RESEARCH ARTICLE

Blinded sample size reestimation for negative binomial regression with baseline adjustment

Antonia Zapf[1,2] | Thomas Asendorf[1] | Christoph Anten[1] | Tobias Mütze[1,3] | Tim Friede[1]

[1]Department of Medical Statistics, University Medical Center Göttingen, Göttingen, Germany

[2]Department of Medical Biometry and Epidemiology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany

[3]Statistical Methodology, Novartis Pharma AG, Basel, Switzerland

Correspondence
Tim Friede, Department of Medical Statistics, Humboldtallee 32, 37073 Göttingen, Germany.
Email: tim.friede@med.uni-goettingen.de

In randomized clinical trials, it is standard to include baseline variables in the primary analysis as covariates, as it is recommended by international guidelines. For the study design to be consistent with the analysis, these variables should also be taken into account when calculating the sample size to appropriately power the trial. Because assumptions made in the sample size calculation are always subject to some degree of uncertainty, a blinded sample size reestimation (BSSR) is recommended to adjust the sample size when necessary. In this article, we introduce a BSSR approach for count data outcomes with baseline covariates. Count outcomes are common in clinical trials and examples include the number of exacerbations in asthma and chronic obstructive pulmonary disease, relapses, and scan lesions in multiple sclerosis and seizures in epilepsy. The introduced methods are based on Wald and likelihood ratio test statistics. The approaches are illustrated by a clinical trial in epilepsy. The BSSR procedures proposed are compared in a Monte Carlo simulation study and shown to yield power values close to the target while not inflating the type I error rate.

KEYWORDS
adaptive design, count data, covariates, generalized linear model, likelihood ratio, prognostic factors, sample size reestimation

1 | INTRODUCTION

Randomized clinical trials provide the highest evidence to estimate a treatment effect and are therefore considered the gold standard study design.[1] Even if randomization guarantees comparable treatment arms on average across repeated trials, in individual trials, in particular if they are small, treatment groups might differ with regard to their covariate distributions. If a covariate is predictive, in the sense that it moderates the treatment effect and its distribution differs between the treatment arms, then comparisons between the treatment arms yield incorrect results unless the analyses are appropriately adjusted for the covariate.[2] Therefore, it is important to include predictive baseline covariates in the primary analysis. Furthermore, this leads to a reduced between-subject variation and, thereby, to an increase in statistical power.

This is obvious for normally distributed outcomes, but also holds for nonnormally distributed outcomes such as binary endpoints.[3] Therefore, the sample size planning and the analysis should match to obtain a sample size which maintains a target power.

Depending on the scale of the outcome variable, different sample size estimation methods exist. For normally distributed outcomes, Frison and Pocock[4] proposed a sample size formula for the analysis of covariance (ANCOVA). For studies with dichotomous outcomes, Hernandez et al[3] state that an adjustment for a predictive variable in the analysis may lead to an increase in statistical power. In practice, sample size planning and analysis frequently mismatch as Lyles et al[5] realize. One reason might be that it is often difficult to obtain reliable estimates of the correlations between the baseline variables and the outcome variable. In these situations, designs with internal pilot study (IPS)[6] provide a framework to deal with uncertainties in nuisance parameters such as correlations between baseline and outcome variables. For instance, Friede and Kieser studied such designs considering ANCOVA models with tests for superiority[7] and noninferiority.[8] They proposed blinded procedures as they have favorable characteristics from a regulatory point of view as compared to unblinded procedures.[9,10]

Many clinical trials have endpoints modeled as count data, for example, the number of malignant melanoma lesions or number of seizures in epilepsy patients. Recurrent event data, for example, the number of multiple sclerosis (MS) relapses during follow-up, can also be considered as count data. For the analysis of count data, in general, a Poisson regression or, in case of overdispersion, a negative binomial regression can be used to consider covariates. Under the assumption of constant event rates, these regression models can also be applied to analyze data with varying follow-up times by including the logarithm of the follow-up times as off set in the regression model.

Several authors proposed sample size formulas for trials with count outcomes. While some assume specific distributions such as negative binomial distribution,[11,12] others consider more general situations.[13-15] In addition, some approaches consider covariates, with Signorini[16] being one of the early proposals. Lyles et al[5] review approaches for the period up to 2005 and propose a very flexible approach for the power calculation in generalized linear models (GLM). Since then, further approaches for the sample size planning of specific designs were proposed, for example, for clustered data,[17] variable follow-up times,[18] noninferiority trials,[19] and overdispersed data.[13] However, the general approach from Lyles et al[5] can also be applied to these designs.

Cook et al,[20] Schneider et al,[21,22] as well as Friede and Schmidli[23,24] proposed adaptive designs for nuisance parameter based sample size reestimation with count outcomes, but did not consider covariates. Therefore, the aim of this article is to propose blinded sample size reestimation (BSSR) procedures for clinical trials with count data considering covariates and to systematically investigate their operational characteristics. The proposed methods are based on Wald or likelihood ratio (LR) test statistics and use the expected Fisher information (FI) or the Lyles approach for the sample size estimation.

In the next section, the clinical evaluation of antiepileptic drugs is discussed and the example dataset is described. In Section 3, the statistical model, the applied test statistics, and the sample size calculation approaches are presented, whereas in Section 4, a procedure for sample size reestimation is proposed. The statistical properties of the approaches regarding type I error rate and statistical power are investigated in Section 5, and the approaches are applied to the example dataset in Section 6. We close with a brief discussion of the results in Section 7.

2 | CLINICAL EVALUATION OF ANTIEPILEPTIC DRUGS

The Food and Drug Administration (FDA) and European Medicines Agency (EMA) generally define the efficacy of anti-epileptic drugs in terms of the number of seizures. The EMA recommends the reduction in the frequency of seizures as a primary endpoint for exploratory studies.[25] In contrast, for therapeutic confirmatory studies, the responder rate is recommended as a primary endpoint, whereas a response is defined as a seizure reduction of 50%.[25] For late phase I and phase II, as well as for phase III studies, the FDA recommends the percentage change from baseline in seizure frequency as primary endpoint. This may lead to studies with different primary endpoints for different regions (see, eg, Rosenfeld et al[26]). A crossover study design is mentioned in both guidelines as one possible design (in the EMA guideline only for exploratory studies). However, in the large randomized studies published over the past few years, a parallel group study design was used predominantly.[27-30] Regarding the analysis of seizure counts, the EMA guideline explains that "the distribution of seizure frequencies are usually heavily skewed." Therefore, in phase III studies, the outcome is usually transformed using a rank- or a log-transformation. Most randomized studies analyzed the seizure frequency as percent change from baseline using an ANCOVA with the baseline seizure frequency as covariate.[26,31,32] However, in the mentioned studies, the sample size calculation ignored the baseline covariate effect and therefore did not match
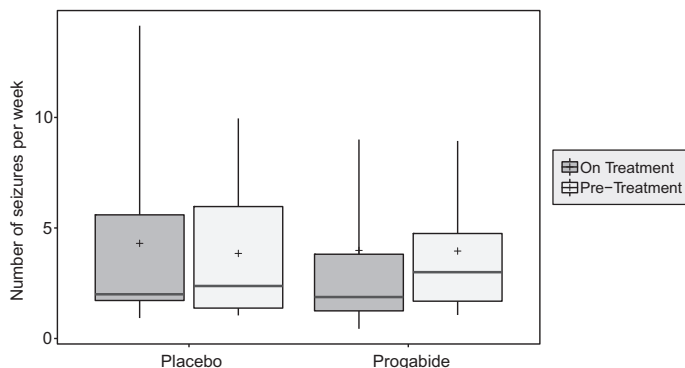
FIGURE 1 Distributions of the number of seizures 8 weeks pretreatment and 8 weeks on treatment in the two treatment groups observed in the study by Leppik et al[34] (boxplot indicates mean +; median line; box: 25%-75% quantile; whisker: 5%-95% quantile)

the analysis approach. We found no phase III trial with a sample size reestimation, although it often would have been helpful, as the assumptions for the sample size calculation differed considerably from the study results in several cases, implying potentially underpowered or overpowered studies.[27,28,32] For example, in Reference 27, for the sample size calculation, the mean log-transformed 28-day seizure rate was estimated to be 2.59 for placebo, 2.32 for pregabalin (PGB) as intervention with dosage of 150 mg, and 2.01 for PGB 300 mg. By contrast, in the end, the adjusted mean values estimated using ANCOVA were 1.96, 1.95, and 1.82 for placebo, PGB 150, and PGB 300, respectively. For confirmatory studies, the EMA and FDA guidelines recommend at least 12 weeks as maintenance period, which is also standard in large-scale confirmatory trials.[28,29,32,33] For the baseline period, no specific duration is recommended, but mostly a 4-week interval is used.[27-29,32,33] The sample size in recently published phase III trials ranged from 100 to 300 patients per group.[27,28,30,32] The median weekly seizure rates in these trials ranged from 1.75 to 3.6. Because only the median number of seizures in the baseline period was reported in the considered articles, it was not possible to calculate the underlying overdispersion.

To motivate our work, we consider the double-blind randomized study for the comparison of proabide and placebo from Leppik et al[34] as an example. This study has already served as an example in other methodological articles.[35,36] The sample size of 59 epileptic patients was small compared to the phase III trials mentioned above. The primary endpoint in the study was the weekly seizure rate in an 8-week interval on treatment, and as a covariate the seizure count in the preceding 8-week interval was used. The median number of seizures per week decreased from 2.3 to 2 in the placebo arm and from 3 to 1.9 in the proabide arm. These seizure rates are comparable to those of the above mentioned trials and follow a skewed distribution, as shown in Figure 1. This example is revisited and analyzed by negative binomial regression in Section 6.

3 | STATISTICAL MODEL, HYPOTHESIS TESTING, AND SAMPLE SIZE PLANNING

Prior to presenting methods for BSSR, the focus of this article, we recapitulate the negative binomial regression model, discuss two group comparisons in the negative binomial regression model, and present sample size planning for the discussed hypothesis tests.

3.1 | Statistical model

Let the random variable $Y_i \in \mathbb{N}_0$ denote the number of events of subject $i = 1, \ldots, n$ after some subject specific follow-up time $t_i \in [0, T]$ and let $\mathbf{X}_i \in \mathbb{R}^p$ be the corresponding p-dimensional vector of baseline covariates and group affiliation, in the following referred to as explanatory variables. We model the number of events Y_i by a negative binomial regression model.[37] In detail, the number of events Y_i has the probability mass function

$$\mathbb{P}_{\text{NB}}(y_i, \lambda_i, \eta) := \mathbb{P}(Y_i = y_i | \mathbf{X}_i = \boldsymbol{x}_i) = \frac{\Gamma(\eta + y_i)}{\Gamma(\eta) y_i!} \left(\frac{\lambda_i}{\eta + \lambda_i} \right)^{y_i} \left(\frac{\eta}{\eta + \lambda_i} \right)^{\eta},$$

where λ_i is the expected number of events conditioned on the explanatory variable, that is,

$$\lambda_i = \mathbb{E}\left[Y_i | \mathbf{X}_i = \mathbf{x}_i\right] = t_i \exp\left(\mathbf{x}_i^\top \boldsymbol{\beta}\right), \quad i = 1, \ldots, n.$$

The conditional variance is given by

$$\mathrm{Var}\left[Y_i | \mathbf{X}_i = \mathbf{x}_i\right] = \lambda_i \left(1 + \frac{\lambda_i}{\eta}\right), \quad i = 1, \ldots, n.$$

We refer to η as the shape parameter. For increasing η, the variance converges to λ_i, the variance of a Poisson distribution. In this manuscript, we focus on the case in which the subjects are allocated to two treatment arms: the experimental treatment group and the control group. The treatment effect can be modeled through the explanatory variable $\mathbf{X}_i$ by setting its first entry to one, that is, $X_{i1} = 1$, and its second entry to an indicator variable which is zero if subject i is in the control group, that is, $X_{i2} = 0$, or one if subject i is in the experimental treatment group, that is, $X_{i2} = 1$. Thus, the first entry β_1 of the parameter vector $\boldsymbol{\beta}$ corresponds to the intercept and the second entry $\theta := \beta_2$ is the treatment effect in terms of the log rate ratio. We assume that smaller rates of λ_i are better and, therefore, the smaller θ, the more efficient is the experimental treatment compared to the control. The statistical hypothesis testing problem of interest is given by

$$H_0 : \theta = \delta \quad \text{versus} \quad H_1 : \theta \neq \delta,$$

where $\delta \in R$, commonly $\delta = 0$ is chosen. In the following subsection, we present several statistical tests for the hypothesis H_0.

3.2 | Hypothesis testing

In this subsection, we discuss two common likelihood-based tests, a Wald and the LR test, for the negative binomial regression model. Both tests are conditional tests, that is, even though we modeled the explanatory variables $\mathbf{X}_i$ ($i = 1, \ldots, n$) as random, the hypothesis tests are defined conditioned on the explanatory variables.

3.2.1 | Wald test

A Wald statistic is defined as the ratio of an estimator for the parameter of interest and its standard error. Here, we focus on estimating the log rate ratio θ through the maximum likelihood approach. Let $\hat{\boldsymbol{\beta}}$ be the maximum likelihood estimator of the parameter vector $\boldsymbol{\beta}$, $\hat{\eta}$ be the maximum likelihood estimator of the shape parameter η, and let $\mathbf{I}(\boldsymbol{\beta}, \eta)$ be the corresponding FI matrix. The conditional FI matrix $\mathbf{I}(\boldsymbol{\beta}, \eta | \mathbf{X})$, where $\mathbf{X}$ is a placeholder for $\mathbf{X}_1, \ldots, \mathbf{X}_n$, can be split into diagonal parts relating to $\boldsymbol{\beta}$ and η and zeroes otherwise, that is,

$$\mathbf{I}(\boldsymbol{\beta}, \eta | \mathbf{X}) = \begin{pmatrix} \mathbf{I}_\beta(\boldsymbol{\beta}, \eta | \mathbf{X}) & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_\eta(\boldsymbol{\beta}, \eta | \mathbf{X}) \end{pmatrix}.$$

According to Lawless,[37] the maximum likelihood estimator $\hat{\boldsymbol{\beta}}$ asymptotically follows a multivariate normal distribution, that is,

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) \underset{n \to \infty}{\overset{D}{\sim}} \mathcal{N}\left(\mathbf{0}, \mathbf{I}_\beta(\boldsymbol{\beta}, \eta | \mathbf{X})^{-1}\right).$$

Since the treatment effect θ is the second entry of the parameter vector $\boldsymbol{\beta}$, the maximum likelihood estimator $\hat{\theta}$ is the second entry of the maximum likelihood estimator $\hat{\boldsymbol{\beta}}$. The maximum likelihood estimator $\hat{\theta}$ asymptotically follows a normal distribution,

$$\sqrt{n}\left(\hat{\theta} - \theta\right) \underset{n \to \infty}{\overset{D}{\sim}} \mathcal{N}\left(\mathbf{0}, \mathbf{c}^\top \mathbf{I}_\beta(\boldsymbol{\beta}, \eta | \mathbf{X})^{-1} \mathbf{c}\right),$$

where **c** is a vector that has a one as the second entry and zeros otherwise. With these preliminaries, we define the Wald statistic as

$$T_W = n \frac{\left(\hat{\theta} - \delta\right)^2}{\mathbf{c}^\top \mathbf{I}_\beta(\hat{\beta}, \hat{\eta}|\mathbf{X})^{-1}\mathbf{c}}. \tag{1}$$

The maximum likelihood estimators $\hat{\beta}$ and $\hat{\eta}$ are consistent and with Slutsky's theorem, it follows that $\mathbf{c}^\top \mathbf{I}_\beta(\hat{\beta}, \hat{\eta}|\mathbf{X})^{-1}\mathbf{c}$ is a consistent estimator of $\mathbf{c}^\top \mathbf{I}_\beta(\beta, \eta|\mathbf{X})^{-1}\mathbf{c}$. Furthermore, under the null hypothesis H_0, the Wald statistic T_W is asymptotically χ^2 distributed with one degree of freedom and an asymptotic level α test of H_0 is obtained by rejecting H_0 when the Wald statistic T_W is larger than the $(1 - \alpha)$-quantile $\chi^2_{1,1-\alpha}$ of a χ^2 distribution with one degree of freedom. It is worthwhile noting that the asymptotic distribution of the maximum likelihood estimators and the Wald statistic hold only conditioned on the explanatory variables.

We defined the test statistic such that it is asymptotically χ^2 distributed. However, the test statistic can also be defined such that it is asymptotically standard normally distributed. Since a squared standard normal distribution corresponds to a χ^2 distribution, the two definitions are equivalent for testing H_0. Furthermore, defining the test statistic such that it is standard normally distributed would also allow for one-sided testing of H_0, whereas the tests based on the χ^2 distribution are inherently two-sided.

3.2.2 | LR test

Let $\Omega = \mathbb{R}^p \times \mathbb{R}_{>0}$ be the parameter space associated with the parameter vector (β, η). The parameter space $\Omega_0 \subset \Omega$ of the null hypothesis H_0 is equal to $\mathbb{R} \times \{\delta\} \times \mathbb{R}^{p-2} \times \mathbb{R}_{>0}$, that is, it is obtained by imposing the null hypothesis $\theta = \delta$ to the parameter space. Then, the LR statistic is defined by

$$T_{\mathrm{LR}} = -2 \log \left(\frac{\sup_{\Omega_0} \mathcal{L}(\beta, \eta|Y_1, \ldots, Y_n, \mathbf{X}_1, \ldots, \mathbf{X}_n)}{\sup_\Omega \mathcal{L}(\beta, \eta|Y_1, \ldots, Y_n, \mathbf{X}_1, \ldots, \mathbf{X}_n)} \right),$$

where $\mathcal{L}(\cdot, \cdot|Y_1, \ldots, Y_n, \mathbf{X}_1, \ldots, \mathbf{X}_n)$ denotes the likelihood function of the negative binomial regression model.[37] Under the null hypothesis H_0 and conditional on the explanatory variables, the LR statistic T_{LR} is asymptotically χ^2 distributed with one degree of freedom.[38] Thus, an asymptotic level α test of the null hypothesis H_0 is obtained by rejecting the null hypothesis of the LR statistic T_{LR} that is larger than the $(1 - \alpha)$-quantile $\chi^2_{1,1-\alpha}$.

3.3 | Power and sample size calculation

An integral step in the planning of a clinical trial is the determination of the sample size. When covariates are to be incorporated in the analysis, it is important to consider them in the sample size determination, even though they are not known prior to the trial, since they can affect the statistical power of the trial. There are two common approaches to define the power and the respective sample size to account for covariates in sample size planning, a *conditional* and an *unconditional* approach. The conditional power is the power of the statistical test for a given set of the explanatory variable and the unconditional power is the expected value of the conditional power over the distribution of the explanatory variable.[39] In practice, the unconditional power is of greater interest than the conditional power. However, versions of the conditional power can be used to approximate the unconditional power as we will see below.

For both the conditional and unconditional approaches, we need to approximate the distribution of the test statistic under the alternative. Therefore, we start with approximating the distribution of the Wald statistic T_W and the LR statistic T_{LR} for a parameter $\theta_1 \neq \delta$, conditional on the covariates. Let β_{H_1} be the parameter vector with θ_1 as a second entry. Under the alternative hypothesis, the distribution of the Wald statistic T_W can be approximated by a noncentral χ^2 distribution with one degree of freedom and noncentrality parameter

$$\lambda_W = \frac{n(\delta - \theta_1)^2}{\mathbf{c}^\top \mathbf{I}_\beta(\beta_{H_1}, \eta)^{-1}\mathbf{c}}.$$

The distribution of the LR statistic T_{LR} under the alternative hypothesis can be approximated by a noncentral χ^2 distribution with one degree of freedom and noncentrality parameter

$$\lambda_{\text{LR}} = -2 \left(\log \mathcal{L}_0 \left(\boldsymbol{\beta}_{H_1}, \eta | Y_1, \ldots, Y_n, \mathbf{X}_1, \ldots, \mathbf{X}_n \right) - \log \mathcal{L} \left(\boldsymbol{\beta}_{H_1}, \eta | Y_1, \ldots, Y_n, \mathbf{X}_1, \ldots, \mathbf{X}_n \right) \right),$$

where $\mathcal{L}_0$ denotes the likelihood function restricted to the null hypothesis.

In the following, we present two approaches for approximating the unconditional power based on which the sample size can be calculated using iterative methods.

3.3.1 | Expanded dataset

Lyles et al[5] proposed to approximate the unconditional power by a conditional power for a given set of explanatory variables $x_1, \ldots, x_n$ on the basis of an expanded dataset (ED). In the following, we briefly recapitulate this approach for the negative binomial regression model. We assume the parameters $\boldsymbol{\beta}$ and η to have certain values under the planning alternative and denote these as $\boldsymbol{\beta}_{H_1}$ and η_{H_1}. An ED is created with J entries for each fictitious explanatory variable x_i ($i = 1, \ldots, n$) where for each value $j = 1, \ldots, J$ the weights

$$w_{ij} = \mathbb{P} \left(Y_i = y_j | \mathbf{X}_i = x_i \right)$$

are calculated assuming $\boldsymbol{\beta}_{H_1}$ and η_{H_1} are true. The parameter J is chosen sufficiently large such that the weights sum up to approximately one for each $i = 1, \ldots, n$:

$$\sum_{j=1}^{J} w_{ij} = \sum_{j=1}^{J} \mathbb{P} \left(Y_i = y_j | \mathbf{X}_i = x_i \right) \approx 1. \tag{2}$$

For example, one might select J such that the above sum is no smaller than a fixed threshold τ, for example, $\tau = 0.999$. The ED consists of $n \times J$ tuples (y_j, x_i, w_{ij}) with $i = 1, \ldots, n$ and $j = 1, \ldots, J$.

After setting up the ED, the weighted log-likelihood

$$\sum_{i=1}^{n} \sum_{j=1}^{J} w_{ij} \log \left(\mathbb{P} \left(Y_i = y_j | \mathbf{X}_i = x_i \right) \right)$$

is maximized with respect to $\boldsymbol{\beta}$ and η. We denote the resulting maximum likelihood estimators by $\hat{\boldsymbol{\beta}}_{\text{ED}}, \hat{\eta}_{\text{ED}}$, and the effect size, equal to the second entry of $\hat{\boldsymbol{\beta}}_{\text{ED}}$, by $\hat{\theta}_{\text{ED}}$. As stated in section 2.2 in Lyles et al,[5] the maximum likelihood estimators are equal to the assumed parameters $\boldsymbol{\beta}_{H_1}$ and η_{H_1} and, more importantly, the inverse observed FI $\mathbf{I}_{\beta}(\hat{\boldsymbol{\beta}}_{\text{ED}}, \hat{\eta}_{\text{ED}} | \mathbf{X})^{-1}$, calculated from the ED, can be used as an estimator for the matrix $\text{Var}(\hat{\boldsymbol{\beta}}_{\text{ED}})$. Specifically of interest for us is the variance of the effect size, that is, $\text{Var}\left(\hat{\theta}_{\text{ED}}\right)$, which is estimated by taking the corresponding entry from the inverse observed FI, that is, $\widehat{\text{Var}}\left(\hat{\theta}_{\text{ED}}\right) = \mathbf{c}^{\top} \mathbf{I}_{\beta}(\hat{\boldsymbol{\beta}}_{\text{ED}}, \hat{\eta}_{\text{ED}} | \mathbf{X})^{-1} \mathbf{c}$. Plugging in the estimate $\widehat{\text{Var}}\left(\hat{\theta}_{\text{ED}}\right)$ for its true value, the approximated noncentrality parameter $\lambda_{\text{W,ED}}$ for the distribution of the Wald statistic under the alternative is given by:

$$\hat{\lambda}_{\text{W,ED}} = \frac{n(\delta - \theta_1)^2}{\widehat{\text{Var}}(\hat{\theta}_{\text{ED}})},$$

and the approximated statistical power of the Wald test (WT) is given by

$$\mathbb{P} \left(\chi_1^2(\hat{\lambda}_{\text{W,ED}}) \geq \chi_{1,1-\alpha}^2 \right), \tag{3}$$

where $\chi_1^2(\hat{\lambda}_{\text{W,ED}})$ denotes a random variable following a noncentral χ^2 distribution with one degree of freedom and noncentrality parameter $\hat{\lambda}_{\text{W,ED}}$. Since the ED and, consequently, the corresponding approximate statistical power depend on the sample size n used for creating the ED, the sample size required to obtain a prespecified target power for the WT must be determined iteratively, for example, by using the bisection method. When determining the sample size required for

the LR test, one would proceed similarly by approximating the noncentrality parameter λ_{LR} through the -2 weighted log-likelihood values from the unrestricted fit and the fit restricted to the null hypothesis.

The performance of this procedure relies on sufficiently large J to be obtained.[5] However, in some situations, J can become very large to reach the threshold τ, resulting in a large ED and high computation times. To reduce required computational resources, we replace w_{ij} by $\tilde{w}_{ij} = \mathbb{P}\left(Y_i \in \{y_{j_1}, \ldots, y_{j_{c(i,j)}}\}|\mathbf{X}_i = \mathbf{x}_i\right)$, where an individual value y_j is replaced by a set of $c(i,j)$ values. For instance, $c(i,j)$ is chosen such that $\tilde{w}_{ij}$ is larger than a set value, for example, 0.001. In the ED, we replace y_j by $\tilde{y}_j$, where $\tilde{y}_j$ is a representative of the set $\{y_{j_1}, \ldots, y_{j_{c(i,j)}}\}$, for example, the median or the value associated with the largest w_{ij}. We refer to this procedure as the binning procedure. For a numerical implementation of this algorithm, we refer to the R-code submitted as supplementary material.

3.3.2 | Expected information

For a given set of explanatory variables $\mathbf{x}_1, \ldots, \mathbf{x}_n$, the sample size of the WT and the LR test in the negative binomial regression model can be expressed in dependency of the FI matrix $\mathbf{I}_\beta(\boldsymbol{\beta}, \eta|\mathbf{X})$. Fisher's information matrix $\mathbf{I}_\beta(\boldsymbol{\beta}, \eta|\mathbf{X})$ is defined conditional on the explanatory variable, which we will emphasize by referring to it as the conditional information. In the following, we provide the sample size formula for the negative binomial regression based on the conditional information. Then, we illustrate how the unconditional power can be attained by numerical approximation.

For conditional power, assuming given parameters $\boldsymbol{\beta}_{H_1}$ and η_{H_1} in the alternative H_1 and a given set of explanatory variables $\mathbf{x}_1, \ldots, \mathbf{x}_n$, under the normal approximation of the Wald statistic T_W from Equation (1), the total sample size of the WT in the negative binomial regression model to yield a power of $1 - \beta$ can be approximated by

$$n = \frac{(q_{1-\alpha/2} + q_{1-\beta})^2}{(\theta_1 - \delta)^2} \mathbf{c}^\top \mathbf{I}_\beta(\boldsymbol{\beta}_{H_1}, \eta_{H_1}|\mathbf{X})^{-1} \mathbf{c}.$$

The explanatory variables $\mathbf{x}_1, \ldots, \mathbf{x}_n$, which include baseline covariates, are not available when planning a clinical trial, and therefore, it is not possible to calculate $\mathbf{I}_\beta(\boldsymbol{\beta}_{H_1}, \eta_{H_1}|\mathbf{X})^{-1}$ at the planning stage of the clinical trial. Instead, we focus on calculating a sample size n for which the unconditional power holds. Using the normal approximation of T_W, it follows that the unconditional power is equal to

$$\int \mathbb{P}\left(T_W \geq q_{1-\alpha/2}|\mathbf{X}\right) d\mathbb{P}^{(\mathbf{X}_1, \ldots, \mathbf{X}_n)} = \int \Phi\left(-q_{1-\alpha/2} + \sqrt{n} \cdot \frac{\theta_1 - \delta}{\sqrt{\mathbf{c}^\top \mathbf{I}_\beta(\boldsymbol{\beta}_{H_1}, \eta_{H_1}|\mathbf{X})^{-1}\mathbf{c}}}\right) d\mathbb{P}^{(\mathbf{X}_1, \ldots, \mathbf{X}_n)}.$$

As only the FI depends on $\mathbf{X}$, we suggest to approximate the integral by replacing $\mathbf{I}_\beta(\boldsymbol{\beta}_{H_1}, \eta_{H_1}|\mathbf{X})$ by its expectation, that is,

$$\tilde{\mathbf{I}}_\beta(\boldsymbol{\beta}_{H_1}, \eta_{H_1}) = \mathbb{E}_\mathbf{X}\left[\mathbf{I}_\beta(\boldsymbol{\beta}_{H_1}, \eta_{H_1}|\mathbf{X})\right] = \int \mathbf{I}_\beta(\boldsymbol{\beta}_{H_1}, \eta_{H_1}|\mathbf{X}) d\mathbb{P}^{(\mathbf{X}_1, \ldots, \mathbf{X}_n)}.$$

Thus, the approximate sample size required to yield an unconditional power of $1 - \beta$ with the WT in the negative binomial regression model is given by

$$\tilde{n} = \frac{(q_{1-\alpha/2} + q_{1-\beta})^2}{(\theta_1 - \delta)^2} \mathbf{c}' \tilde{\mathbf{I}}_\beta(\boldsymbol{\beta}_{H_1}, \eta_{H_1})^{-1} \mathbf{c}. \tag{4}$$

4 | PROPOSED PROCEDURE FOR BSSR

To test the null hypothesis $H_0 : \theta = \delta$, we introduced two tests, namely, a WT (Section 3.2.1) and an LR test (Section 3.2.2). For approximating the unconditional power of these tests under a given alternative, we also introduced two approaches, using the ED (Section 3.3.1) and using the expected information (Section 3.3.2). One possibility, described in Section 3.3.1, is given by creating an ED. This approach requires prior information on nuisance parameters $\boldsymbol{\beta}$ and η, as well as the

support of the covariates, to create an ED, from which the power of the WT or LR test can be calculated for a given sample size. Calculating the sample size for a specific power is then possible by iterating the sample size until the target power is met. We will refer to these sample size calculation methods for the WT and LR test, using the ED, as WT-ED and LR-ED, respectively. The second possibility for calculating the power or sample size, as described in Section 3.3.2, is given by calculating the expected FI matrix. Based on assumption on $\boldsymbol{\beta}$ and η, as well as the distribution of the covariates, the expected FI can be calculated numerically and plugged into the sample size formula (1). This approach is applicable to the described WT. We will refer to this sample size calculation method as WT-FI. The aim of this section is to demonstrate how these three methods for sample size calculation can be applied to construct BSSR procedures.

Assume we have carefully considered a suitable time point for the sample size review. At this time point, the gathered data are still blinded to maintain trial integrity. We therefore observe outcomes $y_1, \ldots, y_m$ with corresponding explanatory variables $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m$. As the trial is still running, follow-up times $t_1, \ldots, t_m$ are likely to vary as recruitment is usually gradual over time. It is likely that some patients have not completed follow-up, and therefore, only partial follow-up information is available for these patients. However, these incomplete observations should be included in the sample size review, as the resulting sample size is expected to be less variable.[21,40]

All three sample size estimation methods, WT-ED, LR-ED, and WT-FI, are based on assumptions on the nuisance parameters $\boldsymbol{\beta}$ and η. Therefore, for a BSSR of the sample size, we estimate these nuisance parameters from the blinded data. Further relevant information required for the estimation of nuisance parameters is the intended length of the study T and the allocation ratio $r = n_E/n_C$, where n_E and n_C denote the initial sample sizes of the experimental treatment and control group, respectively. As the group allocation is unknown, an observation from the blinded dataset follows a mixture distribution of two negative binomial distributions. Assuming that the true treatment effect θ_1 holds, a blinded estimator of $\boldsymbol{\beta}(\theta_1) = (\beta_1, \theta_1, \beta_3, \ldots, \beta_p)^\top$ (regression coefficients with fixed effect size θ_1) and η is given by maximizing the likelihood of a mixture distribution of two negative binomial distributions, with known weight r, which is given by

$$
\sum_{i=1}^{m} \log \left(\frac{r}{r+1} \cdot \mathbb{P}NB \left(y_i, \exp \left((x_{i1}, 1, x_{i3}, \ldots, x_{ip}) \boldsymbol{\beta}(\theta_1) \right) \cdot \frac{t_i}{T}, \eta \frac{t_i}{T} \right) \right. \tag{5}
$$
$$
\left. + \frac{1}{r+1} \cdot \mathbb{P}NB \left(y_i, \exp \left((x_{i1}, 0, x_{i3}, \ldots, x_{ip}) \boldsymbol{\beta}(\theta_1) \right) \cdot \frac{t_i}{T}, \eta \frac{t_i}{T} \right) \right).
$$

By maximizing the likelihood (0), blinded estimates $\hat{\boldsymbol{\beta}}_{\text{blind}}$ and $\hat{\eta}_{\text{blind}}$ for the nuisance parameters $\boldsymbol{\beta}(\theta_1)$ and η are obtained. These are then used to recalculate the sample size by plugging in $\hat{\boldsymbol{\beta}}_{\text{blind}}$ for $\boldsymbol{\beta}$ and $\hat{\eta}_{\text{blind}}$ for η in the corresponding sample size calculation method. For the methods WT-ED and LR-ED, the required sample size to yield the target power, given the blinded estimates, is calculated iteratively. Using the bisection method, for example, the following steps are necessary to obtain the reestimated sample size.

1. Choose an initial maximum total sample size $n_{\max}$ and minimal total sample size $n_{\min}$.
2. Set $n = (n_{\max} + n_{\min})/2$ and create an ED plugging in $\hat{\boldsymbol{\beta}}_{\text{blind}}$ and $\hat{\eta}_{\text{blind}}$ for $\boldsymbol{\beta}$ and η to calculate weights ω_{ij}.
3. From the ED, attain maximum likelihood estimators $\hat{\boldsymbol{\beta}}_{\text{ED}}$, $\hat{\eta}_{\text{ED}}$, $\hat{\theta}_{\text{ED}}$, and $\widehat{\text{Var}(\hat{\theta}_{\text{ED}})}$ to calculate the noncentrality parameters $\lambda_{\text{W,ED}}$ or λ_{LR}, depending on the desired test.
4. Calculate the power of the WT or LR test for n subjects, using the noncentrality parameter $\lambda_{\text{W,ED}}$ for the WT or λ_{LR} for the log-ratio test.
5. If the calculated power is higher than the desired power, set $n_{\max} := n$, else set $n_{\min} := n$.
6. Iterate steps 2-5 until the calculated power is approximately equal to the desired power.

For the method WT-FI, the calculation of the reestimated sample size is computationally less expensive and can be summarized in the following steps:

1. Calculate the expected FI $\tilde{I}_{\boldsymbol{\beta}}(\boldsymbol{\beta}, \eta)$ plugging in $\hat{\boldsymbol{\beta}}_{\text{blind}}$ and $\hat{\eta}_{\text{blind}}$ for $\boldsymbol{\beta}$ and η.
2. Plug in the calculated expected FI $\tilde{I}_{\boldsymbol{\beta}}(\boldsymbol{\beta}_{\text{blind}}, \eta_{\text{blind}})$ for the true FI $I_{\boldsymbol{\beta}}(\boldsymbol{\beta}_{H_1}, \eta_{H_1})$ into the sample size formula (4) to attain n.

Having followed these steps, the total sample size in the study can now be adjusted to be n, where the group specific sample sizes are $n_E = \frac{n \cdot r}{1+r}$ for the treatment group and $n_C = \frac{n}{1+r}$ for the control group. When calculating the expected FI, the covariates distribution is also required. However, because the covariates distribution is assumed to be equal over

Parameters		Assumed value	True values
Intercept log rate	β_1	0	$-1, -0.8, \ldots, 1$
Log rate ratio	β_2	-0.2	0
Covariate rate	β_3	2.5	$0, 0.5, \ldots, 5$
Shape parameter	η	3	$2, 2.2, \ldots, 4$

TABLE 1 Simulation settings for the comparison of type I error rates between the fixed design and the blinded sample size reestimation procedures

all treatment groups, a maximum likelihood estimation of the parameters of the covariates distribution on the blinded data is possible and sufficient. The reestimated sample size should maintain the power even when initial assumptions on nuisance parameters prove to be wrong. This is shown in the following section.

5 | OPERATING CHARACTERISTICS OF THE BSSR PROCEDURES

In this section, we investigate the influence of a BSSR on type I error rate, as a regulatory requirement of adaptive designs is the control of the type I error rate, and study the power to see if the BSSR methods can mitigate overpowering and underpowering when the nuisance parameters are misspecified. The binning procedure, described in Section 3.3.1, is only applied in selected settings, that is, Appendix A.1.

5.1 | Type I error rate

Statistical analyses performed within a running trial, which may result in changes to the trial design, can impact the type I error rate. To investigate whether the BSSR procedures influence the type I error rate, we conducted a simulation study including all three reestimation methods introduced in Section 4, under a variety of simulation settings displayed in Table 1. All simulations are conducted with one normally distributed covariate, which is standardized for the BSSR using sample mean and standard deviation (SD). Group sizes are assumed to be equal and total observation time is one, that is, $r = 1$ and $T = 1$.

For calculating the sample size of the fixed design, the assumed values from Table 1 were considered. The required sample size for rejecting the null hypothesis $H_0 : \theta = 0$ under the alternative $\theta_1 = \beta_2 = -0.2$ with a power of 80% at level $\alpha = 5\%$ for each of the three methods is $n_{\text{WT-FI}} = 379$, $n_{\text{WT-ED}} = 381$, and $n_{\text{LR-ED}} = 383$, per group.

Data were then simulated with values differing from the ones assumed for the sample size planning. First, covariates of the first half of the planned sample were generated from a standard normal distribution. Then, 25% of the initial sample sizes observations were generated as complete observations, before a further 25% of the initial sample sizes observations were generated with random follow-up times t_i, uniformly distributed on $[0, T]$, to simulate varying follow-up times. The generated data were then taken to calculate $\hat{n}_{\text{WT-FI}}$, $\hat{n}_{\text{WT-ED}}$, and $\hat{n}_{\text{LR-ED}}$. Having calculated the reestimated sample size, the 25% data with incomplete follow-up times were completed and further observations were generated accordingly.

Inference was then performed on $n_{\text{WT-FI}}$, $n_{\text{WT-ED}}$, and $n_{\text{LR-ED}}$ observations per group for the fixed design and $\hat{n}_{\text{WT-FI}}$, $\hat{n}_{\text{WT-ED}}$, and $\hat{n}_{\text{LR-ED}}$ observations per group for the designs with BSSR. The results are visualized in Figure 2.

The mean and median type I error rate in the three methods, summarized over all settings depicted in Figure 2, were 0.046 and 0.050, respectively. SD of the type I error rate was 0.015 in all three methods. These summary results were equal compared between the fixed design and the BSSR procedure, up to the third decimal. Therefore, we conclude that the BSSR does not inflate the type I error rate and can be used without the need of a type I error rate adjustment.

5.2 | Power

To demonstrate the benefits of a BSSR, we conduct a simulation study that compares the power of the fixed design and the power of a sample size adjustment, under the settings displayed in Table 1, with the only difference being that the true log rate ratio is equal to the assumed log rate ratio. The observed covariate is assumed to be from a normal distribution and standardized prior to the statistical analysis. Assumptions on the duration of the trial, sample size allocation, and

FIGURE 2 Type I error rates of the three blinded sample size reestimation procedures plotted against the type I error rate of the corresponding fixed design. Every combination of true values from Table 1 corresponds to one simulation setting resulting in a total of 175 distinct settings. For each setting, 10 000 trials were simulated
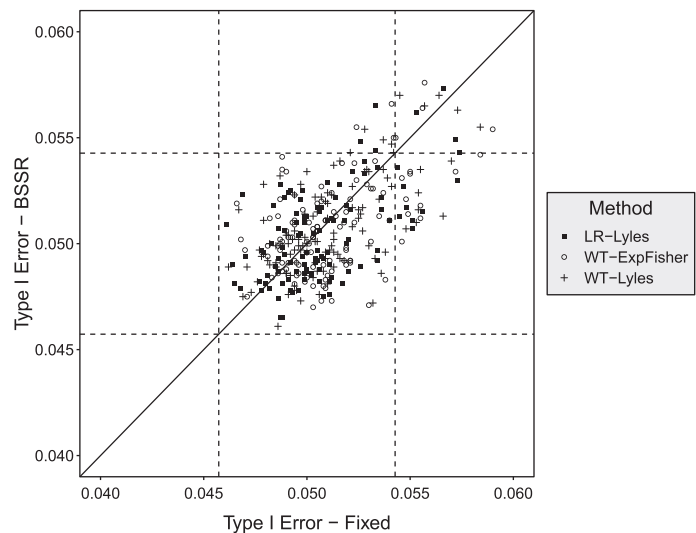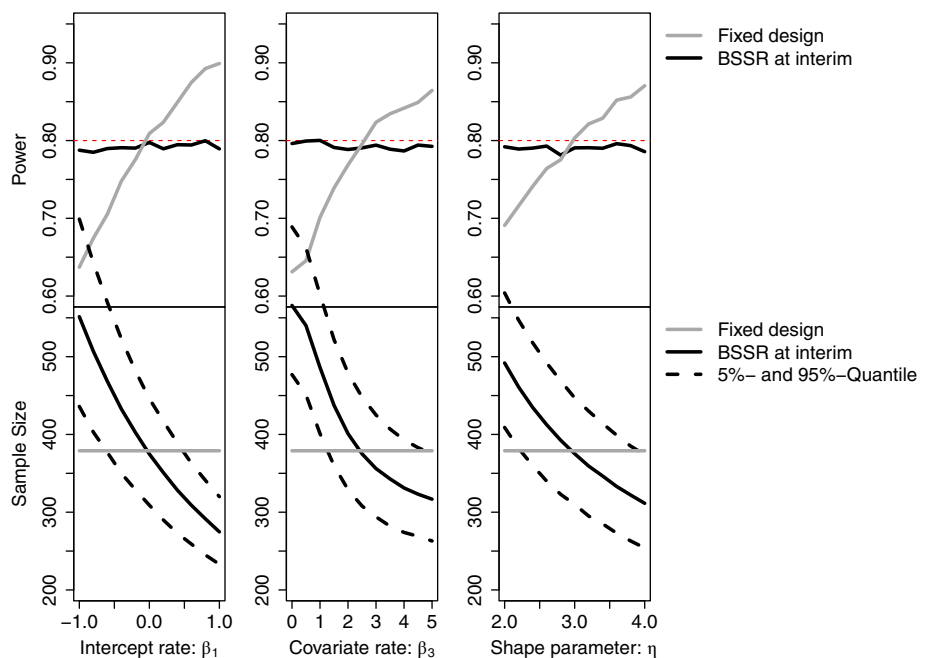

FIGURE 3 Power comparison of the fixed design and blinded sample size reestimation for the method WT-FI. Nuisance parameter altered sequentially, one at a time. Mean sample size reestimates with 5%- and 95%-quantiles are additionally displayed. Tabular results are given in Appendix A.2 [Colour figure can be viewed at wileyonlinelibrary.com]

covariate distribution are the same as in Section 5.1. Therefore, the sample sizes for the fixed design are also equal, that is, $n_{\text{WT-FI}} = 379$, $n_{\text{WT-ED}} = 381$, and $n_{\text{LR-ED}} = 383$ per group. The data of the simulation study are generated analogously as described in Section 5.1. The results of the comparison of the fixed design with BSSR procedures are shown in Figure 3.

Figure 3 compares the sample size and implied power of the fixed design and BSSR for the WT-FI procedure. Apart from a small set of simulation settings, the results from WT-ED and LR-ED are very similar and therefore displayed in Figures A1 and A2 within the Appendix. We can conclude that BSSR corrects the sample size when the true parameters differ from the assumed and the reestimated sample size maintains 80% power, while the sample size from the fixed design does not achieve the desired power when true parameters are different from the assumed. In the regarded settings, these changes are visible for misspecifications of the intercept rate β_1, covariate rate β_3, and shape parameter η. However, the impact of these parameters is specific to the setting and may be different for other parameter combinations. In addition to changes of the nuisance parameters, the effect size may also be misspecified in the BSSR. While the presented BSSR procedure does not correct the sample size due to a misspecified effect size, a possible misspecification did not result in any relevant influence on the sample size. Simulation results are reported in Appendix A.3.

TABLE 2 Results of the negative binomial regression for the example study from Leppik et al[34]

Parameter		Without covariate (AIC=538.2)			With covariates (AIC=472.3)		
		Estimate	Standard error	*p*-value	Estimate	Standard error	*p*-value
Intercept	β_1	3.5389	0.1820	< 0.001	3.2323	0.1280	< 0.001
Treatment	β_2	−0.0772	0.2512	0.759	−0.2677	0.1483	0.071
Baseline seizures	β_3	—			0.7668	0.0755	<0.001
Age category	β_4	—			0.1374	0.1496	0.358
Shape parameter	η	1.114	0.193		3.691	0.798	

TABLE 3 Nuisance parameters from blinded estimation

Parameter	Blinded estimates
Intercept	$\hat{\beta}_1 = 3.1849$
Baseline seizures	$\hat{\beta}_3 = 0.7500, \hat{\mu} = 0, \hat{\sigma} = 1$
Age Category	$\hat{\beta}_4 = 0.1590, \hat{p} = 0.4406$
Shape parameter	$\hat{\eta} = 3.6455$

Notes: Note that the relevant treatment effect β_2 is not estimated from the data, but defined by the user.

6 | EXAMPLE STUDY REVISITED

The results of the negative binomial regression model applied to the double-blinded randomized study by Leppik et al,[34] who compare proabide and placebo in patients with epilepsy, are displayed in Table 2. For illustration, we consider one normally and one binomially distributed covariate. Ignoring the covariate effects (on the left hand side of the table), the treatment effect is quite small (rate ratio equal to 0.93), resulting in a large corresponding *p*-value. When the baseline seizure counts on the log-scale and the age category (< 30 years vs ≥ 30 years) are included in the model, it becomes evident that the patient history has a large effect on outcomes. However, the treatment effect, which is now defined conditional on the covariates, is also larger (rate ratio equal to 0.77), and the corresponding *p*-value becomes smaller. Taking into account the standard errors of the treatment effect estimates, differences between the marginal and conditional treatment effect are not as large anymore, as confidence intervals ([−0.569, 0.415] and [−0.558, 0.022], respectively) overlap. The included covariates explain some of the variance, which is reflected on the one hand in a smaller AIC, which decreases from 538.2 to 472.3, and on the other hand, in a larger shape parameter in the conditional model, which increases from 1.11 to 3.69. Leppik et al[34] concluded that there was no statistically significant difference between proabide and placebo regarding the seizure rates and that proabide is not a potent antiepileptic drug in the investigated population.

For illustrative purposes, let us assume that the example study is an IPS of a phase III trial and the collected data are used for a BSSR. This means that we cannot conduct an unblinded analysis as the one presented in Table 2, as the data would still be blinded. We merely have the number of seizure counts and pretreatment seizure counts for each patient but not the treatment group allocations. Assuming a treatment effect of a log rate ratio of −0.2 (resulting in a rate ratio of 0.82), the nuisance parameters can be estimated in a blinded manner. Table 3 displays the estimated nuisance parameters for all different methods.

With the nuisance parameter estimates from Table 3, a sample size of about 128 patients per group is estimated with the WT-FI approach to prove the assumed effect with a power of 80% at two-sided level $\alpha = 5\%$. Not considering covariates, using the method from Friede and Schmidli,[23] we would have required a sample size of about 353 patients per group.

7 | DISCUSSION

Negative binomial regressions are popular to model count outcomes in clinical trials for various reasons including the following. First, between-patient heterogeneity in the susceptibility to the event of interest is modeled through a Gamma

mixture distribution of individual rates. Second, varying follow-up times can be accommodated by including the logarithms of the follow-up times as an offset into the model. Third, baseline covariates can be easily included into the model by linking the rates through the natural logarithm to a linear predictor.

Sample size calculations for trials with negative binomial regressions as primary analyses are often based on guestimates of the nuisance parameters, that is, control or overall rate, shape parameter, and regression parameters of the covariates, as information on these parameters is often scarce in the design phase. This is partly due to poor reporting of some of the relevant nuisance parameters.[41,42] Here we suggested BSSR procedures based on different methodological approaches. It turned out that these approaches lead to very similar results in practice. They all have in common (a) that they did not inflate the type I error rate in any practically relevant way under the null hypothesis and (b) that their power was close to the target under the planning alternative.

In the presented procedures, the blinded data are treated as observations from a mixture distribution, restricted to the alternative hypothesis $\beta_2 = \theta_1$, and assuming that the allocation ratio r is known. This approach has been used before, for example, by Asendorf et al[40] and Schneider et al.[22] The approach is an alternative to the so-called lumping approach,[43] employed by Friede and Schmidli[23,24] as well as Schneider et al.[21] The latter treats the blinded data as if it were from one common negative binomial distribution, which is also feasible by assuming the expected efficacy parameter holds. Both approaches have been shown to differ only marginally.[40] A discussion of the mixture and lumping approaches applied to count data is also included in Friede et al.[44]

There were no notable differences between the three procedures, apart from a specific number of settings. For settings in which the covariate parameter was large ($\beta_1 = 0$, $\beta_2 = -0.2$, $\beta_3 \geq 2.5$, and $\eta = 3$), the procedure by Lyles tended to overestimate the sample size. This is due to computational limitations encountered with this procedure. During the simulation procedure, it became evident that, to adequately calculate the power within these settings, J in Equation (2) would need to be increased substantially, say from 1000 to 10 000 or even 50 000. This drawback was overcome by the proposed binning procedure, which reduces the computational effort considerably and at the same time improves the results. To further stabilize numerical calculations, we also considered standardizing the normally distributed covariate. Leaving covariates on their original scale in theory will not change the results from the BSSR, but may lead to numerical instability when inverting the expected FI in Equation (4), in particular when the parameter values are of different order.

The considered model and the proposed BSSR procedures could be extended in various ways. For instance, time-varying rates are not uncommon; see, for example, Nicholas et al[45] for an example in relapsing MS or Mütze et al[46] for an example in paediatric MS. Furthermore, the assumption of a common shape parameter across treatment groups is the default in standard software, but might need to be relaxed in some applications. Also, alternatives to including the pre-trial event rate as covariate have been discussed.[47]

CONFLICT OF INTEREST

The authors declare no potential conflict of interests.

DATA ACCESSIBILITY

Data sharing is not applicable to this article as no new data were created or analyzed in this study. R functions for attaining analysis results are openly available in figshare at https://doi.org/10.6084/m9.figshare.11973579.

ORCID
Antonia Zapf https://orcid.org/0000-0001-5339-2472
Thomas Asendorf https://orcid.org/0000-0003-1317-2138
Christoph Anten https://orcid.org/0000-0002-2556-3542
Tobias Mütze https://orcid.org/0000-0002-4111-1941
Tim Friede https://orcid.org/0000-0001-5347-7441

REFERENCES

1. Barton S. Which clinical studies provide the best evidence? the best RCT still trumps the best observational study. *BMJ*. 2000;321(7256):255-256. https://doi.org/10.1136/bmj.321.7256.255.

2. European Medicines Agency (EMA). Guideline on adjustment for baseline covariates in clinical trials. https://www.ema.europa.eu/documents/scientific-guideline/guideline-adjustment-baseline-covariates-clinical-trials_en.pdf. Accessed February 01, 2020; Published September 2015.

3. Hernandez AV, Steyerberg EW, Habbema JD. Covariate adjustment in randomized controlled trials with dichotomous outcomes increases statistical power and reduces sample size requirements. *J Clin Epidemiol*. 2004;57(5):454-460.

4. Frison L, Pocock SJ. Repeated measures in clinical trials: analysis using mean summary statistics and its implications for design. *Stat Med*. 1992;11(13):1685-1704. https://doi.org/10.1002/sim.4780111304.

5. Lyles RH, Lin HM, Williamson JM. A practical approach to computing power for generalized linear models with nominal, count, or ordinal responses. *Stat Med*. 2007;26(7):1632-1648.

6. Wittes J, Brittain E. The role of internal pilot studies in increasing the efficiency of clinical trials. *Stat Med*. 1990;9(1-2):65-72. https://doi.org/10.1002/sim.4780090113.

7. Friede T, Kieser M. Blinded sample size recalculation for clinical trials with normal data and baseline adjusted analysis. *Pharm Stat*. 2011;10(1):8-13.

8. Friede T, Kieser M. Sample size reassessment in non-inferiority trials: internal pilot study designs with ANCOVA. *Methods Inf Med*. 2011;50(3):237-243.

9. European Medicines Agency (EMA). Reflection paper on methodological issues in confirmatory clinical trials planned with an adaptive design. http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003616.pdf. Accessed February 01, 2020; Published October 2007.

10. Food & Drug Administration (FDA). Adaptive designs for clinical trials of drugs and biologics. https://www.fda.gov/downloads/drugs/guidances/ucm201790.pdf. Accessed February 01, 2020; Published November 2019.

11. Keene O, Jones MRK, Lane PW, Anderson J. Analysis of exacerbation rates in asthma and chronic obstructive pulmonary disease: example from the TRISTAN study. *Pharm Stat*. 2007;6:89-97. https://doi.org/10.1002/pst.250.

12. Li D, Zhang S, Cao J. Sample size calculation for clinical trials with correlated count measurements based on the negative binomial distribution. *Stat Med*. 2019;38(28):5413-5427. https://doi.org/10.1002/sim.8378.

13. Matsui S. Sample size calculations for comparative clinical trials with over-dispersed Poisson process data. *Stat Med*. 2005;24(9):1339-1356. https://doi.org/10.1002/sim.2011.

14. Igeta M, Takahashi K, Matsui S. Power and sample size calculation incorporating misspecifications of the variance function in comparative clinical trials with over-dispersed count data. *Biometrics*. 2018;74(4):1459-1467. https://doi.org/10.1111/biom.12878.

15. Tang Y, Fitzpatrick R. Sample size calculation for the Andersen-Gill model comparing rates of recurrent events. *Stat Med*. 2019;38(24):4819-4827. https://doi.org/10.1002/sim.8335.

16. Signorini DF. Sample size for Poisson regression. *Biometrika*. 1991;78(2):446-450. https://doi.org/10.1093/biomet/78.2.446.

17. Amatya A, Bhaumik D, Gibbons RD. Sample size determination for clustered count data. *Stat Med*. 2013;32(24):4162-4179. https://doi.org/10.1002/sim.5819.

18. Li H, Wang L, Wei L, Quan H. Sample size calculation for count data in comparative clinical trials with nonuniform patient accrual and early dropout. *J Biopharm Stat*. 2015;25(1):1-15.

19. Stucke K, Kieser M. Sample size calculations for noninferiority trials with Poisson distributed count data. *Biom J*. 2013;55(2):203-216.

20. Cook RJ, Bergeron PJ, Boher JM, Liu Y. Two-stage design of clinical trials involving recurrent events. *Stat Med*. 2009;28(21):2617-2638.

21. Schneider S, Schmidli H, Friede T. Robustness of methods for blinded sample size re-estimation with overdispersed count data. *Stat Med*. 2013;32(21):3623-3635. https://doi.org/10.1002/sim.5800.

22. Schneider S, Schmidli H, Friede T. Blinded sample size re-estimation for recurrent event data with time trends. *Stat Med*. 2013;32(30):5448-5457. https://doi.org/10.1002/sim.5977.

23. Friede T, Schmidli H. Blinded sample size reestimation with count data: methods and applications in multiple sclerosis. *Stat Med*. 2010;29(10):1145-1156.

24. Friede T, Schmidli H. Blinded sample size reestimation with negative binomial counts in superiority and non-inferiority trials. *Methods Inf Med*. 2010;49(6):618-624.

25. European Medicines Agency (EMA). Guideline on clinical investigation of medicinal products in the treatment of epileptic disorders. http://www.ema.europa.eu/ema/pages/includes/document/open_document.jsp?webContentId=WC500070043. Accessed February 01, 2020; Published July 2010.

26. Rosenfeld W, Conry J, Lagae L, et al. Efficacy and safety of perampanel in adolescent patients with drug-resistant partial seizures in three double-blind, placebo-controlled, phase III randomized clinical studies and a combined extension study. *Eur J Paediatr Neurol*. 2015;19:435-445.

27. French J, Brandt C, Friedman D, et al. Adjunctive use of controlled-release pregabalin in adults with treatment-resistant partial seizures: a double-blind, randomized, placebo-controlled trial. *Epilepsia*. 2014;55:1220-1228.

28. Zaccara G, Almas M, Pitman V, Knapp L, Posner H. Efficacy and safety of pregabalin versus levetiracetam as adjunctive therapy in patients with partial seizures: a randomized, double-blind, noninferiority trial. *Epilepsia*. 2014;55:1048-1057.

29. Krauss GL, Serratosa JM, Villanueva V, et al. Randomized phase III study 306: adjunctive perampanel for refractory partial-onset seizures. *Neurology*. 2012;78:1408-1415.

30. Baulac M, Brodie MJ, Patten A, Segieth J, Giorgi L. Efficacy and tolerability of zonisamide versus controlled-release carbamazepine for newly diagnosed partial epilepsy: a phase 3, randomised, double-blind, non-inferiority trial. *Lancet Neurol*. 2012;11:579-588.

31. Vazquez B, Yang H, Williams B, Zhou S, Laurenza A. Perampanel efficacy and safety by gender: subanalysis of phase III randomized clinical studies in subjects with partial seizures. *Epilepsia*. 2015;56:e90-e94.

32. French JA, Krauss GL, Biton V, et al. Adjunctive perampanel for refractory partial-onset seizures: randomized phase III study 304. *Neurology*. 2012;79:589-596.

33. French JA, Krauss GL, Steinhoff BJ, et al. Evaluation of adjunctive perampanel in patients with refractory partial-onset seizures: results of randomized global phase III study 305. *Epilepsia*. 2013;54:117-125.

34. Leppik IE, Dreifuss FE, Bowman T, et al. A double-blind crossover evaluation of progabide in partial seizures. *Neurology*. 1985;35:285.

35. Thall PF, Vail SC. Some covariance models for longitudinal count data with overdispersion. *Biometrics*. 1990;46:657-671.

36. Fitzmaurice G, Laird NM, Ware JH. *Applied Longitudinal Analysis*. 2nd ed. Hoboken, NJ: John Wiley & Sons; 2011.

37. Lawless JF. Negative binomial and mixed Poisson regression. *Can J Stat*. 1987;15(3):209-225.

38. Silvapulle MJ, Sen PK. *Constrained Statistical Inference: Inequality, Order, and Shape Restrictions*. Hoboken, NJ: John Wiley & Sons, Inc.; 2011.

39. Glueck DH, Muller KE. Adjusting power for a baseline covariate in linear models. *Stat Med*. 2003;22(16):2535-2551.

40. Asendorf T, Henderson R, Schmidli H, Friede T. Modelling and sample size reestimation for longitudinal count data with incomplete follow up. *SMMR*. 2019;28(1):117-133.

41. Röver C, Andreas S, Friede T. Evidence synthesis for count distributions based on heterogeneous and incomplete aggregated data. *Biom J*. 2016;58(1):170-185. https://doi.org/10.1002/bimj.201300288.

42. Holzhauer B, Wang C, Schmidli H. Evidence synthesis from aggregate recurrent event data for clinical trial design and analysis. *Stat Med*. 2018;37(6):867-882.

43. Proschan MA. Sample size re-estimation in clinical trials. *Biom J*. 2009;51(2):348-357. https://doi.org/10.1002/bimj.200800266.

44. Friede T, Häring DA, Schmidli H. Blinded continuous monitoring in clinical trials with recurrent event endpoints. *Pharm Stat*. 2019;18(1):54-64. https://doi.org/10.1002/pst.1907.

45. Nicholas R, Straube S, Schmidli H, Pfeiffer S, Friede T. Time-patterns of annualized relapse rates in randomized placebo-controlled clinical trials in relapsing multiple sclerosis: a systematic review and meta-analysis. *Mult Scler*. 2012;18(9):1290-1296.

46. Mütze T, Salem S, Benda N, Schmidli H, Friede T. Blinded continuous information monitoring of recurrent event endpoints with time trends in clinical trials; 2019. arXiv e-prints:arXiv:1903.02538.

47. Zheng H, Kimber A, Goodwin VA, Pickering RM. A comparison of different ways of including baseline counts in negative binomial models for data from falls prevention trials. *Biom J*. 2018;60(1):66-78. https://doi.org/10.1002/bimj.201700103.

> **How to cite this article:** Zapf A, Asendorf T, Anten C, Mütze T, Friede T. Blinded sample size reestimation for negative binomial regression with baseline adjustment. *Statistics in Medicine*. 2020;39:1980–1998. https://doi.org/10.1002/sim.8525

APPENDIX A: SUPPLEMENTARY RESULTS

A.1 Power of the BSSR procedures WT-ED and LR-ED

In addition to the results for the procedure WT-FI, we present the results for WT-ED and LR-ED.

There were no notable differences between the three procedures, apart from specific number of settings. For settings in which the covariate parameter was large ($\beta_1 = 0$, $\beta_2 = -0.2$, $\beta_3 \geq 2.5$, and $\eta = 3$), the procedure from Lyles tended to overestimate the sample size. This was partially compensated for by raising the parameter J from Equation (2) to $J = 10\,000$ in order to maintain the specified precision, that is, $\sum_j w_{ij} \geq 0.999$ (resulting in a small change in the fixed sample size required). However, this was not sufficient for larger values of β_3. The change observed in the required fixed sample size in the center figure is due to increasing J from $J = 1000$ to $J = 10\,000$. In settings where $\beta_3 = 4$, 4.5, and 5, the binning procedure outlined in Section 3.3.1 yields more appropriate sample sizes resulting in power values of 0.7835, 0.7895, and 0.7890 for WT-ED and 0.7841, 0.7932, and 0.7915 for LR-ED, respectively. Tabular results of Figures A1 and A2 are displayed in Appendix A.2.

A.2 Tabular results of the BSSR procedures

In addition to the graphical representation, we list results of the power calculations from Figures 3, A1, and A2 for improved readability. The power results for WT-FI, WT-ED and LR-ED are given in Tables A1, A2 and A3.

FIGURE A1 Power comparison of the fixed design and blinded sample size reestimation for the method WT-ED. The simulation settings correspond to Table 1, where nuisance parameter altered sequentially, one at a time [Colour figure can be viewed at wileyonlinelibrary.com]
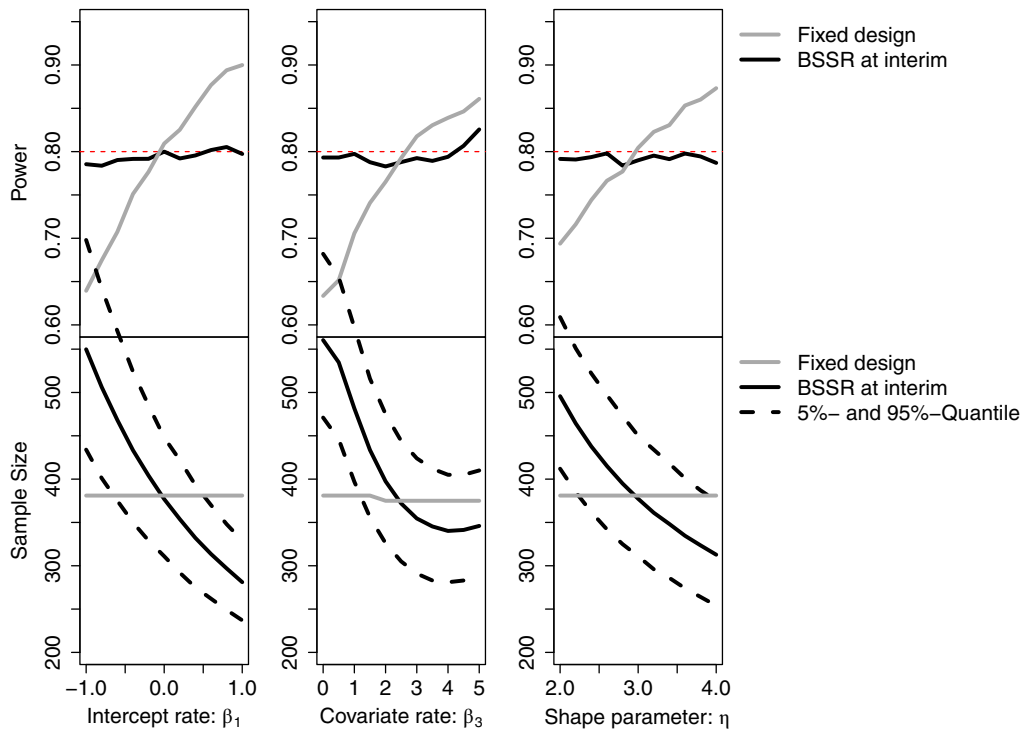

FIGURE A2 Power comparison of the fixed design and blinded sample size reestimation for the method LR-Lyles. The simulation settings correspond to Table 1, where nuisance parameter altered sequentially, one at a time [Colour figure can be viewed at wileyonlinelibrary.com]
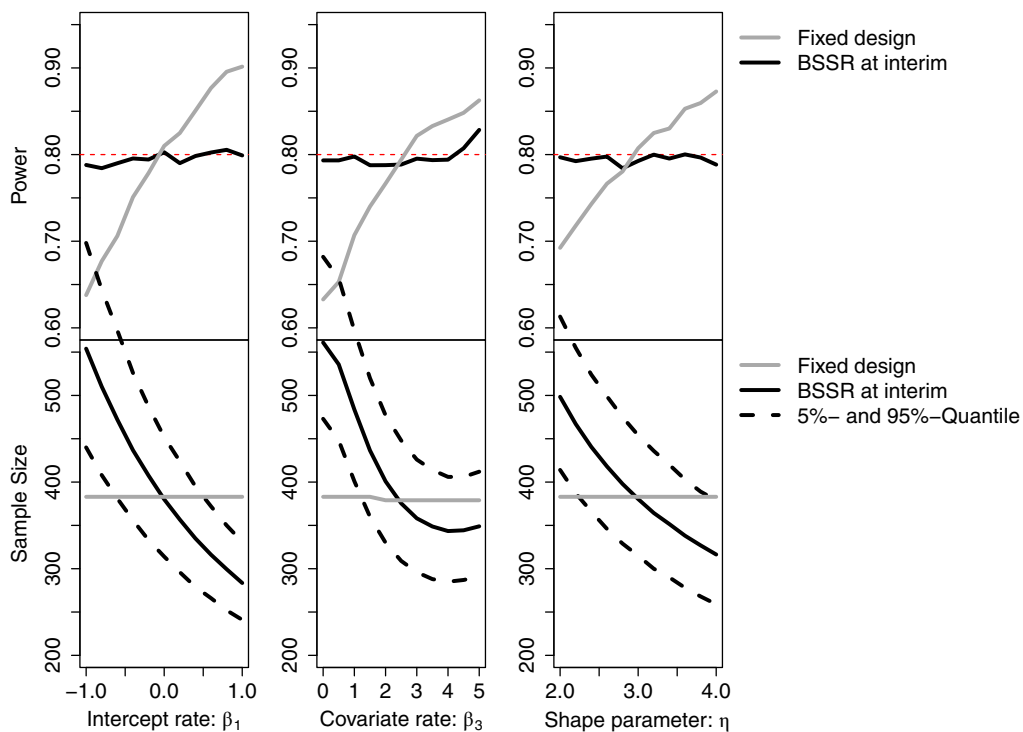
TABLE A1 Power results for varying nuisance parameters using the method WT-FI

Intercept rate: β_1	Covariate rate: β_3	Shape parameter: η	BSSR-Power	Fixed-Power
−1.0	2.5	3.0	0.7876	0.6371
−0.8	2.5	3.0	0.7849	0.6743
−0.6	2.5	3.0	0.7897	0.7053
−0.4	2.5	3.0	0.7908	0.7478
−0.2	2.5	3.0	0.7903	0.7755
0.0	2.5	3.0	0.7976	0.8091
0.2	2.5	3.0	0.7894	0.8235
0.4	2.5	3.0	0.7947	0.8491
0.6	2.5	3.0	0.7943	0.8748
0.8	2.5	3.0	0.7998	0.8926
1.0	2.5	3.0	0.7893	0.8991
0.0	0.0	3.0	0.7962	0.6312
0.0	0.5	3.0	0.7994	0.6454
0.0	1.0	3.0	0.8002	0.7012
0.0	1.5	3.0	0.7911	0.7389
0.0	2.0	3.0	0.7885	0.7687
0.0	2.5	3.0	0.7906	0.7954
0.0	3.0	3.0	0.7941	0.8236
0.0	3.5	3.0	0.7889	0.8344
0.0	4.0	3.0	0.7866	0.8417
0.0	4.5	3.0	0.7942	0.8491
0.0	5.0	3.0	0.7925	0.8645
0.0	2.5	2.0	0.7920	0.6909
0.0	2.5	2.2	0.7891	0.7160
0.0	2.5	2.4	0.7903	0.7410
0.0	2.5	2.6	0.7929	0.7641
0.0	2.5	2.8	0.7813	0.7753
0.0	2.5	3.0	0.7906	0.8033
0.0	2.5	3.2	0.7908	0.8214
0.0	2.5	3.4	0.7901	0.8287
0.0	2.5	3.6	0.7960	0.8520
0.0	2.5	3.8	0.7935	0.8561
0.0	2.5	4.0	0.7857	0.8705

Abbreviations: BSSR, blinded sample size reestimation; FI, Fisher information; WT, Wald test.

TABLE A2 Power results for varying nuisance parameters using the method WT-ED

Intercept rate: β_1	Covariate rate: β_3	Shape parameter: η	BSSR-Power	Fixed-Power
−1.0	2.5	3.0	0.7855	0.6394
−0.8	2.5	3.0	0.7837	0.6746
−0.6	2.5	3.0	0.7903	0.7074
−0.4	2.5	3.0	0.7916	0.7511
−0.2	2.5	3.0	0.7917	0.7767
0.0	2.5	3.0	0.8000	0.8093
0.2	2.5	3.0	0.7921	0.8252
0.4	2.5	3.0	0.7956	0.8518
0.6	2.5	3.0	0.8018	0.8768
0.8	2.5	3.0	0.8053	0.8939
1.0	2.5	3.0	0.7973	0.8999
0.0	0.0	3.0	0.7932	0.6335
0.0	0.5	3.0	0.7932	0.6516
0.0	1.0	3.0	0.7974	0.7057
0.0	1.5	3.0	0.7878	0.7408
0.0	2.0	3.0	0.7829	0.7651
0.0	2.5	3.0	0.7879	0.7926
0.0	3.0	3.0	0.7925	0.8176
0.0	3.5	3.0	0.7893	0.8305
0.0	4.0	3.0	0.7940	0.8390
0.0	4.5	3.0	0.8067	0.8464
0.0	5.0	3.0	0.8256	0.8609
0.0	2.5	2.0	0.7916	0.6938
0.0	2.5	2.2	0.7909	0.7161
0.0	2.5	2.4	0.7937	0.7441
0.0	2.5	2.6	0.7980	0.7664
0.0	2.5	2.8	0.7838	0.7770
0.0	2.5	3.0	0.7900	0.8044
0.0	2.5	3.2	0.7954	0.8228
0.0	2.5	3.4	0.7913	0.8306
0.0	2.5	3.6	0.7978	0.8532
0.0	2.5	3.8	0.7944	0.8601
0.0	2.5	4.0	0.7870	0.8732

Abbreviations: BSSR, blinded sample size reestimation; ED, expanded dataset; WT, Wald test.

TABLE A3 Power results for varying nuisance parameters using the method LR-ED

Intercept rate: β_1	Covariate rate: β_3	Shape parameter: η	BSSR-Power	Fixed-Power
−1.0	2.5	3.0	0.7880	0.6378
−0.8	2.5	3.0	0.7843	0.6770
−0.6	2.5	3.0	0.7900	0.7061
−0.4	2.5	3.0	0.7955	0.7509
−0.2	2.5	3.0	0.7943	0.7787
0.0	2.5	3.0	0.8027	0.8100
0.2	2.5	3.0	0.7901	0.8248
0.4	2.5	3.0	0.7983	0.8506
0.6	2.5	3.0	0.8024	0.8768
0.8	2.5	3.0	0.8055	0.8956
1.0	2.5	3.0	0.7990	0.9015
0.0	0.0	3.0	0.7933	0.6328
0.0	0.5	3.0	0.7933	0.6530
0.0	1.0	3.0	0.7978	0.7070
0.0	1.5	3.0	0.7877	0.7400
0.0	2.0	3.0	0.7878	0.7666
0.0	2.5	3.0	0.7885	0.7941
0.0	3.0	3.0	0.7953	0.8216
0.0	3.5	3.0	0.7936	0.8329
0.0	4.0	3.0	0.7941	0.8404
0.0	4.5	3.0	0.8073	0.8482
0.0	5.0	3.0	0.8285	0.8626
0.0	2.5	2.0	0.7969	0.6924
0.0	2.5	2.2	0.7924	0.7179
0.0	2.5	2.4	0.7953	0.7429
0.0	2.5	2.6	0.7977	0.7664
0.0	2.5	2.8	0.7842	0.7805
0.0	2.5	3.0	0.7926	0.8073
0.0	2.5	3.2	0.7999	0.8249
0.0	2.5	3.4	0.7954	0.8301
0.0	2.5	3.6	0.8002	0.8529
0.0	2.5	3.8	0.7966	0.8595
0.0	2.5	4.0	0.7884	0.8728

Abbreviations: BSSR, blinded sample size reestimation; ED, expanded dataset; LR, likelihood ratio.

A.3 Misspecification of the effect size

An additional simulation study was conducted to investigated whether a misspecified effect size has an influence on the reestimated sample size. The BSSR does not reestimate the effect size. Therefore, the BSSR procedure cannot be expected to adjust the sample size to meet the predefined power. The results are given in Figure A3.

In Figure A3, we consider the same settings as described in Section 5.1. All nuisance parameters are equal to the assumed nuisance parameters. The effect size, however, is different to that of the assumed effect size and ranges from −0.3 to −0.1 in steps of 0.01. The simulation reveals that under a misspecfied treatment effect, the BSSR performs similar to the fixed design. All settings were simulated 10 000 times.
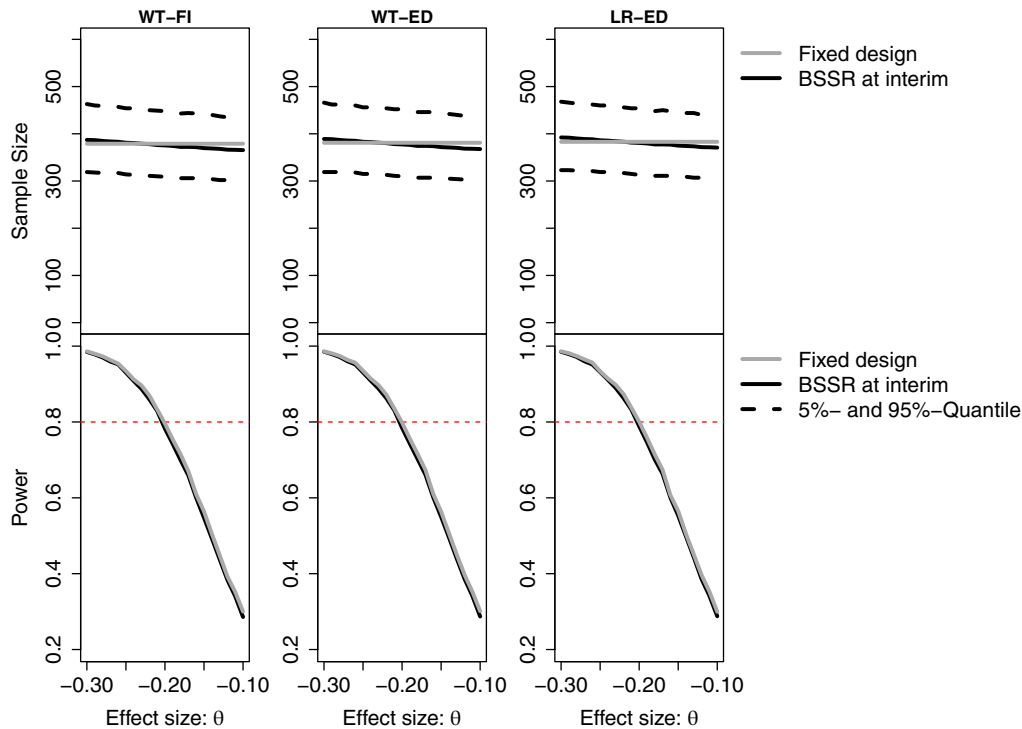

FIGURE A3 Sample size comparison of the fixed design and blinded sample size reestimation for the methods WT-FI, WT-ED, and LR-ED. The simulation settings correspond to Table 1, where nuisance parameters are kept equal to the expected but the effect size is altered sequentially [Colour figure can be viewed at wileyonlinelibrary.com]