## 4.2 Sample size calculation

If there are  $m$  and  $n$  individuals randomly allocated to the treatment and control condition in each of the  $k$  centres, the following expression for the variance of the average *INMB* estimator has been derived based on ML estimation as covered by Verbeke and Molenberghs (2000), Manju et al. (2015)

$$\text{var}(\hat{\beta}_1) = \frac{v'_c}{k} \left\{ A' + \left( \frac{1}{m} + \frac{1}{n} \right) B' \right\}, \quad (5)$$

where  $A' = \frac{\phi\theta_e}{\sqrt{(1-\theta_e)(1-\theta_c)\phi}} + \theta_c - 2\rho_{u1}\sqrt{\theta_e\theta_c\phi}$ ,  $B' = \frac{\phi(1-\theta_e)}{\sqrt{(1-\theta_e)(1-\theta_c)\phi}} + (1-\theta_c) - 2\rho_{e1}\sqrt{\theta_e\theta_c\phi}$  and  $v'_c = \sigma_{c1}^2 + (b-a)^{-2}\sigma_{e1}^2$ . Note that, the random slope and random error variances for *NMB* as specified in the previous section are  $\sigma_1^2 = \text{var}(\lambda u_{1j}^e - u_{1j}^c) = A'v'_c$  and  $\sigma^2 = \text{var}(\lambda \epsilon_{ij}^e - \epsilon_{ij}^c) = (b-a)^2 B'v'_c$  respectively. Further,  $v'_c = \sigma_{c1}^2 + (b-a)^{-2}\sigma_{e1}^2$  is analogous to the total cost variance  $v_c$  in Eq. (2), except that intercept variance is replaced with slope variance.

To find the optimal design, we need a function that relates sample size to costs. Assume that the cost per included centre is  $c > 0$ , whereas the costs per included individual in the treatment arm is  $s_t > 0$  and in the control arm is  $s_c > 0$ . The budget  $C$  needed for including  $k$  centres of  $m+n$  individuals each is then (Liu 2003; Manju et al. 2015):

$$C = k(c + ms_t + ns_c). \quad (6)$$

Similar to what has been discussed for CRTs, the optimal design as well as *MMD* for the multicentre cost-effectiveness trials have been derived by Manju et al. (2015) and the equations for these designs (optimal and *MMD*) are given in Appendix B. Analogously to what was seen in the section on cluster randomized trials, these equations for multicenter trials can be used to maximize the power for a given budget, or to minimize the budget needed for a given power, or to evaluate the power of a given design. And again, the effect size upon which the power depends is expressed in a shape close to the classical Cohen's  $d$ , (Cohen 1992) (see Appendix B for details).



## 5 SamP2Cet program

The computer program **SamP2CeT** (**S**ample size and **P**ower calculation for **2**-level **C**ost-effectiveness **T**rials) is an easy to use tool to design cost-effectiveness trials. **SamP2CeT** runs within the  $R(\geq 3.2.4)$  system for statistical computing (R Development Core Team 2015). The reason for choosing R is that R is a free software environment for statistical computing. Also, R is a popular software program within the field of medical science for the design and analysis of clinical trials (Bornkamp et al. 2009; Peace and Chen 2010). The present computer program **SamP2CeT** asks the user to specify various design characteristics as well as input parameters through a Graphical User Interface (GUI). The GUI uses the  $R$ -Tk/Tk(Welch 1995) interface implemented in the R package: `tk` (Dalggaard 2001).

This section describes the GUI of the program which calculates the optimal (or maximin) number of clusters and number of individuals per cluster for a cost-effectiveness CRT or a multicentre cost-effectiveness trial. The program also calculates the minimum budget for a desired level of power, the maximum power for a given budget, or the power for a user-specified design. The program is available as supplementary material from the journal webpage. The program **SamP2CeT** was developed and tested on a Windows 7 operating system.

To start the program the user types `library(SamP2CeT)` and then `SamP2CeT()` in the command window of R and presses the enter key. The directory containing **SamP2CeT** should be the current directory on the R desktop. The main window of the **SamP2CeT** will appear as presented in Fig. 1. The main menu requires the user to make three different choices (see Fig. 1): (a) between a cluster randomized and a multicentre trial, and (b) between minimizing the study costs, or maximizing the study power, or computing the power of some user-specified design, and (c) between specific values and ranges for the outcome variances and correlations. Pushing the button `Next` displays one of 12 different windows (submenus), depending on the user's choices concerning level of randomization, purpose and knowledge of parameters. The input parameters for these 12 different submenus are discussed in the **SamP2CeT** user's manual, which is included as supplementary material.

## 6 Illustrations

The program **SamP2CeT** will be illustrated for a cost-effectiveness CRT known as the SPHERE study and for a multicentre cost-effectiveness trial known as the EVALUATE trial. The user can calculate the sample sizes for both trials either for a fixed power or for a fixed budget. Because of the similarity of the input and result windows for fixed power and fixed budget, we will only consider a fixed power in the following illustrations, but we will consider the case of a fixed budget in Sect. 7. In addition to calculating the sample sizes, the user can evaluate the power of the actual design of these trials. For the sake of illustration, we assume imprecise knowledge of variances and correlations for the SPHERE study, leading to a maximin design and precise knowledge for the EVALUATE trial, leading to an optimal design.

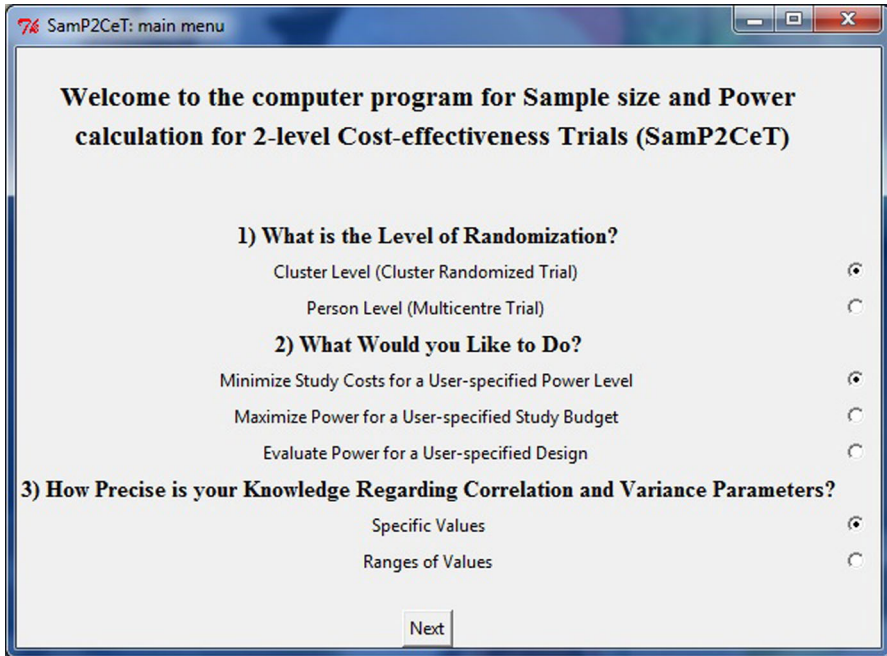


Fig. 1 Main menu of the **SamP2CeT** program

## 6.1 Example of a cost-effectiveness CRT: the SPHERE study

We illustrate the program **SamP2CeT** on the SPHERE study which assessed the cost-effectiveness of a secondary prevention strategy for patients with coronary heart disease in general practice (Ng et al. 2016). General practices were randomised to control (usual care) or treatment (tailored care). The data were collected from 48 general practices (24 treatment clusters and 24 control clusters) with, on the average, 18 and 19 patients in each of the treatment and control clusters respectively. The main endpoints, costs and effects, were expressed in terms of Euros (€) and quality-adjusted life-years (QALYs), respectively.

### 6.1.1 Maximin sample sizes

It is assumed that the user wants to calculate the *MMD* for a cost-effectiveness CRT with the same research question as the SPHERE study, and that the power level is fixed (so that the optimization minimizes the study costs). In that case, the user needs to specify the upper bounds for the intraclass correlation (ICC) for effects ( $\rho_e$ ) and the intraclass correlation (ICC) for costs ( $\rho_c$ ). The ICC upper bounds are based on Gomes et al. (2012)), and the cluster and subject-specific costs were chosen by us for the sake of illustration lacking any information on the actual costs of the SPHERE trial. All input values are shown in Fig. 2, which yield the maximin design as shown in Fig. 3.

Parameter	Value
Upper Bound of Intraclass Correlation for Effects	0.3
Upper Bound of Intraclass Correlation for Costs	0.3
Level of Significance	0.05
Effect Size (ES)	0.5
Power Level	0.8
Cluster-Specific Cost in the Treatment Condition ( $c_{t>0}$ )	600
Cluster-Specific Cost in the Control Condition ( $c_{c>0}$ )	400
Subject-Specific Cost in the Treatment Condition ( $s_{t>0}$ )	200
Subject-Specific Cost in the Control Condition ( $s_{c>0}$ )	100

**Fig. 2** Entries in **SamP2CeT** submenu to calculate the *MMD* for the SPHERE study when the power is fixed

Note that the small number of subjects per cluster is due to the large upper bound for the ICCs (see appendix A for details).

### 6.1.2 Power calculation for a user-specified design

Suppose that the user wants to check the power of the SPHERE study. Choosing in Fig. 1 for “Evaluate Power for a User-specified Design”, and then filling in the parameter values and actual design as in Fig. 4 gives a power of 0.83 or 83%. Note that, for a power that is slightly larger than 80% this actual design requires a budget of 156000 euro, more than twice as much as the *MMD* in Sect. 6.1.1.

## 6.2 Example of a multicentre cost-effectiveness trial: the EVALUATE trial

This section illustrates the program **SamP2CeT** for a multicentre cost-effectiveness trial known as the EVALUATE trial which compared laparoscopic-assisted hysterectomy (treatment) and standard hysterectomy (control) (Sculpher et al. 2004; Manca et al. 2005). The data were collected from 25 centres with, on the average, 23 treated and 11 control patients per centre.

### 6.2.1 Optimal sample sizes

Suppose that the user wants to calculate the optimal design for the EVALUATE trial for a fixed power level. For this illustration, we use as parameter values the estimates from the EVALUATE trial as published in Manca et al. (2005). These input values are displayed in Fig. 5. Furthermore, costs per centre and costs per individual in each treatment arm are assumed by us in the absence of empirical information on the

**Maximin Design for a Cost-effectiveness Cluster Randomized Trial (CRT):**

Maximin number of subjects per treatment cluster ( $m_{MMD}$ )	3
Maximin number of subjects per control cluster ( $n_{MMD}$ )	4
Maximin number of treatment clusters ( $k_{t\_MMD}$ )	30
Maximin number of control clusters ( $k_{c\_MMD}$ )	37
Total number of subjects	238
Total Costs (C)	65600
Power	0.8

OK

**Fig. 3** Result window of **SamP2CeT** appearing for the *MMD* for the SPHERE study when the power is fixed

planning costs from the EVALUATE trial. Figure 6, shows the optimal design for the EVALUATE trial for a given power level. To run this trial at a power level of 80% a budget of €209000 is required.

### 6.2.2 Power calculation for a user-specified design

To evaluate the power for the actual design ( $k = 25$ ,  $m = 23$  and  $n = 11$ ) of the EVALUATE trial the user again needs to supply the information on correlation and variance parameters ( $\theta_e$ ,  $\theta_c$ ,  $\rho_{u1}$ ,  $\rho_e$  and  $\phi$ ) which is displayed in Fig. 7. The resulting power of the actual design is close to 1.0, which is very high as the actual design of the EVALUATE trial consists of a large number of centres and the price of this high power is of course a high study budget, as the actual design requires  $C = \text{€}812500$  when assuming the same costs as in the previous section.

## 7 Benefits of using optimal and maximin sample sizes

The benefits of optimal and maximin designs can be assessed by considering the percentage budget gain (for fixed power) or the percentage power gain (for fixed budget) of these designs compared with the actual design of a study. This comparison is made for the SPHERE study and the EVALUATE trial in Table 1, using the cost ratios in Manju et al. (2014), Manju et al. (2015), and the parameter values in Ng et al. (2016), and Sculpher et al. (2004) and Manca et al. (2005) for the optimal designs,

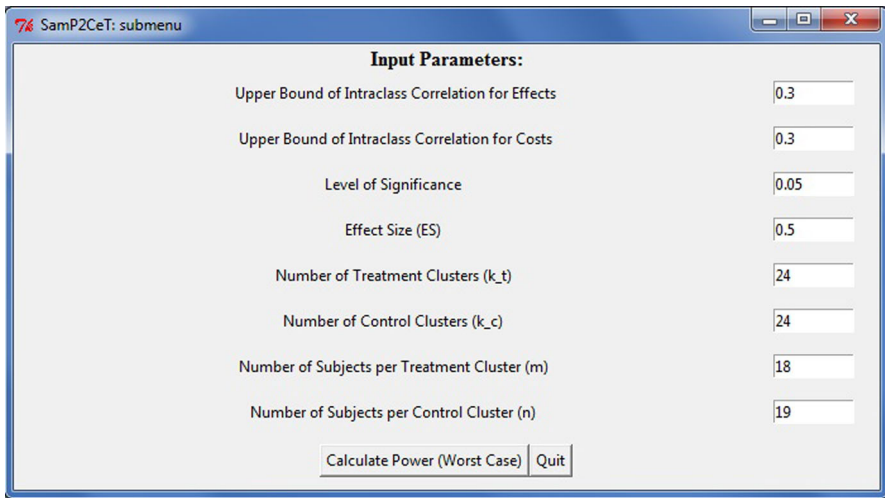


Fig. 4 Entries in **SamP2CeT** submenu to calculate power for the actual design of the SPHERE study when the information on correlation and variance parameters is not precise

and the parameter boundaries in Gomes et al. (2012) and Manju et al. (2015) for the maximin designs. The parameters values are  $\rho_e = 0.001$ ,  $\rho_c = 0.007$ ,  $\rho_{u0} = -0.18$ ,  $\rho_\epsilon = -0.04$ ,  $\varphi = 0.232$  for the SPHERE study, and  $\theta_e = 0.26$ ,  $\theta_c = 0.05$ ,  $\rho_{u1} = 0.66$ ,  $\rho_\epsilon = -0.083$ ,  $\phi = 0.60$  for the EVALUATE trial, and the chosen parameter boundaries are the ICC upper bounds for effects = 0.30 and for costs = 0.30 for the SPHERE study and the quasi-ICC upper bounds for effects = 0.30 and for costs = 0.30 for the EVALUATE trial. In the power comparison of the designs we took an ES of 0.20 which implies a power of 0.79 and 0.21 for the actual design of the SPHERE study and of 0.59 and 0.25 for the actual design of the EVALUATE study for the parameter values as obtained in these studies and as assumed for the maximin scenarios respectively. In contrast, the ES does not play a role in the budget comparison between actual, optimal and maximin designs. This is because the ratio of the budgets needed for any two designs to have the same power for a given ES only depends on the relative efficiency of these two designs, that is, on the ratio of their  $\text{var}(\hat{\beta}_1)$ . The latter in turn does not depend on the ES. Further, for given cost ratios, the actual design defines the available budget, that is: from the assumed costs and the actual design (nr of clusters, nr of individuals per cluster) the budget needed for the actual design was computed. The optimal and maximin designs are then calculated for that budget and the same cost ratios. The percentage power gains of these optimal and maximin designs are shown in the last two columns for each set of cost ratios. The budget is thus the same within each row (and thus allows for comparison of powers within a row), but may differ from one row to the other.

Table 1 shows that using optimal and maximin designs can substantially reduce the total study costs and increase the power of the study as compared with the actual design, depending on the cost ratios. Table 1 shows that in some cases the *MMD* reduces the budget and increases the power more than does the optimal design. This happens

**Input Parameters:**

Quasi-ICC for Effects ( $0 \leq \theta_e \leq 1$ )	0.26
Quasi-ICC for Costs ( $0 \leq \theta_c \leq 1$ )	0.05
Correlation between Random Slope for Costs and Effects ( $-1 \leq \rho_{u1} \leq 1$ )	0.66
Correlation between Costs and Effects at Individual Level ( $-1 \leq \rho_{\epsilon} \leq 1$ )	-0.083
Quasi-Variance Ratio ( $\phi > 0$ )	0.60
Level of Significance	0.05
Effect Size (ES)	0.5
Power Level	0.80
Costs per Centre ( $c > 0$ )	4000
Costs per Individual in the Treatment Arm ( $s_t > 0$ )	1000
Costs per Individual in the Control Arm ( $s_c > 0$ )	500

Buttons: Calculate Optimal Design, Quit

**Fig. 5** Entries in **SamP2CeT** submenu to calculate the optimal design for the EVALUATE trial when the power is fixed

**Optimal Design for a Multicentre Cost-effectiveness Trial:**

Optimal number of treatment subjects per centre ( $m_{opt}$ )	9
Optimal number of control subjects per centre ( $n_{opt}$ )	12
Optimal number of centres ( $k_{opt}$ )	11
Total number of subjects	231
Total Costs (C)	209000
Power	0.8

Button: OK

**Fig. 6** Result window of **SamP2CeT** appearing for the optimal design for the EVALUATE trial when the power is fixed



The screenshot shows a window titled "SamP2CeT: submenu" with a blue header. The main area is titled "Input Parameters:" and contains several input fields with their corresponding values:

Parameter	Value
Quasi-ICC for Effects ( $0 \leq \theta_e \leq 1$ )	0.26
Quasi-ICC for Costs ( $0 \leq \theta_c \leq 1$ )	0.05
Correlation between Random Slope for Costs and Effects ( $-1 \leq \rho_{u1} \leq 1$ )	0.66
Correlation between Costs and Effects at Individual Level ( $-1 \leq \rho_{\epsilon} \leq 1$ )	-0.083
Quasi-Variance Ratio ( $\phi \geq 0$ )	0.60
Level of Significance	0.05
Effect Size (ES)	0.5
Number of Centres (k)	25
Number of Treatment Subjects per Centre (m)	23
Number of Control Subjects per Centre (n)	11

At the bottom of the window, there are two buttons: "Calculate Power" and "Quit".

**Fig. 7** Entries in **SamP2CeT** submenu to calculate power for the actual design of the EVALUATE trial when the information on correlation and variance parameters is precise

because the optimal design and the *MMD* were computed based on two different sets of parameters, making these two designs not directly comparable with each other in Table 1, but only with the actual study designs of the SPHERE and EVALUATE trials. Actually, the maximin design is the optimal design for those parameter values within the chosen range of parameter values that give the largest (worst case) sampling variance of the treatment effect (for details, see Manju et al. (2014), Manju et al. (2015)). In the present examples, these worst case parameter values are  $\rho_e = 0.3$  and  $\rho_c = 0.30$  for the SPHERE study, and  $\theta_e = 0.3$  and  $\theta_c = 0.30$  for the EVALUATE trial. Choosing these values for the optimal design would have given the same design as the maximin design.

## 8 Discussion

In this paper, we have briefly reviewed recent methodological developments on sample size and power calculation for both cost-effectiveness CRTs and multicentre cost-effectiveness trials, and we have introduced a software implementation with a graphical user interface **SamP2CeT**. The sample sizes concern the numbers of clusters, and subjects per cluster, in each treatment arm for a cost-effectiveness CRT, or the numbers of centres and subjects per treatment arm per centre for a multicentre cost-effectiveness trial. This paper describes the use of the program **SamP2CeT** to compute the smallest budget needed for a user-specified power level, or the largest power attainable with a user-specified budget, as well as the corresponding sample sizes. In addition, the program allows for power calculation for a user-specified design. Calculating either



**Table 1** Budget and power gain (%) of optimal and maximin versus actual samples

Cost ratio for treatment arm ( $c_t/s_t$ )	Cost ratio for control arm ( $c_c/s_c$ )	Subject specific cost ratio ( $s_t/s_c$ )	Budget gain (optimal vs actual) (%)	Budget gain (maximin vs actual) (%)	Power gain (optimal vs actual) (%)	Power gain (maximin vs actual) (%)
Cluster randomized trial: the SPHERE study						
2	2	1	0.080	62.51	0.08	128.86
2	10	1	5.60	47.23	2.48	72.77
2	50	1	33.44	33.56	17.10	42.68
10	2	1	5.60	47.23	2.48	72.77
10	10	1	9.01	31.47	3.03	38.04
10	50	1	32.40	19.17	15.62	20.22
50	2	1	33.44	33.56	17.10	42.68
50	10	1	32.40	19.17	15.62	20.22
50	50	1	40.97	4.16	19.00	3.80
2	2	4	8.73	65.76	4.86	145.82
2	10	4	8.23	54.46	3.78	95.18
2	50	4	20.30	37.50	9.97	49.67
10	2	4	18.64	49.32	10.01	79.20
10	10	4	17.24	37.67	8.12	49.85

Table 1 continued

Cost ratio for treatment arm ( $c_t/s_t$ )	Cost ratio for control arm ( $c_c/s_c$ )	Subject specific cost ratio ( $s_t/s_c$ )	Budget gain (%) (optimal vs actual)	Budget gain (%) (maximin vs actual)	Power gain (%) (optimal vs actual)	Power gain (%) (maximin vs actual)
10	50	4	24.39	20.72	10.99	21.85
50	2	4	49.26	38.54	23.35	52.89
50	10	4	47.31	29.34	22.48	35.27
50	50	4	46.64	13.36	21.70	13.20
Cost ratio ( $c_t/s_t$ )	Cost ratio ( $c_c/s_c$ )	Cost ratio ( $s_t/s_c$ )	Budget gain (%) (optimal vs actual)	Budget gain (%) (maximin vs actual)	Power gain (%) (optimal vs actual)	Power gain (%) (maximin vs actual)
Multicentre trial: the EVALUATE trial						
2	2	1.0	21.23	65.49	21.41	113.29
2	5	2.5	31.41	67.00	31.27	124.00
2	10	5.0	39.10	68.93	39.32	136.25
5	2	0.4	9.16	56.31	10.06	79.26
5	5	1.0	14.32	53.15	13.88	72.56
5	10	2.0	21.89	53.49	20.59	75.83
10	2	0.2	5.88	51.98	6.99	67.26
10	5	0.5	4.84	43.34	5.18	48.82
10	10	1.0	9.97	40.43	8.85	44.74

Power gain (%) = (Power of optimal or maximin design - Power of actual design) \* 100% / (Power of actual design)

the sample sizes, or the power for a user-specified design, requires prior knowledge on correlation and variance parameters. Therefore, the computer program allows the user to either specify values for these parameters, or plausible ranges for them, depending on the precision of the user's knowledge on these parameters.

Following Manju et al. (2014), Manju et al. (2015) this paper and software use maximin design to handle uncertainty about those parameter values on which the optimal design depends. Two other approaches in the literature are the Bayesian and the sequential approach. The Bayesian approach (Spiegelhalter et al. 2007) maximizes the expected instead of minimum efficiency, which is a more optimistic but also less safe approach than maximin design. Apart from this, Bayesian design has two drawbacks that need consideration before preferring it to maximin design. First, maximizing expected efficiency, or maximizing expected power, or minimizing expected sampling variance, give different designs because efficiency, power and sampling variance are nonlinearly related. In contrast, maximin design gives the same design for all three criteria because they are monotonically related. Secondly, Bayesian design depends on the chosen prior for the parameters on which the design depends, whereas maximin design only depends on the range for those parameters. The sequential procedure handles parameter uncertainty by updating the design based on parameter estimates at interim analyses (Wu 1985). This second approach could be combined with maximin design, for instance by choosing the maximin design as initial design, which can then be updated at interim.

**Acknowledgements** This research was entirely funded by the Netherlands Organisation for Health Research and Development (ZonMw) [Grant Number 80-82310-97-12201].

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## Appendix A

### Cluster randomized trial (CRT)

#### Optimal sample sizes

Finding an optimal design means finding the  $m$ ,  $n$ ,  $k_t$  and  $k_c$  which minimize  $\text{var}(\hat{\beta}_1)$  in Eq. (2), and thus maximizes the statistical power, given the budget constraint in Eq. (3). It has been shown by Manju et al. (2014) that this gives the following optimal design:

$$m_{opt} = \sqrt{\frac{c_t B}{s_t A}} \quad \text{and} \quad n_{opt} = \sqrt{\frac{c_c B}{s_c A}},$$

$$k_{t_{opt}} = \frac{C\sqrt{A}}{\sqrt{c_t}\{(\sqrt{c_c A} + \sqrt{s_c B}) + (\sqrt{c_t A} + \sqrt{s_t B})\}} \quad \text{and} \quad k_{c_{opt}} = \frac{k_{t_{opt}}\sqrt{c_t}}{\sqrt{c_c}},$$
(A.1)

where  $A = \varphi\rho_e + \rho_c - 2\rho_{u0}\sqrt{\rho_e\rho_c\varphi}$  and  $B = \varphi(1 - \rho_e) + (1 - \rho_c) - 2\rho_e\sqrt{(1 - \rho_e)(1 - \rho_c)\varphi}$ , and the resulting optimal (= minimal) variance for the vector of parameters  $\Pi = (\rho_e, \rho_c, \rho_{u0}, \rho_e, \varphi, v_c)$  is then given by:

$$\text{var}(\hat{\beta}_1|\Pi, \text{opt}) = \frac{\{(\sqrt{c_t} + \sqrt{c_c})\sqrt{A} + (\sqrt{s_t} + \sqrt{s_c})\sqrt{B}\}^2 v_c}{C}. \tag{A.2}$$

Note that, the optimal design in Eq. (A.1) does not depend on the coding of the treatment indicator variable  $x_j$  in any way, since changing the coding would only affect the fixed intercepts in Eq. (1), but none of the model parameters on which the optimality criterion in Eq. (2) depends. In this section the budget  $C$  was fixed and the power was then maximized. However, the present equations can also be used to minimize the budget needed for a given power, or to evaluate the power of a given design, as will be seen in the section on power calculation.

**Maximin sample sizes**

Equation (A.1) shows that computing the optimal design requires knowledge about the model parameters  $\rho_e, \rho_c, \rho_{u0}, \rho_e, \varphi$  and  $v_c$ . As described in the previous section, since specifying these parameter values may not be possible at the planning stage of a trial, a *MMD* can be determined as an efficient design over a plausible range of values for these unknown model parameters. The *MMD* for planning a cost-effectiveness CRT is given by the following expression (Manju et al. 2014):

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

$$k_{tMMD} = \frac{C\sqrt{\rho}}{\sqrt{c_t}\{(\sqrt{c_c} + \sqrt{c_t})\sqrt{\rho} + (\sqrt{s_c} + \sqrt{s_t})\sqrt{(1 - \rho)}\}} \text{ and } k_{cMMD} = \frac{k_{tMMD}\sqrt{c_t}}{\sqrt{c_c}}, \tag{A.3}$$

where  $\rho$  is the maximum of the upper bounds of  $\rho_e$  and  $\rho_c$ . The maximin (= worst case) variance of the ML estimator  $\hat{\beta}_1$  is then:

$$\text{var}(\hat{\beta}_1|\Pi_1, MMD) = \frac{\{(\sqrt{c_t} + \sqrt{c_c})\sqrt{\rho} + (\sqrt{s_t} + \sqrt{s_c})\sqrt{(1 - \rho)}\}^2(\varphi^U + 1 - 2r\sqrt{\varphi^U})v_c^U}{C}, \tag{A.4}$$

where  $r(-1 \leq r < 0)$  is the minimum of the two lower boundaries of  $\rho_{u0}$  and  $\rho_e$ , and  $\varphi^U$  and  $v_c^U$  are the upper bounds for plausible ranges of  $\varphi$  and  $v_c$ , respectively, therefore,  $\Pi_1 = (\rho, r, \varphi^U, v_c^U)$  denotes the vector of the maximin parameter boundaries. Just as in the previous section, the present equations can be used to maximize the power, or to minimize the budget, or to evaluate the power of a given design. This is explained in more detail in the next section.

## Power calculation

The optimal design in Eq. (A.1) and the *MMD* in Eq. (A.3) have been derived for a fixed budget. Therefore, Eqs. (A.2) and (A.4) give the smallest variances of the treatment effect estimator and the largest precision, given the budget and the model parameters (optimal design) or plausible ranges of these model parameters (maximin design). In designing a cost-effectiveness study, however, one will often aim at achieving a certain power for the treatment effect test instead of fixing the budget beforehand. It will now be shown how the equations in the preceding sections can be used to compute the budget needed to achieve the desired power. To understand this it is important to note that the optimal respectively maximin cluster sizes  $m$  and  $n$  do not depend on the budget, whereas the optimal respectively maximin numbers of clusters do. So if a given budget does not give sufficient power, then the budget must be increased and this will lead to an increase of the number of clusters  $k_t$  and  $k_c$ , not of the cluster sizes  $m$  and  $n$ .

To test the null hypothesis  $H_0 : \beta_1 = 0$ , that is, that the treatment and control are equally cost-effective, against the alternative hypothesis  $H_A : \beta_1 \neq 0$ , we have the test statistic  $f = \frac{\hat{\beta}_1^2}{\text{var}(\hat{\beta}_1)}$  (Liu 2003). Under the alternative hypothesis, this test statistic  $f$  follows the non-central  $F$  distribution, that is,  $F(1, k_t + k_c - 2; \delta)$ , with numerator and denominator degrees of freedom 1 and  $k_t + k_c - 2$ , respectively, and with non-centrality parameter  $\delta = \frac{\beta_1^2}{\text{var}(\hat{\beta}_1)}$ . Under  $H_0$ ,  $\delta = 0$  and so the critical value  $f_0$  for the test statistic is the  $(1 - \alpha)$ th-quantile of the central  $F(1, k_t + k_c - 2; \delta = 0)$  distribution and the power of the test for the treatment effect is:

$$1 - \gamma = 1 - Pr[f < f_0]. \quad (\text{A.5})$$

where  $Pr[f < f_0]$  obeys the non-central  $F$ -distribution. The power depends on the non-centrality parameter  $\delta$ , which can be rewritten in terms of the effect size ( $ES$ ) measure introduced by Cohen's  $d$ , (Cohen 1992) that is,  $ES = |\beta_1| / \sqrt{\text{var}(NMB_{ij})}$ , where the variance of  $NMB_{ij}$ ,  $\text{var}(NMB_{ij}) = (A + B)v_c$ . More specifically,

$$\delta = \frac{k_t k_c m n ES^2 (A + B)}{k_t m (nA + B) + k_c n (mA + B)}. \quad (\text{A.6})$$

The researcher may now choose an effect size ( $ES$ ) based on clinical considerations (e.g., smallest relevant effect size) and then calculate the maximum power for the optimal design (or the *MMD*) in Eq. (A.1) (or Eq. (A.3)). Given this  $ES$ , the non-centrality parameter  $\delta$  in Eq. (A.6) does not depend on  $v_c$  and so the researcher does not need to specify  $v_c$ . Further, for the *MMD* in Eq. (A.3) the expression for the non-centrality parameter  $\delta$  can be written as:

$$\delta = \frac{k_t k_c m n ES^2}{m n \rho (k_t + k_c) + (1 - \rho) (k_t m + k_c n)}, \quad (\text{A.7})$$

where  $ES = |\beta_1| / \sqrt{(\varphi^U + 1 - 2r\sqrt{\varphi^U})v_c^U}$ . Note that, the expression in Eq. (A.7) is obtained by substituting in  $A$  and  $B$  in Eq. (A.6)  $\rho$  for  $\rho_e$  and  $\rho_c$ , where  $\rho$  is the maximum of the upper bounds of  $\rho_e$  and  $\rho_c$ ,  $r$  for  $\rho_{u0}$  and  $\rho_\epsilon$ , where  $r$  is the minimum of the lower bounds of  $\rho_{u0}$  and  $\rho_\epsilon$ , and  $\varphi^U$  for  $\varphi$ , where  $\varphi^U$  is the upper bound of  $\varphi$ . The non-centrality parameter  $\delta$  in Eq. (A.7) is independent of  $r$ ,  $\varphi^U$  and  $v_c^U$  once the  $ES$  has been chosen. As a consequence, the researcher can get rid of  $r$ ,  $\varphi^U$  and  $v_c^U$  in calculating the power for the  $MMD$  in Eq. (A.3). It is to be noted that, the denominator of the  $ES$  for the  $MMD$  is the worst case (maximum) variance for the  $NMB$  and so the  $ES$  for the  $MMD$  is the worst case  $ES$ .

The procedure as to how to obtain the power of the optimal design for a given budget using Eq. (A.5) is as follows. For a given budget  $C$ , model parameters and sampling costs, one can compute the optimal  $m$ ,  $n$ ,  $k_t$  and  $k_c$  with Eq. (A.1). Using these optimal  $m$ ,  $n$ ,  $k_t$  and  $k_c$ , the non-centrality parameter  $\delta$  in Eq. (A.6) can be calculated for a given effect size  $ES$ , and  $A$  and  $B$ . Also, given the optimal  $k_t$  and  $k_c$ , the critical  $F$ -value,  $f_0$ , in Eq. (A.5) can be obtained from the central  $F$ -distribution for a given type I error risk  $\alpha$ . The power  $1 - Pr[f < f_0]$  is then obtained from the non-central  $F$ -distribution. This power is the largest possible power, given the model parameters, sampling costs,  $ES$ , type I error rate and budget. If this power is too low (or high), then we have to increase (or decrease) the budget and repeat the process of computing the optimal design,  $\delta$  in Eq. (A.6), and  $f_0$ , and the power, using Eq. (A.5) until the desired level of power is attained. This budget is then the smallest possible budget, given the model parameters, sampling costs,  $ES$ , and type I error rate yielding a pre-specified power level. Note that, in a similar way, we can calculate the  $MMD$ , either when fixing the budget or the power, this time making use of Eq. (A.3) instead of Eq. (A.1). The same procedure as described above for the optimal design is followed, the only difference being that the computation of  $\delta$  is based on Eq. (A.7) instead of Eq. (A.6).

## Appendix B

### Multicentre trial

#### Optimal sample sizes

The following optimal design of multicentre cost-effectiveness trials has been derived from Eqs. (5) and (6), consisting of the number of centres ( $k_{opt}$ ) and the number of individuals in each treatment arm within each centre ( $m_{opt}$ ,  $n_{opt}$ ) (Manju et al. 2015):

$$m_{opt} = \sqrt{\frac{cB'}{s_t A'}}, n_{opt} = \sqrt{\frac{cB'}{s_c A'}} \text{ and } k_{opt} = \frac{C}{\sqrt{\frac{cB'}{A'}} (\sqrt{s_t} + \sqrt{s_c}) + c}. \tag{B.1}$$

where  $A' = \phi\theta_e + \theta_c - 2\rho_{u1}\sqrt{\theta_e\theta_c\phi}$ ,  $B' = \phi(1 - \theta_e) + (1 - \theta_c) - 2\rho_\epsilon\sqrt{(1 - \theta_e)(1 - \theta_c)\phi}$ . The optimal (= minimal) variance of the average  $INMB$  esti-

mator, given a vector of parameters  $\Pi' = (\theta_e, \theta_c, \rho_{u1}, \rho_\epsilon, \phi, v'_c)$ , is then:

$$\text{var}(\hat{\beta}_1 | \Pi', \text{opt}) = \frac{\{\sqrt{B'}(\sqrt{s_t} + \sqrt{s_c}) + \sqrt{A'c}\}^2 v'_c}{C}. \quad (\text{B.2})$$

Analogously to what was seen in the section on cluster randomized trials, the present equations for multicenter trials can be used to maximize the power for a given budget, or to minimize the budget needed for a given power, or to evaluate the power of a given design.

### Maximin sample sizes

If precise knowledge of the model parameters  $\theta_e, \theta_c, \rho_{u1}, \rho_\epsilon, \phi$  and  $v'_c$  is lacking, which is usually the case, then a *MMD* can again be chosen. We assume finite intervals for each of the model parameters  $\theta_e, \theta_c, \rho_{u1}, \rho_\epsilon, \phi$  and  $v'_c$ , where the lower bounds can be negative for  $\rho_{u1}$  and  $\rho_\epsilon$ , but not for the other three parameters. Four different *MMDs* were derived based on the signs of the lower bounds for the correlations  $\rho_{u1}$  and  $\rho_\epsilon$  (Manju et al. 2015). Here we consider the case where both correlations can be negative and up to  $-1$ , which is the worst case in terms of the maximum possible sampling variance of the average *INMB* estimator ( $\hat{\beta}_1$ ). This gives the following *MMD* (Manju et al. 2015):

$$m_{MMD} = \sqrt{\frac{c(1-\theta)}{s_t\theta}}, n_{MMD} = \sqrt{\frac{c(1-\theta)}{s_c\theta}} \text{ and } k_{MMD} = \frac{C}{\sqrt{\frac{c(1-\theta)}{\theta}}(\sqrt{s_t} + \sqrt{s_c}) + c}, \quad (\text{B.3})$$

where  $\theta$  is the maximum of the upper bounds of  $\theta_e$  and  $\theta_c$ . The maximin (= worst case) variance of the ML estimator  $\hat{\beta}_1$  is then:

$$\text{var}(\hat{\beta}_1 | \Pi'_1, MMD) = \frac{\{\sqrt{(1-\theta)}(\sqrt{s_t} + \sqrt{s_c}) + \sqrt{\theta c}\}^2 (\phi^U + 1 - 2r'\sqrt{\phi^U}) v'_c{}^U}{C}, \quad (\text{B.4})$$

where  $r' (-1 \leq r' < 0)$  is the minimum of the two lower boundaries of  $\rho_{u1}$  and  $\rho_\epsilon$ , and  $\phi^U$  and  $v'_c{}^U$  are the upper bounds for plausible ranges of  $\phi$  and  $v'_c$ , respectively, therefore,  $\Pi'_1 = (\theta, r', \phi^U, v'_c{}^U)$  denotes the vector of the maximin parameter boundaries.

### Power calculation

A design minimizing the costs of the study,  $C$ , for given power  $(1 - \gamma)$ , or maximizing the power  $(1 - \gamma)$  for given total costs  $C$ , can be found if one uses the test statistic  $f = \frac{\hat{\beta}_1^2}{\text{var}(\hat{\beta}_1)}$ . The test statistic  $f$  follows the non-central  $F$  distribution, that is  $F(1, k - 1; \delta)$ , with the numerator and denominator degrees of freedom 1 and  $k - 1$ , respectively, and with the non-centrality parameter  $\delta = \frac{kmnES^2}{mnES_v + m + n}$ . Here,  $ES = \frac{|b-a|\beta_1}{\sigma}$  is the effect size, (Cohen 1992) and  $ES_v = \frac{(b-a)^2\sigma_1^2}{\sigma^2}$  is the effect size variability (Raudenbush and



Liu 2000). Note that, in case of CRTs, the  $ES$  as defined in the Power calculation section for CRT uses as denominator the square root of the total outcome variance of  $NMB$ , whereas the  $ES$  in the multicentre trial uses as denominator the square root of the individual level  $NMB$  variance only. Also note that by specifying  $\theta_e$ ,  $\theta_c$ ,  $\rho_{u1}$ ,  $\rho_\epsilon$  and  $\phi$  the effect size variability ( $ES_v$ ) is already fixed, and need not be specified anymore. The power function for a two-sided test becomes (Manju et al. 2015):

$$1 - \gamma = 1 - Pr[f < f_0], \quad (\text{B.5})$$

where  $f_0$  is the  $(1 - \alpha)$ th-quantile of the central  $F(1, k - 1; \delta = 0)$  distribution. When fixing a clinically relevant effect size ( $ES$ ) in calculating the optimal design in Eq. (B.1) for a multicentre trial, similar to a cost-effectiveness CRT, the non-centrality parameter  $\delta$  does not depend on  $v'_c$ . Furthermore, when fixing  $ES$  in calculating the  $MMD$  in Eq. (B.3), since the effect size variability ( $ES_v$ ),  $m$ ,  $n$  and  $k$  do not depend on  $r'$ ,  $\phi^U$  and  $v'_c^U$ , also the non-centrality parameter  $\delta$  no longer depends on these model parameters. Following the same procedure as described in the Power calculation section for CRTs, but now using the corresponding equations from the Multicentre trials section, Eq. (B.5) can be used to obtain the smallest possible budget for a desired level of power, or to obtain the largest power for a given budget, or to compute the power for a user-specified design in case of a multicentre trial.

## References

- Atkinson AC, Donev AN, Tobias RD (2007) Optimum experimental designs, with SAS. Oxford University Press, Oxford
- Berger MPF, Wong WK (2009) An introduction to optimal designs for social and bio-medical research. Wiley, Chichester
- Bornkamp B, Pinheiro J, Bretz F (2009) MCPMod: an R package for the design and analysis of dose-finding studies. *J Stat Softw* 29(7):1–23
- Bosker RJ, Snijders TAB, Guldemond H (2003) PINT, (Power IN Two-level Designs): Users Manual. URL <http://www.stats.ox.ac.uk/~snijders/multilevel.htm>. Accessed 21 Dec 2015
- Cohen J (1992) A power primer. *Psychol Bull* 112(1):155
- Cools W, Van den Noortgate W, Onghena P (2008) ML-DEs: a program for designing efficient multilevel studies. *Behav Res Methods* 40(1):236–249
- Dalgaard P (2001) The **R-tcl/tk** interface. In: DSC 2001 proceedings of the 2nd international workshop on distributed statistical computing, Vienna, Austria, p 2
- Gomes M, Ng ESW, Grieve R, Nixon R, Carpenter J, Thompson SG (2012) Developing appropriate methods for cost-effectiveness analysis of cluster randomized trials. *Med Dec Mak* 32(2):350–361
- Grieve R, Hutton J, Bhalla A, Rastenyte D, Ryglewicz D, Sarti C, Lamassa M, Giroud M, Dundas R, Wolfe CDA et al (2001) A comparison of the costs and survival of hospital-admitted stroke patients across europe. *Stroke* 32(7):1684–1691
- Grieve R, Nixon R, Thompson SG, Cairns J (2007) Multilevel models for estimating incremental net benefits in multinational studies. *Health Econ* 16(8):815–826
- Grieve R, Nixon R, Thompson SG (2010) Bayesian hierarchical models for cost-effectiveness analyses that use data from cluster randomized trials. *Med Dec Mak* 30:163–175
- Liu X (2003) Statistical power and optimum sample allocation ratio for treatment and control having unequal costs per unit of randomization. *J Educ Behav Stat* 28(3):231–248
- Manca A, Rice N, Sculpher MJ, Briggs AH (2005) Assessing generalisability by location in trial-based cost-effectiveness analysis: the use of multilevel models. *Health Econ* 14(5):471–485
- Manju MA (2016) Optimal sample sizes for cost-effectiveness cluster randomized and multicentre trials. Ph.D. thesis, Maastricht University

- Manju MA, Candel MJJM, Berger MPF (2014) Sample size calculation in cost-effectiveness cluster randomized trials: optimal and maximin approaches. *Stat Med* 33(15):2538–2553
- Manju MA, Candel MJJM, Berger MPF (2015) Optimal and maximin sample sizes for multicentre cost-effectiveness trials. *Stat Methods Med Res* 24(5):513–539
- Morrell CJ, Slade P, Warner R, Paley G, Dixon S, Walters SJ, Brugha T, Barkham M, Parry GJ, Nicholl J (2009) Clinical effectiveness of health visitor training in psychologically informed approaches for depression in postnatal women: pragmatic cluster randomised trial in primary care. *Br Med J* 338:276–280
- Ng ESW, Diaz-Ordaz K, Grieve R, Nixon RM, Thompson SG, Carpenter JR (2016) Multilevel models for cost-effectiveness analyses that use cluster randomised trial data: an approach to model choice. *Stat Methods Med Res* 25(5):2036–2052
- Peace K, Chen DG (2010) Clinical trial data analysis using R. Chapman and Hall, Boca Raton
- Raudenbush SW, Liu X (2000) Statistical power and optimal design for multisite randomized trials. *Psychol Methods* 5(2):199–213
- R Development Core Team (2015) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, URL <http://www.R-project.org>
- Sculpher M, Manca A, Abbott J, Fountain J, Mason S, Garry R (2004) Cost effectiveness analysis of laparoscopic hysterectomy compared with standard hysterectomy: results from a randomised trial. *Br Med J* 328(7432):134–139
- Senn S (2014) A note regarding “random effects”. *Stati Med* 33(16):2876–2877
- Snijders TAB, Bosker RJ (1993) Standard errors and sample sizes for two-level research. *J Educ Behav Stat* 18(3):237–259
- Spiegelhalter DJ, Abrams KR, Myles JP (2007) Bayesian approaches to clinical trials and health-care evaluation. Wiley, Chichester
- Spybrook J, Bloom H, Congdon R, Hill C, Martinez A, Raudenbush SW (2011) Optimal Design for Longitudinal and Multilevel Research: Documentation for the Optimal Design Software Version 3.0. URL [http://sitemaker.umich.edu/group-based/optimal\\_design\\_software](http://sitemaker.umich.edu/group-based/optimal_design_software). Accessed 13 Mar 2015
- Stinnett AA, Mullahy J (1998) Net health benefits: a new framework for the analysis of uncertainty in cost-effectiveness analysis. *Med Dec Mak* 18(2):S68–S80
- Tokola K, Laroque D, Nevalainen J, Oja H (2011) Power, sample size and sampling costs for clustered data. *Stat Probab Lett* 81(7):852–860
- Verbeke G, Molenberghs G (2000) Linear mixed models for longitudinal data. Springer, New York
- Welch BB (1995) Practical programming in Tcl and Tk, vol 391. Prentice Hall PTR, New Jersey
- Willan AR, Briggs AH (2006) Statistical analysis of cost-effectiveness data. Wiley, Hoboken
- Wu CFJ (1985) Efficient sequential designs with binary data. *J Am Stat Assoc* 85:974–984

## Affiliations

Md Abu Manju<sup>1</sup>  · Math J. J. M. Candel<sup>1</sup> · Gerard J. P. van Breukelen<sup>1</sup>

<sup>1</sup> Department of Methodology and Statistics, CAPHRI Care and Public Health Research Institute, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands