

RESEARCH ARTICLE

Open Access



# Sample size calculation in multi-centre clinical trials

Markus Harden\*  and Tim Friede

## Abstract

**Background:** Multi-centre randomized controlled clinical trials play an important role in modern evidence-based medicine. Advantages of collecting data from more than one site are numerous, including accelerated recruitment and increased generalisability of results. Mixed models can be applied to account for potential clustering in the data, in particular when many small centres contribute patients to the study. Previously proposed methods on sample size calculation for mixed models only considered balanced treatment allocations which is an unlikely outcome in practice if block randomisation with reasonable choices of block length is used.

**Methods:** We propose a sample size determination procedure for multi-centre trials comparing two treatment groups for a continuous outcome, modelling centre differences using random effects and allowing for arbitrary sample sizes. It is assumed that block randomisation with fixed block length is used at each study site for subject allocation. Simulations are used to assess operation characteristics such as power of the sample size approach. The proposed method is illustrated by an example in disease management systems.

**Results:** A sample size formula as well as a lower and upper boundary for the required overall sample size are given. We demonstrate the superiority of the new sample size formula over the conventional approach of ignoring the multi-centre structure and show the influence of parameters such as block length or centre heterogeneity. The application of the procedure on the example data shows that large blocks require larger sample sizes, if centre heterogeneity is present.

**Conclusion:** Unbalanced treatment allocation can result in substantial power loss when centre heterogeneity is present but not considered at the planning stage. When only few patients by centre will be recruited, one has to weigh the risk of imbalance between treatment groups due to large blocks and the risk of unblinding due to small blocks. The proposed approach should be considered when planning multi-centre trials.

**Keywords:** Block randomisation, Linear mixed model, Random effects

## Background

When planning a randomized controlled clinical trial, sample size considerations are necessary to assess how many subjects are needed, e.g. to demonstrate a beneficial effect of a new treatment. These considerations usually are based on initial assumptions regarding a clinically meaningful treatment effect, the variability in the data and prespecified type I and type II error rates. Patients are often recruited in more than one centre, for example to account for a low incidence of the disease [1]. Since these centres might differ in some ways, between-centre heterogeneity at baseline needs to be

accounted for in the analysis and therefore sample size planning.

When analysing continuous outcomes, baseline differences between centres can be accounted for using either a linear fixed-effects or a linear mixed model. Due to the central limit theorem, sample size calculation can often be based on the normal approximation

$$N = \frac{\sigma^2(k+1)^2}{k} \left( \frac{q_{1-\alpha/2} + q_{1-\beta}}{\mu^*} \right)^2 \quad (1)$$

where  $N$  denotes the total sample size,  $\mu^*$  the assumed treatment effect,  $\sigma^2$  the variance of the observations,  $k$  the allocation ratio between treatment groups and  $q_\gamma$  the  $\gamma$ -quantile of the standard normal distribution [2].

\*Correspondence: [markus.harden@med.uni-goettingen.de](mailto:markus.harden@med.uni-goettingen.de)  
Department of Medical Statistics, University Medical Centre Göttingen,  
Humboldtallee 32, 37073 Göttingen, Germany



There have been various attempts to extend Formula (1) to multi-centre trials, e.g. by including a multiplicative factor to account for deviations from the standard design. Gallo as well as Ruvuna suggested such an inefficiency factor to account for centre size imbalances in the fixed effects setting [3, 4]. Both methods rely on the proportion of the treatment effects' variances between balanced and imbalanced centre sizes, where the balanced case gives optimal power.

Van Breukelen and colleagues introduced an inefficiency factor for the mixed model [5]. This factor is based on the relative efficiency of unequal versus equal cluster sizes for the weighted least squares estimator, assuming a linear mixed effects model with an interaction between study site and treatment effect. Fedorov and Jones consider sample size formulas for balanced multi-centre designs and suggest simulations for more complex situations [6]. Vierron and Giradeau suggested a *design-effect* to adjust Eq. (1) for different study designs [7, 8].

All of the approaches mentioned above assume balanced treatment allocation by centre. Randomisation techniques such as block randomisation do not guarantee equal group sizes in all centres, especially if centres are small and block lengths are large. The normal approximation in (1) gives a lower boundary of the necessary sample size, but underpowered trials could occur, especially when between centre heterogeneity is large. We believe that this assumption is too strict for real trials and therefore suggest a sample size formula that accounts for unequal sample sizes.

It has been demonstrated that mixed models tend to yield better results compared to fixed effects models, especially when the number of patients per centre is small [9, 10]. For a small number of centres, however, the fixed effects design might result in better results, because the between-centre variation is likely to be estimated with bias in mixed models in that situation. We therefore aim to construct a sample size formula that accounts for baseline heterogeneity between study centres, assuming a linear mixed model for multi-centre designs.

## Methods

### Statistical model and estimators

We assume a linear mixed-effects model with a fixed intercept  $\mu_0$ , random effects  $u_j$ ,  $j = 1, \dots, c$  to account for centre heterogeneity at baseline, and a fixed treatment effect  $\mu$ . The data are assumed to follow some continuous distribution allowing for unequal sample sizes. The statistical model is given by

$$Y_{ijk} = \mu_0 + u_j + \mu \cdot x_i + \epsilon_{ijk} \tag{2}$$

for pairwise independent  $u_j, \epsilon_{ijk}$  with  $E(u_j) = 0$ ,  $Var(u_j) = \tau^2 < \infty$ ,  $E(\epsilon_{ijk}) = 0$ ,  $Var(\epsilon_{ijk}) = \sigma^2 < \infty$ , treatment

indicator  $x_i = 1_{\{i=2\}}$  for treatment groups  $i = 1, 2$ , centres  $j = 1, \dots, c$  and individuals  $k = 1, \dots, n_{ij}$  for each treatment-centre combination. The shared random effect  $u_j$  within centres induces the following covariance matrix  $Cov(Y_{111}, \dots, Y_{2cn_{2c}}) = \bigoplus_{j=1}^c [\sigma^2 \mathbf{I}_{n_j} + \tau^2 \mathbf{J}_{n_j}]$ , including all  $N$  observations with  $N = \sum_{i=1}^2 \sum_{j=1}^c n_{ij}$ . Here,  $\mathbf{I}_{n_j}$  denotes the  $n_j$ -dimensional identity matrix and  $\mathbf{J}_{n_j}$  the  $n_j$ -dimensional matrix consisting of ones only with  $n_j = n_{1j} + n_{2j}$ . We assume zero risk of contamination of the control group.

We are interested in differences between treatment groups and test the null hypothesis  $H_0 : \mu = 0$  against the two-sided alternative  $H_A : \mu \neq 0$ . The distribution of the estimated treatment effect  $\hat{\mu}$  can be approximated by a normal distribution, if the sample size is sufficiently large (say sample sizes larger 30). It follows

$$T = \frac{\hat{\mu}}{\sqrt{Var(\hat{\mu})}} \stackrel{H_0}{\sim} N(0, 1). \tag{3}$$

The null hypothesis can be rejected if the test statistic  $|T|$  exceeds the quantile  $q_{1-\alpha/2}$  of the reference distribution for some fixed type I error rate  $\alpha \in (0, 1)$ . In order to apply the statistical test, suitable estimators for the unknown parameters have to be chosen.

We choose  $\hat{\mu} = \bar{Y}_{2..} - \bar{Y}_{1..}$  to measure treatment group differences, where  $\bar{Y}_{i..} = \frac{1}{N_i} \sum_{j=1}^c \sum_{k=1}^{n_{ij}} Y_{ijk}$  denotes the group mean in treatment group  $i$ . This estimator is unbiased, even if centres recruited patients for one treatment group only. The variance of  $\hat{\mu}$  can be written as

$$Var(\hat{\mu}) = \sigma^2 \frac{N}{N_1 N_2} + \tau^2 \sum_{j=1}^c \left( \frac{n_{1j}}{N_1} - \frac{n_{2j}}{N_2} \right)^2. \tag{4}$$

Details on the derivation can be found elsewhere [8].

$Var(\hat{\mu})$  depends on the overall sample size  $N$ , the variances  $\sigma^2$  and  $\tau^2$ , and additionally the sample sizes by treatment group ( $N_1, N_2$ ), number of study centres ( $c$ ), and the sample sizes within study centres ( $n_{1j}, n_{2j}$ ). In case of a perfectly balanced randomisation, the differences between centres cancel out and the treatment effect's variance only depends on sample sizes  $N, N_1, N_2$  and variance  $\sigma^2$ , resulting in a sample size formula similar to (1).

The unknown variance parameters  $\tau^2$  and  $\sigma^2$  can be estimated using the following quadratic forms

$$\hat{\sigma}^2 = \frac{1}{2c} \sum_{i=1}^2 \sum_{j=1}^c \frac{1}{n_{ij} - 1} \sum_{k=1}^{n_{ij}} (Y_{ijk} - \bar{Y}_{ij.})^2$$

$$\hat{\tau}^2 = \frac{1}{2} \sum_{i=1}^2 \frac{1}{c - 1} \sum_{j=1}^c (\bar{Y}_{ij.} - \bar{Y}_{i..})^2.$$

**Additional assumptions for sample size calculation**

$N_i$  and  $n_{ij}$  are determined by recruitment and treatment allocation. We aim to replace all  $N_i, n_{ij}$  in (4) by their expectations that can be calculated based on the randomisation procedure and planned allocation proportion between treatment groups.

In the following, we want to calculate the overall sample size  $N$  and assume

- 1 a block randomisation stratified by centre with fixed block length  $b$ ,
- 2  $k : 1$  allocation ratio at each study site for  $k \in \mathbb{N}$ ,
- 3 proportion of overall sample sizes between treatment groups according to allocation:  $N_1 = kN_2$ .

**Block randomisation**

Since randomisation will not always result exactly in the planned allocation, we take a closer look at the randomisation process. Block randomisation with fixed block length  $b$  is a procedure where every  $b$  subjects get randomised between treatment groups at a time [11]. Complete blocks do always fulfil the planned  $k : 1$  allocation ratio. The block size should be unknown to investigators to strengthen the blinding in the trial.

In this article, we assume patients to be assigned to treatment groups  $i = 1, 2$  within centres, for a fixed  $k : 1$  allocation ratio. This means that in each randomisation block  $b$  patients are randomized between treatment groups 1 and 2 in a way that for each patient receiving treatment 2,  $k$  patients will receive treatment 1. The set of randomisation tuples  $\Pi_b^k$  depends on block length  $b$  and allocation parameter  $k$ . It is defined as follows

$$\Pi_b^k = \left\{ (x_1, \dots, x_b) \in \Pi_b \mid \sum_{\ell=1}^b 1_{\{x_\ell=1\}} = \frac{kb}{k+1} = b - \sum_{\ell=1}^b 1_{\{x_\ell=2\}} \right\} \tag{5}$$

where  $\Pi_b := \{(x_1, \dots, x_b) \mid x_\ell \in \{1, 2\}\}$ .

Treatment allocation imbalances can only occur in incomplete blocks with an upper boundary of  $kb/(k+1)$ . The choice of  $k$  can be based on several assumptions as ethics, costs and other factors and will not be discussed further in this article. This topic is covered in more detail in a review by Dumville and colleagues [12]. It is available in many software packages and is therefore easy to apply [13, 14]. Further advantages and disadvantages of block randomisation are considered in the Discussion.

**Derivation of the sample size formula**

The underlying idea of sample size calculation is to find the overall sample size  $N$ , such that the quantile  $q_{1-\alpha/2}$  of the reference distribution under the null hypothesis equals

the quantile  $q_\beta^*$  of the reference distribution under a fixed alternative for type I and II error rates  $\alpha$  and  $\beta$ .

Since we do assume a normally distributed test statistic,  $q_\beta^*$  can be approximated by a shifted  $N(0, 1)$ -quantile  $q_\beta^* \approx q_\beta + \frac{\mu}{\sqrt{\text{Var}(\mu)}}$  resulting in the following equation to construct a sample size formula

$$(q_{1-\alpha/2} + q_{1-\beta})^2 = \frac{\mu^2}{\text{Var}(\mu)}. \tag{6}$$

By isolating the sample size  $N$ , which is part of  $\text{Var}(\mu)$ , one can derive a sample size formula. In case of an ideal allocation, i. e.,  $n_{1j} = kn_{2j}$  for all centres, (6) is equal to (1). Since this is unlikely to be observed, unbalanced designs are taken into account by incorporating expectations with respect to the randomisation procedure.

We derive the sample size formula for the general case of a  $k : 1$  allocation ratio and assume the overall treatment group sample sizes to fulfil  $N_1 = kN_2$  and therefore  $N = (k+1)N_2$ . This leads to the set of randomisation tuples in (5). By taking assumptions 1-3, the variance of  $\hat{\mu}$  given in (4) simplifies to

$$\text{Var}(\hat{\mu}) = \frac{\sigma^2(k+1)^2}{kN} + \frac{\tau^2(k+1)^2}{N^2} \sum_{j=1}^c \Delta_j^2 \tag{7}$$

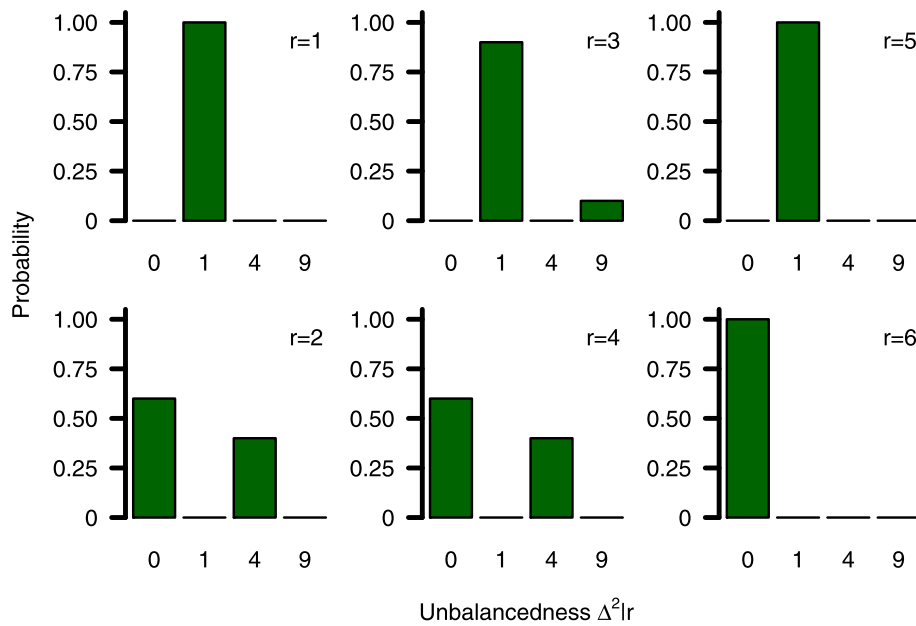
where  $\Delta_j^2 := \left(\frac{n_{1j}}{k} - n_{2j}\right)^2 \in [0, m^*]$  describe each centre's imbalance that will result from incomplete blocks with  $m^* = b^2/(k+1)^2$ . The (discrete) probability distribution of  $\Delta_j^2$  depends on  $\Pi_b^k$  and the number of patients in the last block which equals the remainder of the Euclidean division  $r_j = n_j \bmod b$ . An example of  $\Delta_j^2 | r_j$  for a single centre is illustrated in Fig. 1.

The probability distribution of  $\Delta_j^2 | r_j$  is fully described by block length  $b$ , allocation parameter  $k$  and  $r_j$ . For planning purposes it therefore seems reasonable to replace  $\Delta_j^2 | r_j$  by its expectation  $E(\Delta_j^2 | r_j)$  to eliminate sample sizes  $n_{1j}$  and  $n_{2j}$  from (7). The expectation of the probability distribution can easily be derived as

$$E(\Delta_j^2 | r_j) = \frac{1}{m^*} \sum_{\ell=0}^{m^*} p(\ell | r_j) \cdot \ell \tag{8}$$

where  $p(\cdot | r_j)$  denotes the conditional density function of  $\Delta_j^2 | r_j$ . These expectations are shown in Fig. 2 for a single randomization block,  $k = 1, 2, 3$  and various block lengths  $b$ . Since the expected imbalance is the largest for  $k = 1$  we will restrict simulations to this case.

The expected imbalance between treatment groups  $E(\Delta_j^2 | r_j)$  increases with block length. This happens due to the fact that the probability to receive an incomplete randomisation block increases with increasing block length  $b$ . It is maximised when the last randomisation



**Fig. 1** Probability distribution. Conditional probability distributions of  $\Delta^2|r$  for varying numbers of randomized subjects  $r = 1, \dots, b = 6$

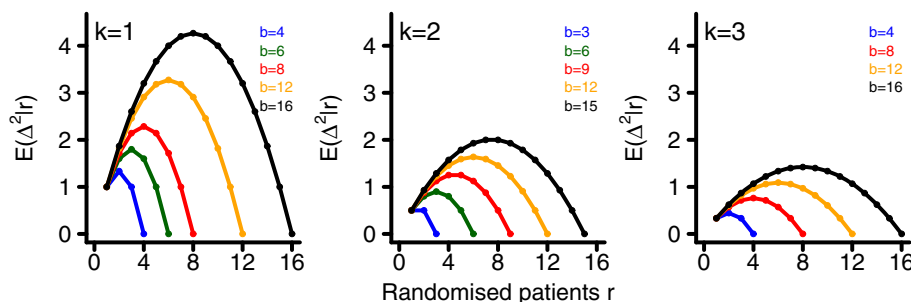
block only consists of  $\frac{bk}{k+1}$  patients receiving treatment 1 or  $\frac{b}{k+1}$  patients receiving treatment 2, respectively. If we replace  $\Delta_j^2|r_j$  by  $E(\Delta_j^2|r_j)$ , we can transform (6) into the following sample size formula for multi-centre trials (derivation is given in the appendix, see Additional file 1)

$$N_{MC}^k = \frac{\sigma^2(k+1)^2}{2k} \left( \frac{q_{1-\alpha/2} + q_{1-\beta}}{\mu} \right)^2 + \sqrt{\frac{\sigma^4(k+1)^4}{4k^2} + \frac{\tau^2(k+1)^2\mu^2 \sum_{j=1}^c E(\Delta_j^2|r_j)}{(q_{1-\alpha/2} + q_{1-\beta})^2}} \cdot \left( \frac{q_{1-\alpha/2} + q_{1-\beta}}{\mu} \right)^2. \tag{9}$$

**Simulations**

**General settings**

We perform a simulation study to assess the accuracy of the sample size formulas in terms of statistical power using R, version 3.3.1 [15]. For each scenario, we repeat  $n_{sim} = 10,000$  independent simulation runs. The R package `blockrand` is used for block randomisation [13]. All data are generated based on the statistical model described in (2), assuming a normal distribution for  $u_j$  and  $\epsilon_{ijk}$ . The test statistic  $T$  given in (3) is used for all power simulations and it is approximated by a standard normal distribution under the null hypothesis. The effect of block randomisation is strongest for  $k = 1$ , we therefore present simulation results for this setting only. The assumed effect size  $\mu^*$  and values for variance components  $\sigma^2$  and  $\tau^2$  are based on an example trial described in the next section.



**Fig. 2** Expectation of  $\Delta^2|r$ . Conditional expected imbalance between treatment groups for allocation parameter  $k = 1, 2, 3$  and various numbers of subjects  $r = 1, \dots, b$  and block lengths  $b$

All code used for simulations is available in the Appendix, see Additional files 2, 3, 4, 5, 6, 7 and 8.

### Subject-to-centre allocation

We consider equally as well as unequally sized study centres based on the following methods.

- 1 *Equally sized study centres*: Only in this situation, we can predict the sample size very precisely, since we can specify  $r_j$  correctly prior to recruitment. Here, study centres are assumed to be equal. Since this assumption is limited to the fixed overall sample size  $N$ , we distribute the overall sample size to centres as follows

$$\text{Equal: } n_j = \left\lfloor \frac{N}{c} \right\rfloor + 1_{\{j \leq m\}}$$

for  $m = N \bmod c$  and  $j = 1, \dots, c$ .

- 2 *Unequally sized study centres*: In most experiments, the true allocation of patients cannot be foreseen. Recruitment rates can be estimated beforehand, but since only the number of patients in the last randomisation block affects the presented sample size formulas, no precise estimation of unbalanced treatment allocation can be made. To model this situation, study centres are assumed to be unequal but limited to the fixed overall sample size  $N$ . We use a multinomial distribution to generate unequal sample sizes by centre, assuming the following scenarios

$$\begin{aligned} \text{Unequal 1: } (n_1, \dots, n_c)' &\sim \text{Multi}_c(N, p_1, \dots, p_c) \\ \text{with } p_j &= \frac{1}{c} \end{aligned}$$

$$\begin{aligned} \text{Unequal 2: } (n_1, \dots, n_c)' &\sim \text{Multi}_c(N, p_1^*, \dots, p_c^*) \\ \text{with } p_j^* &= \frac{p_j}{\sum_{k=1}^c p_k} \text{ and } p_j \stackrel{iid}{\sim} U[0; 1]. \end{aligned}$$

### Example: The COMPETE II trial

Multi-centre trials are applied in many different disease areas. We present an example in the setting of disease management systems and use this trial to illustrate the sample size approach proposed.

Holbrook and colleagues conducted a randomized, multi-centre trial to investigate the benefit of an individualized electronic decision support system in adult patients diagnosed with type 2 diabetes [16]. This new intervention provided patient specific summaries and recommendations based on electronic medical records, aiming to improve the quality of diabetes management between patients and general practitioners. The tool was integrated into the practice work flow and offered web-based access by patients. In addition, an automated telephone reminder

system was provided and all patients received a colour-coded printout quarterly. The control treatment consisted of usual care without use of this tool.

Primary outcome was a composite score difference compared to baseline. The composite score measured process quality on a scale from 0 to 10, based on the following parameters: blood pressure, cholesterol, glycated haemoglobin, foot check, kidney function, weight, physical activity, and smoking behaviour. The clinical targets are described in the original article. It was assessed at baseline and 6 months after randomization.

For this trial, 511 patients from 46 primary care providers were randomly assigned to intervention or control. At planning stage, the investigators aimed to recruit 508 patients to achieve 80% power to detect a difference of 1 for the primary outcome between treatment groups using a two-sided t-test with a type-1 error rate of  $\alpha = 0.05$ . No information on the assumed standard deviation is given in the article. Block-randomisation was stratified by study site in blocks of six, following a 1:1 allocation scheme.

The absolute measured improvement of composite scores between treatment groups was 1.26 (95% confidence interval (CI) 0.79-1.75;  $p$ -value  $< 0.001$ ) favouring the new intervention.

## Results

### Approaches to sample size calculation

As long as no values for  $r_j$  are assumed, Formula (9) cannot be used for sample size calculation. We consider a setting, where each centre will have at most one incomplete randomisation block. In the following, we present different ways to specify values of  $E(\Delta_j^2 | r_j)$  for each centre prior to recruitment.

The resulting sample size formulas are listed in Table 1. More detailed explanations are provided in the following subsections.

### Lower boundary

Given an ideal treatment allocation ( $n_{1j} = kn_{2j}$  for all centres), potential centre differences cancel out and do not affect the statistical power of the trial. In this case, sample size formula (1) would be suitable to plan the trial. However, the trial still had to be analysed with a mixed effects model to get an unbiased estimate for  $\text{Var}(\mu)$ , since  $\text{Var}(Y_{ijk}) = \sigma^2 + \tau^2$ . This formula is the standard approach to sample size calculation and is often used to plan multicentre trials. It therefore serves as one reference for sample size calculation and power simulations.

### Equal centre sizes

The exact number of patients by study centre will be unknown at the planning stage, but in some situations, it might be reasonable to assume centres to be equally large.

**Table 1** Overview of sample size formulas

Lower boundary	$N_{\text{lower}}^k = \left(\frac{q_{1-\alpha/2} + q_{1-\beta}}{\mu}\right)^2 \left(\frac{\sigma^2(k+1)^2}{k}\right)$
Equal centres	$N_{\text{MC,E}}^k = \left(\frac{q_{1-\alpha/2} + q_{1-\beta}}{\mu}\right)^2 \left(\frac{\sigma^2(k+1)^2}{2k} + \sqrt{\frac{\sigma^4(k+1)^4}{4k^2} + \frac{\tau^2(k+1)^2 \mu^2 c E(\Delta_1^2   r_1)}{(q_{1-\alpha/2} + q_{1-\beta})^2}}\right)$
Unequal centres	$N_{\text{MC,U}}^k = \left(\frac{q_{1-\alpha/2} + q_{1-\beta}}{\mu}\right)^2 \left(\frac{\sigma^2(k+1)^2}{2k} + \sqrt{\frac{\sigma^4(k+1)^4}{4k^2} + \frac{\tau^2(k+1)^2 \mu^2 c E(\Delta_1^2   \cdot)}{(q_{1-\alpha/2} + q_{1-\beta})^2}}\right)$
Upper boundary	$N_{\text{upper}}^k = \left(\frac{q_{1-\alpha/2} + q_{1-\beta}}{\mu}\right)^2 \left(\frac{\sigma^2(k+1)^2}{2k} + \sqrt{\frac{\sigma^4(k+1)^4}{4k^2} + \frac{\tau^2(k+1)^2 \mu^2 c E\left(\Delta_1^2 \middle  \frac{b}{k+1}\right)}{(q_{1-\alpha/2} + q_{1-\beta})^2}}\right)$

Lower and upper boundaries as well as sample size formulas for equal and unequal centre sizes

In this scenario, the sum of expected differences simplifies to a single quantity

$$\sum_{j=1}^c E(\Delta_j^2 | r_j) \approx c \cdot E(\Delta_1^2 | r_1) \tag{10}$$

which needs to be specified for sample size estimation.

Since the overall sample size  $N$  and the number of subjects in the last randomisation block depend on each other,  $r_1$  still has to be specified. The unspecified value of  $E(\Delta_1^2 | r_1)$  can be determined by calculating the sample size  $N_{\text{MC,E}}^k$  for each  $r_1 \in \{1, \dots, b\}$ . Keep  $N_{\text{MC,E}}^k(r_1^*)$  with

$$r_1^* = \underset{r_1}{\operatorname{argmin}} \left| \left( \frac{N_{\text{MC,E}}^k(r_1)}{c} \bmod b \right) - r_1 \right|. \tag{11}$$

**Unequal centre sizes**

For the general case, we suggest using the average of the expected imbalance for each centre

$$E(\Delta_j^2 | r_j) \approx \frac{1}{b} \sum_{\ell=1}^b E(\Delta_1^2 | \ell) =: \overline{E(\Delta_1^2 | \cdot)}. \tag{12}$$

This basically assumes a univariate distribution of  $r_j$  on  $[1, \dots, b]$ .

**Upper boundary**

We can identify an upper boundary of the sample size, given that all parameters are specified correctly at planning stage. The maximal imbalance between treatments would occur, if each centre recruited and allocated  $r_j = b/(k + 1)$  or  $r_j = kb/(k + 1)$  subjects to a single treatment group resulting in  $\Delta_j^2 = m^*$ . Therefore, the most conservative sample size calculation will be performed with the following approximation

$$\sum_{j=1}^c E(\Delta_j^2 | r_j) \approx c \cdot E\left(\Delta_1^2 \middle| \frac{b}{k+1}\right). \tag{13}$$

**Example: Sample size calculation**

To give an example of the application of the proposed sample size formula, we demonstrate the effects of between-centre heterogeneity combined with incomplete block randomisation based on the COMPETE II trial.

The number of patients per study site is reported in [17]. Based on those numbers, we know the completeness of all randomisation blocks by centre as shown in Fig. 3.

We use these numbers to illustrate the influence of unbalanced treatment group allocation on sample size and statistical power based on the assumed model. In total, 40 incomplete randomisation blocks ( $r < 6$ ) out of 46 study sites occurred. Based on the assumed model, statistical power of the analysis might be reduced due to those 40 incomplete randomisation blocks, compared to a trial, where the same amount of patients would have been recruited at a single centre.

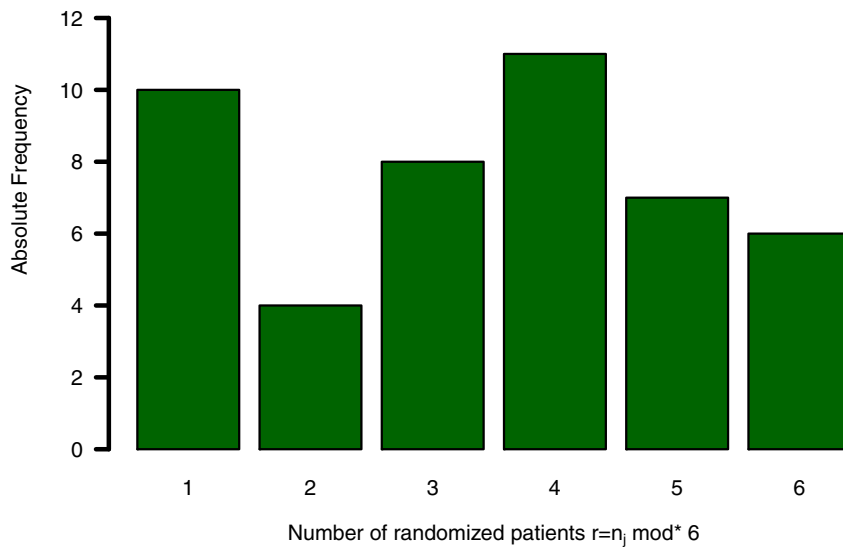
If we were to plan a new trial with similar features ( $\mu = 1, \sigma = 4$ , type 1 and type 2 error rates  $\alpha = 0.05$  and  $\beta = 0.2$ , respectively) we could plug these values into the sample size formula  $N_{\text{MC,U}}^1$  for unequal centre sizes. The assumption of uniformly distributed values of  $r_j$  on  $[1, \dots, b]$  could be underpinned by a  $\chi^2$ -test for goodness-of-fit ( $p = 0.4159$ ). The influence of intra-class correlation  $\rho \in [0, 0.5]$ , number of centres  $c \in \{23, 46, 92\}$  and block lengths  $b \in \{6, 8, 16\}$  on the sample size is shown in Fig. 4.

For a block length of  $b = 6$ , as chosen in the trial, no substantial influence of the intraclass correlation  $\rho$  on the overall sample size can be observed. The reported value of  $\rho = 0.08$  in the trial would not require a sample size adjustment compared to the standard approach ( $N_{\text{lower}}^k$ ). For larger block lengths, however, a strong increase of the estimated sample size can be seen, especially for  $\rho > 0.2$  and an increasing number of centres.

**Power simulations**

In addition to sample size calculations, we present some power simulation results for parameter settings based on the COMPETE II trial. We specify a treatment effect of  $\mu = 1$ , standard deviation  $\sigma = 4$  and intraclass-correlation coefficient  $\rho = 0.5$ . Data was generated for various block lengths, numbers of centres and subject-to-centre allocation schemes. Resulting sample sizes and associated statistical power are given in Table 2 and Fig. 5.

Analyses based on the lower boundary formula do not achieve the planned power of 0.8. The deviance from the nominal level increases with block length and number of



**Fig. 3** Example: Block randomisation. Number of final randomisation blocks by centre with block length  $b = 6$  based on the COMPETE II trial [17]

centres. The upper boundary formula results in power levels that exceed the nominal level slightly.

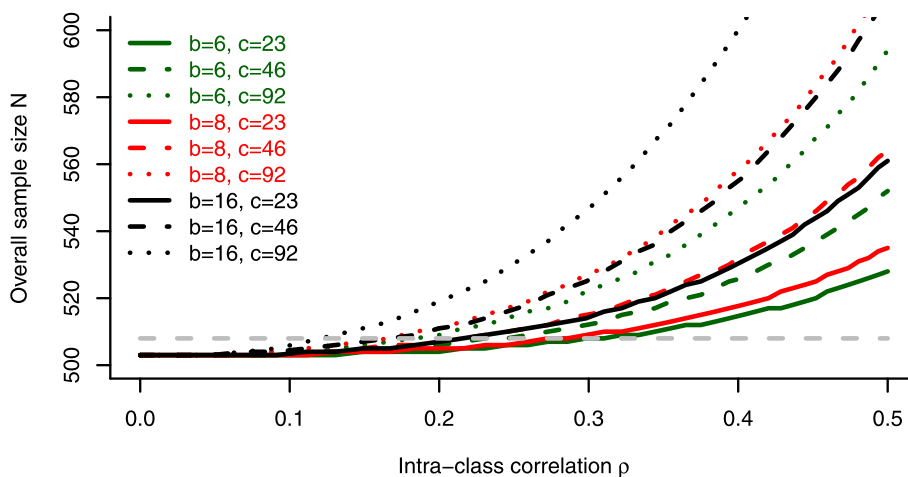
The new sample size formulas for equal and unequal centre sizes lead to reasonable power simulation results. Even if subject allocation is performed randomly in each simulation run (Unequal 2), adequate statistical power can be obtained.

Subject allocation might, by chance, result in more complete randomisation blocks for  $N_{MC,U}^1$  than  $N_{upper}^1$ . This leads to some situations where the formula with smaller sample size ( $N_{MC,U}^1$ ) achieves higher (estimated) power. This observation underlines the necessity to take block length and patient recruitment into consideration when planning large multi-centre trials.

**Discussion**

Patient enrolment and treatment allocation are key elements of every successful clinical trial. Randomisation techniques are used to achieve comparable treatment groups minimizing the risk of selection bias. Unfortunately, these randomisation procedures can result in unequal treatment group sizes and therefore a power loss compared to a balanced trial. Such imbalances cannot be determined prior to the trial, but we presented a way to estimate these values based on expectations.

The ICH E9 Guideline encourages the use of block randomisation and states the following on the choice of block sizes [1]: “Care should be taken to choose block lengths that are sufficiently short to limit possible imbalance, but that



**Fig. 4** Example: Sample size based on  $N_{MC,U}^1$ . Derived for  $\mu = 1, \sigma = 4$ , varying block length  $b$ , number of centres  $c$  and intra-class correlation  $\rho$ . Dashed grey line represents the planned sample size of the trial ( $N = 508$ )

**Table 2** Example: Comparison of sample size formulas

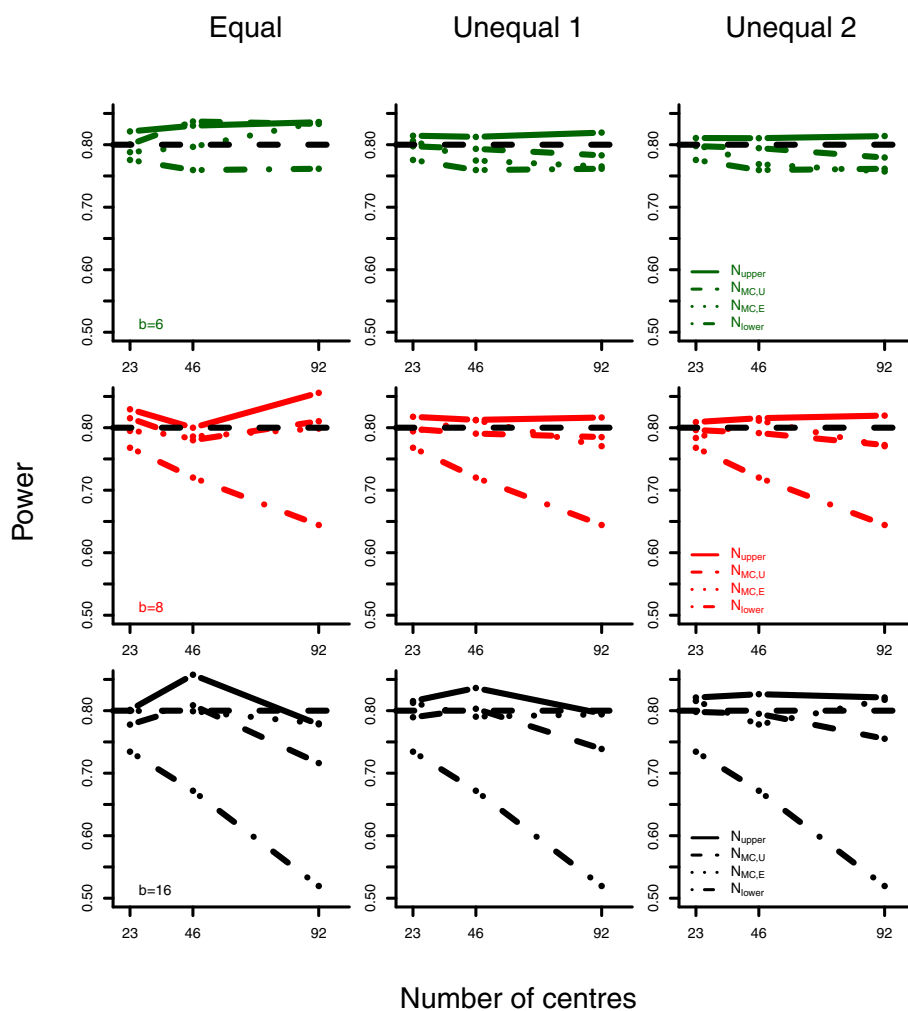
Block length	Formula	Number of centres		
		c=23	c=46	c=92
6	$N_{lower}^k$	503	503	503
	$N_{MC,E}^k$	525	524	569
	$N_{MC,U}^k$	528	552	594
	$N_{upper}^k$	541	575	634
8	$N_{lower}^k$	503	503	503
	$N_{MC,E}^k$	525	587	606
	$N_{MC,U}^k$	535	564	616
	$N_{upper}^k$	551	592	662
16	$N_{lower}^k$	503	503	503
	$N_{MC,E}^k$	586	603	762
	$N_{MC,U}^k$	561	610	692
	$N_{upper}^k$	587	654	762

Derived for  $\mu = 1, \sigma = 4, \rho = 0.5$ , varying block lengths and varying numbers of centres

are long enough to avoid predictability towards the end of the sequence in a block. Investigators and other relevant staff should generally be blind to the block length; the use of two or more block lengths, randomly selected for each block, can achieve the same purpose.”

The results shown remain valid when using variable block sizes, since incomplete blocks can occur using either method. Only the determination of expected values  $E(\Delta^2|r)$  is more complicated for variable blocks, because it depends on the range of block sizes used.

In a recent systematic review on prevalence and reporting of recruitment, randomisation and treatment errors in phase III randomized, controlled trials, stratified block randomisation was identified as randomisation technique of choice in 50% of 82 included studies published in *New England Journal of Medicine*, *Lancet*, *Journal of the American Medical Association*, *Annals of Internal*



**Fig. 5** Example: Power simulations. Simulated power based on the planned sample sizes for  $\mu = 1, \sigma = 4, \rho = 0.5$ , varying numbers of centres and varying block lengths. Dashed black line represents the targeted power of 0.8



Medicine, or British Medical Journal between January and March 2015 [18]. The median number of participants per trial was 650 (range 40–84,496) and the number of centres varied between 1 and 1,161 with a median of 24. There are a number of trials that used fixed block lengths greater than 10 to allocate subjects between two treatment groups [19, 20]. Trials using random block randomisation almost never report the underlying block sizes, therefore we can not compare fixed versus random blocks any further. Overall, these observations support our idea to account for incomplete blocks to plan a multi-centre trial using either method for randomisation.

One limitation of our approach is a lack of knowledge on centre heterogeneity at the planning stage. There have been various articles with estimates of intraclass-correlation coefficients (ICC) derived from cluster-randomized trials. These estimates can be used to get an initial guess for centre heterogeneity in multi-centre trials. A nice overview is given in the following text book [21]. Also, the implementation of an adaptive sample size reestimation procedure could account for this problem as it has been applied for cluster randomized trials and the fixed effects multi-centre trial design [22–24]. The development of sample size reestimation strategies based on the approach proposed here is subject to ongoing research. When planning an individually-randomized multi-centre trial there is a risk of control group contamination. This can partly be handled in placebo-controlled pharmacological trials or when proper blinding of patients and researchers is implemented [25]. Alternatively a cluster-randomized trial could be used to prevent contamination of treatment groups. This would, however, be associated with a higher sample size compared to the multi-centre design [26].

Here, we assumed a constant treatment effect across the centres. This is in line with the ICH E9 Guideline, which demands to avoid treatment-by-centre interactions in the primary analysis [1]. Therefore, this is at least for the planning of a trial an adequate assumption. Nevertheless, sensitivity analyses might explore treatment-by-centre interactions. Extending the sample size approach to a model including treatment-by-centre interactions is subject to future research.

## Conclusion

Imbalances in treatment allocation will lead to a power loss in multi-centre trials, if baseline heterogeneity is present. This risk can be accounted for when using appropriate methods for sample size calculation. To reduce uncertainty of sample size calculation, we recommend to calculate lower and upper boundaries in addition to the sample size.

## Additional files

**Additional file 1:** This file contains the calculation of the sample size formula. (PDF 24 kb)

**Additional file 2:** This file contains all source code used for simulations. (R 13 kb)

**Additional file 3:** This file contains the code used for Fig. 1. (R 3 kb)

**Additional file 4:** This file contains the code used for Fig. 2. (R 13 kb)

**Additional file 5:** This file contains the code used for Fig. 3. (R 1 kb)

**Additional file 6:** This file contains the code used for Fig. 4. (R 4 kb)

**Additional file 7:** This file contains the code used for Fig. 5. (R 12 kb)

**Additional file 8:** This file contains the code used for Table 2. (R 1 kb)

## Abbreviations

CI: Confidence interval; ICC: Intraclass-correlation coefficient

## Acknowledgements

The authors are grateful to Christian Röver for his helpful comments.

## Funding

This work was partly funded by the the German Federal Ministry of Education and Research project "Biostatistical Methods for efficient Evaluation of Individualised Therapies (BIMIT)", BMBF-05M13MGE.

## Availability of data and materials

All data generated or analysed during this study are included in this published article and its supplementary information files.

## Authors' contributions

MH and TF conceived the concept of this study, MH carried out the simulations and drafted the manuscript. TF critically reviewed and made substantial contributions to the manuscript. All authors commented on and approved the final manuscript.

## Ethics approval and consent to participate

Not applicable.

## Consent for publication

Not applicable.

## Competing interests

The authors declare that they have no competing interests.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 9 May 2018 Accepted: 1 November 2018

Published online: 29 November 2018

## References

1. Ich: Ich harmonized tripartite guideline e9: Statistical principles for clinical trials. *Stat Med.* 1999;18:1905–42.
2. Julious SA, Owen RJ. Sample size calculations for clinical studies allowing for uncertainty about the variance. *Pharm Stat.* 2006;5(1):29–37.
3. Gallo PP. Practical issues in linear models analyses in multicenter clinical trials. *Biopharm Rep.* 1998;6:2–9.
4. Ruvuna F. Unequal center sizes, sample size, and power in multicenter clinical trials. *Drug Inf J.* 2004;38(4):387–94.
5. van Breukelen GJ, Candel MJ, Berger MP. Relative efficiency of unequal versus equal cluster sizes in cluster randomized and multicentre trials. *Stat Med.* 2007;26(13):2589–603.
6. Fedorov V, Jones B. The design of multicentre trials. *Stat Methods Med Res.* 2005;14(3):205–48.
7. Vierron E, Giraudeau B. Sample size calculation for multicenter randomized trial: taking the center effect into account. *Contemp Clin Trials.* 2007;28(4):451–8.

8. Vierron E, Giraudeau B. Design effect in multicenter studies: gain or loss of power? *BMC Med Res Methodol*. 2009;9(1):39.
9. Pickering RM, Weatherall M. The analysis of continuous outcomes in multi-centre trials with small centre sizes. *Stat Med*. 2007;26(30):5445–56.
10. Kahan BC, Morris TP. Analysis of multicentre trials with continuous outcomes: when and how should we account for centre effects? *Stat Med*. 2013;32(7):1136–49.
11. Matts JP, Lachin JM. Properties of permuted-block randomization in clinical trials. *Control Clin Trials*. 1988;9(4):327–44.
12. Dumville J, Hahn S, Miles J, Torgerson D. The use of unequal randomisation ratios in clinical trials: a review. *Contemp Clin Trials*. 2006;27(1):1–12.
13. Snow G. Blockrand: Randomization for Block Random Clinical Trials. 2013. R package version 1.3. <https://CRAN.R-project.org/package=blockrand>.
14. Asghari-Jafarabadi M, Sadeghi-Bazargani H. Randomization: techniques and software-aided implementation in medical studies. *J Clin Res & Governance*. 2015;4(2).
15. R Development Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2008. R Foundation for Statistical Computing. ISBN 3-900051-07-0. <http://www.R-project.org>.
16. Holbrook A, Thabane L, Keshavjee K, Dolovich L, Bernstein B, Chan D, Troyan S, Foster G, Gerstein H, Investigators CI, et al. Individualized electronic decision support and reminders to improve diabetes care in the community: Compete ii randomized trial. *Can Med Assoc J*. 2009;181(1-2):37–44.
17. Chu R, Thabane L, Ma J, Holbrook A, Pullenayegum E, Devereaux PJ. Comparing methods to estimate treatment effects on a continuous outcome in multicentre randomized controlled trials: a simulation study. *BMC Med Res Methodol*. 2011;11(1):21.
18. Yelland LN, Kahan BC, Dent E, Lee KJ, Voysey M, Forbes AB, Cook JA. Prevalence and reporting of recruitment, randomisation and treatment errors in clinical trials: a systematic review. *Clin Trials*. 2018;15(3):278. <https://journals.sagepub.com/doi/10.1177/1740774518761627>.
19. Yaxley JW, Coughlin GD, Chambers SK, Occhipinti S, Samaratunga H, Zajdlewicz L, Dungleison N, Carter R, Williams S, Payton DJ, et al. Robot-assisted laparoscopic prostatectomy versus open radical retropubic prostatectomy: early outcomes from a randomised controlled phase 3 study. *The Lancet*. 2016;388(10049):1057–66.
20. Döring G, Meisner C, Stern M, Group FVTS, et al. A double-blind randomized placebo-controlled phase iii study of a pseudomonas aeruginosa flagella vaccine in cystic fibrosis patients. *Proc Natl Acad Sci*. 2007;104(26):11020–25.
21. Teerenstra S, Moerbeek M. *Power Analysis of Trials with Multilevel Data*. Boca Raton: Chapman and Hall/CRC; 2015.
22. Lake S, Kammann E, Klar N, Betensky R. Sample size re-estimation in cluster randomization trials. *Stat Med*. 2002;21(10):1337–50.
23. Jensen K, Kieser M. Blinded sample size recalculation in multicentre trials with normally distributed outcome. *Biom J*. 2010;52(3):377–99.
24. van Schie S, Moerbeek M. Re-estimating sample size in cluster randomised trials with active recruitment within clusters. *Stat Med*. 2014;33(19):3253–68.
25. Moerbeek M. Randomization of clusters versus randomization of persons within clusters: which is preferable? *Am Stat*. 2005;59(2):173–9.
26. Torgerson DJ. Contamination in trials: is cluster randomisation the answer? *Bmj*. 2001;322(7282):355–7.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

