

# Blinded Sample Size Reestimation with Negative Binomial Counts in Superiority and Non-inferiority Trials

T. Friede<sup>1</sup>; H. Schmidli<sup>2</sup>

<sup>1</sup>Abteilung Medizinische Statistik, Universitätsmedizin Göttingen, Göttingen, Germany;

<sup>2</sup>Statistical Methodology, Novartis Pharma AG, Basel, Switzerland

## Keywords

Internal pilot study, sample size, clinical trial, chronic obstructive pulmonary disease, COPD

## Summary

**Background:** In the planning of clinical trials with count outcomes such as the number of exacerbations in chronic obstructive pulmonary disease (COPD) often considerable uncertainty exists with regard to the overall event rate and the level of overdispersion which are both crucial for sample size calculations.

**Objectives:** To develop a sample size reestimation strategy that maintains the blinding of the trial, controls the type I error rate and is robust against misspecification of the nuisance parameters in the planning phase in that the actual power is close to the target.

**Methods:** The operation characteristics of the developed sample size reestimation procedure are investigated in a Monte Carlo simulation study.

**Results:** Estimators of the overall event rate and the overdispersion parameter that do not require unblinding can be used to effectively adjust the sample size without inflating the type I error rate while providing power values close to the target.

**Conclusions:** If only little information is available regarding the size of the overall event rate and the overdispersion parameter in the design phase of a trial, we recommend the use of a design with sample size reestimation as the one suggested here. Trials in COPD are expected to benefit from the proposed sample size reestimation strategy.

## Correspondence to:

Tim Friede  
Abteilung Medizinische Statistik  
Universitätsmedizin Göttingen  
Humboldtallee 32  
37073 Göttingen  
Germany  
E-mail: tim.friede@med.uni-goettingen.de

Methods Inf Med 2010; 49: 618–624

doi: 10.3414/ME09-02-0060

received: December 17, 2009

accepted: May 11, 2010

prepublished: August 5, 2010

## 1. Introduction

In trials with event counts as primary outcomes the sample sizes depend on the overall event rate and the level of overdispersion found in such data. Specification of these quantities for the purpose of sample size calculation in the planning phase of a trial is often difficult since only limited information is available from previous studies or trial characteristics such as the study population or the definition of events vary from trial to trial. As a result the sample size calculation is subject to some uncertainty that can lead to inappropriately sized trials.

Designs with so-called internal pilot study aim to minimize this uncertainty by reestimating the required sample size based on interim estimates of the nuisance parameters (e.g. overall event rate, level of overdispersion) [1]. Such designs with blinded sample size reestimation (BSSR) have been proposed for a variety of situations including various types of endpoints (e.g. continuous, binary) and objectives (superiority, non-inferiority). We refer to recent reviews on this subject for an overview of the relevant literature [2, 3].

Very recently Cook et al. [4] and Friede and Schmidli [5] proposed blinded sample

size reestimation procedures for overdispersed count data. Whereas Cook et al. suggested procedures based on a likelihood approach assuming negative binomial counts and utilizing EM algorithms, Friede and Schmidli developed a procedure for Poisson counts which they extended to overdispersed counts by applying quasi-likelihood ideas ([6], Section 6.2.3). The procedure by Friede and Schmidli is computationally much less demanding than the EM algorithms suggested by Cook et al., but does not utilize all information under sampling from negative binomial distributions. Here we develop a likelihood approach for blinded sample size reestimation for negative binomial distributed counts while maintaining the simplicity of the previously proposed quasi-likelihood procedure. Furthermore, we extend the procedures to non-inferiority trials which have not been considered for count data.

The paper is organized as follows. In Section 2 we describe trials in chronic obstructive pulmonary disease (COPD) that motivated the investigations reported here. Previously proposed models, analyses and sample size calculation are briefly reviewed in Section 3 before a procedure for blinded sample size reestimation is proposed and its characteristics are explored by means of a simulation study in Section 4. In Section 5 the approach is extended to non-inferiority trials and properties of the procedure are investigated in a simulation study. We close by summarizing the conclusions.

## 2. Clinical Trials with Exacerbation Endpoints in COPD

Chronic obstructive pulmonary disease is a lung disease that affects about 210 million

people worldwide [7]. COPD exacerbations can be life-threatening, and hence drugs that prevent exacerbations will provide meaningful benefit to patients. In clinical trials that aim to show prevention of exacerbations, the primary efficacy endpoint will usually be the exacerbation rate, and the duration of the studies will be at least one year [8]. In such clinical trials, the number of exacerbations and the follow-up time is recorded for each patient. Due to the heterogeneity of patients, the number of exacerbations cannot be assumed to follow a Poisson distribution, and overdispersion has to be taken into account for valid statistical inference [9, 10].

The negative binomial distribution is often used to model overdispersed count data, and appears to describe COPD exacerbation counts appropriately [11]. A detailed statistical analysis of exacerbation counts in a large randomized clinical trial was provided by [12]. Using a negative binomial model, the exacerbation rates were estimated as 1.71 events per year ( $N = 361$ ) in placebo, and 1.16 events per year ( $N = 356$ ) in the combination treatment arm. The exacerbation rate ratio (combination/placebo) was estimated as 0.68 with a 95% confidence interval of 0.57 to 0.83. The shape parameter of the negative binomial distribution was estimated as  $\phi = 0.46$ . This corresponds to an overdispersion (variance/mean) of 1.8 and 1.5 for the placebo and combination treatment group, respectively.

Sample size calculations for clinical trials with overdispersed count data require specification of two nuisance parameters: the event rate in the control group, and the overdispersion parameter. Incorrect assumptions on the nuisance parameters can lead to under- or over-powered studies. This is illustrated by two recent examples of randomized two-group superiority trials in COPD patients, both with a treatment duration of 12 months. In a trial with 782 COPD patients [13], the primary endpoint was the exacerbation rate, and a negative binomial model was used for the primary analysis. For sample size calculations, event rates in the control group were assumed to be 1.9 events per year, and overdispersion (variance/mean) was assumed to be 1.5. However, the observed event rate in the

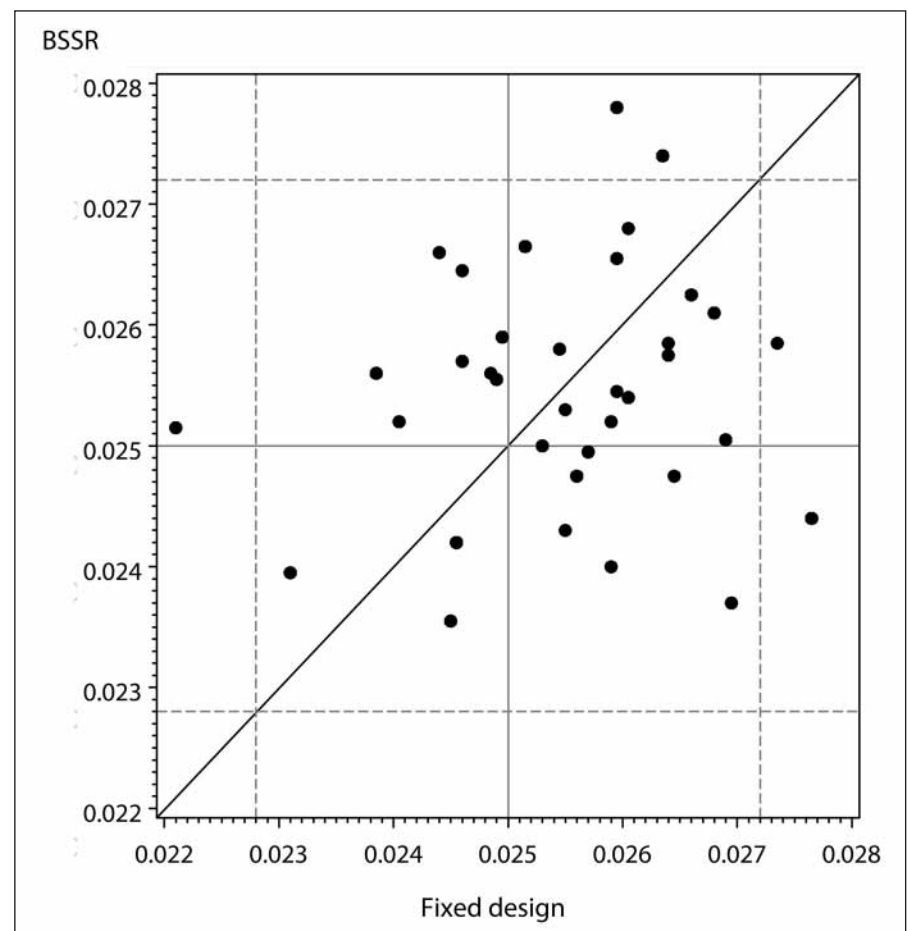
trial was only 1.5 events per year (no information on the actual overdispersion was provided). In another randomized clinical trial in 797 COPD patients [14] with a very similar study design and analysis, the event rate in the control group was assumed to be 1.9 events, while the observed event rate at the end of the trial was 1.6 events per year (no details on overdispersion provided). Recruitment duration in the two studies was 13 and 18 months, respectively.

As treatments for the prevention of COPD exacerbations are approved, an active comparator rather than placebo may be used as a control group, with the aim to show non-inferiority. Such trial designs in COPD are possible, when a well-defined, reproducible treatment effect for the established comparator is available [8]. One important design aspect in such trials is the choice of the non-inferiority margin. De-

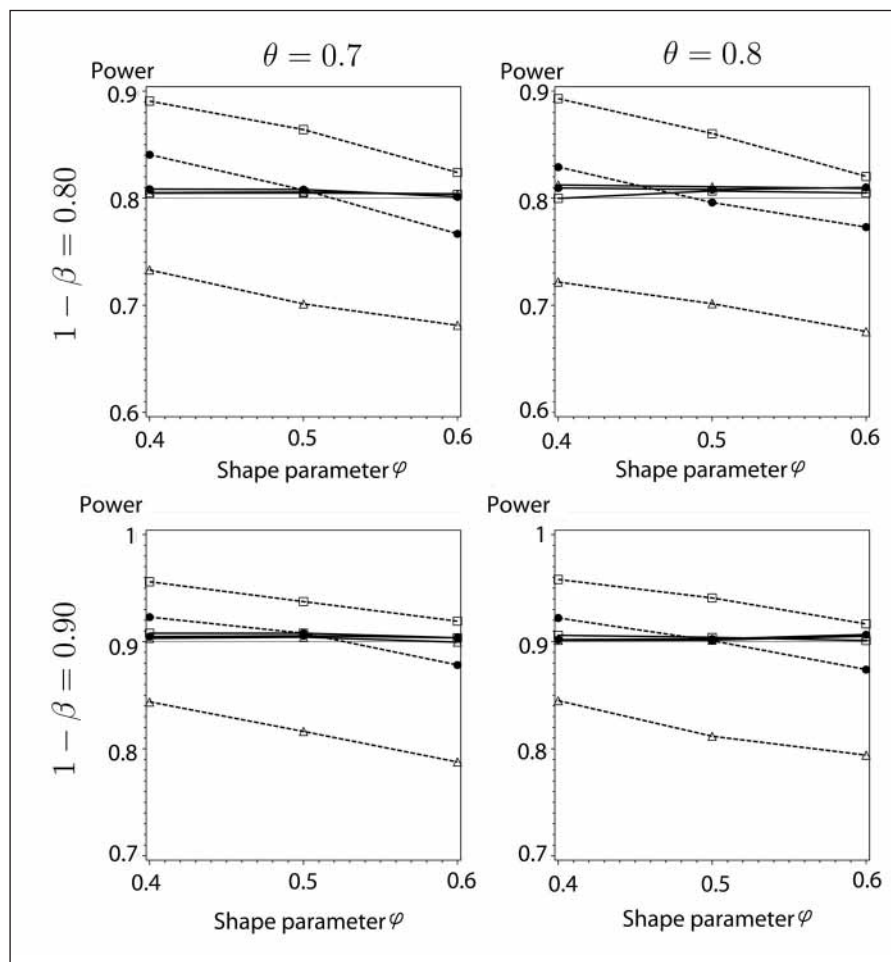
fining a non-inferiority margin is complex, involving both medical and statistical issues. For illustrative purposes below, we will consider a non-inferiority margin of 15%. This may possibly be an adequate choice when the combination treatment discussed in [12] is chosen as the active control: the observed reduction compared to placebo was 32%, i.e. more than double the non-inferiority margin, and also the upper 95% confidence limit was 18%, i.e. larger than the non-inferiority margin.

### 3. Statistical Model, Analysis and Sample Size Calculation

We consider a clinical trial in COPD patients where patients are randomly allo-



**Fig. 1** Type I error rate of the blinded sample size reestimation design (BSSR) against the type I error rates of the fixed sample size designs for a total of 36 scenarios (see text for details). The dashed reference lines indicate the area within two simulation standard errors from  $\alpha = 0.025$ .



**Fig. 2** Power in blinded sample size reestimation design (solid lines) depending on the shape parameter for overall event rates  $\lambda = 1$  ( $\Delta$ ),  $\lambda = 1.5$  ( $\bullet$ ), and  $\lambda = 2$  ( $\square$ ). For comparison the power in the fixed sample size design (dashed lines) with sample sizes per group of 147 (196) and 370 (496) for  $\theta = 0.7$  and  $\theta = 0.8$  and target power  $1 - \beta = 0.80$  (0.90) respectively are also shown.

cated to a treatment and a control group, with an allocation ratio of  $k:1$ . For each patient  $i$ , the number of exacerbations  $y_i$  and the follow-up time  $t_i$  is available at the end of the trial. It is assumed that  $y_i$  is an observation from a negative binomial distribution with event rate per time unit  $\lambda_1$  (treatment) or  $\lambda_0$  (control), respectively, and common positive shape parameter  $\varphi$ . The mean (expected value) of the number of exacerbations for patient  $i$  is then given by  $\mu_{m(i)} = t_i \lambda_{m(i)}$  and its variance is given by  $\mu_{m(i)} (1 + \varphi \mu_{m(i)})$ , where  $m(i) = 1$  if patient  $i$  was allocated to the treatment group, and  $m(i) = 0$  otherwise. The overdispersion defined as the ratio of variance to mean is given by  $(1 + \varphi \mu_{m(i)})$ , and approaches 1 as the shape parameter  $\varphi$  approaches 0, cor-

responding to the Poisson distribution. The maximum likelihood (ML) estimates of the unknown parameters  $(\lambda_1, \lambda_0, \varphi)$  can be obtained by an iterative algorithm (Newton-Raphson), using standard generalized linear model software with a logarithmic link function, and the logarithm of the follow-up time as an offset variable ([15], chapter 5).

Denoting the rate ratio by  $\theta = \lambda_1/\lambda_0$ , interest is in testing the null hypothesis  $H_0: \theta = 1$  against the one-sided alternative  $H_1: \theta < 1$ . A testing procedure (Wald test) at one-sided significance level  $\alpha$  is obtained by first deriving the one-sided upper  $(1 - \alpha)$  Wald confidence limit  $\hat{U}$  for the rate ratio  $\theta$  from the ML estimates, based on the asymptotic normal distribution of the

logarithm of the rate estimates. If  $\hat{U}$  is less than 1, then the null hypothesis is rejected.

For the balanced case ( $k = 1$ ), Keene et al. [12] provided an approximate formula to determine the required sample size to achieve a power of  $1 - \beta$  for a given alternative rate ratio  $\theta^*$ . For the unbalanced case, the required sample size in the control group is given by:

$$n_0 = \frac{\{(z_{1-\beta} + z_{1-\alpha})/\log(\theta^*)\}^2 \times \{(\lambda_0 + k\lambda_1^*)/(k\lambda_0\lambda_1^*) + \varphi(1 + 1/k)\}}{\quad} \quad (1)$$

where  $z_\gamma$  denotes the  $\gamma$  quantile of the standard normal distribution, and  $\lambda_1^* = \lambda_0 \theta^*$ . The sample size formula is based on a normal approximation of the logarithm of the rate estimates, and use of the delta method. It is assumed that all patients have the same follow-up time  $t_i = 1$ .

## 4. Blinded Sample Size Reestimation

The sample size formula depends on two nuisance parameters: the event rate in the control group  $\lambda_0$  and the shape parameter  $\varphi$ . When planning the sample size of a clinical trial, the values of these parameters have to be specified, which may be difficult as seen in two recent large COPD trials [13, 14]. The sample size formula (1) can be rewritten as a function of the overall event rate  $\lambda$  and the shape parameter  $\varphi$  as follows:

$$n_0 = \frac{\{(z_{1-\beta} + z_{1-\alpha})/\log(\theta^*)\}^2 \times \{1/\lambda \times (1 + k\theta^*)^2/\theta^*k(1+k) + \varphi(1 + 1/k)\}}{\quad} \quad (2)$$

This can be derived by noting that  $\lambda = (\lambda_0 + k\lambda_1)/(k + 1)$ , and, for an assumed alternative rate ratio  $\theta^*$ , that  $\lambda_0 = \lambda(1 + k)/(1 + k\theta^*)$  and  $\lambda_1 = \lambda\theta^*(1 + k)/(1 + k\theta^*)$ . Blinded sample size reestimation estimates the nuisance parameters  $\lambda$  and  $\varphi$  based on blinded data from the ongoing trial, and uses this information to adjust the sample size. In a blinded interim review, a negative binomial distribution is fitted to the available count data of the pooled treatment groups, providing an estimate of the overall event rate  $\lambda = (\lambda_0 + k\lambda_1)/(k + 1)$ , and of the shape parameter  $\varphi$ .

These estimates are then plugged into the sample size formula (2) above, providing a reestimated sample size.

In a simulation study we investigated type I error rate, power and sample size distribution of the blinded sample size reestimation procedure over a range of scenarios motivated by COPD trials as described in Section 2 above. For comparison we also computed the type I error rate and power for the fixed sample size design. In both designs an initial sample size calculation was carried out based on assumed parameter values. Whereas this constituted the final sample size for the fixed design, the sample size was adjusted in the design with blinded sample size reestimation based on data of half of the initially calculated sample size. All 72 combinations of the following parameter values were considered: assumed overall event rate  $\lambda^* = 1.5$ ; true overall event rate  $\lambda = 1, 1.5, 2$ ; assumed shape pa-

rameter  $\varphi^* = 0.5$ ; shape parameter  $\varphi = 0.4, 0.5, 0.6$ ; assumed rate ratio  $\theta^* = 0.7, 0.8$ ; true rate ratio  $\theta = \theta^*$  (under the alternative) or  $\theta = 1$  (under the null hypothesis). Furthermore, the designs were balanced, i.e.  $k = 1$ , and the significance level and target power were the usual  $\alpha = 0.025$  (one-sided) and  $1 - \beta = 0.80, 0.90$ , respectively. The fixed design sample sizes are 147 (196) and 370 (496) patients per group and the sample sizes per group of the internal pilot study are 74 (98) and 185 (248) for rate ratio  $\theta = 0.7$  and  $0.8$ , respectively, and for  $1 - \beta = 0.80$  (0.90). For each scenario under the alternative 10,000 trials were simulated whereas 20,000 trials were simulated for each scenario under the null hypothesis. The simulation standard errors are therefore 0.001 and 0.004 (0.003) assuming probabilities 0.025 and 0.80 (0.90) under the null hypothesis and alternative, respectively.

► Figure 1 shows the type I error rates of the blinded sample size reestimation procedure against the type I error rates of the fixed sample size design for the 36 scenarios considered under the null hypothesis. The type I error rates of the fixed design are within two simulation standard errors of  $\alpha = 0.025$  for all but three scenarios and the type I error rates of the BSSR procedure are within these margins for all but two scenarios. The type I error rates for the fixed and the BSSR design are scattered around the diagonal indicating no inflation of the type I error rate due to the interim sample size adjustment.

The simulated power values are displayed in ► Figure 2. Whereas the power of the sample size reestimation procedure is close to the target of 80% or 90% independent of the overall event rate and the shape parameter (level of overdispersion), the power of the fixed sample size design is only

**Table 1** Power and distribution of the sample size in the design with blinded sample size reestimation depending on the rate ratio  $\theta$ , shape parameter  $\varphi$ , overall event rate  $\lambda$  and target power  $1 - \beta$ . The sample size distribution is summarized by mean, standard deviation (SD) and 95% percentile (95% P.). For comparison the sample sizes  $n_0$  required in a fixed design are given.

Scenario			$1 - \beta = 0.80$					$1 - \beta = 0.90$				
			$n_0$	Power	Sample size distribution			$n_0$	Power	Sample size distribution		
$\theta$	$\varphi$	$\lambda$	$n_0$		Mean	SD	95% P.	$n_0$		Mean	SD	95% P.
0.7	0.4	1	177	0.805	182.3	27.3	230.0	237	0.903	244.2	31.6	299.0
		1.5	134	0.809	139.2	19.3	173.0	180	0.905	186.5	22.5	226.0
		2	113	0.805	117.5	15.6	145.0	151	0.908	157.7	18.4	190.0
	0.5	1	189	0.805	195.0	29.6	247.0	253	0.904	260.9	34.0	321.0
		1.5	147	0.808	151.8	21.3	189.0	196	0.906	203.2	24.7	246.0
		2	125	0.806	130.0	17.5	160.0	168	0.908	174.3	20.3	210.0
	0.6	1	201	0.804	207.5	31.2	262.0	270	0.900	277.8	35.9	341.0
		1.5	159	0.801	164.0	23.0	204.0	213	0.904	220.1	26.8	267.0
		2	138	0.804	142.5	19.2	176.0	184	0.904	191.1	22.1	230.0
0.8	0.4	1	445	0.813	450.5	41.9	522.0	596	0.901	603.8	48.8	687.0
		1.5	339	0.810	343.7	30.1	395.0	454	0.902	460.8	34.9	520.0
		2	286	0.800	290.1	24.5	331.0	382	0.906	388.8	28.4	437.0
	0.5	1	477	0.811	482.6	45.5	560.0	638	0.902	646.6	52.7	737.0
		1.5	370	0.808	375.4	33.2	432.0	496	0.903	503.1	38.4	569.0
		2	317	0.807	321.9	27.4	368.0	425	0.904	431.3	31.8	486.0
	0.6	1	508	0.809	514.1	48.7	597.0	681	0.905	689.0	56.3	785.0
		1.5	402	0.810	406.7	36.1	470.0	538	0.907	545.6	41.8	617.0
		2	349	0.805	353.3	30.0	404.0	467	0.901	473.8	35.1	534.5

**Table 2** Type I error rates and sample size distributions under the null hypothesis in non-inferiority trials with blinded sample size reestimation depending on the shape parameter  $\varphi$ , the overall event rate  $\lambda$  and the noninferiority margin  $\delta$ . The sample size distribution is summarized by mean, standard deviation (SD) and 95% percentile (95% P.). For comparison the sample sizes  $n_0$  required in a fixed design are given.

Scenario				Type I error rate	Sample size distribution		
$\delta$	$\varphi$	$\lambda$	$n_0$		Mean	SD	95% P.
1.15	0.4	1	1125	0.0253	1130.6	66.9	1244
		1.5	857	0.0255	862.0	48.5	944
		2	723	0.0269	728.0	39.3	794
	0.5	1	1205	0.0227	1211.2	71.9	1334
		1.5	938	0.0241	942.6	52.8	1033
		2	804	0.0243	808.4	43.7	882
	0.6	1	1286	0.0235	1291.4	77.9	1423.5
		1.5	1018	0.0259	1023.3	57.9	1122
		2	884	0.0259	888.8	48.4	970
1.2	0.4	1	661	0.0259	666.9	51.1	754
		1.5	504	0.0273	508.9	36.8	572
		2	425	0.0257	429.7	29.9	480.5
	0.5	1	708	0.0246	714.3	55.3	810
		1.5	551	0.0259	556.2	40.3	625
		2	472	0.0260	477.0	33.5	534
	0.6	1	756	0.0236	762.0	59.2	862
		1.5	598	0.0248	603.8	44.0	678
		2	519	0.0260	524.4	36.8	587

close to the target if the assumed values of the nuisance parameters are close to the true values. Otherwise the power of the fixed designs deviates at times substantially from the target with some power values being as low as 68% (79%) for a target of 80% (90%). The corresponding sample size distributions are summarized in ► Table 1. The average sample sizes are about five patients higher than the sample sizes per group required in fixed sample size designs based on formula (1) if the true parameter values were known in planning phase. Hence, even compared with the unrealistic situation where the true nuisance parameter values are known, there is only a very small price to pay for the robustness achieved by the BSSR design. The difference of five patients per group is similar to findings reported in [16] considering normally distributed data where it could be shown that the excess in total sample size is

about  $(z_{1-\beta} + z_{1-\alpha})^2/2$  when the true treatment effect is equal to the effect assumed in the sample size calculation. For binomial data an inequality for the deviation of the expected sample size from the sample size required in fixed designs could be specified [17]. Derivations of this kind would be far more complex for negative binomial data where no explicit formulae for the MLE exist. Therefore we rely here on simulations only.

For illustration, we consider how a blinded sample size review could be implemented in a COPD trial. Suppose that a two-arm superiority trial is planned in COPD patients, with the number of exacerbations observed within 12 months as the primary endpoint. An initial total sample size of  $N = 900$  is chosen to provide 90% power to detect a 20% reduction in the exacerbation rate of the experimental versus the control arm, assuming an exacerbation

rate of 2.0 in the control arm (i.e. overall event rate of 1.8), and a shape parameter of 0.5. A blinded sample size review is planned shortly before the end of the recruitment, 18 months after the start of the study, based on a recruitment rate of 50 patients per month. At this timepoint, 12-month data from 300 patients is available. Suppose that an overall event rate of 1.7 and a shape parameter of 0.6 are observed in the blinded review. Assuming these values for the nuisance parameters, approximately  $N = 1,000$  patients would be required to maintain 90% power, i.e. 100 additional patients would be recruited increasing the trial duration by about two months.

## 5. Extension of the BSSR Procedure to Non-inferiority Trials

In a non-inferiority clinical trial in COPD patients, the aim is to demonstrate non-inferiority of an investigational treatment to an established active control treatment. Considering the same setup as for the superiority trials described above, one is now interested in testing the null hypothesis  $H_0: \theta \geq \delta$  against the one-sided alternative  $H_1: \theta < \delta$ , where  $\delta$  is the specified positive non-inferiority margin. A Wald test can again be used for hypothesis testing. If the one-sided upper  $(1 - \alpha)$  Wald confidence limit  $\hat{U}$  for the rate ratio  $\theta$  is less than  $\delta$ , then the null hypothesis is rejected.

Using the same methodology as for superiority trials, an approximate formula for the required sample size in the active control group to achieve a power of  $1 - \beta$  for a given alternative rate ratio  $\theta^*$  is given by:

$$n_0 = \frac{\{(z_{1-\beta} + z_{1-\alpha})/\log(\theta^*/\delta)\}^2 \times \{1/\lambda \times (1 + k\theta^*)^2/\theta^*k(1+k) + \varphi(1+1/k)\}}{\quad} \quad (3)$$

In non-inferiority trials it is very often assumed, that the rate ratio  $\theta^* = 1$ , i.e.  $\lambda = \lambda_0 = \lambda_1$ . Then the required sample size for the control group is simply given by:

$$n_0 = \frac{\{(z_{1-\beta} + z_{1-\alpha})/\log(\delta)\}^2 \times (1 + 1/k) \times (1/\lambda + \varphi)}{\quad} \quad (4)$$

The blinded interim review of a non-inferiority COPD trial is carried out in the same way as for a superiority trial, however using the sample size formula (3) or (4) instead of (2) to reestimate the sample size.

In a simulation study we computed the rejection probabilities and sample size distributions in the BSSR design described above for 18 scenarios each under the null and alternative hypothesis. The same combinations of true and assumed nuisance parameter values as in Section 4 were considered, i.e. assumed overall event rate  $\lambda^* = 1.5$ ; true overall event rate  $\lambda = 1, 1.5, 2$ ; assumed shape parameter  $\varphi^* = 0.5$ ; true shape parameter  $\varphi = 0.4, 0.5, 0.6$ . The non-inferiority margin was set to  $\delta = 1.15$  as motivated in Section 2 for a specific COPD trial and to  $\delta = 1.20$  to study the characteristics of the reestimation procedure with smaller sample sizes. The treatments were assumed to be equally efficacious ( $\theta^* = 1$ ). The hypothesis tests were carried out at a one-sided significance level of  $\alpha = 0.025$  and a power of  $1 - \beta = 0.80$  was targeted in a balanced design ( $k = 1$ ). According to formula (4) the required sample size in a fixed sample size design for the scenario based on the assumed parameter values (i.e.  $\lambda^* = 1.5$  and  $\varphi^* = 0.5$ ) is 938 (551) patients per group for  $\delta = 1.15$  (1.20). Blinded sample size reviews were carried out with half of these patients, i.e. the sample size of the internal pilot study was 469 (276) patients per group for  $\delta = 1.15$  (1.20).

► Table 2 gives the type I error rates and sample size distributions under the null hypothesis depending on the noninferiority margin  $\delta$ , the shape parameter  $\varphi$  and the overall event rate  $\lambda$ . The simulated type I error rates are all but one within two times their simulation standard errors from  $\alpha = 0.025$ , i.e. the sample size reestimation during the ongoing trial does not inflate the type I error rate. Previously small inflations of the type I error rate by blinded sample size reviews had been reported for noninferiority trials with normal data [18]. However, no such inflations were observed in noninferiority trials with binary data [19].

The expected sample sizes under the null hypothesis are about 5 per group larger than the required sample size in a fixed design (► Table 2) and are thereby larger than

the expected sample sizes under the alternative  $\theta = 1$ , which are about the same size as the sample sizes required in a fixed sample size design (► Table 3). Under the null hypothesis, i.e.  $\theta = \delta$ , the shape parameter  $\varphi$  is slightly overestimated since it is estimated from the pooled sample ignoring the existing treatment difference whereas the treatments have the same event rates under the alternative ( $\theta = 1$ ), in which case pooling the data from the two groups gives an unbiased estimate of the shape parameter. As can be seen from ► Table 3 the simulated power values are all very close to the target of 80%. The expected sample sizes are close to the sample sizes required in a fixed sample size for the considered parameter constellations. For instance, the mean of the simulated sample sizes for  $\varphi = 0.5$  and  $\lambda = 1.5$  is 937.8 (551.4) with 938 (551) patients per group being required in a fixed sample size design according to the

approximation formula (4) for noninferiority margin  $\delta = 1.15$  (1.20).

## 6. Conclusions

In this paper we developed a blinded sample size reestimation procedure for overdispersed count data that follow negative binomial distributions and extended the ideas to non-inferiority trials. In simulation studies we investigated the characteristics of these procedures with regard to type I error rate, power and sample size. We found that the type I error rate is not inflated by the blinded reestimation procedure. Therefore these procedures satisfy international regulatory guidelines that require that the blinding is maintained and that the type I error rate is controlled [20–22]. Furthermore, we could demonstrate that the blinded sample size reesti-

**Table 3** Power and sample size distribution in non-inferiority trials with blinded sample size reestimation depending on the shape parameter  $\varphi$ , the overall event rate  $\lambda$  and the noninferiority margin  $\delta$ . The sample size distribution is summarized by mean, standard deviation (SD) and 95% percentile (95% P.). For comparison the sample sizes  $n_0$  required in a fixed design are given.

Scenario				Power	Sample size distribution		
$\delta$	$\varphi$	$\lambda$	$n_0$		Mean	SD	95% P.
1.15	0.4	1	1125	0.799	1125.3	66.2	1239
		1.5	857	0.798	857.5	47.9	939
		2	723	0.805	723.2	38.9	789
	0.5	1	1205	0.807	1206.3	71.6	1329
		1.5	938	0.798	937.8	53.1	1027
		2	804	0.801	803.4	43.5	877
	0.6	1	1286	0.802	1286.7	77.2	1418
		1.5	1018	0.795	1018.2	57.8	1115
		2	884	0.799	884.1	48.0	965
1.2	0.4	1	661	0.799	661.8	50.3	749
		1.5	504	0.801	504.0	36.4	566
		2	425	0.797	424.9	29.6	474
	0.5	1	708	0.802	709.4	55.1	804
		1.5	551	0.804	551.4	39.9	620
		2	472	0.802	472.4	33.0	529
	0.6	1	756	0.800	757.3	58.7	857
		1.5	598	0.801	598.8	43.7	674
		2	519	0.799	520.1	36.3	582

mation designs are robust against misspecification of nuisance parameters in the planning phase in that their power is always close to the target value irrespective whether the true nuisance parameter values are close to the values assumed in planning phase or not.

The investigations reported here were motivated by trials in COPD. Our findings suggest that trials in COPD can benefit from the designs proposed here in that the trials are more likely to be adequately sized and therefore are more likely to be conclusive if a BSSR procedure is implemented. As COPD exacerbations are more common in the cold season, the timing of the blinded sample size reestimation should be chosen such that representative data is available for the interim review or analyses might be adjusted for seasonality. For analyses of covariance (ANCOVA) with normally distributed data blinded sample size reestimation procedures have recently been proposed and investigated in superiority and non-inferiority trials [23, 24].

The application of the designs proposed here is not limited to COPD, and other indications such as asthma are likely to benefit from such procedures, too. For example, samples size reestimation procedures for count data were considered in multiple sclerosis [5] and immune thrombocytopenic purpura [4]. Before using the methodology considered here in indications with count data as the primary outcome, the adequacy of the negative binomial distribution needs to be assessed. This could be based on plausible mechanisms such as the six situations listed in [25]. For example, in COPD the fifth situation (individual differences in patients, i.e. different degrees of severity) may be the most likely reason for the overdispersion seen. The adequacy of the negative binomial distribution could also be based on an evaluation of historical data. For example in multiple sclerosis, the negative binomial distribu-

tion was found to provide the best model among several alternatives to describe lesion counts (see e.g. [26]).

### Acknowledgments

We are grateful to three anonymous referees for their thoughtful comments that helped us to improve the presentation of the material.

### References

1. Wittes J, Brittain E. The role of internal pilot studies in increasing efficiency of clinical trials. *Statistics in Medicine* 1990; 9: 65–72.
2. Friede T, Kieser M. Sample size recalculation in internal pilot study designs: A review. *Biometrical Journal* 2006; 48: 537–555.
3. Proschan MA. Sample size re-estimation in clinical trials. *Biometrical Journal* 2009; 51: 348–357.
4. Cook RJ, Bergeron PJ, Boher JM, Lie Y. Two-stage design of clinical trials involving recurrent events. *Statistics in Medicine* 2009; 28: 2617–2638.
5. Friede T, Schmidli H. Blinded sample size reestimation with count data: Methods and applications in multiple sclerosis. *Statistics in Medicine* 2010; 29: 1145–1156.
6. McCullagh P, Nelder JA. *Generalized Linear Models* (2nd ed). Chapman & Hall: London; 1989.
7. World Health Organization (WHO). Chronic obstructive pulmonary disease (COPD). Fact sheet N°315, May 2008. [www.who.int/mediacentre/factsheets/fs315/en/print.html](http://www.who.int/mediacentre/factsheets/fs315/en/print.html). Accessed November 2, 2009.
8. Food and Drug Administration (FDA). Guidance for industry – Chronic obstructive pulmonary disease: Developing drugs for treatment. Draft guidance, 2007.
9. Suissa S. Statistical treatment of exacerbations in therapeutic trials of chronic obstructive pulmonary disease. *American Journal of Respiratory Critical Care Medicine* 2006; 173: 842–846.
10. Aaron SD, Fergusson D, Marks GB, Suissa S, Vandemheen KL, Doucette S, Maltais F, Bourbeau JF, Goldstein RS, Balter M, O'Donnell D, FitzGerald M for the Canadian Thoracic Society/Canadian Respiratory Clinical Research Consortium. Counting, analyzing and reporting exacerbations of COPD in randomized controlled trials. *Thorax* 2008; 63: 122–128.
11. Keene ON, Calvery PMA, Jones PW, Vestbo J, Anderson JA. Statistical analysis of exacerbation rates in COPD: TRISTAN and ISOLDE revisited. *European Respiratory Journal* 2008; 32: 17–24.
12. Keene ON, Jones MRK, Lane PW, Anderson J. Analysis of exacerbation rates in asthma and chronic obstructive pulmonary disease: example from the TRISTAN study. *Pharmaceutical Statistics* 2007; 6: 89–97.
13. Ferguson GT, Anzueto A, Fei R, Emmett A, Knobil K, Kalberg C. Effect of fluticasone propionate/salmeterol (250/50 µg) or salmeterol (50 µg) on COPD exacerbations. *Respiratory Medicine* 2008; 102: 1099–1108.
14. Anzueto A, Ferguson GT, Feldman G, Chinsky K, Seibert A, Emmett A, Knobil K, O'Dell D, Kalberg C, Crater G. Effect of fluticasone propionate/salmeterol (250/50) on COPD exacerbations and impact on patient outcomes. *Journal of Chronic Obstructive Pulmonary Disease* 2009; 6: 320–329.
15. Hilbe JM. *Negative Binomial Regression*. Cambridge University Press: Cambridge, 2007.
16. Friede T, Kieser M. A comparison of methods for adaptive sample size adjustment. *Statistics in Medicine* 2001; 20: 3861–3873.
17. Friede T, Kieser M. Sample size recalculation for binary data for internal pilot study designs. *Pharmaceutical Statistics* 2004; 3: 269–279.
18. Friede T, Kieser M. Blinded sample size reassessment in noninferiority and equivalence trials. *Statistics in Medicine* 2003; 22: 995–1007.
19. Friede T, Mitchell C, Müller-Velten G. Blinded sample size reestimation in non-inferiority trials with binary endpoints. *Biometrical Journal* 2007; 49: 903–916.
20. International Conference on Harmonisation (ICH). ICH Harmonised Tripartite Guideline E9: Statistical Principles for Clinical Trials. *Statistics in Medicine* 1999; 18: 1905–1942.
21. European Medicines Agency (EMA). Reflection paper on methodological issues in confirmatory clinical trials planned with an adaptive design. CHMP/EWP/2459/02, 2007.
22. Food and Drug Administration (FDA). Guidance for industry – Adaptive design clinical trials for drugs and biologics. Draft guidance, 2010.
23. Friede T, Kieser M. Blinded sample size recalculation for clinical trials with normal data and baseline adjusted analysis. *Pharmaceutical Statistics* 2010. DOI 10.1002/pst.398.
24. Friede T, Kieser M. Reassessment of sample size in internal pilot study designs for non-inferiority trials evaluated with ANCOVA. *Methods Inf Med* 2010. DOI 10.3414/ME09-01-0063.
25. Crofton HD. A quantitative approach to parasitism. *Parasitology* 1971; 62: 179–193.
26. van den Elskamp IJ, Knol DL, Uitdehaag BMJ, Barkhof F. The distribution of new enhancing lesion counts in multiple sclerosis: further explorations. *Multiple Sclerosis* 2009; 15: 42–49.