
**Blinded Sample Size Re-estimation for Longitudinal
Overdispersed Count Data in Randomized Clinical Trials with
an Application in Multiple Sclerosis**

Dissertation
zur Erlangung des humanwissenschaftlichen Doktorgrades
in der Medizin
der Georg-August-Universität Göttingen



GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN

UNIVERSITÄTSMEDIZIN : UMG
GÖTTINGEN

vorgelegt von
THOMAS ASENDORF
aus Cali, Kolumbien

GÖTTINGEN, 2020

First Supervisor: Prof. Dr. Tim Friede
Department of Medical Statistics
University Medical Center Göttingen
Georg-August-University Göttingen

Second Committee Member: Prof. Dr. Heike Bickeböller
Institute of Genetic Epidemiology
University Medical Center Göttingen
Georg-August-University Göttingen

Third Committee Member: Prof. Dr. Thomas Kneib
Chair of Statistics
Chair of Econometrics
Georg-August-University Göttingen

ABSTRACT

In many clinical experiments, particularly in randomized clinical trials, the sample size required needs to be assessed and justified. For calculating a clinical trial's sample size, assumptions have to be made regarding the clinical trial's outcome data. These assumptions are based on prior clinical trials or merely on expert knowledge and always subject to some degree of uncertainty. To cope with this uncertainty in sample size estimation, adaptive designs were developed to re-estimate the sample size within a running trial. Especially adaptive designs for blinded sample size re-estimation, also referred to as non-comparative adaptive designs, have gained popularity, as these generally do not require an adjustment of the significance level to maintain type I error rates.

In the first part of this thesis, we will consider developing sample size re-estimation methods for longitudinal overdispersed count data. As a first step, such data is modeled by a negative binomial counting process, and techniques for inference, sample size estimation and sample size re-estimation are provided. In a second step, presented methods are extended to handle time trends, which may occur during the course of a clinical trial. These trends are modeled by a gamma frailty model, for which inference, sample size estimation and sample size re-estimation techniques are also described in detail. As an application, we consider lesion counts measured by magnetic resonance imaging (MRI), which play an important role in phase II multiple sclerosis (MS) trials for measuring disease progression. These lesion counts are generally overdispersed and often measured multiple times per patient during a running trial, therefore resembling the statistical model. Methods are kept general to allow for applications outside of MS, whenever similar data arise, and shown to preserve type I error rates while correcting the sample size, such that a desired power level is reached, in extensive simulation runs.

The second part of this thesis will consider univariate negative binomial data with baseline covariates. For example, such data arise in MS when the total number of lesions at the end of a clinical trial, corrected for the number of lesions at baseline or other baseline variables, is taken as an endpoint. Developed sample size re-estimation techniques are also shown to preserve type I error rates while correcting the sample size such that a desired power level is reached. The summarized results are made available as R-functions and extend current methodology in the field of non-comparative adaptive designs.

DEDICATION AND ACKNOWLEDGEMENTS

First and foremost I would like to thank my supervisor Tim Friede for supporting me throughout the development of this dissertation, as well as my second and third supervisors, Heike Bickeböller and Thomas Kneib. I also thank my co-authors Heinz Schmidli and Robin Henderson for their profound input and ideas on the first two papers, as well as Antonia Zapf, Christoph Anten and Tobias Mütze who co-authored the third paper with me. Special thanks also to my colleagues and fellow PhD students, especially Marius Placzek, for the great working atmosphere under which this dissertation was developed. Last but not least, I would like to thank my wife and kids for their emotional support, who were probably even more excited than I was.

STATUTORY DECLARATION

I, Thomas Asendorf, declare that this dissertation titled "Blinded Sample Size Re-estimation for Longitudinal Overdispersed Count Data in Randomized Clinical Trials with an Application in Multiple Sclerosis" and the works presented within are my own and that it was written independently with no other sources and aids than quoted. I further declare that I have not submitted a previous dissertation containing parts of this work.

.....
Date

.....
Signature

TABLE OF CONTENTS

	Page
List of Tables	ix
List of Figures	xi
1 Introduction	1
1.1 Adaptive Designs for Sample Size Re-estimation	1
1.2 Clinical Trials in Multiple Sclerosis	2
1.3 Considerations for Statistical Modeling	4
1.4 Outline of the Presented Research	5
2 Proposed Adaptive Designs	7
2.1 NB-INAR(1) Model	7
2.1.1 Statistical Inference	8
2.1.2 Blinded Sample Size Re-estimation	9
2.1.3 Numerical Evaluation of Type I Error Rate and Power	11
2.1.4 Software Implementation and Availability of Methods	12
2.2 Gamma Frailty Model	13
2.2.1 Statistical Inference	14
2.2.2 Blinded Sample Size Re-estimation	17
2.2.3 Numerical Evaluation of Type I Error Rate and Power	17
2.2.4 Software Implementation and Availability of Methods	19
2.3 Negative Binomial Outcomes with Covariates	20
2.3.1 Statistical Inference	20
2.3.2 Sample Size Estimation	21
2.3.3 Blinded Sample Size Re-estimation	22
2.3.4 Numerical Evaluation of Type I Error Rate and Power	22
2.3.5 Software Implementation and Availability of Methods	25
3 Discussion	27
Bibliography	29

LIST OF TABLES

TABLE	Page
2.1 Functions implemented in the R-package <code>spass</code> for sample size estimation and blinded re-estimation within the NB-INAR(1) model.	13
2.2 Functions implemented in the R-package <code>spass</code> for sample size estimation and blinded re-estimation within the gamma frailty model.	20
2.3 Simulation settings for the comparison of type I error rates between the fixed design and the blinded sample size re-estimation procedures for univariate negative binomial counts with covariates.	23
2.4 Functions uploaded as supplementary material within an R-script for sample size estimation and blinded re-estimation with baseline covariates.	25

LIST OF FIGURES

FIGURE	Page
1.1 Number of enhancing lesions attained from serial brain MRI in relapsing-remitting multiple sclerosis (RRMS) and secondary progressive multiple sclerosis (SPMS). [61]	4
2.1 Possible recruitment scheme of a clinical trial with 6 month follow-up per patient. Each line represents the observation time of an individual patient with interim analysis planned 12 months after trial onset.	10
2.2 Simulated power of BSSR and fixed design with true nuisance parameters (x-axis) altered one by one. Underlying Model: NB-INAR(1). Each parameter combination was simulated 10,000 times.	12
2.3 Comparison of type I error rates of the test statistic Z (2.20) using restricted and unrestricted variance estimates. Each dot represents one simulation setting. Solid lines represent 95% random intervals of simulated type I error rate.	16
2.4 Simulated power of BSSR and fixed design with true nuisance parameters (x-axis) altered one by one. Underlying Model: Gamma Frailty, assuming constant rates. Every setting was simulated 10,000 times.	18
2.5 Simulated power of BSSR and fixed design with true nuisance parameters altered one by one. Underlying Model: Gamma Frailty, assuming exponential trend. Every setting was simulated 10,000 times.	19
2.6 Type I error rates of BSSR and fixed design for all possible combinations of parameters as listed in Table 2.3 while under the null hypothesis. Dashed lines represent 95%-random intervals for simulated type I error rates. Every setting was simulated 10,000 times.	24
2.7 Power simulation of BSSR and fixed design for the WT-FI method, with true nuisance parameters altered sequentially. Underlying Model: Univariate Negative Binomial Counts with Covariates. Every setting was simulated 10,000 times.	24

INTRODUCTION

1.1 Adaptive Designs for Sample Size Re-estimation

When planning clinical trials, the sample size has to be justified for ethical and economical reasons. On the one hand, if a sample size is too low, the power for rejecting the null hypothesis is too low and the planned clinical trial may fail to empirically prove the clinical objective, requiring further clinical trials in the follow-up. On the other hand, a too high sample size will burden unnecessarily many patients with undergoing the clinical trial, potentially putting them at risk for adverse events. Furthermore, a larger trial will likely take more time to complete, delaying the introduction of a potentially safe and efficacious treatment into the market, and requiring more resources in the process. Therefore, attaining a proper sample size for a clinical trial is a necessary task. Formulas for calculating the sample size depend on many different aspects, such as trial design, intended power, significance level for hypothesis testing, but also the minimal clinically relevant effect size and nuisance parameters, which includes the event rate in the control arm and shape parameter, when considering negative binomial count data, or the variability in groups, when considering continuous outcomes [40].

While some of these influencing aspects, such as intended power and significance level, can be determined when planning a clinical trial, nuisance parameters can only be assessed from other sources, such as previous clinical trials. However, assessing nuisance parameters from previous clinical trials can be a difficult task. Previous clinical trials may not report required nuisance parameters or are insufficiently comparable to the planned trial, as they do not consider the same patient population or differ in trial design. In some cases, for example when the endpoint is being studied for the first time, previous trials may not even exist in the first place [13]. Therefore, the nuisance parameters assumed from results of previous clinical trials are always subject to some degree of uncertainty. Consequently, the sample size calculated with these nuisance

parameters is uncertain as well. To overcome this uncertainty in sample size estimation, sample size re-estimation procedures have been proposed, which re-estimate the sample size based on estimates for nuisance parameters within a running trial, ensuring adequate power while maintaining statistical rigor of trial results [39, 51, 53].

Following the introduction above, sample size re-estimation aims to change the planned sample size within a running trial, making it a type of adaptive design. In the guideline Adaptive Designs for Clinical Trials of Drugs and Biologics [21], the United States Food & Drug Administration (FDA) distinguishes between adaptive designs based on comparative data and adaptive designs based on non-comparative data. Sample size re-estimation can be done both on comparative data, i.e. using the group assignment for estimation, and on non-comparative data, i.e. without knowledge of the group assignment. Sample size re-estimation based on comparative data includes designs which adapt the sample size based on an interim result of the effect size, also called unblinded sample size re-estimation. Such designs are employed when there is considerable uncertainty about the true treatment effect size, but require proper adjustment of the significance level, following one of many different methods [9, 12, 17, 18, 20, 48] as they can inflate the type I error probability. For example, Proschan and Hunsberger [54] have shown that the type I error probability can be more than double the significance level if no proper adjustment is employed. This stands in contrast to blinded sample size re-estimation, which, as a non-comparative adaptive design, generally has no effect or a limited effect on the type I error probability and therefore does not require an adjustment of the significance level [21]. Another type of re-estimation method lies in between these two categories, in that they re-estimate the sample size based on nuisance parameters attained from comparative data, but without an interim estimated treatment effect. These methods are also referred to as partially blinded or partially unblinded [51].

Blinded sample size re-estimation procedures, which we will focus on, estimate nuisance parameters required for calculating the sample size based on non-comparative data from a running trial, assuming a relevant effect size used in the planning phase to be present. Estimates for nuisance parameters are then used to calculate the sample size anew and the sample size is adjusted for incorrect assumptions met at the initial sample size calculation. Procedures for blinded sample size re-estimation have been developed for numerous outcome types, including normally distributed data [29], normally distributed data with covariates [23, 65], dichotomous outcomes [24], Poisson distributed count data and overdispersed count data [15, 25, 26, 56, 57].

1.2 Clinical Trials in Multiple Sclerosis

As an application of blinded sample size re-estimation for longitudinal count data, we will consider clinical trials in multiple sclerosis (MS). MS is an autoimmune disorder affecting the brain and spinal cord by damaging the insulating covers of nerve cells. It is the most common inflammatory neurological disorder in young adults with approximately 2.2 million affected people worldwide in

2016 [63]. To this day, no known cure for MS exists, making it an intense field of medical research. Disease progression is often experienced as an increased occurrence of related symptoms, such as fatigue, spasticity, resulting in walking difficulties, and vision problems. These symptoms can occur either in isolated attacks (relapsing and remitting) or steadily increase in severity over time (progressive) [43]. Depending on the pattern of progression experienced by affected patients, the National Multiple Sclerosis Society defined four distinct disease-courses in 1996 and updated these in 2013 [42, 43]:

- Clinically Isolated Syndrome (CIS)
- Relapsing-Remitting Multiple Sclerosis (RRMS)
- Primary Progressive Multiple Sclerosis (PPMS)
- Secondary Progressive Multiple Sclerosis (SPMS)

CIS is the occurrence of symptoms that could be MS, but have yet to fulfill dissemination criteria in time [47]. Once current MS diagnostic criteria are fulfilled, a patient with CIS is categorized to RRMS. Patients with RRMS experience active phases with increased MRI activity and occurrences of relapses, as well as inactive phases with less signs of symptoms. PPMS and SPMS are two distinct progressive forms of MS, in which symptom severity increases gradually in time. While it is important to distinguish different types of MS for treatment, and therefore also when designing clinical trials for specific patient types, the underlying type of MS can not be determined with full certainty and is often assessed in retrospect by considering the patient's medical history [43].

In phase III confirmatory trials, the occurrence of relapses, as a temporary worsening of neurological symptoms, e.g. summarized as annualized relapse rates, are clinical endpoints in RRMS, whereas endpoints in PPMS or SPMS include measures of disability worsening, e.g. based on assessments of the expanded disability status scale (EDSS) [62]. For phase II trials, however, these endpoints are not feasible as they require a high number of subjects and long observation period to discover significant treatment effects. Therefore, numerous surrogate parameters, such as T2-hyperintense lesions, Gadolinium-enhancing T1 lesions and whole brain atrophy, have been established to monitor disease progression in smaller and shorter phase II trials [62] and used in numerous clinical trials [8, 11, 28, 37, 38]. We will focus on the count measures, i.e. Gadolinium-enhancing T1 lesion, also referred to as 'black holes', and T2-hyperintense lesions, also referred to as 'white spots'. These measures are inter-related but different from a medical perspective. Gadolinium-enhancing T1 lesion counts may arise from severe chronic demyelination, inflammation, permanent axonal damage and gliosis through MS. T2-hyperintense lesions can show the total number of lesions as they are a marker of past injury, which only rarely disappear completely [14]. Lesion counts can be measured at several time points during a running trial. For example, Chataway et al. [11], who conducted a phase II clinical trial to examine the efficacy of

simvastatin in progressive MS, examined T2-hyperintense lesion counts at baseline, 12 months after admission and 24 months after admission, resulting in longitudinal count data.

1.3 Considerations for Statistical Modeling

Different possibilities for modeling longitudinal count data exist. The choice of the statistical model we consider is based on two characteristics observed in real data. The first characteristic we observe, is that counts at each time point are overdispersed, i.e. the variance of marginal observations is substantially higher than their mean. For example, Tubridy et al. [61] measure an overdispersion of 3.37 when considering lesions of RRMS patients at month one (mean lesion rate of 1.61 and variance of 5.44). Therefore, a model which allows for overdispersed marginal observations is required. The second characteristic is that observations between different time points are correlated, e.g. Tubridy et al. [61] measure a correlation of 0.48 between month one and month two lesion counts of RRMS patients. Therefore, any statistical model considered should account for dependencies observed between different time points. Figure 1.1 displays the complete data from Tubridy et al. [61] which show the need for a statistical model to handle these two characteristics.

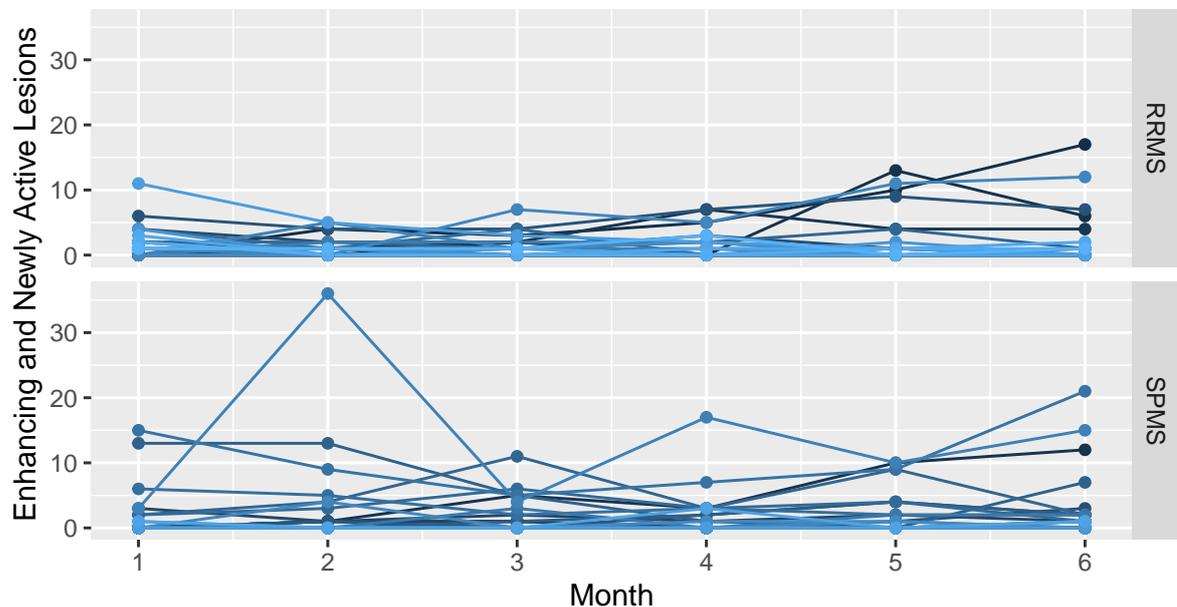


Figure 1.1: Number of enhancing lesions attained from serial brain MRI in relapsing-remitting multiple sclerosis (RRMS) and secondary progressive multiple sclerosis (SPMS). [61]

Although modeling correlation between continuous observations is well explored, this task is incomparably more difficult for count data, as no natural multivariate extension of the Poisson or negative binomial distribution exists. Many different approaches for modeling longitudinal count

data have been developed, including inter time series as Markov Chains [16, 34, 35] parametric models [2, 3, 31], general classes of estimating equations [27, 36], random effects and latent process approaches [10, 19, 32, 33]. Here, we present two different models, which both allow for modeling dependencies between overdispersed count data. Each of these models has their strengths and weaknesses, which we will acknowledge and discuss. The first statistical model used to model lesion counts is a negative binomial integer valued autoregressive process of first order, also NB-INAR(1) model, introduced by McKenzie [46] and re-parametrized for our purposes. The second model we present, is a gamma frailty model, introduced by Henderson and Shimakura [33] and extended by Fiocco et al. [19].

1.4 Outline of the Presented Research

The main purpose of the presented research is to extend the methodology developed for blinded sample size re-estimation of univariate overdispersed count data developed by Friede and Schmidli [25, 26] and Schneider et al. [56, 57] to a multivariate setting. Friede and Schmidli developed blinded sample size re-estimation procedures for comparing two groups with univariate Poisson or negative binomial distributed outcomes. Schneider et al. extended these procedures to cope with incomplete observations at the time point of the interim analysis, as well as time trends within response rates. Extending these methods will allow for sample size re-estimation in clinical trials with longitudinal count end points, observed multiple times at fixed time points. The developed methods are each presented in a similar structure. We begin by introducing and adapting the underlying statistical model. The step of statistical modeling is followed by deriving procedures for statistical inference, from which sample size formulas can be calculated. These sample size formulas are then studied to develop blinded sample size re-estimation techniques. Finally, the performance of the developed sample size re-estimation techniques is examined in simulations. The procedures are shown to properly adjust the sample size of clinical trials while not inflating the type I error probability and therefore fulfilling regulatory requirements. The developed methods are further demonstrated on real data examples and an implementation in R [55] is given.

The outlined research is presented in three different papers and has been published in peer-reviewed journals:

- Asendorf, T, Henderson, R, Schmidli, H, Friede, T. (2019). Modelling and sample size reestimation for longitudinal count data with incomplete follow up. *Statistical Methods in Medical Research*, 28(1), 117-133. <https://doi.org/10.1177/0962280217715664>
- 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, 1503-1528. <https://doi.org/10.1002/sim.8061>

- Zapf, A, Asendorf, T (shared first authorship), Anten, C, Mütze, T, Friede, T. (2020). Blinded sample size reestimation for negative binomial regression with baseline adjustment. *Statistics in Medicine*, 39, 1980-1998. <https://doi.org/10.1002/sim.8525>.

As a first attempt of modeling longitudinal count data, the first paper considers the NB-INAR(1) model for blinded sample size re-estimation, which easily allows for incorporating incomplete follow up data. The second paper uses a gamma frailty model to allow for trend formulations within clinical trials and further evaluates the methodology for non-inferiority trials. The third paper explores the utilization of covariates in blinded sample size re-estimation, but in the simpler context of univariate overdispersed negative binomial counts.

PROPOSED ADAPTIVE DESIGNS

2.1 NB-INAR(1) Model

The negative binomial integer valued autoregressive process of first order (NB-INAR(1) model) presented here, and used in deriving the proposed adaptive designs, is a variation of the binomial thinning model from McKenzie [46] and Al-Osh and Alzaid [1]. The NB-INAR(1) model allows for generating observations which marginally follow a negative binomial distribution, while at the same time allowing for autoregressive dependencies of first order between different observations. This is achieved using an operation called binomial thinning, formally introduced by Steutel and van Harn [60]. Let X be some positive random integer and $B_k(p)$ i.i.d. Bernoulli distributed random variables with parameter p , i.e. $\mathbb{P}(B_k(p) = 1) = p$, independent of X for all $k \in \mathbb{N}$. Then, the binomial thinning operator is defined as

$$(2.1) \quad X \circ p := \sum_{k=1}^X B_k(p).$$

From the definition it becomes clear that the random variables X and $X \circ p$ are positively correlated for $p \in (0, 1]$. Furthermore, both X and $X \circ p$ are integers. Therefore binomial thinning is a viable operation for generating dependent count data, which we will make use of. Let $X_{ij}^{(t)}$ denote the observation of patient j in group i at time point t , where $i = E, C$ (experiment and control group), $j = 1, \dots, n_i$ and $t = 1, \dots, T$. Then the statistical model is given by

$$(2.2) \quad X_{ij}^{(t)} = X_{ij}^{(t-1)} \circ U_{ij}^{(t)} + W_{ij}^{(t)} \quad \text{for } t = 2, \dots, T, j = 1, \dots, n_i, i = E, C,$$

where the random variables observed at the first time point are defined to be negative binomial distributed with mean λ_i and shape parameter η , i.e. $X_{ij}^{(1)} \stackrel{i.i.d.}{\sim} NB(\lambda_i, \eta)$. The parameter used for bi-

nomial thinning is now a random variable $U_{ij}^{(t)} \stackrel{i.i.d.}{\sim} \text{Beta}(a\eta, (1-a)\eta)$ and the resulting term from binomial thinning is added to an independent random variable $W_{ij}^{(t)} \stackrel{i.i.d.}{\sim} \text{NB}((1-a)\lambda_i, (1-a)\eta)$. In Appendix A of Asendorf et al. [6], it was proven that observations $X_{ij}^{(t)}$ then follow a negative binomial distribution with mean λ_i and shape parameter η for $t = 1, \dots, T$. Further, two observations from the same patient were shown to be correlated with $\text{Cov}(X_{ij}^{(t)}, X_{ij}^{(s)}) = a^{|s-t|} \cdot (\lambda_i + \lambda_i^2/\eta)$ (Appendix B, Asendorf et al. [6]). Therefore, we could model a marginal negative binomial distribution and at the same time have an autoregressive dependency structure of first order between time points, a typical assumption in time series analysis. In the following calculations we will further denote the dependency parameter ρ , defined as

$$(2.3) \quad \rho = \frac{1}{T^2} \sum_{t=1}^T \sum_{s=1}^T a^{|t-s|}.$$

The parameter ρ is essentially a geometric sum of the correlation parameter a and will ease the notation in following calculations.

2.1.1 Statistical Inference

Within the NB-INAR(1) model, the main interest lies in testing differences of means between groups. More specifically, we will be testing for differences between groups by using the rate ratio $\theta = \lambda_E/\lambda_C$, where the hypothesis of interest is given by

$$(2.4) \quad H_0 : \theta \geq 1 \quad \text{vs.} \quad H_1 : \theta < 1.$$

For testing the null hypothesis H_0 , we derive a Wald-type test statistic Z using moment estimators $\widehat{\lambda}_i$, $\widehat{\rho}$ and $\widehat{\frac{1}{\lambda_i} + \frac{1}{\eta}}$ for λ_i , ρ and $\frac{1}{\lambda_i} + \frac{1}{\eta}$, respectively. Using the delta method in conjunction with the central limit theorem, it was proven that

$$(2.5) \quad Z = \frac{\log(\widehat{\lambda}_E/\widehat{\lambda}_C)}{\sqrt{\widehat{\rho} \cdot \left(\frac{1}{n_E} \left(\widehat{\frac{1}{\lambda_E} + \frac{1}{\eta}} \right) + \frac{1}{n_C} \left(\widehat{\frac{1}{\lambda_C} + \frac{1}{\eta}} \right) \right)}} \stackrel{\text{approx}}{\sim} N(0, 1) \text{ under } H_0.$$

It is worthwhile to point out, that for the derivation of the test statistic moment estimators were used. This approach is similar to an approach taken by Friede and Schmidli [25] for the case of univariate overdispersed Poisson counts, and chosen because the resulting Wald-type statistic is of a closed form and therefore allows for an explicit sample size formula. An approach with maximum likelihood estimates would have required an explicit formula of the inverse Fisher information, which could not be attained due to the complex likelihood structure, especially for multiple time points. Therefore, using a maximum likelihood approach would only have been feasible with numeric calculations, giving less insight on the influence of specific nuisance

parameters on the sample size. Furthermore, we estimate the full term $\frac{1}{\lambda_i} + \frac{1}{\eta}$ instead of plugging in estimates for η and λ_i . The reason is, that the moment estimator for η can be negative in the case of underdispersed data. While underdispersed lesion counts are not probable in MS, these can occur occasionally during simulation runs and halt calculations. The employed estimators themselves are described in more detail within Asendorf et al. [7].

Using the derived test statistic Z and its approximate normal distribution, the required sample size to reject the null hypothesis with a power of $1 - \beta$ at significance level α can be calculated by

$$(2.6) \quad n_C = \frac{(z_\beta + z_\alpha)^2 \cdot \rho}{\log(\theta^*)^2} \left(\frac{(1 + k\theta^*)^2}{(1 + k)k\theta^*\bar{\lambda}} + \frac{1}{\eta} \left(1 + \frac{1}{k} \right) \right),$$

where n_C denotes the required sample size of the control group, $k = n_E/n_C$ denotes the allocation ratio, θ^* the assumed effect under alternative or clinically relevant effect size, and $\bar{\lambda} = (k\lambda_E + \lambda_C)/(k + 1)$ the overall rate. The closed form sample size formula allows us to study the influence of certain parameters on the sample size. For example, if the correlation parameter α is equal to 1, then ρ is equal to 1 and increasing the number of time points T has no effect on the required sample size. In that case, the sample size formula reduces to the simpler case of a two group comparison and one time point, as considered by Friede and Schmidli [26]. On the other end, if $\alpha = 0$, the parameter ρ is equal to $1/T$ and increasing the number of time points is equivalent to increasing the sample size. The sample size formula was shown to be accurate and hold the desired power in a number of different settings, described in Table 2 of Asendorf et al. [6].

2.1.2 Blinded Sample Size Re-estimation

To perform a blinded sample size re-estimation, it is necessary to estimate all nuisance parameters influencing the sample size in a blinded manner. To accomplish this task, we will estimate the nuisance parameters using a likelihood approach assuming that the blinded observations follow a mixture distribution of both groups with known weights. This approach is an alternative to the so called lumping approach, which assumes data to be from a common distribution, for example employed in Friede and Schmidli [25, 26] and discussed by Proschan [52]. The likelihood necessary to achieve a blinded estimation of nuisance parameters $\bar{\lambda}$, α and η is given by

$$(2.7) \quad L(\bar{\lambda}, \eta, \alpha | x_1^{(1)}, \dots, x_{n_E+n_C}^{(T)}) = \prod_{j=1}^{n_E+n_C} \left(f_{X_j^{(1)}}(\bar{\lambda}, \eta, \alpha | x_j^{(1)}) \cdot \prod_{t=1}^{T-1} f_{X_j^{(t+1)} | X_j^{(t)}}(\bar{\lambda}, \eta, \alpha | x_j^{(t+1)}) \right)$$

where $f_{X_j^{(1)}}(\cdot)$ is the marginal probability function of the first observation and $f_{X_j^{(t+1)} | X_j^{(t)}}(\cdot)$ describes the conditional probabilities of subsequent observations, exploiting the Markov structure of the NB-INAR(1) model. More explicitly, the probability functions are

$$f_{X_j^{(1)}}(\bar{\lambda}, \eta, \alpha | x_j^{(1)}) = \frac{1}{1+k} \cdot \left[k \cdot \mathbb{P}_{NB} \left(x_j^{(1)}, \frac{\bar{\lambda}(1+k)}{k+1/\theta}, \eta \right) + \mathbb{P}_{NB} \left(x_j^{(1)}, \frac{\bar{\lambda}(1+k)}{1+k\theta}, \eta \right) \right]$$

and

$$f_{X_j^{(t+1)}|X_j^{(t)}}(\bar{\lambda}, \eta, a | x_j^{(t+1)}) = \frac{1}{1+k} \sum_{y=0}^{\min(x_j^{(t+1)}, x_j^{(t)})} \binom{x_j^{(t)}}{y} \frac{B(a\eta + y, (1-a)\eta + x_j^{(t)} - y)}{B(a\eta, (1-a)\eta)} \left[k \mathbb{P}_{\text{NB}}\left(x_j^{(t+1)} - y, \frac{\bar{\lambda}(1+k)}{k+1/\theta}, (1-a)\eta\right) + \mathbb{P}_{\text{NB}}\left(x_j^{(t+1)} - y, \frac{\bar{\lambda}(1+k)}{1+k\theta}, (1-a)\eta\right) \right]$$

where $B(x, y) = \Gamma(x)\Gamma(y)/\Gamma(x+y)$ denotes the beta function, $\Gamma(x) = \int_0^\infty t^{x-y} e^{-t} dt$ the gamma function and $\mathbb{P}_{\text{NB}}(x, \lambda, \eta)$ the probability function of the negative binomial distribution with mean λ and variance $\lambda + \lambda^2/\eta$. Note, that the shape parameter is assumed to be equal in both groups. To perform a sample size re-estimation within a running trial, the nuisance parameters are estimated from the likelihood (2.8), assuming that the effect size is equal to the assumed effect size, and plugged into the sample size formula from (2.6) to attain a blinded sample size re-estimate. From a practical perspective, however, another issue appears. Because recruitment capacity of centers is limited and patient numbers are therefore accumulated over time, it is common for recruitment in prospective longitudinal trials to follow a certain scheme. This results in observations of patients not being complete at interim time points, e.g. as displayed in Figure 2.1.

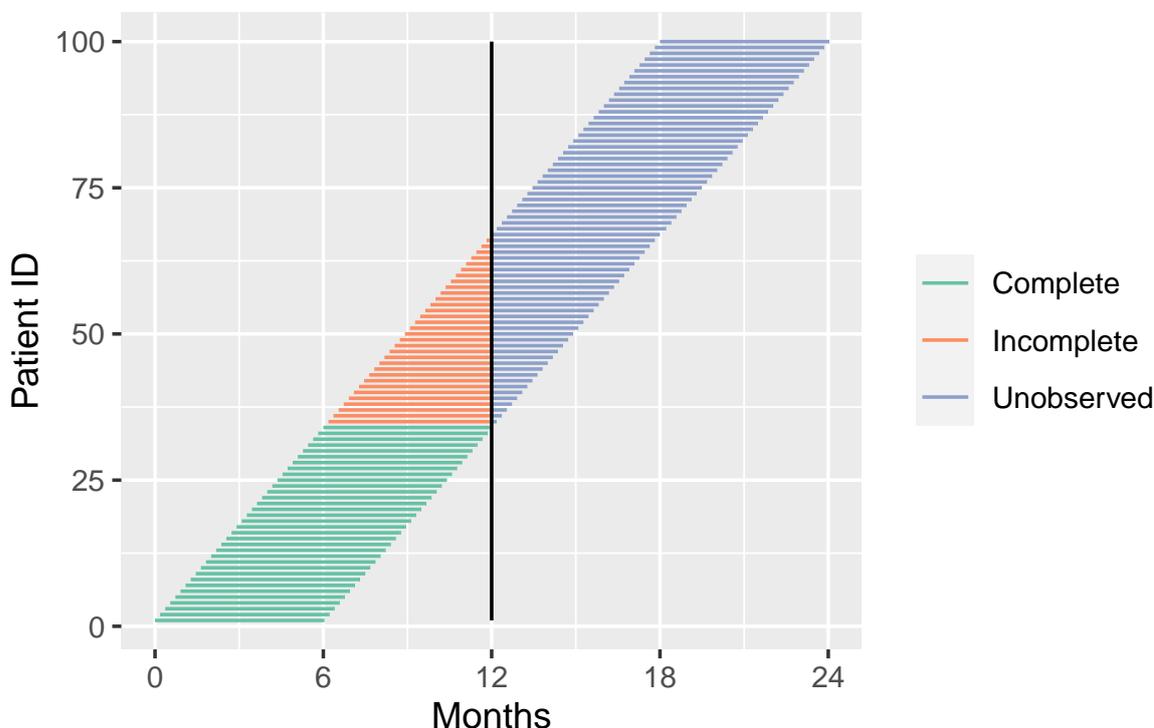


Figure 2.1: Possible recruitment scheme of a clinical trial with 6 month follow-up per patient. Each line represents the observation time of an individual patient with interim analysis planned 12 months after trial onset.

Consequently, some data from patients are available although their follow-up has not been completed. This data can be incorporated into the analysis by replacing the fixed final time point T by a subject-specific final time points T_j for $j = 1, \dots, n_E + n_C$ in Equation (2.7). Depending on the proportion of incomplete data at interim, incorporating incomplete data can substantially reduce the variability of the sample size estimate as shown in Asendorf et al. [6], see in particular Table 4.

2.1.3 Numerical Evaluation of Type I Error Rate and Power

The performance of the proposed BSSR procedure, especially in comparison to the fixed design, was thoroughly investigated for settings realistic for lesion counts in multiple sclerosis. First, the type I error of an experiment with the re-estimated sample size and that of a fixed design was compared. The fixed sample size was calculated assuming an overall rate of $\bar{\lambda} = 2$, correlation parameter of $\alpha = 0.5$ and shape parameter of $\eta = 1$ (equivalent to a variance of 6), resulting in a required sample size of 165 patients per group to prove an alternative of $\theta^* = 0.8$ with a power of 80% at one-sided significance level $\alpha = 2.5\%$. The data, however, was generated with different nuisance parameters and no effect present. A comparison between the fixed design and an experiment with the re-estimated sample size in terms of type I error is given in Asendorf et al. [6] Table 3. The BSSR does not show any form of type I error inflation in the regarded settings. Although this result is somewhat expected, as sample size re-estimation methods for non-comparative designs generally have no or a limited effect on the type I error rate [21], in certain situations an inflation of the type I error rate can be observed, e.g. in Friede and Kieser [22]. For this reason, the type I error rate always needs to be examined when introducing such designs.

To investigate the effect of the BSSR procedure on the power of a trial, a simulation study was performed for complete observations in which data was simulated with nuisance parameters different to those at the initial planning phase, altering each nuisance parameter one by one. Then, half of the data was taken for a blinded sample size re-estimation and the power was compared between recruiting the re-estimated sample size and continuing with the initially planned sample size. Simulation results are displayed in Figure 2.2.

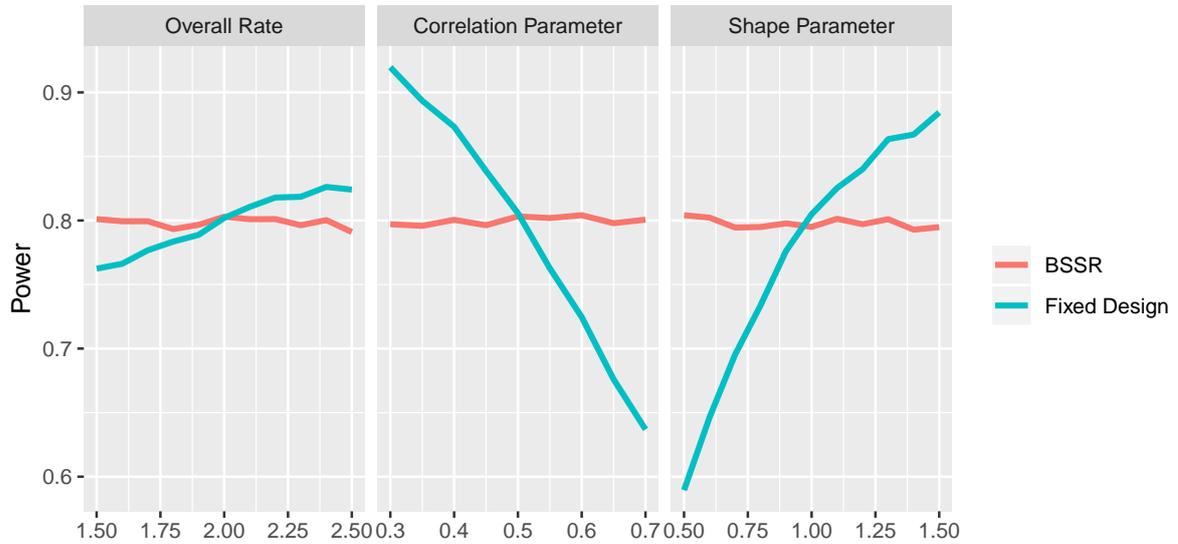


Figure 2.2: Simulated power of BSSR and fixed design with true nuisance parameters (x-axis) altered one by one. Underlying Model: NB-INAR(1). Each parameter combination was simulated 10,000 times.

While the fixed design assumed nuisance parameters $\bar{\lambda} = 2$, $\alpha = 0.5$ and $\eta = 1$, true parameters were simulated to be different one by one, with the overall rate varying between 1.5 and 2.5, the correlation parameter between 0.3 and 0.7 and the shape parameter between 0.5 and 1.5. Simulation results revealed, that the re-estimated sample size achieved the intended power, while the fixed design (with the initially planned sample size) was either over- or underpowered, apart from the situation in which the true parameters and assumed nuisance parameters coincide.

2.1.4 Software Implementation and Availability of Methods

The presented methods have been implemented within the R-package `spass` (study planning and adaptation of sample size [5]) and made available on CRAN. The package contains functions for sample size estimation and blinded sample size re-estimation, as well as a function for generating observations from the NB-INAR(1) model, which may be useful for custom simulations. An overview of the implemented functions is given in Table 2.1.

Table 2.1: Functions implemented in the R-package `spass` for sample size estimation and blinded re-estimation within the NB-INAR(1) model.

Function	Description
<code>rnbinom.inar1</code>	Generate time series with marginal negative binomial distribution and autoregressive correlation structure of order one
<code>fit.nb.inar1</code>	Fitting the NB-INAR(1) model on a given data set
<code>n.nb.inar1</code>	Calculate required initial sample size in planning phase
<code>bssr.nb.inar1</code>	Blinded sample size re-estimation on a given data set
<code>test.nb.inar1</code>	Statistical inference for testing treatment effects

All methods are presented with examples of their usage and explanations of parameters, as required for R-packages, to allow for a simple implementation.

2.2 Gamma Frailty Model

After having explored and used the NB-INAR(1) model for deriving an adaptive design, the aim was to extend the methods for time trends in the group means. The NB-INAR(1) model could not have trends in the rates while at the same time maintaining the property of marginal negative binomial counts, which is why a different model was chosen for this purpose. The gamma frailty model introduced by Henderson and Shimakura [33] and extended by Fiocco et al. [19], to account for arbitrary shape parameters, maintains a marginally negative binomial distributed random variables, while at the same time allowing for arbitrary choices of the marginal means. This advantage over the NB-INAR(1) model, however, comes at the cost of less control over the correlation between time points. Observations within the gamma frailty model are best defined in two steps. First, we assume that for each patient there exists an unobservable multivariate gamma random variable $\mathbf{Z}_{ij} = (Z_{ij}^{(1)}, \dots, Z_{ij}^{(T)})$. In this multivariate random variable, each $Z_{ij}^{(t)}$ follows a gamma distribution with mean one and variance η , i.e. $Z_{ij}^{(t)} \sim \gamma(1, \eta^{-1})$. These within-patient frailties are generated such that they are dependent and it holds that $Cor(Z_{ij}^{(t)}, Z_{ij}^{(s)}) = a^{|s-t|}$. In the second step, we generate the observations $X_{ij}^{(t)}$ by using the gamma frailty terms as arguments within a Poisson distribution. Conditional on the gamma frailty, we assume

$$(2.8) \quad X_{ij}^{(t)} | \mathbf{Z}_{ij}^{(t)} \sim \text{Poisson}(\lambda_i^{(t)} Z_{ij}^{(t)}),$$

from which it follows that unconditionally on the gamma frailty observations $X_{ij}^{(t)} \sim NB(\lambda_i^{(t)}, \eta)$. Similarly to the NB-INAR(1) model we now have a marginal negative binomial distribution, with the difference that the mean of the distribution, i.e. $\lambda_i^{(t)}$, can also depend on the time point t and not only the group affiliation i . However, the covariance of two observations from the same

patients given by

$$(2.9) \quad \text{Cov}(X_{ij}^{(t)}, X_{ij}^{(s)}) = a^{|t-s|} \cdot \frac{\lambda_i^{(t)} \lambda_i^{(s)}}{\eta}$$

is not autoregressive anymore, but merely has an autoregressive appearance inherited from the autoregressive within-patient gamma frailty. This property also limits the maximum possible correlation which can be modeled between observations. For example, assume that $\lambda_i^{(t)} = 2$ and $\eta = 0.5$ for all $t \in \mathbb{N}$. Then, the maximum possible correlation between observations, which can be modeled by setting $a = 1$, is 0.8.

The main advantage of the gamma frailty model is that, because the means can be chosen arbitrarily, we can model time trends occurring within a clinical trial while maintaining the property of marginal negative binomial observations. This allows for different and much more diverse hypotheses to be tested. By defining a set of regression parameters $\beta = (\beta_1, \dots, \beta_d)$ we can view the means $\lambda_i^{(t)}$ as functions of these regression parameters, i.e. $\lambda_i^{(t)} = f_i^{(t)}(\beta)$. For example, if we assume changing rates over time, we could define exponential trends as

$$(2.10) \quad f_E^{(t)}(\beta) = \exp(\beta_1 + (\beta_2 + \beta_3) \cdot t) \quad \text{and} \quad f_C^{(t)}(\beta) = \exp(\beta_1 + \beta_2 \cdot t).$$

This is just one possibility of many thinkable trends, some of which are outlined in more detail within Asendorf et al. [7]. While this specific trend is motivated by observations in MS clinical trials, see e.g. Nicholas et al. [49], a general notation for different trends was maintained throughout the calculations to potentially broaden the application of the presented method.

2.2.1 Statistical Inference

Inference through maximum likelihood is not straightforward within the gamma frailty model. Instead of using standard maximum likelihood theory, we estimate parameters using a pseudo likelihood approach with independent working assumption. Estimates for parameters $\lambda = (\lambda_E^{(1)}, \dots, \lambda_C^{(T)})$ and η are attained by maximizing

$$(2.11) \quad L(\lambda, \eta | \mathbf{x}) = \frac{1}{n_E + n_C} \sum_{i=E,C} \sum_{j=1}^{n_i} L_{ij}(\lambda, \eta | \mathbf{x}_{ij}),$$

where

$$(2.12) \quad L_{ij}(\lambda, \eta | \mathbf{x}_{ij}) = \sum_{t=1}^T \log \left(\mathbb{P}_{\text{NB}} \left(x_{ij}^{(t)}, f_i^{(t)}(\beta), \eta \right) \right)$$

with respect to β and η . Hereby $\mathbf{x} = (x_{E1}^{(1)}, \dots, x_{E1}^{(T)}, \dots, x_{En_E}^{(T)}, \dots, x_{Cn_C}^{(T)})$ and $\mathbf{x}_{ij} = (x_{ij}^{(1)}, \dots, x_{ij}^{(T)})$ denote the observed values. The function $\mathbb{P}_{\text{NB}}(x, \mu, \eta)$ denotes the probability function of a negative binomial random variable with mean μ and variance $\mu + \mu^2/\eta$. The correlation parameter a can be estimated at a second step by maximizing the pairwise composite likelihood defined as

$$(2.13) \quad cl(a|\hat{\eta}, \hat{\beta}, x) = \frac{1}{n_C + n_E} \sum_{i=E, C} \sum_{j=1}^{n_i} \sum_{t=1}^{T-1} \sum_{s=t+1}^T \log \left(\mathbb{P}_{\text{NB}}^{\text{pair}} \left(x_{ij}^{(t)}, x_{ij}^{(s)}; a, f_i^{(t)}(\hat{\beta}), f_i^{(s)}(\hat{\beta}) \right) \right)$$

where $\mathbb{P}_{\text{NB}}^{\text{pair}}(\cdot)$ denotes the bivariate probability function for counts at different times but on the same patient. More details on the pairwise probability function are given in Fiocco et al. [19].

Main interest within the gamma frailty model lies in the regression parameters β_1, \dots, β_d . For example, in the presented trend from Equation (2.10) the main interest lies in testing $H_0 : \beta_3 = 0$ vs. $H_1 : \beta_3 \neq 0$, because β_3 corresponds to the treatment effect. Generally, however, we will be testing the hypothesis

$$(2.14) \quad H_0 : h(\eta, \beta) \geq h_0 \quad \text{vs.} \quad H_1 : h(\eta, \beta) < h_0,$$

where $h : \mathbb{R}^{d+1} \rightarrow \mathbb{R}$ is a twice differentiable and monotone function in all dimensions. Standard asymptotic theory on likelihood estimators (see Asendorf et al. [7] Appendix E for a rigorous derivation) then yields

$$(2.15) \quad \sqrt{n_C + n_E} \cdot (h(\hat{\eta}, \hat{\beta}) - h(\eta, \beta)) \stackrel{\text{approx}}{\sim} N(0, \nabla_{(\eta, \beta)} h(\eta, \beta)^\top H^{-1} J H^{-1} \nabla_{(\eta, \beta)} h(\eta, \beta))$$

for $n_E, n_C \rightarrow \infty$ and $k = n_E/n_C$ constant. The introduced Hessian matrix H and outer gradient matrix J are defined as

$$(2.16) \quad H = \frac{1}{n_C + n_E} \sum_{i=E, C} \sum_{j=1}^{n_i} E \left[\nabla_{(\eta, \beta)(\eta, \beta)} L_{ij}(\lambda, \eta | \mathbf{x}_{ij}) \right]$$

and

$$(2.17) \quad J = \frac{1}{n_C + n_E} \sum_{i=E, C} \sum_{j=1}^{n_i} E \left[\nabla_{(\eta, \lambda)} L_{ij}(\lambda, \eta | \mathbf{x}_{ij}) \cdot \nabla_{(\eta, \lambda)} L_{ij}(\lambda, \eta | \mathbf{x}_{ij})^\top \right].$$

In standard maximum likelihood theory, using the Fisher information, the Hessian and outer gradient matrix would coincide, and the term $H^{-1} J H^{-1}$ would reduce to H^{-1} , i.e. the Fisher information matrix. However, because the pseudo likelihood does not fully describe the data, as observations are not independent, the matrices J and H differ and need to be estimated separately. Estimation of H and J is possible through

$$(2.18) \quad \hat{H} = \frac{1}{n_C + n_E} \sum_{i=E, C} \sum_{j=1}^{n_i} \nabla_{(\eta, \lambda)(\eta, \lambda)} L_{ij}(\hat{\eta}, \hat{\lambda} | \mathbf{x}_{ij})$$

and

$$(2.19) \quad \hat{J} = \frac{1}{n_C + n_E} \sum_{i=E, C} \sum_{j=1}^{n_i} \nabla_{(\eta, \lambda)} L_{ij}(\hat{\eta}, \hat{\lambda} | \mathbf{x}_{ij}) \cdot \nabla_{(\eta, \lambda)} L_{ij}(\hat{\eta}, \hat{\lambda} | \mathbf{x}_{ij})^\top.$$

With these estimators we can define the test statistic Z , which asymptotically follows a standard normal distribution

$$(2.20) \quad Z = \sqrt{n_E + n_C} \cdot \frac{h(\hat{\eta}, \hat{\lambda}) - h_0}{\sqrt{\nabla_{(\eta, \lambda)} h(\hat{\eta}, \hat{\lambda})^\top (\hat{H}^{-1} \hat{J} \hat{H}^{-1}) \nabla_{(\eta, \lambda)} h(\hat{\eta}, \hat{\lambda})}} \stackrel{\text{approx}}{\sim} N(0, 1).$$

It is noteworthy that the asymptotics hold for $n_E, n_C \rightarrow \infty$ and constant ratio of sample sizes $k = n_E/n_C$. In case of large sample sizes ($n_i \geq 100$), a simulation study showed that the approximation controls the type I error rate for a set of realistic settings in MS. However, for small sample sizes, the test statistic showed a simulated type I error rate of 0.03 to 0.04, while the nominal type I error rate was 0.025. To improve the testing procedure, a method was implemented which estimates the variance $\hat{\sigma}^2 = \nabla_{(\eta, \lambda)} h(\hat{\eta}, \hat{\lambda})^\top (H^{-1} J H^{-1}) \nabla_{(\eta, \lambda)} h(\hat{\eta}, \hat{\lambda})$ restricted to the parameter space of the null hypothesis, resulting in an estimate $\hat{\sigma}_r^2$. A simulation study to compare the type I error rates of both procedures was conducted and results are depicted in Figure 2.3.

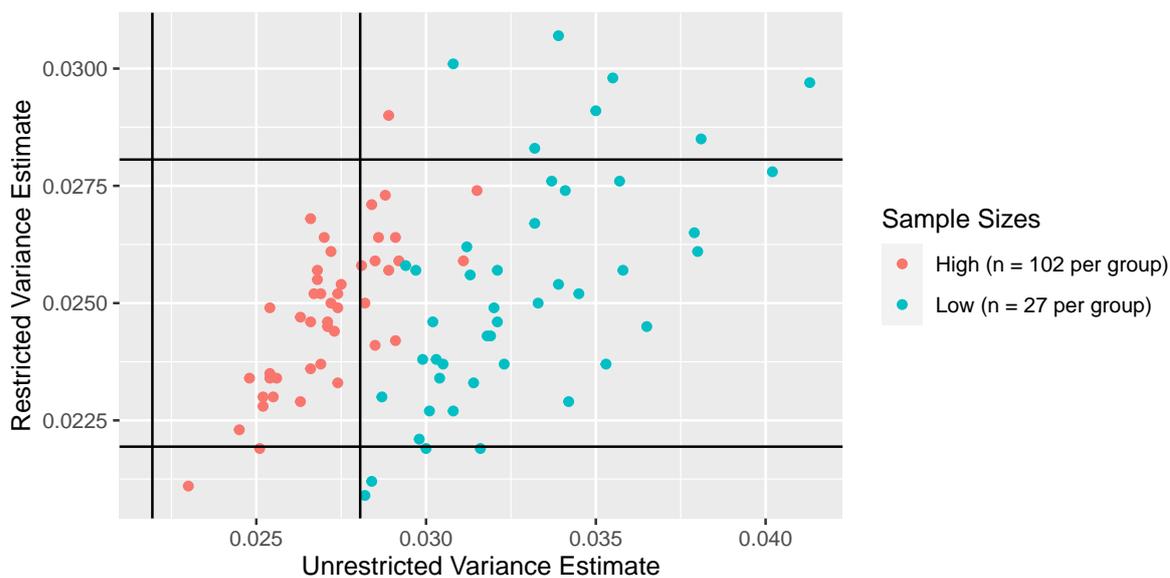


Figure 2.3: Comparison of type I error rates of the test statistic Z (2.20) using restricted and unrestricted variance estimates. Each dot represents one simulation setting. Solid lines represent 95% random intervals of simulated type I error rate.

In Figure 2.3 we can see that the unrestricted variance estimates shows higher type I error rates than the intended significance level of 2.5%, while using the restricted variance estimate controls the type I error rate even for small sample sizes ($n_i = 27$).

2.2.2 Blinded Sample Size Re-estimation

Using the normal approximation of Z , a sample size formula was derived which requires information on the nuisance parameters and effect size under the alternative. Denoting the shape parameter, further nuisance parameters and the effect size under alternative as η^* , β^* and $\theta^* = h(\eta^*, \beta^*) - h_0$, respectively, the sample size required to reject the null hypothesis under alternative to a predefined power $1 - \beta$ and significance level α is approximated by

$$(2.21) \quad n_C = \frac{(z_\beta + z_\alpha)^2 \cdot \sigma^2}{(1+k) \cdot \theta^{*2}}.$$

This sample size formula can be used at the planning phase of a clinical trial and also when re-estimating the sample size. Estimating the nuisance parameters from blinded data was achieved by assuming the blinded data to be from a mixture distribution of two negative binomial distributions. Similar to the inference procedure, estimation of the nuisance parameters is done in two steps. First, nuisance parameters β_1, \dots, β_d are estimated by maximizing the mixture likelihood

$$(2.22) \quad L_{\text{blind}}(\eta, \beta | \mathbf{x}) = \frac{1}{n_E + n_C} \sum_{j=1}^{n_E + n_C} \sum_{t=1}^T \ln \left(\frac{k}{1+k} \cdot \mathbb{P}_{\text{NB}} \left(x_j^{(t)}, f_E^{(t)}(\beta), \eta \right) + \frac{1}{1+k} \cdot \mathbb{P}_{\text{NB}} \left(x_j^{(t)}, f_C^{(t)}(\beta), \eta \right) \right)$$

assuming the effect θ^* from the alternative hypothesis being present. In a second step, the correlation parameter a is estimated through a mixture distribution of the pairwise likelihood function, i.e. by maximizing

$$(2.23) \quad cl_{\text{blind}}(a | \hat{\eta}, \hat{\beta}, \mathbf{x}) = \frac{1}{n_C + n_E} \sum_{j=1}^{n_C + n_E} \sum_{t=1}^{T-1} \sum_{s=t+1}^T \ln \left(\frac{k}{k+1} \cdot \mathbb{P}_{\text{NB}}^{\text{pair}} \left(x_j^{(t)}, x_j^{(s)}; a, f_E^{(t)}(\hat{\beta}), f_E^{(s)}(\hat{\beta}), \hat{\eta} \right) + \frac{1}{1+k} \cdot \mathbb{P}_{\text{NB}}^{\text{pair}} \left(x_j^{(t)}, x_j^{(s)}; a, f_C^{(t)}(\hat{\beta}), f_C^{(s)}(\hat{\beta}), \hat{\eta} \right) \right)$$

where estimates $\hat{\beta}$ and $\hat{\eta}$ are plugged in from the estimation in the first step. The resulting nuisance parameter estimates are then plugged into the sample size formula from Equation (2.21) to attain a re-estimated sample size.

2.2.3 Numerical Evaluation of Type I Error Rate and Power

Operational characteristics were investigated for numerous settings realistic for lesion counts in MS, assuming constant means (i.e. the same setting as investigated by the NB-INAR(1) model) and the exponential trend described in Equation (2.10). An extensive type I error simulation (Figure 3 and Figure 4 from Asendorf et al. [7]) over numerous parameters (Table 3 from Asendorf et al. [7]) revealed no type I error inflation induced by the blinded sample size re-estimation

procedure. For the constant trend, additionally to investigating the type I error for superiority tests, the type I error for non-inferiority tests was also examined and equally revealed no type I error inflation attributable to the BSSR.

Power Evaluation of the BSSR procedure revealed, that the BSSR is able to correct the power in the case of misspecified nuisance parameters. For better readability we first summarize the results for constant rates and then the results for an underlying exponential trend.

2.2.3.1 Constant Rates

For the sample size of the fixed design, parameters were assumed to be $\beta_1 = 0$, $\eta = 1$ and $\rho = 0.5$ with an effect size of $\beta_2 = -0.3$, resulting in a sample size at planning stage of $n_E = n_C = 102$ to attain a power of 80% when testing one-sided at 2.5% significance level. Similar to prior power simulations, the data was simulated with different nuisance parameters to those assumed, with nuisance parameters altered sequentially as displayed in Figure 2.4. The intercept rate was altered between -1.0 and 1.0 , the correlation parameter between 0.0 and 1.0 , and the shape parameter between 0.5 and 1.5 .

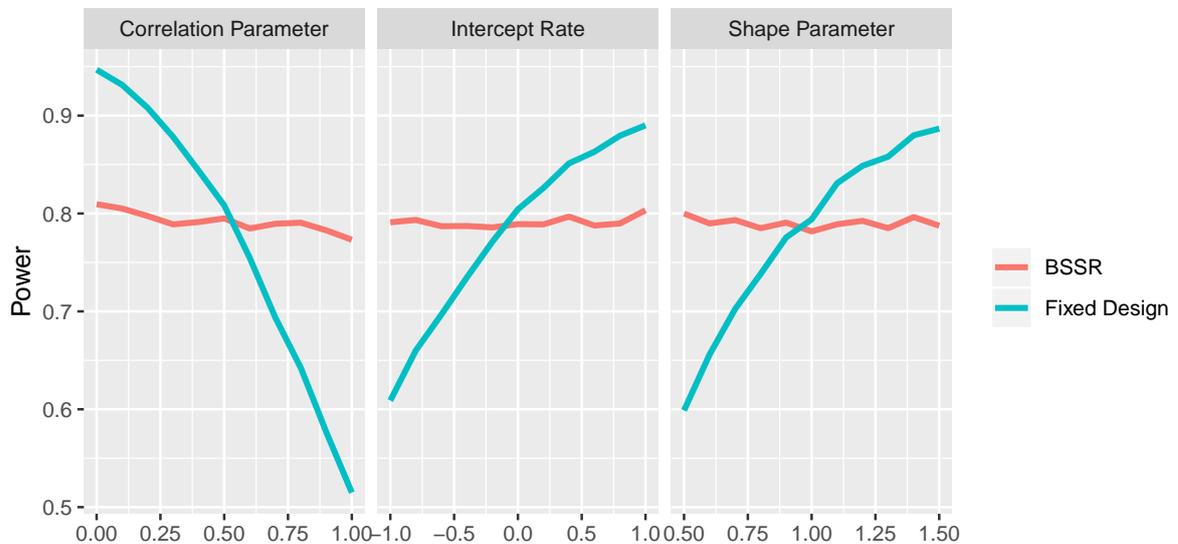


Figure 2.4: Simulated power of BSSR and fixed design with true nuisance parameters (x-axis) altered one by one. Underlying Model: Gamma Frailty, assuming constant rates. Every setting was simulated 10,000 times.

From Figure 2.4 we can conclude that the blinded sample size re-estimation procedure is able to correct the sample size under misspecified nuisance parameters accordingly, to reach the pre-specified power of 80%.

2.2.3.2 Exponential Trends in Event Rates

In the simulation inspecting the blinded sample size re-estimation procedure with an underlying exponential trend, parameters for the fixed sample size were assumed to be $\beta_1 = 0$, $\beta_2 = 0$, $\eta = 1$ and $\rho = 0.5$, resulting in a required sample size of $n_C = n_E = 229$ for rejecting the null hypothesis with 80% power at 2.5% one-sided significance level, assuming an effect of $\beta_3 = 0.05$. Data were simulated with the true treatment effect being present, but nuisance parameters were altered sequentially. The power comparison between keeping to the initial sample size and switching to the re-estimated sample size are given in Figure 2.5.

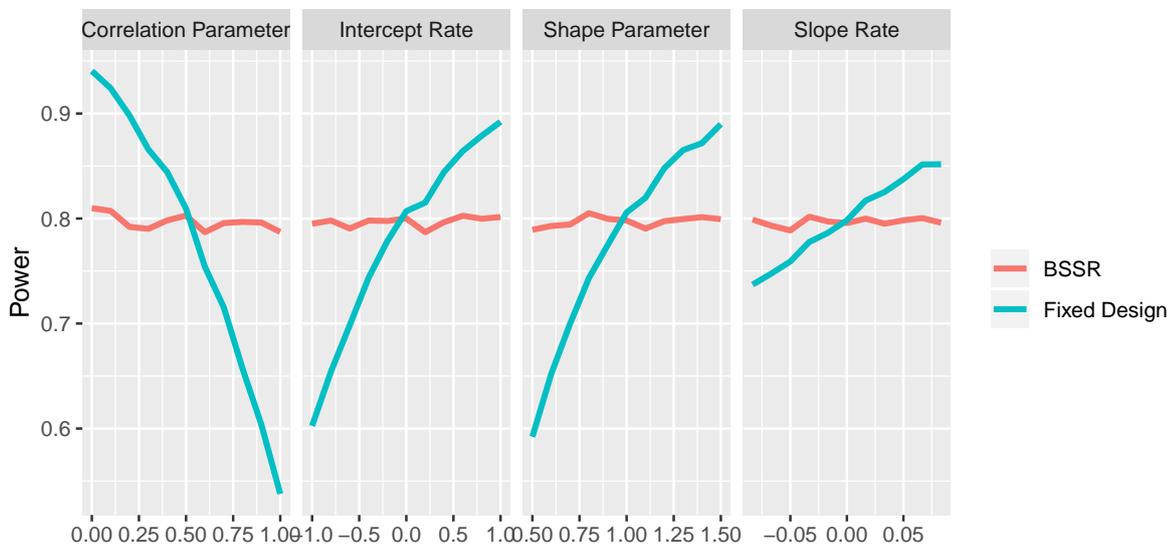


Figure 2.5: Simulated power of BSSR and fixed design with true nuisance parameters altered one by one. Underlying Model: Gamma Frailty, assuming exponential trend. Every setting was simulated 10,000 times.

In Figure 2.5 we see that the blinded sample size re-estimation procedure is able to correct the sample size such that the targeted power is reached. Thus, we conclude, that the blinded sample size re-estimation is capable of correcting the sample size even in the presence of time trends. The general notation allows for the implementation of flexible time trends as required, broadening the potential applications.

2.2.4 Software Implementation and Availability of Methods

The presented methods for constant and exponential trends have been implemented within the R-package `spass` and made available on CRAN. The package contains functions for sample size estimation and blinded sample size re-estimation, as well as a function for generating

observations from the gamma frailty model. An overview of the implemented functions is given in Table 2.2.

Table 2.2: Functions implemented in the R-package *spass* for sample size estimation and blinded re-estimation within the gamma frailty model.

Function	Description
<code>rnbinom.gf</code>	Generate time series with marginal negative binomial distribution and unobservable multivariate gamma frailty
<code>fit.nb.gf</code>	Fitting the gamma frailty model on a given data set
<code>n.nb.gf</code>	Calculate required initial sample size in planning phase
<code>bssr.nb.gf</code>	Blinded sample size re-estimation on a given data set during trial
<code>test.nb.gf</code>	Statistical inference for testing treatment effects

All methods are presented with examples of their usage and explanations of parameters. Furthermore, input and output of these functions is kept similar to the functions written for the INAR(1) to allow for model comparisons.

2.3 Negative Binomial Outcomes with Covariates

So far, the presented methods modeled a treatment and possible time effect on the outcomes. However, treatment effects can also be influenced by further sources, such as baseline observations or center effects. These effects are usually incorporated by modeling observations conditional on the covariates. In Zapf et al. [64] two approaches were developed which allow for sample size estimation and blinded sample size re-estimation of univariate negative binomial counts with baseline covariates. The statistical model is as follows. Assume we have only one time point, e.g. the last observation of the trial, and observations of two groups, denoted by X_{ij} , with $i = E, C$ and $j = 1, \dots, n_i$. Further we observe a covariate Y_{ij} (or a set of covariates) for each observation X_{ij} . Then, the conditional model is given by

$$(2.24) \quad X_{ij}|Y_{ij} \sim NB(\lambda_{ij}, \eta)$$

where λ_{ij} and η are fixed parameters. Similarly to the gamma frailty model, the mean λ_{ij} can be defined as a function of regression parameters, usually expressed through a log-link, e.g. for the case of one covariate $\lambda_{ij} = \exp(\beta_1 + \beta_2 \cdot \mathbb{1}_{\{i=E\}} + \beta_3 \cdot y_{ij})$, where $\mathbb{1}_{\{\cdot\}}$ denotes the indicator function.

2.3.1 Statistical Inference

For two groups, main interest lies in testing the null hypothesis $H_0 : \mathbf{c}^\top \beta = \delta$ vs $H_1 : \mathbf{c}^\top \beta \neq \delta$, where \mathbf{c} is usually chosen to reflect the treatment effect, i.e. $\mathbf{c}^\top \beta = \beta_2$ referring to the example

above. To test this null hypothesis, two different methods were used, namely the likelihood ratio test statistic and a Wald-type test based on maximum likelihood estimators. Denoting the observed Fisher information conditional on the covariates as $\mathbf{I}_\beta(\hat{\beta}, \hat{\eta}|\mathbf{y})$, where $\mathbf{y} = (\mathbf{y}_{E1}, \dots, \mathbf{y}_{Cn_C})$ denotes the collection of all covariates and $\mathbf{y}_{ij} = (y_1, \dots, y_p)$ the covariates of an individual, the Wald-type test statistic is given by

$$(2.25) \quad T_W = n \frac{(\mathbf{c}^\top \beta - \delta)^2}{\mathbf{c}^\top \mathbf{I}_\beta(\hat{\beta}, \hat{\eta}|\mathbf{y})^{-1} \mathbf{c}}.$$

Furthermore, the employed likelihood ratio test statistic is given by

$$(2.26) \quad T_{LR} = -2 \log \left(\frac{\sup_{\Omega_0} L(\beta, \eta | x_{E1}, x_{E2}, \dots, x_{Cn_C}, \mathbf{y}_{E1}, \dots, \mathbf{y}_{Cn_C})}{\sup_{\Omega} L(\beta, \eta | x_{E1}, x_{E2}, \dots, x_{Cn_C}, \mathbf{y}_{E1}, \dots, \mathbf{y}_{Cn_C})} \right).$$

Hereby, $L(\cdot, \cdot | x_{E1}, x_{E2}, \dots, x_{Cn_C}, \mathbf{y}_{E1}, \dots, \mathbf{y}_{Cn_C})$ denotes the likelihood function of the negative binomial regression model. It can be shown that under the null hypothesis, $T_W \stackrel{\text{approx}}{\sim} N(0, 1)$ and $T_{LR} \stackrel{\text{approx}}{\sim} \chi_{1, 1-\alpha}^2$ as $n_E \rightarrow \infty$ and $n_C \rightarrow \infty$, assuming the ratio n_E/n_C to be constant. Both test statistics are commonly used in practice with the Wald-Test being computationally simpler and the likelihood ratio test having some advantages when computing confidence intervals, as these are scale invariant and only contain valid values.

2.3.2 Sample Size Estimation

The difficulty when deriving a sample size formula for these test statistics lies in incorporating the baseline observations. Inference is performed conditional on the baseline observations, however, as these are not known at the beginning of a trial, the initial sample size calculation needs to be performed unconditionally of the observed covariates. To achieve this, Lyles et al. [44] proposed a method which creates an artificial data set, also referred to as expanded data set, which depends on the distribution of the covariates and outcomes, imitating the unknown covariates. The observed Fisher Information of the expanded data set can be used for attaining an estimate of the variance of the effect size, denoted by $\widehat{\text{Var}}_{ED}$, which in turn is required to calculate the power for specific total sample size $n = n_E + n_C$ using the formula

$$(2.27) \quad \mathbb{P} \left(\chi_1^2(\hat{\lambda}_{ED}) \geq \chi_{1, 1-\alpha}^2 \right),$$

where $\hat{\lambda}_{ED} = n(\delta - \theta_1)^2 / \widehat{\text{Var}}_{ED}(\hat{\theta})$ and $\chi_1^2(\hat{\lambda}_{ED})$ denotes a chi-squared distributed random variable with one degree of freedom and non-centrality parameter $\hat{\lambda}_{ED}$. By using an iterative method, such as interval halving, a total sample size n (at fixed allocation ratio $k = n_E/n_C$) is calculated which fulfills the power requirement. Although this method of sample size estimation was shown

to be adequate, the iterative approach was computationally expensive, consequently limiting simulation runs. Therefore, a second approach for calculating the sample size was considered, which is based on the Wald-test. When testing with the Wald-test, the required total sample size for attaining a specific conditional power under a given alternative $\theta^* = \mathbf{c}^\top \beta^* - \delta$ and significance level α can be approximated by

$$(2.28) \quad n = \frac{(q_{1-\alpha/2} + q_{1-\beta})^2}{\theta^{*2}} \mathbf{c}^\top \mathbf{I}_\beta(\beta^*, \eta^* | \mathbf{y})^{-1} \mathbf{c}.$$

In Equation (2.28), the Fisher information is calculated conditional on the covariates. As these are not known at the planning stage of a clinical trial, the conditional Fisher information is replaced by the expected Fisher information, defined as $\mathbb{E}_\mathbf{Y}(\mathbf{I}_\beta(\beta^*, \eta^* | \mathbf{Y}))$. This approach gives an approximation of the sample size required for attaining the unconditional power. More details as to why this is not an exact approach are given in Section 3.3.2 in Zapf et al. [64]. However, the presented heuristic was shown to adequately calculate the required sample size in numerical evaluations.

2.3.3 Blinded Sample Size Re-estimation

Extending the methodology for blinded sample size re-estimation is possible by estimating nuisance parameters from the blinded data using a mixture approach. The likelihood of a mixture of to negative binomial distribution, now incorporating covariates, is given by

$$(2.29) \quad \sum_{j=1}^n \ln \left(\frac{k}{k+1} \cdot \mathbb{P}_{\text{NB}} \left(x_j, \exp((y_{i1}, 1, y_{i3}, \dots, y_{ip}) \beta(\theta^*)) \cdot \frac{t_i}{T}, \eta \frac{t_i}{T} \right) + \frac{1}{k+1} \cdot \mathbb{P}_{\text{NB}} \left(x_j, \exp((y_{i1}, 0, y_{i3}, \dots, y_{ip}) \beta(\theta^*)) \cdot \frac{t_i}{T}, \eta \frac{t_i}{T} \right) \right).$$

Within Equation (2.29), the parameter $\beta(\theta^*)$ denotes the regression parameters β with the effect size under alternative θ^* fixed. For example, if the null hypothesis is $H_0 : \beta_2 = 0$ and $\lambda_{ij} = \exp(\beta_1 + \beta_2 \cdot \mathbb{1}_{\{i=E\}} + \beta_3 \cdot y_{ij})$ then $\beta(\theta^*) = (\beta_1, \beta_2^*, \beta_3)^\top$. Blinded estimates for parameters β and η resulting from estimation with the likelihood in Equation (2.29) are then plugged into the sample size formulas from Equation (2.28) to attain a re-estimated sample size. Under circumstances, it is also necessary to estimate parameters specific to the distribution of the covariates. Because we assume that patients were randomized to one of the two groups, parameters corresponding to the covariate distribution can be estimated by pooling together all observations, regardless of their group assignment.

2.3.4 Numerical Evaluation of Type I Error Rate and Power

Both methods for blinded sample size re-estimation were evaluated in an extensive simulation. Furthermore, for the expanded data set approach, the likelihood ratio approach and the Wald-

type test statistic were compared. The type I error was evaluated for one normally distributed covariate, assuming an intercept log rate of $\beta_1 = 0$ a log rate ratio of $\theta^* = \beta_2 = -0.2$, a covariate effect of $\beta_3 = 2.5$ and shape parameter of $\eta = 3$. Under these assumptions, the required sample size for rejecting the null hypothesis $H_0 : \beta_2 = 0$ at 80% power and 5% two-sided significance level is equal to approximately 380 patient per group. The chosen parameters were comparable to those observed in a clinical trial in epilepsy by Leppik et al. [41]. For the simulation, true values of parameters were chosen differently to those of the assumed as displayed in Table 2.3.

Table 2.3: Simulation settings for the comparison of type I error rates between the fixed design and the blinded sample size re-estimation procedures for univariate negative binomial counts with covariates.

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

The data was simulated in multiple steps, to best resemble realistic scenarios. First, 25% of the initial sample size was simulated as complete observations. Second, another 25% of the initial sample size was simulated, but with observation times uniformly distributed between 0 and T , i.e. $t_i \stackrel{i.i.d.}{\sim} U(0, T)$, where T denotes the total follow-up time per patient. The combined data set from the first and second step was taken to perform a sample size re-estimation. Inference was then performed with sample sizes $n_{\text{WT-FI}}$, $n_{\text{WT-ED}}$ and $n_{\text{LR-ED}}$ for the fixed design and with re-estimated sample sizes $\hat{n}_{\text{WT-FI}}$, $\hat{n}_{\text{WT-ED}}$ and $\hat{n}_{\text{LR-ED}}$ for the blinded sample size re-estimation design. Hereby, WT-FI refers to the method using the expected Fisher information plugged into the sample size formula from Equation (2.28), LR-ED to the method using the expanded data set from Lyles et al. [44] and performing inference with the likelihood ratio test, and WT-ED to the method using the expanded data set and the Wald-type test statistic for inference. Type I error rates of all three sample size estimation and re-estimation methods are displayed in Figure 2.6.

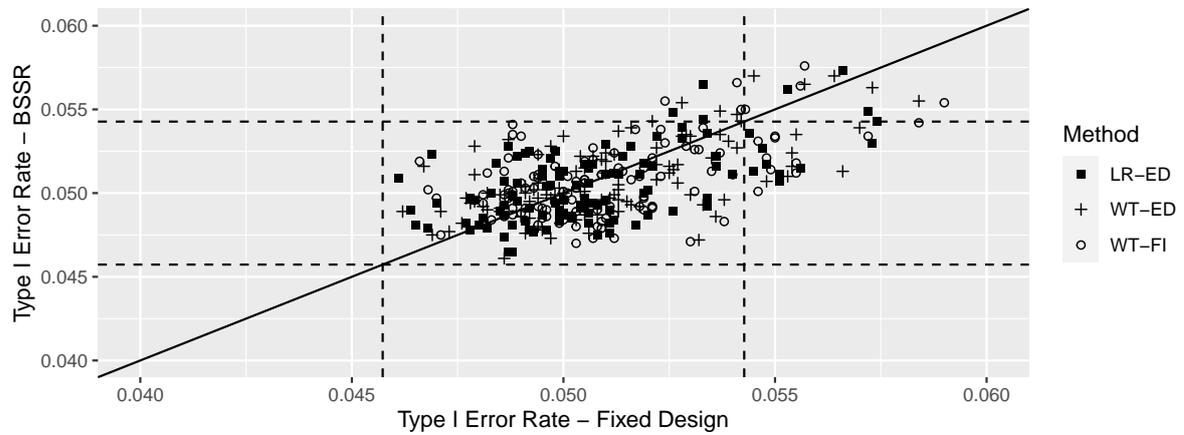


Figure 2.6: Type I error rates of BSSR and fixed design for all possible combinations of parameters as listed in Table 2.3 while under the null hypothesis. Dashed lines represent 95%-random intervals for simulated type I error rates. Every setting was simulated 10,000 times.

The type I error simulation did not reveal any notable differences in type I error rates between the fixed design and the blinded sample size re-estimation procedure for all three procedures. Additionally to the type I error, the power was also examined in an extensive simulation. For this, the same simulation settings as in Table 2.3 were considered, with the only difference being that the true treatment effect was chosen to equal the assumed treatment effect. Figure 2.7 displays the results from the power simulation for the WT-FI method.

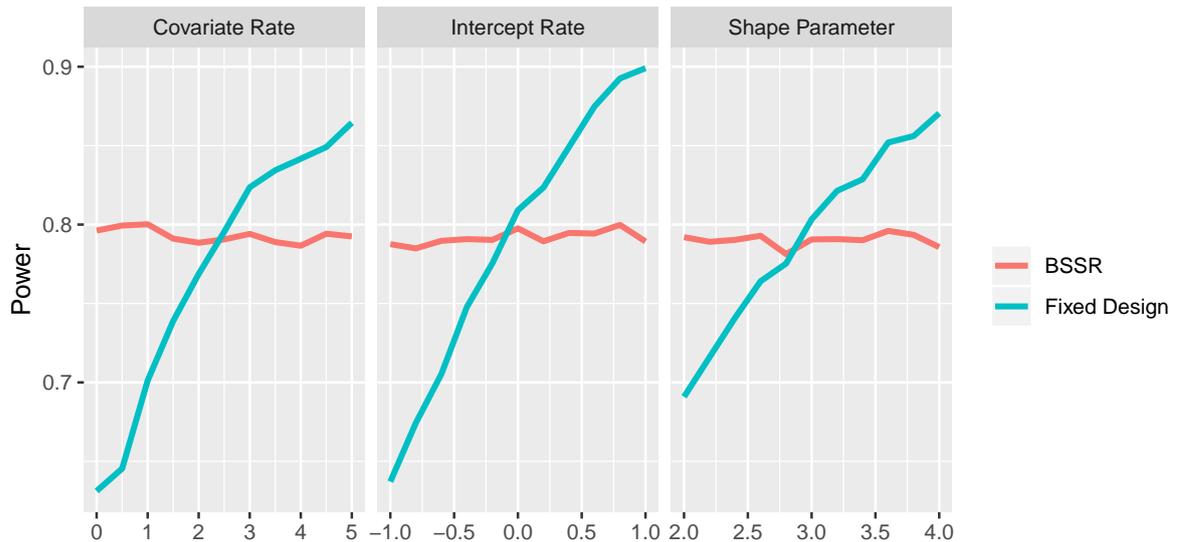


Figure 2.7: Power simulation of BSSR and fixed design for the WT-FI method, with true nuisance parameters altered sequentially. Underlying Model: Univariate Negative Binomial Counts with Covariates. Every setting was simulated 10,000 times.

For the power simulation, true nuisance parameters were altered one by one and the influence of the blinded sample size re-estimation procedure examined. In case of a misspecified nuisance parameters, the blinded sample size re-estimation was capable of correcting the required sample size to a level required for attaining the pre-specified power. Results of the WT-ED and LR-ED methods were comparable and are reported in the Appendix of Zapf et al. [64].

2.3.5 Software Implementation and Availability of Methods

The presented methods for sample size estimation and blinded re-estimation with baseline covariates have been implemented in R and made available as supplementary material to the publication on figshare under <https://doi.org/10.6084/m9.figshare.11973579>. An overview of the implemented functions is given in Table 2.4.

Table 2.4: Functions uploaded as supplementary material within an R-script for sample size estimation and blinded re-estimation with baseline covariates.

Function	Description
<code>wt.fi</code>	Initial sample size calculation for the WT-FI method
<code>wt.ed</code>	Initial sample size calculation for the WT-ED method
<code>lr.ed</code>	Initial sample size calculation for the LR-ED method
<code>sse</code>	Initial sample size calculation for all three presented methods
<code>bssr</code>	Blinded sample size re-estimation for all three presented methods

Additionally to these R-functions, examples are given to demonstrate their usage and parameters are explained in detail. Functions for inference are not provided as inference in this case is possible using standard negative binomial regression.

DISCUSSION

The summarized research extends the available methodology for blinded sample size re-estimation for longitudinal negative binomial counts and univariate negative binomial counts with baseline covariates. Methods have been shown to not inflate type I error rates, therefore fulfilling regulatory requirements. In the presence of the expected treatment effect, but misspecified nuisance parameters, the methods were able to correct the required sample size to attain the necessary power for the final analysis.

For longitudinal modeling of negative binomial counts two different models were used, the NB-INAR(1) model and the gamma frailty model. It was pointed out, that the gamma frailty model allows for arbitrary means at different time points, therefore allowing for trends within clinical trials. However, this came at the cost of limiting the modeling correlation between observations. Specifically, the correlation between observations within the NB-INAR(1) model can lie between 0 and 1, while the maximum correlation in the gamma frailty model depends on the mean and shape parameter at a given time point. This may not pose a problem in the presence of intermediate correlation, however, neither model can cope with trends and high correlation between time points.

For statistical inference, the NB-INAR(1) model used method of moment estimators as an analytical derivation of the Fisher information was not feasible for a likelihood approach. However, a numerical approach may be possible and beneficial. Furthermore, from a practical point of view, data following an NB-INAR(1) or gamma frailty model, can not be analysed using standard negative binomial regression as implemented in common statistical software. To support statisticians interested in applying these methods, R-functions were written and made available. However, these are limited to the cases as described within the publications and extensions, for example to further trends, are not as easily implemented.

While missing data due to the recruitment scheme was addressed in all three methods, other types of missing data were not fully addressed. The procedures as described here can still be employed for situations where data are missing completely at random or missing at random [58]. There exists literature on coping with such missingness [4, 45], however, the described methods have not been implemented and need to be discussed in light of the ICH E9 Addendum on Estimands and Sensitivity Analysis in Clinical Trials [30].

One crucial point, which is frequently discussed, is the timing of interim analyses at which a blinded sample size re-estimation is to be performed. While the methods do not give an answer to an optimal time point, it was shown within Asendorf et al. [6], that conducting the blinded sample size re-estimation at later time points yields less variable sample size estimates. Therefore, it is desirable to conduct a blinded sample size re-estimation at the latest possible time point which allows for a continuation of the recruitment. However, this may bear the risk that the sample size at interim may already be larger than required. There have been numerous proposals considering this trade off between low variability of the sample size estimate and risking a too large clinical trial, with a summary of these given in Section 6 of Friede and Schmidli [25].

Another difficulty when working with time trends in running clinical trials, is defining the type of underlying trend. The presented methods allow for a very flexible trend definition. However, assessing the correct underlying trend for a clinical trial is a difficult task, especially in the given setting where little data is available before the clinical trial. Furthermore, trends with small effects require a high number of observations to be detected. A possible solution would be to also allow for switching the underlying trend depending on the goodness of a model fit during a running trial. However, this would also require a switch of the null hypothesis, which might influence type I error rates. This aspect has so far not been explored but may pose a substantial improvement in practice.

Adaptive designs have generally received increased attention over the past decades in all phases of drug development [59], with the pace of uptake in clinical trials, however, staying well behind that of statistical literature [50]. Among others, contributing factors are speculated to be unfamiliarity with advantages and limitations of these procedures within the scientific community, as well as inadequate implementation of methods [50]. This supports the importance of not only thoroughly explaining the used methods and their limits, but also to provide an implementation of these, such that the proposed methods may find their way into clinical trials.

BIBLIOGRAPHY

- [1] M. AL-OSH AND A. ALZAID, *First-order integer-valued autoregressive (INAR(1)) process.*, J. Time Ser. Anal., 8 (1987), pp. 261–275.
- [2] R. M. ALTMAN AND A. J. PETKAU, *Application of hidden Markov models to multiple sclerosis lesion count data*, Stat Med, 24 (2005), pp. 2335–2344.
- [3] R. M. ALTMAN, A. J. PETKAU, D. VRECKO, AND A. SMITH, *A longitudinal model for magnetic resonance imaging lesion count data in multiple sclerosis patients*, Stat Med, 31 (2012), pp. 449–469.
- [4] J. ANDERSSON AND D. KARLIS, *Treating missing values in INAR(1) models: an application to syndromic surveillance data.*, J. Time Ser. Anal., 31 (2010), pp. 12–19.
- [5] T. ASENDORF, R. GERA, S. ISLAM, M. HARDEN, AND M. PLACZEK, *spass: Study Planning and Adaptation of Sample Size*, 2020.
R package version 1.3.
- [6] T. ASENDORF, R. HENDERSON, H. SCHMIDLI, AND T. FRIEDE, *Modelling and sample size reestimation for longitudinal count data with incomplete follow up*, Statistical Methods in Medical Research, 28 (2019), pp. 117–133.
- [7] T. ASENDORF, R. HENDERSON, H. SCHMIDLI, AND T. FRIEDE, *Sample size re-estimation for clinical trials with longitudinal negative binomial counts including time trends*, Statistics in Medicine, 38 (2019), pp. 1503–1528.
- [8] F. BARKHOF, H. E. HULST, J. DRULOVIC, B. M. UITDEHAAG, K. MATSUDA, AND R. LANDIN, *Ibudilast in relapsing-remitting multiple sclerosis: a neuroprotectant?*, Neurology, 74 (2010), pp. 1033–1040.
- [9] P. BAUER AND K. KOHNE, *Evaluation of experiments with adaptive interim analyses*, Biometrics, 50 (1994), pp. 1029–1041.
- [10] A. CANALE AND D. DUNSON, *Nonparametric bayes modelling of count processes*, Biometrika, 100 (2013), pp. 801–816.

BIBLIOGRAPHY

- [11] J. CHATAWAY, N. SCHUERER, A. ALSANOUSI, D. CHAN, D. MACMANUS, ET AL., *Effect of high-dose simvastatin on brain atrophy and disability in secondary progressive multiple sclerosis (MS-STAT): a randomised, placebo-controlled, phase 2 trial*, *Lancet*, 383 (2014), pp. 2213–2221.
- [12] S.-C. CHOW AND M. CHANG, *Adaptive Design Methods in Clinical Trials*, CRC Press, Taylor and Francis Group, New York, second ed., 2011.
- [13] C. CHUANG-STEIN, K. ANDERSON, P. GALLO, AND S. COLLINS, *Sample size reestimation: A review and recommendations*, *Drug Information Journal*, 40 (2006), pp. 475–484.
- [14] J. COHEN AND R. RUDICK, *Multiple sclerosis therapeutics, fourth edition*, Cambridge University Press, 2011.
- [15] R. J. COOK, P.-J. BERGERON, J.-M. BOHER, AND Y. LIU, *Two-stage design of clinical trials involving recurrent events*, *Statistics in Medicine*, 28 (2009), pp. 2617–2638.
- [16] D. COX AND H. MILLER, *The theory of stochastic processes.*, Methuen and Co. Ltd., London, 1965.
- [17] L. CUI, H. M. HUNG, AND S. J. WANG, *Modification of sample size in group sequential clinical trials*, *Biometrics*, 55 (1999), pp. 853–857.
- [18] J. S. DENNE, *Sample size recalculation using conditional power*, *Statistics in Medicine*, 20 (2001), pp. 2645–2660.
- [19] M. FIOCCO, H. PUTTER, AND J. C. VAN HOUWELINGEN, *A new serially correlated gamma-frailty process for longitudinal count data*, *Biostatistics*, 10 (2009), pp. 245–57.
- [20] L. D. FISHER, *Self-designing clinical trials*, *Statistics in Medicine*, 17 (1998), pp. 1551–1562.
- [21] FOOD AND DRUG ADMINISTRATION (FDA), *Adaptive designs for clinical trials of drugs and biologics, guidance for industry*.
<https://www.fda.gov/regulatory-information/search-fda-guidance-documents/adaptive-design-clinical-trials-drugs-and-biologics-guidance-industry>, 2019.
Accessed: 2020-12-21.
- [22] T. FRIEDE AND M. KIESER, *Blinded sample size reassessment in non-inferiority and equivalence trials*, *Statistics in Medicine*, 22 (2003), pp. 995–1007.
- [23] T. FRIEDE AND M. KIESER, *Blinded sample size recalculation for clinical trials with normal data and baseline adjusted analysis*, *Pharm Stat*, 10 (2011), pp. 8–13.

-
- [24] T. FRIEDE, C. MITCHELL, AND G. MULLER-VELTEN, *Blinded sample size reestimation in non-inferiority trials with binary endpoints*, *Biom J*, 49 (2007), pp. 903–916.
- [25] T. FRIEDE AND H. SCHMIDLI, *Blinded sample size reestimation with count data: methods and applications in multiple sclerosis*, *Stat Med*, 29 (2010), pp. 1145–1156.
- [26] T. FRIEDE AND H. SCHMIDLI, *Blinded sample size reestimation with negative binomial counts in superiority and non-inferiority trials*, *Methods Inf Med*, 49 (2010), pp. 618–624.
- [27] C. GENEST AND J. NESLEHOVA, *A primer on copulas for count data*, *Astin Bulletin*, 37 (2007), pp. 475–515.
- [28] G. GIOVANNONI, G. COMI, S. COOK, K. RAMMOHAN, P. RIECKMANN, ET AL., *A placebo-controlled trial of oral cladribine for relapsing multiple sclerosis*, *N. Engl. J. Med.*, 362 (2010), pp. 416–426.
- [29] A. L. GOULD AND W. J. SHIH, *Sample size re-estimation without unblinding for normally distributed outcomes with unknown variance*, *Communications in Statistics - Theory and Methods*, 21 (1992), pp. 2833–2853.
- [30] I. E. W. GROUP, *Addendum on estimands and sensitivity analysis in clinical trials*.
https://database.ich.org/sites/default/files/E9-R1_Step4_Guideline_2019_1203.pdf, 2019.
Accessed: 2020-08-14.
- [31] G. GRUNWALD, S. BRUCE, L. JIANG, M. STRAND, AND N. RABINOVITCH, *A statistical model for under- or overdispersed clustered and longitudinal count data*, *Biometrical Journal*, 53 (2011), pp. 578–594.
- [32] P. HALL, H. MUELLER, AND F. YAO, *Modelling sparse generalised longitudinal observations with latent gaussian processes*, *Journal of the Royal Statistical Society, Series B*, 70 (2008), pp. 703–723.
- [33] R. HENDERSON AND S. SHIMAKURA, *A serially correlated gamma frailty model for longitudinal count data.*, *Biometrika*, 90 (2003), pp. 355–366.
- [34] P. JACOBS AND P. LEWIS, *Discrete time series generated by mixtures. I: Correlational and runs properties.*, *J. R. Stat. Soc., Ser. B*, 40 (1978), pp. 94–105.
- [35] P. A. JACOBS AND P. A. LEWIS, *Discrete time series generated by mixtures II: Asymptotic properties.*, *J. R. Stat. Soc., Ser. B*, 40 (1978), pp. 222–228.
- [36] V. JOWAHEER AND B. SUTRADHAR, *Analysing longitudinal count data with overdispersion*, *Biometrika*, 89 (2002), pp. 389–399.

BIBLIOGRAPHY

- [37] L. KAPPOS, E. W. RADUE, P. O’CONNOR, C. POLMAN, R. HOHLFELD, ET AL., *A placebo-controlled trial of oral fingolimod in relapsing multiple sclerosis*, N. Engl. J. Med., 362 (2010), pp. 387–401.
- [38] O. KHAN, P. RIECKMANN, A. BOYKO, K. SELMAJ, R. ZIVADINOV, ET AL., *Three times weekly glatiramer acetate in relapsing-remitting multiple sclerosis*, Ann. Neurol., 73 (2013), pp. 705–713.
- [39] W. LEHMACHER AND G. WASSMER, *Adaptive sample size calculations in group sequential trials*, Biometrics, 55 (1999), pp. 1286–1290.
- [40] R. V. LENTH, *Some practical guidelines for effective sample size determination*, The American Statistician, 55 (2001), pp. 187–193.
- [41] I. LEPPIK, F. DREIFUSS, T. BOWMAN, N. SANTILLI, M. JACOBS, ET AL., *A Double-blind Crossover Evaluation of Progabide in Partial Seizures*, Neurology, 35 (1985), p. 285.
- [42] F. D. LUBLIN AND S. C. REINGOLD, *Defining the clinical course of multiple sclerosis: results of an international survey*, Neurology, 46 (1996), pp. 907–911.
- [43] F. D. LUBLIN, S. C. REINGOLD, J. A. COHEN, G. R. CUTTER, ET AL., *Defining the clinical course of multiple sclerosis*, Neurology, 83 (2014), pp. 278–286.
- [44] R. H. LYLES, H.-M. LIN, AND J. M. WILLIAMSON, *A practical approach to computing power for generalized linear models with nominal, count, or ordinal responses*, Statistics in Medicine, 26 (2007), pp. 1632–1648.
- [45] B. MCCABE AND G. MARTIN, *Bayesian predictions of low count time series*, International Journal of Forecasting, 21 (2005), pp. 315 – 330.
- [46] E. MCKENZIE, *Autoregressive moving-average processes with negative-binomial and geometric marginal distributions.*, Adv. Appl. Probab., 18 (1986), pp. 679–705.
- [47] D. MILLER, F. BARKHOF, X. MONTALBAN, A. THOMPSON, AND M. FILIPPI, *Clinically isolated syndromes suggestive of multiple sclerosis, part I: natural history, pathogenesis, diagnosis, and prognosis*, Lancet Neurol, 4 (2005), pp. 281–288.
- [48] H. H. MULLER AND H. SCHAFER, *Adaptive group sequential designs for clinical trials: combining the advantages of adaptive and of classical group sequential approaches*, Biometrics, 57 (2001), pp. 886–891.
- [49] R. NICHOLAS, S. STRAUBE, H. SCHMIDLI, S. PFEIFFER, AND T. FRIEDE, *Time-patterns of annualized relapse rates in randomized placebo-controlled clinical trials in relapsing multiple sclerosis: a systematic review and meta-analysis*, Mult. Scler., 18 (2012), pp. 1290–1296.

-
- [50] P. PALLMANN, A. W. BEDDING, B. CHOODARI-OSKOOEI, M. DIMAIRO, L. FLIGHT, ET AL., *Adaptive designs in clinical trials: why use them, and how to run and report them*, BMC Medicine, 16 (2018), p. 29.
- [51] Y. L. PRITCHETT, S. MENON, O. MARCHENKO, Z. ANTONIJEVIC, E. MILLER, ET AL., *Sample size re-estimation designs in confirmatory clinical trials, A current state, statistical considerations, and practical guidance*, Statistics in Biopharmaceutical Research, 7 (2015), pp. 309–321.
- [52] M. A. PROSCHAN, *Two-stage sample size re-estimation based on a nuisance parameter: A review*, Journal of Biopharmaceutical Statistics, 15 (2005), pp. 559–574.
- [53] M. A. PROSCHAN, *Sample size re-estimation in clinical trials*, Biometrical Journal, 51 (2009), pp. 348–357.
- [54] M. A. PROSCHAN AND S. A. HUNSBERGER, *Designed extension of studies based on conditional power*, Biometrics, 51 (1995), pp. 1315–1324.
- [55] R CORE TEAM, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2019.
- [56] S. SCHNEIDER, H. SCHMIDLI, AND T. FRIEDE, *Blinded sample size re-estimation for recurrent event data with time trends*, Statistics in Medicine, 32 (2013), pp. 5448–5457.
- [57] S. SCHNEIDER, H. SCHMIDLI, AND T. FRIEDE, *Robustness of methods for blinded sample size re-estimation with overdispersed count data*, Statistics in Medicine, 32 (2013), pp. 3623–3635.
- [58] S. SEAMAN, J. GALATI, D. JACKSON, AND J. CARLIN, *What is meant by 'missing at random'?*, Statist. Sci., 28 (2013), pp. 257–268.
- [59] N. STALLARD, L. HAMPSON, N. BENDA, W. BRANNATH, T. BURNETT, T. FRIEDE, ET AL., *Efficient adaptive designs for clinical trials of interventions for covid-19*, Statistics in Biopharmaceutical Research, 0 (2020), pp. 1–15.
- [60] F. STEUTEL AND K. VAN HARN, *Discrete analogues of self-decomposability and stability.*, Ann. Probab., 7 (1979), pp. 893–899.
- [61] N. TUBRIDY, H. J. ADER, F. BARKHOF, A. J. THOMPSON, AND D. H. MILLER, *Exploratory treatment trials in multiple sclerosis using MRI: sample size calculations for relapsing-remitting and secondary progressive subgroups using placebo controlled parallel groups*, J. Neurol. Neurosurg. Psychiatr., 64 (1998), pp. 50–55.
- [62] C. E. VAN MUNSTER AND B. M. UITDEHAAG, *Outcome Measures in Clinical Trials for Multiple Sclerosis*, CNS Drugs, 31 (2017), pp. 217–236.

BIBLIOGRAPHY

- [63] M. T. WALLIN, W. J. CULPEPPER, E. NICHOLS, Z. A. BHUTTA, T. T. GEBREHIWOT, ET AL., *Global, regional, and national burden of multiple sclerosis 1990-2016: a systematic analysis for the global burden of disease study 2016*, *The Lancet Neurology*, 18 (2019), pp. 269–285.
- [64] A. ZAPF, T. ASENDORF, C. ANTEN, T. MÜTZE, AND T. FRIEDE, *Blinded sample size reestimation for negative binomial regression with baseline adjustment*, *Statistics in Medicine*, 39 (2020), pp. 1980–1998.
- [65] G. ZIMMERMANN, M. KIESER, AND A. C. BATHKE, *Sample size calculation and blinded recalculation for analysis of covariance models with multiple random covariates*, *Journal of Biopharmaceutical Statistics*, 0 (2019), pp. 1–17.