Sample size recalculation in multicenter randomized controlled clinical trials based on noncomparative data

Markus Harden¹ | Tim Friede¹,²

¹Department of Medical Statistics, University Medical Centre Göttingen, Göttingen, Germany
²DZHK (German Center for Cardiovascular Research), Partner Site Göttingen, Göttingen, Germany

Correspondence
Markus Harden, Department of Medical Statistics, University Medical Centre Göttingen, Humboldallee 32, 37073 Göttingen, Germany. Email: markus.harden@med.uni-göttingen.de

This article has earned an open data badge “Reproducible Research” for making publicly available the code necessary to reproduce the reported results. The results reported in this article could fully be reproduced.

Abstract
Many late-phase clinical trials recruit subjects at multiple study sites. This introduces a hierarchical structure into the data that can result in a power-loss compared to a more homogeneous single-center trial. Building on a recently proposed approach to sample size determination, we suggest a sample size recalculation procedure for multicenter trials with continuous endpoints. The procedure estimates nuisance parameters at interim from noncomparative data and recalculates the sample size required based on these estimates. In contrast to other sample size calculation methods for multicenter trials, our approach assumes a mixed effects model and does not rely on balanced data within centers. It is therefore advantageous, especially for sample size recalculation at interim. We illustrate the proposed methodology by a study evaluating a diabetes management system. Monte Carlo simulations are carried out to evaluate operation characteristics of the sample size recalculation procedure using comparative as well as noncomparative data, assessing their dependence on parameters such as between-center heterogeneity, residual variance of observations, treatment effect size and number of centers. We compare two different estimators for between-center heterogeneity, an unadjusted and a bias-adjusted estimator, both based on quadratic forms. The type 1 error probability as well as statistical power are close to their nominal levels for all parameter combinations considered in our simulation study for the proposed unadjusted estimator, whereas the adjusted estimator exhibits some type 1 error rate inflation. Overall, the sample size recalculation procedure can be recommended to mitigate risks arising from misspecified nuisance parameters at the planning stage.

KEYWORDS
adaptive design, hierarchical model, internal pilot study, linear mixed model

1 INTRODUCTION

Multicenter randomized controlled clinical trials are a key element of modern evidence-based medicine and the number of publications of such trials is constantly increasing (Danielsen, Okholm, Pommergaard, Burchardt, & Rosenberg, 2014). The need to recruit subjects at more than one location often arises from the need to reach a sufficient sample size in limited time.
Also, conducting a multicenter trial with multiple sites offers some form of replication within the trial, which is advantageous compared to monocenter trials to “provide a better basis for the subsequent generalization of its findings,” see ICH (1998), Section 3.2. Compared to single-center trials, multicenter trials should account for potential heterogeneity between study sites. This heterogeneity can be considered as an additional nuisance parameter for planning and analysis purposes. The statistical model of such a trial can be implemented in a Gaussian linear fixed-effects or random-effects model, if a continuous outcome is considered. An advantage of a random-effects model is that only a single parameter has to be estimated compared to a fixed-effects model or a stratified analysis that will include at least one additional model parameter for each center. A fixed effects model should be used when only few centers are included, to avoid bias in estimation of between-center variation (Kahan & Morris, 2013).

Sample size formulas for multicenter trials have been described for both fixed and random-effects models. An overview of existing approaches and a new sample size formula is given in Harden and Friede (2018). In contrast to other approaches, this formula does not assume balanced numbers of observations across the treatment groups within centers. Since exact balance is unlikely to be observed in multicenter trials in practice, we believe that this is a useful approach to sizing multicenter trials.

Sample size calculations are typically informed by results from previous trials. Therefore, fixed study designs, for example, studies with an a priori fixed sample size are at risk to be planned with incorrect values when differences between previous trials and the new one occur. Adaptive study designs have been developed to reduce the risk of false negative study results, due to initial misspecification and have been adopted to various trial designs (Wassmer & Brannath, 2016, e.g., Chapters 2, 6, 9). One option is the implementation of an internal pilot study (Wittes & Brittain, 1990). This means that assumed values of nuisance parameters for the initial sample size calculation will be replaced by estimates from the accruing data during the course of the trial. They can be calculated in a blinded fashion using an estimator of the overall variance based on the data lumped across treatment groups, as proposed by Gould (1992), Kieser and Friede (2003), Zucker, Wittes, Schabenberger, and Brittain (1999), or based on unblinded data, for example, Wittes and Brittain (1990), Denne and Jennison (1999), Coffey and Muller (1999), Miller (2005). An overview to this kind of study design is given in Friede and Kieser (2006) and Proschan (2009). Since treatments cannot always be blinded in clinical trials but adaptations can still be based on the data pooled across the treatment groups, the terms “blinded” and “unblinded” are nowadays often replaced by “noncomparative” and “comparative,” respectively, to indicate that adaptive designs are not limited to blinded trials. In the following, we will call these procedures comparative or noncomparative.

There are several guidance documents on the use of adaptive designs and sample size recalculation in particular. “Whenever possible, methods for blinded sample size reassessment that properly control the type 1 error rate should be used” as stated in EMA (2007), Section 4.2.2. It is mentioned further that “sufficient justification should be made,” whenever unblinded data need to be reassessed. In a recent draft guidance on “Adaptive designs for clinical trials of drugs and biologics” by the U.S. Food and Drug Administration, it is said that “adequately prespecified adaptations based on noncomparative data have a negligible effect in the type 1 error probability. This makes them an attractive choice in many settings, particularly when uncertainty about event probabilities or endpoint variability is high,” see FDA (2018). Regarding comparative data, it is stated that “adaptations based on comparative data generally do directly increase the type 1 error probability and induce bias in treatment effect estimates. Therefore, statistical methods should take into account the adaptive trial design” (Section V, lines 414–416). The adaptation of sample sizes based on comparative data is suggested, when “there is considerable uncertainty about the true treatment effect size” (Section V.B). The use of adaptive designs is of course not limited to pharmacological interventions and specific guidance was released for example for device trials (FDA, 2016). Following these recommendations, we will present an approach for sample size recalculation based on noncomparative data to deal with uncertainties in nuisance parameters.

To the best of our knowledge, only few approaches for sample size recalculation in multicenter trials have been suggested so far. Shih and Long (1998) consider multicenter trials with unequal variances between treatment groups but equal group sizes within centers. Jensen and Kieser (2010) presented a sample size recalculation procedure based on a sample size formula by Ruvuna (2004) for the fixed effects model. In this article, we aim to apply the fixed sample size formula for the mixed effects model described in Harden and Friede (2018) to an internal pilot study approach to allow for sample size recalculation during the trial.

The manuscript is organized as follows. We introduce the COMPETE II trial as the motivating example in Section 2. Section 3 defines the statistical model and describes the sample size formula for the fixed study design. We extend the sample size formula to a sample size recalculation procedure in Section 4. For the sake of completeness, we will look at sample size recalculation based on comparative, as well as noncomparative data. In Section 5, we assess operation characteristics of the new approach based on a simulation study with parameters similar to the COMPETE II trial. We discuss the results and close with a conclusion in Section 6.
MOTIVATING EXAMPLE: THE COMPETE II TRIAL

Holbrook et al. (2009) investigated the benefit of an individualized electronic decision support system in adult patients diagnosed with type 2 diabetes. Primary outcome was a composite score difference compared to baseline measuring 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 behavior. The clinical targets are described in the original article. It was assessed twice, at baseline and 6 months after randomization.

For this multicenter trial, 511 patients from 46 primary care providers (here referred to as centers) were randomly assigned to intervention or control. Block-randomization was stratified by study site in blocks of six, following a 1:1 allocation scheme. The number of patients by study center are published in another article by Chu et al. (2011) and displayed in Figure 1. At the planning stage, the investigators aimed to recruit 508 patients to achieve 80% power assuming a difference of 1 for the primary outcome between treatment groups using a two-sided $t$-test with a significance level of $\alpha = 0.05$.

We will illustrate ideas and simulations using this example. Obvious parameters that can be reestimated from interim data are the within- and the between-group variance. Additionally, information on subject recruitment can be retrieved at the interim stage. We want to explore whether this information can be used to improve sample size recalculation.

STATISTICAL MODEL AND FIXED SAMPLE SIZE CALCULATION

The statistical model is based on a linear mixed-effects model described as follows

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

with fixed intercept $\mu_0$, treatment effect $\mu$, treatment indicator $x_1 = 0$ and $x_2 = 1$, and pairwise independent $u_j, \epsilon_{ijk}$ satisfying $E(u_j) = 0, \text{Var}(u_j) = \tau^2 < \infty, E(\epsilon_{ijk}) = 0, \text{Var}(\epsilon_{ijk}) = \sigma^2 < \infty \text{ and } \text{Cov}(u_j, \epsilon_{i'j'k}) = 0$. Indices refer to treatment groups $i = 1, 2$, centers $j, j' = 1, \ldots, c$ and subjects within centers and treatment groups $k = 1, \ldots, n_{ij}$, sample sizes are defined as $n_j = n_{1j} + n_{2j}, N_i = \sum_{j=1}^c n_{ij}$ and $N = N_1 + N_2$. Subjects within centers share a common random effect that creates a block-diagonal covariance matrix given by $\text{Cov}(Y_{111}, \ldots, Y_{2c_2c_2}) = \bigoplus_{j=1}^c (\sigma^2 I_{n_{1j}} + \tau^2 J_{n_{2j}})$ with the $n_{2j}$-dimensional identity matrix $I_{n_{2j}}$ and the $n_{2j}$-dimensional matrix $J_{n_{2j}}$ consisting of ones only. The structure of the study design is shown in Table 1.

In addition to the statistical model, the following assumptions are necessary to use the closed sample size formula as presented in Harden and Friede (2018):

1. Block randomization with fixed block length $b$, locally applied at each center for treatment allocation,
2. fixed $R : 1$ allocation ratio for each randomization block with $R \in \mathbb{N}$,
3. the proportion of overall sample sizes between treatment groups fulfills the assumed allocation ratio $N_1 = RN_2$.

The unknown parameter of interest $\mu$ can be estimated in an unbiased fashion based on overall treatment group means $\hat{\mu} = \bar{Y}_2 - \bar{Y}_1$, with $\bar{Y}_i = \sum_{j,k} Y_{ijk} / N_i$ for $i = 1, 2$. The variance of $\hat{\mu}$ has been calculated elsewhere, see Vierron and Giraudreau.
TABLE 1  Study design of a multicenter trial for a two treatment comparison

| Center | Treatment | Observations | Sample sizes by center |
|--------|-----------|--------------|-----------------------|
| 1      | 1         | $Y_{111}$   | $n_{11}$ = : $n_1$   |
|        | 2         | $Y_{121}$   | $n_{21}$ = : $n_2$   |
|        | $\vdots$ | $Y_{1j1}$   | $n_{1j}$ = : $n_j$   |
|        | $\vdots$ | $Y_{1c1}$   | $n_{1c}$ = : $n_c$   |
| $j$    | 1         | $Y_{211}$   | $n_{21}$ = : $n_2$   |
|        | 2         | $Y_{221}$   | $n_{22}$ = : $n_2$   |
|        | $\vdots$ | $Y_{2j1}$   | $n_{2j}$ = : $n_j$   |
|        | $\vdots$ | $Y_{2c1}$   | $n_{2c}$ = : $n_c$   |
| $c$    | 1         | $Y_{11}$    | $n_{11}$ = : $n_1$   |
|        | 2         | $Y_{12}$    | $n_{21}$ = : $n_2$   |
|        | $\vdots$ | $Y_{1n_11}$ | $n_{1n_1}$ = : $n_1$ |
|        | $\vdots$ | $Y_{1n_1j}$ | $n_{1n_1j}$ = : $n_j$|
|        | $\vdots$ | $Y_{1n_1c}$ | $n_{1n_1c}$ = : $n_c$|

(2009), Appendix I, and is given by

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

The unknown nuisance parameters $\sigma^2$ and $\tau^2$ can be estimated by

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

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

resulting in the following test statistic, which follows asymptotically a normal distribution

$$T = \frac{\hat{\mu}}{\sqrt{\hat{\sigma}^2 \frac{N}{N_1 N_2} + \hat{\tau}^2 \sum_{j=1}^{c} \left( \frac{n_{1j}}{N_1} - \frac{n_{2j}}{N_2} \right)^2}} \approx N(\delta, 1).$$

Under the null hypothesis of no treatment effect $H_0 : \mu = 0$, $T$ follows asymptotically a standard normal distribution, while $\delta = \sqrt{N \mu / \sigma}$ under the two-sided alternative $H_A : \mu \neq 0$. The null hypothesis will be rejected, if $|T| > q_{1-\alpha/2}$ for $\alpha \in (0, 1)$, where $q_{\gamma}$ denotes the $\gamma$-quantile of the standard normal distribution. We determined a sample size formula for $\hat{\mu}$ in a recent article (Harden & Friede, 2018) given by

$$N_{MC} = \left( \frac{q_{1-\alpha/2} + q_{1-\beta}}{\mu^*} \right) \frac{\sigma^2 (R+1)^2}{2 R} + \sqrt{\frac{\sigma^4 (R+1)^4}{4 R^2} + \left[ \frac{\tau (R+1) \mu^*}{2} \sum_{j=1}^{c} \Delta_j^2 \right] \left( q_{1-\alpha/2} + q_{1-\beta} \right)^2},$$

where $\mu^*$ describes the assumed treatment effect and $\Delta_j^2 := (n_{1j}/R - n_{2j})^2$ the deviation between planned and observed allocation ratio within centers. Nuisance parameters $\sigma^2$ and $\tau^2$ as well as deviations $\Delta_j^2$ are unknown at the planning stage and must be replaced by reasonable guesses prior to the trial. For instance, $\Delta_j^2$ could be replaced by expected values as shown in Harden and Friede (2018). If sample sizes within centers match the planned allocation ratio, that is, $\Delta_j^2 = 0$ for all $j$, between-center heterogeneity will not decrease the statistical power of the trial and (4) reduces to

$$N_{\text{lower}} = \left( \frac{q_{1-\alpha/2} + q_{1-\beta}}{\mu^*} \right) \frac{\sigma^2 (R+1)^2}{R}.$$  

An upper boundary of the required sample size is given, if every center would contribute an incomplete randomization block with $b/(R+1)$ subjects receiving the same treatment. Here, we assume that every center can at most contribute one incomplete block.
4 | SAMPLE SIZE RECALCULATION

We use Formula (4) to construct a sample size recalculation procedure for multicenter trials with continuous outcomes and arbitrary sample sizes per treatment arm and center. The goal of an internal pilot study is to gain information on nuisance parameters of the trial during recruitment to improve sample size calculations.

4.1 | What can be learned from noncomparative interim data?

In order to achieve the planned statistical power, we will analyze what can be learned based on interim data to improve nuisance parameter estimation. In the considered multicenter design, we assume values for the variability between observations $\sigma^2$ and the dependency of observations within study sites $\tau^2$. The influence of $\tau^2$ on the statistical power is influenced by sample size deviations within centers $\Delta_j^2$. These values require knowledge of the treatment group allocation, that is, unblinding of the data.

In a sample size review based on noncomparative data, we can, however, estimate the distribution of $\Delta_j^2$ at the interim stage and impute $E(\Delta_j^2)$ into Formula (4) for sample size recalculation. In this article, the number of centers $c$ is assumed to be fixed.

4.2 | Nuisance parameter estimation

We suggest the following noncomparative quadratic forms to estimate unknown nuisance parameters $\sigma^2$ and $\tau^2$ at interim

$$\hat{\sigma}^2 = \frac{1}{N-c} \sum_{i=1}^{N} \sum_{j=1}^{c} \sum_{k=1}^{n_{ij}} \left( Y_{ijk} - \bar{Y}_j \right)^2$$ and

$$\hat{\tau}^2 = \frac{1}{c-1} \sum_{j=1}^{c} \left( \bar{Y}_j - \bar{Y} \right)^2.$$ (5)

When a comparative estimation of nuisance parameters seems appropriate, estimators described in Formulas (1) and (2) can be applied. All these estimators do not assume any distribution function and can be calculated as long as $n_j > 2 \forall j$ $(n_{ij} > 2 \forall j)$ based on noncomparative (comparative) data. Some of these estimators are biased with the following expected values

$$E(\hat{\sigma}^2) = \sigma^2,$$

$$E(\hat{\sigma}^2_b) = \sigma^2 + \frac{\mu^2}{N-c} \sum_{j=1}^{c} \frac{n_{1j}n_{2j}}{n_j},$$

$$E(\hat{\tau}^2) = \tau^2 + \frac{\sigma^2}{2c} \sum_{i=1}^{c} \sum_{j=1}^{c} \frac{1}{n_{ij}}$$ and

$$E(\hat{\tau}^2_b) = \tau^2 + \frac{\sigma^2}{c} \sum_{j=1}^{c} \frac{1}{n_j} + \frac{\mu^2}{c-1} \sum_{j=1}^{c} \left( \frac{n_{2j}}{n_j} - \frac{1}{c} \sum_{i=1}^{c} \frac{n_{ij}}{n_i} \right)^2.$$ (6)

Detailed derivations are given in the Appendix. Friede and Kieser (2001) and others showed that the bias of the noncomparative variance estimator $\hat{\sigma}^2_b$ is neglectable in typical situations for clinical trials. The bias of estimators for $\tau^2$ can partly be adjusted using estimators

$$\hat{\tau}^2 = \text{max} \left\{ \hat{\tau}^2 - \frac{\hat{\sigma}^2}{2c} \sum_{i=1}^{c} \sum_{j=1}^{c} \frac{1}{n_{ij}} ; 0 \right\}$$ (7)

$$\hat{\tau}^2_b = \text{max} \left\{ \hat{\tau}^2_b - \frac{\hat{\sigma}^2_b}{c} \sum_{j=1}^{c} \frac{1}{n_j} ; 0 \right\}.$$ (8)
In the following, we will refer to estimators $\hat{\tau}^2$ and $\hat{\tau}_b^2$ as unadjusted estimators, while referring to $\tilde{\tau}^2$ and $\tilde{\tau}_b^2$ as adjusted estimators of $\tau^2$. We will assess how these biases affect the sample size determination process and if corrections shown in Formulas (7) and (8) improve results using simulation studies.

### 4.3 Incomplete randomization blocks within centers

In addition to nuisance parameters, the unknown dispersion by center $\Delta_j^2$ has to be determined during sample size review. Since the quantity is defined by sample sizes $n_{ij}$ which are unknown at interim, ideas similar to Harden and Friede (2018) can be used to calculate the expectation of the distribution of $\Delta_j^2$ which only depends on block length $b$, allocation ratio $R$ and the number of patients in the last randomization block $r_j$, all of which are available from noncomparative data. In order to investigate how well the distribution of $\Delta_j^2$ can be estimated at an interim stage, we use the COMPETE II trial-based center sizes to compare the distribution of block lengths at interim stages round\(\{\rho \cdot (n_1, \ldots, n_c)\}\), for $\rho \in \{0.2, 0.3, 0.4, 0.5\}$ to the ones displayed in Figure 1. The results are shown in Figure 2.

It can be seen that the distribution of block lengths at any interim stage does not look alike the final distribution after including all 511 subjects. We therefore recommend to assume a uniform distribution of $r_j$ on values $\{1, \ldots, b\}$ for sample size calculation in general, leading to the following approximation

$$\sum_{j=1}^c \Delta_j^2 \approx \frac{1}{b} \sum_{k=1}^b \mathbb{E}(\Delta_j^2 | r_j = k).$$

### 4.4 Recalculation procedure

All parameters in Formula (4) are either known or can be estimated as described above. For sample size recalculation noncomparative, as well as comparative nuisance parameter estimators as listed in Formulas (1), (2), (5), (6), (7), and (8) can be applied. The sample size recalculation procedure is executed as follows:

1. Calculate initial sample size $N_{\text{init}}$ using formula $N_{\text{MC}}$ based on initial assumptions on $\alpha$, $\beta$, $\mu^*$, $\sigma^2$, $\tau^2$, and $\Delta_j^2$ and specify a number of subjects $N_{BSSR} = \rho \cdot N_{\text{init}}$ which will be used to recalculate nuisance parameters with $\rho \in (0, 1)$.
2. Estimate nuisance parameters $\sigma^2$, $\tau^2$ based on the first $N_{BSSR}$ subjects recruited.
3. Calculate sample size $N_1$ using formula $N_{MC}$ based on initial $\alpha$, $\beta$, and $\mu$ and estimated $\sigma^2$, $\tau^2$.
4. Recruit additional subjects into the study until $N_{\text{final}} = \max(N_1; N_{BSSR})$ is reached. If an upper limit $N_{\text{max}}$ is given for recruitment, $N_{\text{final}} = \min\{\max(N_1; N_{BSSR}); N_{\text{max}}\}$.
5. Perform final analysis based on $N_{\text{final}}$ subjects.

## 5 SIMULATION STUDY

### 5.1 General settings

A simulation study is carried out to assess the operation characteristics of the sample size recalculation procedure. All analyses are performed using R, version 3.6.1 (R Development Core Team, 2008). Each simulation scenario consists of $n_{\text{sim}} = 10,000$ simulation runs to estimate the type 1 and 2 error rates. For data generation, the R-package blockrand is used for block randomization (Snow, 2013). A normal distribution is assumed for both the random effects $u_j$ and observation errors $\epsilon_{ijk}$. Source code to reproduce the results is available as Supporting Information online at https://onlinelibrary.wiley.com/doi/10.1002/bimj.201900138 at the end of the article.

The parameter settings of the generated data are motivated by the COMPETE II trial. Block length is chosen to be large to show the benefit of the new approach, which comes into play for unbalanced data. If not stated otherwise, we set the treatment effect to $\mu = 1$, nuisance parameters $\sigma^2 = \tau^2 = 16$, block length $b = 16$ and the allocation ratio $R = 1$. The number of centers $c$, timing of the sample size recalculation $\rho$ and the assumed treatment effect under the alternative for type 1 error simulations will vary. We generate equally sized study centers distributing a fixed overall sample size $N$ to $c$ centers based on a multinomial
**FIGURE 2** COMPETE II trial-based center sizes for different time points of interim analysis. We define $b \mod^* b = b$, since zero subjects refer to a previous block with $b$ subjects, assuming all centers are not empty.
Table 2  Overview of nuisance parameter estimates used for sample size recalculation

|                      | Noncomparative | Adjusted | Estimator for | \( \tau^2 \) |
|----------------------|----------------|----------|---------------|---------------|
| No                   | No             | No       | \( \hat{\sigma}^2 \) | \( \hat{\tau}^2 \) |
| Yes                  | No             | Yes      | \( \hat{\sigma}^2 \) | \( \hat{\tau}^2 \) |
| No                   | Yes            | No       | \( \hat{\sigma}^2 \) | \( \hat{\tau}^2 \) |
| Yes                  | Yes            | Yes      | \( \hat{\sigma}^2 \) | \( \hat{\tau}^2 \) |

Table 3  Sample sizes without sample size recalculation for nuisance parameters \( \sigma^2 = \tau^2 = 16 \) and varying treatment effects \( \mu^* \) and number of centers \( c \)

| Centers | Treatment effect \( \mu^* \) | 0.82 | 0.9 | 1 | 1.11 | 1.22 | 1.35 | 1.49 | 1.65 |
|---------|-------------------------------|------|-----|---|------|------|------|------|------|
| 1       | 750                           | 624  | 506 | 410| 340  | 278  | 230  | 188  |
| 10      | 775                           | 640  | 530 | 435| 364  | 302  | 252  | 210  |
| 20      | 800                           | 673  | 554 | 459| 387  | 324  | 274  | 230  |

| Centers | Treatment effect \( \mu^* \) (continued) | 1.82 | 2.01 | 2.23 | 2.46 | 2.72 | 3 | 3.32 |
|---------|---------------------------------------------|------|------|------|------|------|---|------|
| 1       | 154                                        | 128  | 104  | 86  | 70  | 58  | 48 |
| 10      | 177                                        | 149  | 125  | 106 | 90  | 77  | 66 |
| 20      | 196                                        | 167  | 142  | 122 | 105 | 91  | 79 |

The different nuisance parameter estimators given in Section 4.2 are compared and we will present the difference between sample size recalculation based on noncomparative and comparative data. We consider different estimators for nuisance parameters as listed in Table 2. Also, the test statistic described in (3) will be calculated using the adjusted as well as the unadjusted estimator for \( \tau^2 \). We will refer to the test statistic in accordance to the nuisance estimator of \( \tau^2 \).

5.2 Type 1 error rate

In this section, we present simulation results for the estimated type 1 error rate. We simulate data from \( c = 10 \) and \( c = 20 \) centers with varying sample sizes determined by \( \mu^* \in \{0.82, \ldots, 3.32\} \) and a varying proportion of data used for sample size recalculation \( \rho \in \{0.3, \ldots, 0.8\} \). Initial sample sizes before sample size recalculation for varying center sizes are shown in Table 3. For the purpose of comparison, we included sample sizes for a balanced t-test in that table (centers = 1). The results for a prespecified two-sided type 1 error rate of \( \alpha = 0.05 \) are shown in Figures 3 and 4. Each subfigure contains four panels that represent results for noncomparative and comparative, as well as adjusted and unadjusted estimators for the nuisance parameters \( \sigma^2 \) and \( \tau^2 \) described in (1) – (8), which were applied for sample size recalculation.

The unadjusted test statistic controls the type 1 error rate and shows some conservative behavior for larger treatment effects and an increasing number of centers. The use of noncomparative or comparative nuisance estimators does not seem to affect the estimated type 1 error rate in situations considered for this simulation study. The adjusted estimator of \( \tau^2 \) at interim does not affect the estimated type 1 error rate either.

When using the adjusted test statistic for the final analysis, some inflation of the estimated type 1 error rate can be observed for fewer centers and larger treatment effects. Again, no influence of noncomparative/comparative or adjusted/unadjusted nuisance parameters used for sample size recalculation can be seen for the simulated type 1 error. Due to the possible inflation of the type 1 error rate of the adjusted test statistic, we do not consider it for power analyses.
In this section, we explore whether the sample size recalculation procedure achieves the pre-specified statistical power for varying parameter settings. The simulation results refer to the parameter settings described earlier. The desired statistical power is set to 0.8. Simulation results are presented in Figure 5, which shows the behavior of the unadjusted test statistic for varying values of initially (mis)-specified nuisance parameters, varying number of centers and treatment effects.

The power simulation results show in general that an initial misspecification of nuisance parameters will be corrected by the sample size recalculation. For a larger treatment effects, we can see, however, that the adjusted estimator for $\tau^2$ can lead to underpowered trials. The unadjusted estimator for $\tau^2$ will at least lead to the planned power level. No difference between noncomparative and comparative estimators can be observed in the simulation settings considered here.

The distribution of nuisance parameters estimated at interim and final sample sizes is shown in Figures 6 and 7. Here we only presents results for one scenario, where both nuisance parameters are correctly assumed at the initial planning stage of the trial.

For a small treatment effect of $\mu = 1$ we observe only little differences regarding estimated nuisance parameters and resulting sample sizes, as seen Figure 5. For a larger treatment effect ($\mu = 2$) we can see, however, that resulting sample sizes are substantially increased for unadjusted nuisance parameter estimated or $\tau^2$. This increase seems to be influenced by the number of centers and overall number of subjects. Noncomparative variances estimated are slightly larger than comparative estimators, but this difference does not seem to affect sample size recalculation for the parameters considered here.

**Figure 3** Simulated type 1 error rate for the nonadjusted test statistic. Light-colored regions describe the 99% confidence band of the simulation error. Left (right)-panel figures show results for unadjusted (adjusted) estimators of $\tau^2$ at interim. Comparative (noncomparative) nuisance parameters for $\sigma^2$ and $\tau^2$ are used at interim in the top (bottom) panels, respectively.

### 5.3 Power

In this section, we explore whether the sample size recalculation procedure achieves the pre-specified statistical power for varying parameter settings. The simulation results refer to the parameter settings described earlier. The desired statistical power is set to 0.8. Simulation results are presented in Figure 5, which shows the behavior of the unadjusted test statistic for varying values of initially (mis)-specified nuisance parameters, varying number of centers and treatment effects.

The power simulation results show in general that an initial misspecification of nuisance parameters will be corrected by the sample size recalculation. For a larger treatment effects, we can see, however, that the adjusted estimator for $\tau^2$ can lead to underpowered trials. The unadjusted estimator for $\tau^2$ will at least lead to the planned power level. No difference between noncomparative and comparative estimators can be observed in the simulation settings considered here.

The distribution of nuisance parameters estimated at interim and final sample sizes is shown in Figures 6 and 7. Here we only presents results for one scenario, where both nuisance parameters are correctly assumed at the initial planning stage of the trial.

For a small treatment effect of $\mu = 1$ we observe only little differences regarding estimated nuisance parameters and resulting sample sizes, as seen Figure 5. For a larger treatment effect ($\mu = 2$) we can see, however, that resulting sample sizes are substantially increased for unadjusted nuisance parameter estimated or $\tau^2$. This increase seems to be influenced by the number of centers and overall number of subjects. Noncomparative variances estimated are slightly larger than comparative estimators, but this difference does not seem to affect sample size recalculation for the parameters considered here.
Assumed treatment effect under alternative
Proportion of Ninit used for sample size review

ρ
0.3
0.4
0.5
0.6
0.7
0.8

0.82 0.9 1 1.11 1.22 1.35 1.49 1.65 1.82 2.01 2.23 2.46 2.72 3 3.32

unadjusted, non−comparative
adjusted, non−comparative

unadjusted, comparative
adjusted, comparative

FIGURE 4 Simulated type 1 error rate for the adjusted test statistic. Light-colored regions describe the 99% confidence band of the simulation error. Left (right)-panel figures show results for unadjusted (adjusted) estimators of \( \tau^2 \) at interim. Comparative (noncomparative) nuisance parameters for \( \sigma^2 \) and \( \tau^2 \) are used at interim in the top (bottom) panels, respectively.

6 | DISCUSSION AND CONCLUSIONS

The specification of nuisance parameters is crucial for sample size calculation. Today, it is standard practice to use sample size recalculation procedures based on noncomparative data to correct inappropriate initial guesses of such parameters, generally without any practically relevant inflation of the type 1 error rate.

In this article, we presented a sample size recalculation procedure based on noncomparative data for a random-effects model of multicenter trials. The underlying sample size formula was described previously and depends, in addition to treatment effect, variance, and type 1 and 2 error rates, on the number of centers and heterogeneity between study sites (Harden & Friege, 2018). Whereas other approaches assume balanced data within centers, we relax this assumption here. However, we consider block randomization to approximate the imbalance.

Based in simulation results, we suggest to use the unadjusted nuisance parameter of \( \tau^2 \) for sample size recalculation as well as the final analysis to control the type 1 error rate and achieve the desired power. Power simulations confirm that a recalculation of the sample size based on noncomparative data is an adequate method to correct initial misspecification of nuisance parameters. This is especially helpful in multicenter trials, since the heterogeneity between centers is often unknown at the planning stage and since it is barely reported in publications of clinical trials. In terms of type 1 error rate and power, we cannot detect any differences between sample size recalculation based on noncomparative or comparative data. We think that this observation is mainly due to the rather large sample sizes typical for multicenter trials and therefore considered in the presented simulations.

An alternative variance estimator specifically for block-randomized trials has been suggested (Ganju & Xing, 2009; Xing & Ganju, 2005). This estimator is calculated for each randomization block, therefore blindness of the data can be preserved.
The estimator is unbiased, if all randomization blocks are complete. This estimation technique has also been extended to $\tau^2$ in cross-over as well as cluster randomized trials (Grayling, Mander, & Wason, 2018a, 2018b). Since this estimator relies on balanced data, we did not consider it here. Also, it has been demonstrated that the estimator’s variance is larger compared to the one-sample variance estimator in situations that are common in clinical trials (Friede & Kieser, 2013).

In our motivating example, the COMPETE II trial, no sample size recalculation method was applied. However, we do believe that multicenter trials such as the COMPETE II trial can benefit from sample size recalculation methods. In practice, their implementation is not too different from previously proposed approaches. Data of multicenter trials are usually stored in a central trial database and all calculations can be based on a data export at interim. An ideal time point for the sample size

**FIGURE 5** Simulated power for the nonadjusted test statistic. Light-colored regions describe the 99% confidence band of the simulation error. Left (right)-panel figures show results for unadjusted (adjusted) estimators of $\tau^2$ at interim. Comparative (noncomparative) nuisance parameters for $\sigma^2$ and $\tau^2$ are used at interim in the top (bottom) panels, respectively. Comparative is abbreviated as comp.
FIGURE 6  Distribution of reestimated nuisance parameters and sample sizes based on $\mu = 1$ and varying center sizes. Results based on a block length of $b = 16$ and nuisance parameters $\sigma^2 = \tau^2 = 16$ and a sample size recalculation after half of the initial sample size is recruited. Red line represents true nuisance parameters and sample size of 530 (554) subjects for $c = 10$ ($c = 20$) centers based on true values. Comparative is abbreviated as comp.

recalculation depends on several factors, such as pace of recruitment, trial duration, as well as uncertainty on initial choices of nuisance parameters (Chuang-Stein, Anderson, Gallo, & Collins, 2006). Given that not all patients enrolled at interim might have completed follow-up, approaches combining short- and long-term data have been proposed to increase precision of the estimated variance components, which might be transferred to multicenter trials, see, for example, Asendorf, Henderson, Schmidli, and
The integration of between-center heterogeneity into sample size considerations is a necessary step to control statistical power in multicenter trials. The use of a sample size recalculation procedure based on noncomparative data is a helpful tool to account for uncertainty in nuisance parameters at the planning stage of a trial.

Fig. 7  Distribution of reestimated nuisance parameters and sample sizes based on \( \mu = 2 \) and varying center sizes. Results based on a block length of \( b = 16 \) and nuisance parameters \( \sigma^2 = \tau^2 = 16 \) and a sample size recalculation after half of the initial sample size is recruited. Red line represents true nuisance parameters and sample size of 150 (168) subjects for \( c = 10 \) (\( c = 20 \)) centers based on true values. Comparative is abbreviated as comp.

Friede (2019) for an application to longitudinal counts and Friede and Kieser (2006) for an overview. For further reading on practical guidance, we refer to the summary by Pritchett et al. (2015).
CONFLICT OF INTEREST

The authors have declared no conflict of interest.

OPEN RESEARCH BADGES

This article has earned an Open Data badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. The data is available in the Supporting Information section.

This article has earned an open data badge “Reproducible Research” for making publicly available the code necessary to reproduce the reported results. The results reported in this article could fully be reproduced.

ORCID

Markus Harden https://orcid.org/0000-0003-2533-9396
Tim Friede https://orcid.org/0000-0001-5347-7441

REFERENCES

Asendorf, T., Henderson, R., Schmidli, H., & Friede, T. (2019). Sample size re-estimation for clinical trials with longitudinal negative binomial counts including time trends. Statistics in Medicine, 38(9), 1503–1528.

Chu, R., Thabane, L., Ma, J., Holbrook, A., Pullenayagum, E., & Devereaux, P. J. (2011). Comparing methods to estimate treatment effects on a continuous outcome in multicentre randomised controlled trials: A simulation study. BMC Medical Research Methodology, 11(1), 21.

Chuang-Stein, C., Anderson, K., Gallo, P., & Collins, S. (2006). Sample size reestimation: A review and recommendations. Drug Information Journal, 40(4), 475–484.

Coffey, C. S., & Muller, K. E. (1999). Exact test size and power of a Gaussian error linear model for an internal pilot study. Statistics in Medicine, 18(10), 1199–1214.

Danielsen, A. K., Okholm, C., Pomergaard, H.-C., Burchardt, J., & Rosenberg, J. (2014). Number of published randomized controlled multi center trials testing pharmacological interventions or devices is increasing in both medical and surgical specialties. PloS One, 9(7), e101383.

Denne, J. S., & Jennison, C. (1999). Estimating the sample size for at-test using an internal pilot. Statistics in Medicine, 18(13), 1575–1585.

EMA (2007). Reflection paper on methodological issues in confirmatory clinical trials planned with an adaptive design. London: EMEA.

FDA (2016). Adaptive designs for medical device clinical studies. Washington DC: Food and Drug Administration.

FDA (2018). Guidance for industry: Adaptive design clinical trials for drugs and biologics. Washington DC: Food and Drug Administration.

Friede, T., & Kieser, M. (2001). A comparison of methods for adaptive sample size adjustment. Statistics in Medicine, 20(24), 3861–3873.

Friede, T., & Kieser, M. (2006). Sample size recalculation in internal pilot study designs: A review. Biometrical Journal, 48(4), 537–555.

Friede, T., & Kieser, M. (2013). Blinded sample size re-estimation in superiority and noninferiority trials: Bias versus variance in variance estimation. Pharmaceutical Statistics, 12(3), 141–146.

Ganju, J., & Xing, B. (2009). Re-estimating the sample size of an on-going blinded trial based on the method of randomization block sums. Statistics in Medicine, 28(1), 24–38.

Gould, A. L. (1992). Interim analyses for monitoring clinical trials that do not materially affect the type I error rate. Statistics in Medicine, 11(1), 55–66.

Grayling, M. J., Mander, A. P., & Wason, J. M. (2018a). Blinded and unblinded sample size reestimation in crossover trials balanced for period. Biometrical Journal, 60(5), 917–933.

Grayling, M. J., Mander, A. P., & Wason, J. M. (2018b). Blinded and unblinded sample size reestimation procedures for stepped-wedge cluster randomized trials. Biometrical Journal, 60(5), 903–916.

Harden, M., & Friede, T. (2018). Sample size calculation in multi-centre clinical trials. BMC Medical Research Methodology, 18(1), 156.

Holbrook, A., Thabane, L., Keshavjee, K., Dolovich, L., Bernstein, B., Chan, D.... COMPETE II Investigators (2009). Individualized electronic decision support and reminders to improve diabetes care in the community: COMPETE II randomized trial. Canadian Medical Association Journal, 181(1-2), 37–44.

ICH (1998). ICH Topic E9 Statistical principles for clinical trials (Report No. CPMP/ICH/363/96). London: European Medicines Agency.

Jensen, K., & Kieser, M. (2010). Blinded sample size recalculation in multicentre trials with normally distributed outcome. Biometrical Journal, 52(3), 377–399.

Kahan, B. C., & Morris, T. P. (2013). Analysis of multicentre trials with continuous outcomes: When and how should we account for centre effects? Statistics in Medicine, 32(7), 1136–1149.

Kieser, M., & Friede, T. (2003). Simple procedures for blinded sample size adjustment that do not affect the type I error rate. Statistics in Medicine, 22(23), 3571–3581.

Miller, F. (2005). Variance estimation in clinical studies with interim sample size reestimation. Biometrics, 61(2), 355–361.

Pritchett, Y. L., Menon, S., Marchenko, O., Antonijevic, Z., Miller, E., Sanchez-Kam, M..., Prucka, W. R. (2015). Sample size re-estimation designs in confirmatory clinical trials–current state, statistical considerations, and practical guidance. Statistics in Biopharmaceutical Research, 7(4), 309–321.
SUPPORTING INFORMATION
Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Harden M, Friede T. Sample size recalculation in multicenter randomized controlled clinical trials based on noncomparative data. Biometrical Journal. 2020;62:1284–1299. https://doi.org/10.1002/bimj.201900138

APPENDIX: EXPECTED VALUES OF THE NUISANCE PARAMETER ESTIMATORS
Given the model in Section 3 and formulating every estimator as a quadratic form, the expectations of variance components given in (1), (2), (5), and (8) can be calculated as shown below.

Let \( Y_{nij} = (Y_{ij1}, \ldots, Y_{ijn_{ij}})' \) denote the vector of all observations in treatment group \( i \) and center \( j \) and \( P_{nij} = I_{nij} - n_{ij}^{-1}J_{nij} \) the \( n_{ij} \)-dimensional centering matrix with the \( n_{ij} \)-dimensional identity matrix \( I_{nij} \) and the \( n_{ij} \times n_{ij} \)-dimensional matrix \( J_{nij} \) consisting of ones only.

\[
E(\hat{\sigma}^2) = \frac{1}{2c} \sum_{i=1}^{2} \sum_{j=1}^{c} \frac{1}{n_{ij} - 1} E\left( \sum_{k=1}^{n_{ij}} (Y_{ijk} - \bar{Y}_{ij})^2 \right) \\
= \frac{1}{2c} \sum_{i=1}^{2} \sum_{j=1}^{c} \frac{1}{n_{ij} - 1} E\left( \hat{Y}_{nij}' P_{nij} \hat{Y}_{nij} \right) \\
= \frac{1}{2c} \sum_{i=1}^{2} \sum_{j=1}^{c} \frac{1}{n_{ij} - 1} \text{trace} \left( P_{nij} \cdot \left( \sigma^2 \cdot I_{nij} + \sigma_c^2 \cdot J_{nij} \right) \right) \\
= \sigma^2
\]

A similar calculation can be done for \( \hat{\tau}^2 \) when denoting \( \bar{Y}_i = (\bar{Y}_{i1}, \ldots, \bar{Y}_{ic})' \) the vector of group means for treatment group \( i \) at centers \( j = 1, \ldots, c \) and \( P_c \) the \( c \)-dimensional centering matrix.

\[
E(\hat{\tau}^2) = \frac{1}{2(c - 1)} \sum_{i=1}^{2} E\left( \sum_{j=1}^{c} (\bar{Y}_{ij} - \bar{Y}_i)^2 \right) \\
= \frac{1}{2(c - 1)} \sum_{i=1}^{2} E\left( \bar{Y}_i' P_c \bar{Y}_i \right)
\]
\[
= \frac{1}{2(c - 1)} \sum_{i=1}^{2} \text{trace } \left[ P_c \cdot \text{Diag} \left( \frac{\sigma^2}{n_{ij}} + \tau^2 \right) \right]_{j=1,...,c} \\
= \frac{1}{2(c - 1)} \sum_{i=1}^{2} \sum_{j=1}^{c} \frac{c - 1}{c} \left( \frac{\sigma^2}{n_{ij}} + \tau^2 \right) \\
= \tau^2 + \frac{\sigma^2}{2c} \sum_{i=1}^{c} \sum_{j=1}^{c} \frac{1}{n_{ij}}
\]

For the noncomparative variance estimator denote \( Y_{nj} = (Y'_{nj1}, Y'_{nj2}, ...)' \) the vector of observations at center \( j \), \( \mu_j = \text{E}(Y_{nj}) = (\mu_{0j}, \mu_{1j}, \mu_{2j})' \) the vector of expectations and \( P_{nj} \) the \( n_j \)-dimensional centering matrix.

\[
E(\hat{\sigma}_b^2) = \frac{1}{c} \sum_{j=1}^{c} \frac{1}{n_j - 1} E \left( \sum_{i=1}^{n_j} Y_{ijk} - \bar{Y}_{j.} \right)^2 \\
= \frac{1}{c} \sum_{j=1}^{c} \frac{1}{n_j - 1} E \left( Y'_{nj} P_{nj} Y_{nj} \right) \\
= \frac{1}{c} \sum_{j=1}^{c} \frac{1}{n_j - 1} \left( \text{trace } \left[ P_{nj} \cdot \text{Diag} \left( \sigma^2 \cdot I_{n_j} + \sigma^2 \cdot J_{n_j} \right) \right] + \mu_j' P_{nj} \mu_j \right) \\
= \sigma^2 + \frac{1}{c} \sum_{j=1}^{c} \frac{1}{n_j - 1} \cdot \mu_j' P_{nj} \mu_j \\
= \sigma^2 + \frac{\mu^2}{N - c} \sum_{j=1}^{c} \frac{n_{1j}n_{2j}}{n_j}
\]

And with a similar calculation

\[
E(\hat{\tau}_b^2) = \frac{1}{c - 1} E \left( \sum_{j=1}^{c} (\bar{Y}_{j.} - \bar{Y}_{..})^2 \right) \\
= \frac{1}{c - 1} E \left( \bar{Y}'_{.} P_{.} \bar{Y}'_{.} \right) \\
= \frac{1}{c - 1} \left( \text{trace } \left[ P_{.} \cdot \text{Diag} \left( \frac{\sigma^2}{n_j} + \frac{\sigma^2}{n} \right) \right]_{j=1,...,c} \right) + \mu'_{\cdot} P_{.} \mu_{\cdot} \\
= \frac{1}{c - 1} \left( \sum_{j=1}^{c} \frac{c - 1}{c} \left( \frac{\sigma^2}{n_j} + \tau^2 \right) + \mu'_{\cdot} P_{.} \mu_{\cdot} \right) \\
= \tau^2 + \frac{\sigma^2}{c} \sum_{j=1}^{c} \frac{1}{n_j} + \frac{1}{c - 1} \cdot \mu'_{\cdot} P_{.} \mu_{\cdot} \\
= \tau^2 + \frac{\sigma^2}{c} \sum_{j=1}^{c} \frac{1}{n_j} + \frac{\mu^2}{c - 1} \sum_{j=1}^{c} \left( \frac{n_{2j}}{n_j} - \frac{1}{c} \sum_{\ell=1}^{c} \frac{n_{2\ell}}{n_{\ell}} \right)^2
\]