Estimation after blinded sample size reassessment

Martin Posch,1 Florian Klinglmueller,1,2 Franz König1 and Frank Miller3

Abstract
Blinded sample size reassessment is a popular means to control the power in clinical trials if no reliable information on nuisance parameters is available in the planning phase. We investigate how sample size reassessment based on blinded interim data affects the properties of point estimates and confidence intervals for parallel group superiority trials comparing the means of a normal endpoint. We evaluate the properties of two standard reassessment rules that are based on the sample size formula of the \( z \)-test, derive the worst case reassessment rule that maximizes the absolute mean bias and obtain an upper bound for the mean bias of the treatment effect estimate.

Keywords
Adaptive design, interim analysis, internal pilot study, point estimate, sample size reassessment

1 Introduction
In clinical trials for the comparison of means of normally distributed observations, the sample size to achieve a specific target power depends on the true effect size and variance. For the purpose of sample size planning, the effect size is usually assumed to be equal to a minimally clinically relevant effect while the variance is often estimated from historical data. For situations where only little prior knowledge on the variance is available, clinical trial designs with sample size reassessment based on interim estimates of this nuisance parameter have been proposed. Stein1 developed a two-stage procedure, where the second stage sample size is calculated based on a first stage variance estimate aiming to achieve a pre-specified target power. In the context of clinical trials, several extensions of Steins two-stage procedure have been considered2–8 that all require an unblinding of the interim data for the computation of the variance estimate. However, regulatory agencies generally prefer blinded interim analyses as they entail less potential for bias.9–13

Gould and Shih14 proposed to estimate the variance from the blinded interim data by computing the variance from the total sample (pooling the observations from both groups), instead. Although this interim estimate is not consistent15 and has a positive bias if the alternative hypothesis holds, the bias is negligible for effect sizes typically observed in clinical trials.16 Furthermore, sample size reassessment based on the total variance has no relevant impact on the type I error rate in parallel group superiority trials17 (see also the literature18–21) and achieves the target power well. Similar results on sample size reassessment based on blinded nuisance parameter estimates were obtained for binary data,22 count data23–25 longitudinal data,26 and for fully sequential sample size reassessment.27 For non-inferiority trials with normal endpoints, a minor inflation of the type I error rate for small sample sizes has been observed.28,29 However, if the sample size reassessment rule is not only based on the primary endpoint but
also on blinded secondary or safety endpoint data, the type I error rate for the test of the primary endpoint may be substantially inflated.30

While for adaptive clinical trials with sample size reassessment based on the unblinded interim treatment effect estimate, it is well known that unadjusted point estimates of the effect size and confidence intervals may be biased,31,32 the properties of point estimates and confidence intervals computed at the end of an adaptive clinical trial with blinded sample size reassessment have received less attention. In this paper, we investigate the bias and standard error of the treatment effect point estimate and compare the absolute mean bias due to blinded clinical trial with blinded sample size reassessment have received less attention. In this paper, we investigate the properties of point estimates and confidence intervals computed at the end of an adaptive clinical trial with blinded sample size reassessment rules that are based on the blinded interim variance estimate. Previously, this bias had been investigated for a corresponding unblinded reassessment rule (which had no upper sample size bound), sharp bounds for the bias had been derived and an additive bias correction had been suggested.8 Here, in this paper, we derive corresponding bounds for the bias of the final variance estimate in the blinded case. Furthermore, we assess the coverage probability of one- and two-sided confidence intervals.

In Section 2, we introduce adaptive designs with blinded sample size reassessment. Theoretical results on the bias of estimates are presented in Section 3. In Section 4, we report a simulation study quantifying the bias and coverage probabilities in a variety of scenarios. The impact of the results is discussed in the context of a case study in Section 5. Section 6 concludes the paper with a discussion and recommendations. Technical proofs are given in Appendix 1.

2 Blinded sample size reassessment

Consider a parallel group comparison of the means \( \mu_a, \mu_b \) of a normally distributed endpoint with common unknown variance \( \sigma^2 \). The one-sided null hypothesis \( H_0: \delta \leq 0 \), is tested against \( H_1: \delta > 0 \) at level \( \alpha \), where \( \delta = \mu_b - \mu_a \) denotes the true effect size. Let \( \delta_0 \) denotes the alternative for which the trial is powered and assume that in the planning phase a first stage per group sample size \( n_1 \geq 2 \) is chosen based on an a priori variance estimate \( \sigma_0^2 \). Note, however, that all results below depend only on the chosen \( n_1 \) and not the way it has been determined. As a consequence, for given \( n_1 \), they also apply if it has been determined with other justifications. After the endpoints of the first stage subjects are observed, in a blinded interim analysis, the one-sample variance estimate is computed, given by

\[
S^2_{1,OS} = \frac{1}{2n_1 - 1} \left[ \sum_{i=a,b} \sum_{k=1}^{n_i} (X_{1ik} - \bar{X}_1)^2 \right]
\]

where \( X_{jk} \) is the observation \( k = 1, \ldots, n_i \) in stage \( j = 1, 2 \) for treatment \( i = a, b \) and \( \bar{X}_1 \) is the mean of the pooled first stage samples. Based on \( S^2_{1,OS} \), the second stage sample size is chosen with a pre-specified sample size function \( n_2(S^2_{1,OS}) \). Below we drop the argument of the function \( n_2 \) for convenience if the meaning is clear from the context. Then, the overall sample size is \( n = n_1 + n_2 \) per group. After the study is completed and unblinded, we assume that the point estimates for the mean difference and the variance

\[
\Delta = \bar{X}_b - \bar{X}_a, \quad S^2 = \frac{1}{2n - 2} \left[ \sum_{i=a,b} \sum_{j=1}^{2} \sum_{k=1}^{n_i} (X_{ijk} - \bar{X}_i)^2 \right]
\]

are computed, where \( \bar{X}_i = \sum_{j=1}^{2} \sum_{k=1}^{n_i} X_{ijk}/n_i, i = a, b \). Furthermore, (i) the standard fixed sample lower confidence bound corresponding to the one-sided null hypothesis \( H_0 \) given by \( \bar{X} - t_{2n-2,1-\alpha} \sqrt{S^2/2}/n \), where \( t_{2n-2,1-\alpha} \) denotes the \( 1 - \alpha \) quantile of the central \( t \)-distribution with \( 2n - 2 \) degrees of freedom, (ii) the upper confidence bound—corresponding to the complementary null hypotheses \( H_0^\prime: \delta \geq 0 \) given by \( \bar{X} + t_{2n-2,1-\alpha} \sqrt{S^2/2}/n \) and (iii) the two-sided confidence interval given by \( \bar{X} \pm t_{2n-2,1-\alpha/2} \sqrt{S^2/2}/n \) is determined.

As example consider a sample size reassessment rule aiming to control the power under a pre-specified absolute treatment effect. It is derived from the standard sample size formula for the comparison of means of normally distributed observations with a one-sided \( z \)-test using the variance estimate \( S^2_{1,OS} \)

\[
n_2^2(S^2_{1,OS}) = \min \left\{ n_{2,\text{max}}, \max \left\{ n_{2,\text{min}}, 2(z_{1-\alpha} + z_{1-\beta})^2 \frac{S^2_{1,OS}}{\delta^2} - n_1 + 1 \right\} \right\}
\]
where $1 - \beta$ is the desired power and $n_{2\text{min}}$ and $n_{2\text{max}}$ are pre-specified minimal and maximal second stage sample sizes with $0 \leq n_{2\text{min}} < n_{2\text{max}}$. The $+1$ in formula (1) was added for mathematical convenience to derive a lower bound for the variance estimate in Theorem 4 below. Given that we anyway have to round the sample size to an integer, this modification is of minor practical importance. While $S^2_{1,OS}$ is an unbiased variance estimator for $\sigma^2$ when the true effect size is zero, it is positively biased by $\delta^2 n_1/(4n_1 - 2)$ for effect sizes $\delta \neq 0$.5 A sample size reassessment rule that is based on an adjusted variance estimate which is unbiased under the effect size $\delta_0$ for which the trial is powered is obtained from equation (1) by replacing $S^2_{1,OS}$ by $S^2_{1,OS} - \delta^2 n_1/(4n_1 - 2)$ and is given by

$$
n^2_2(S^2_{1,OS}) = \min \left\{ n_{2\text{max}}, \max \left\{ n_{2\text{min}}, 2(z_{1-\alpha} + z_{1-\beta})^2 \left( \frac{S^2_{1,OS}}{\delta_0^2} - \frac{n_1}{4n_1 - 2} \right) - m + 1 \right\} \right\}
$$

(2)

3 Theoretical results on the bias of the final mean and variance estimators

In this section, we consider adaptive two-stage designs with general sample size reassessment rules where the second stage sample size is given as some non-constant function $n_2 : \mathbb{R}^+ \to \mathbb{N}_0$ of the blinded interim variance estimate $S^2_{1,OS}$. We derive the bias of the final effect size and variance estimators $\hat{\Omega}$ and $S^2$.

Theorem 1. Consider an adaptive two-stage design with a general sample size reassessment rule $n_2(S^2_{1,OS})$.

(a) The bias $E(\hat{\Omega} - \delta)$ is given by

$$
\int_0^\infty \int_0^\infty \frac{r(2n_1 - 2)(r(q, \delta - t) - r(q, \delta + t))}{\sigma^2[1 + r(q, \delta + t)][1 + r(q, \delta - t)]} \Omega_{q, \sqrt{2\delta/n_1}}(r) \chi^2_{2n_1 - 2} \left( \frac{q(2n_1 - 2)}{\sigma^2} \right) dq dt
$$

where

$$r(x, y) = n_2 \left( \frac{2(n_1 - 1)}{2n_1 - 1} x + \frac{n_1}{2(2n_1 - 1)} y \right)^2 / n_1
$$

(b) Under the null hypothesis $\delta = 0$, the final mean estimator $\hat{\Omega}$ is unbiased.

(c) If $n_2(S^2_{1,OS})$ is increasing and $\delta > 0$ ($\delta < 0$) the estimator $\hat{\Omega}$ is negatively (positively) biased, respectively. Furthermore, the absolute bias is symmetric in $\delta$ around 0.

(d) If the sample size reassessment rule is increasing and $\lim_{S^2_{1,OS} \to \infty} n_2(S^2_{1,OS}) = \infty$, the bias of $\hat{\Omega}$ converges to 0 for $\delta \to \pm \infty$.

Note that the sample size functions $n_2^m$, $n_2^u$ are increasing such that the property (c) applies. If $n_2^m = \infty$, $n_2^m$, $n_2^u$ tend to infinity for $S^2_{1,OS} \to \infty$ such that the property (d) applies.

In Theorem 2, we give an upper bound for the bias of the mean in adaptive two-stage designs with a general sample size reassessment rule $n_2(S^2_{1,OS})$ such that $n_2^m \leq n_2(S^2_{1,OS}) \leq n_2^u$, for lower and upper bounds where $n_2^m, n_2^u$, $n_2^u$ may be infinite (corresponding to the case of unrestricted sample size reassessment). For general unblinded sample size reassessment rules that may depend on the fully unblinded interim data, the upper bound for the bias is

$$
\left( \frac{1}{n_1 + n_2^m} - \frac{1}{n_1 + n_2^u} \right) 0.4 \sigma \sqrt{2n_1}
$$

(3)

and is realized for the sample size reassessment rule that sets $n_2 = n_2^m$ if $\hat{\Omega} \leq \delta$ and $n_2 = n_2^u$ otherwise. This upper bound obviously also applies for the case of blinded sample size reassessment based on the total variance.
estimate because the set of unblinded sample size reassessment rules for which \( n_2 \) may depend on the unblinded interim data in any way, includes all blinded sample size reassessment rules. This bound is not sharp if the sample size depends on \( S_{1,OS}^2 \), only. In the theorem below, we derive the sample size reassessment rule that maximizes the bias and compute the corresponding maximal bias which is a sharp upper bound for the bias that can occur under any sample reassessment based on sample size rules \( n_2(S_{1,OS}^2) \). We show that for increasing effect sizes, the bias approaches the bias of unblinded sample size reassessment.

**Theorem 2.** Consider an adaptive two-stage design with a general sample size reassessment rule \( n_2(S_{1,OS}^2) \), such that \( 0 \leq n_2 \leq n_2(S_{1,OS}^2) \leq n_2 \max \), where \( n_2 \max \) may be infinite. Then,

(a) the expected first stage treatment effect estimate conditional on the variance estimate \( S_{1,OS}^2 \) is given by

\[
E(\Delta_1|S_{1,OS}^2) = \int_{-b}^{b} \frac{x}{\sqrt{2\pi/\sigma^2}} f(x) dx.
\]

where \( \Delta_1 \) denotes the unblinded first stage estimate of the mean treatment effect, and \( b = \sqrt{2S_{1,OS}^2(2n_1 - 1)/n_1} = f_{-b} f_{+b} \sqrt{2\pi/\sigma^2} f(x)^2 dx \). Then, for \( n_2 \) depending on \( S_{1,OS}^2 \) and \( n_2 \min \),

(b) the bias of \( \Delta \) is maximized for the sample size rule

\[
n_2(S_{1,OS}^2) = \begin{cases} 
n_2 \min & \text{if } E(\Delta_1|S_{1,OS}^2) > \delta \\
n_2 \max & \text{otherwise} \end{cases}
\]

(c) the maximum bias of \( \Delta \) depends on the true effect size \( \delta \) and is given by

\[
\left( \frac{n_1}{n_1 + n_2 \min} - \frac{n_1}{n_1 + n_2 \max} \right) E\left( E(\Delta_1 - \delta|S_{1,OS}^2) 1_{E(\Delta_1|S_{1,OS}^2) > \delta} \right)
\]

(d) For \( \delta \to \infty \) and \( \sigma \) fixed, the maximum bias of \( \Delta \) caused by blinded sample size reassessment rules based on \( S_{1,OS}^2 \) converges to the maximum bias caused by unblinded sample size reassessment rules given by equation (3).

Note that, by symmetry, the negative bias of \( \Delta \) is maximized by the sample size rule equation (4) and is given by equation (5) with the inequality signs reversed. If \( n_2 \min = 0 \) and \( n_2 \max = \infty \), the first factor in equation (5) reduces to 1.

In Theorem 3, we derive the bias of the variance estimator computed at the end of the trial for general sample size reassessment rules \( n_2(S_{1,OS}^2) \) and show that it is symmetric in the true treatment effect \( \delta \) and converges to 0 for increasing \( \delta \) under suitable conditions (which are satisfied for \( n_2, n_2 \max \) if \( n_2 \max = \infty \)).

**Theorem 3.** Consider an adaptive two-stage design with a general sample size reassessment rule \( n_2(S_{1,OS}^2) \). Then,

(a) the bias of the variance \( S^2 \) is given by

\[
\int_{0}^{\infty} \int_{-\infty}^{\infty} \frac{1}{(n_1 - 1 + n_1 r(q, t + \delta))} \left( (n_1 - 1)(q - \sigma^2) + \frac{r^2 n_1/2 - \sigma^2}{2 + 2r(q, t + \delta)} \right) \varphi_{q, \sqrt{2\pi/\sigma^2}}(t) K_{2m-2}(qy) dy dt
\]

where \( \gamma = (2n_2 - 1)/\sigma^2 \) and \( r \) is defined in Theorem 1;

(b) the bias of \( S^2 \) is symmetric in \( \delta \) around 0;

(c) if \( \lim_{S_{1,OS}^2 \to \infty} n_2(S_{1,OS}^2) = \infty \), the bias of \( S^2 \) converges to 0 for \( \delta \to \pm \infty \).
In Theorem 4, we derive a lower bound for the bias of the variance estimator $S^2$ (which is negative) in an adaptive two-stage design with sample size reassessment rule $n^*_0$.

**Theorem 4.** Consider a design with blinded sample size reassessment based on $S^2_{1,OS}$, where the second stage sample size is set to $n^*_0(S^2_{1,OS})$. Under the null hypothesis $\delta = 0$, the bias of the variance estimator $S^2$ at the end of the study has the following bounds

$$\frac{-2n_1 - 1}{2n_1 - 3} \cdot \frac{1}{\nu} < ES^2 - \sigma^2 < 0$$

where $\nu = 2(z_{1-\alpha} + z_{1-\beta})^2/\delta_0^2$. If $n^*_{\text{max}} = \infty$ (no upper restriction for the sample size), the bias converges to the lower bound for large variances

$$ES^2 - \sigma^2 \rightarrow -\frac{2n_1 - 1}{2n_1 - 3} \cdot \frac{1}{\nu} \quad \text{for } \sigma^2 \rightarrow +\infty$$

Note that this lower bound has a similar form as the bound derived for the bias in adaptive two-stage designs with sample size reassessment based on the unblinded variance estimate. In the latter setting, the bound is given by $-(n_1 - 1)/[(n_1 - 2)\nu]$. The bound derived in Theorem 4 does not apply to designs where the adjusted rule $n^*_0$ is applied. The numerical results below suggest that for the adjusted rule the absolute bias is maximized for a finite $\sigma^2$ and not for $\sigma^2 \rightarrow \infty$ as for the unadjusted rule.

### 4 Simulation study of the properties of point estimates and confidence intervals

To quantify the bias of mean and variance estimates as well as the actual coverage of corresponding one- and two-sided confidence intervals in adaptive two-stage designs with blinded sample size reassessment, a simulation study was performed. We assumed that a trial is planned for an effect size $\delta_0 = 1$ and a priori variance estimate $\sigma_0$ at a one-sided significance level $\alpha = 0.025$. The preplanned sample size of a two-armed parallel group trial was chosen to provide $80\%$ power (based on the normal approximation) for $\sigma_0 \in \{1, 1.5, 2\}$. The first stage sample size was set to half of the preplanned total sample size leading to first stage per group sample sizes of $n_1 \in \{8, 18, 32\}$. For each first stage sample size, first stage data were simulated from normal distributions for effect sizes $\delta$ ranging from $-2$ to $2$ (in steps of $0.1$) and common standard deviations $\sigma$, ranging from $0.5$ to $2$ in steps of $0.5$. The second stage sample sizes were then reassessed to $n^*_0, n^*_2$ (with $n^*_{\text{min}} = 0$). Below we refer to the latter reassignment rules as the unadjusted and adjusted sample size reassessment rules. For each scenario $5 \times 10^7$ trials were simulated with R. The source code is available in the R-package blindConfidence which can be downloaded and installed from github. Details on the implementation and additional results are presented in the Supplementary Material. Figures 1 and 2 show the mean bias of the final estimates of the mean and variance based on the total sample. If the alternative holds, the estimates may be substantially biased for small first stage sample sizes.

For both sample size reassessment rules, the bias of the mean is zero under the null hypothesis and has the opposite sign as the true effect size, otherwise (in accordance with Theorem 1). For very large positive and negative effect sizes, the bias is close to zero. This is due to the fact that a large effect size results in a large positive bias of the blinded interim variance estimates which in turn results in very large second stage sample sizes. As a consequence, the overall estimates are essentially equal to the (unbiased) second stage estimates, and the bias becomes negligible. In the considered scenarios, the absolute bias is decreasing in the first stage sample size. This also holds for the upper bound of the bias for general sample size reassessment rules. However, while the bias for the adjusted and unadjusted reassessment rules is small for larger sample sizes, the upper bound for general sample size reassessment rules still exceeds $10\%$ in the considered scenarios even in the case where $n_1 = 32$ subjects per group are recruited in the first stage. In the considered scenarios, we observe that the upper bound for the bias of the mean estimate increases in the true effect size $\delta$ and the true standard deviation $\sigma$. The numerical results confirm that for increasing $|\delta|$ the maximum bias under blinded sample size reassessment approaches the maximum bias under unblinded sample size reassessment as shown in Theorem 2.

The variance estimate is negatively biased for both considered sample size reassessment rules. For increasing variances, the absolute bias under the null hypothesis $\delta = 0$ for the rule $n^*_0$ approaches the lower bound derived in Theorem 4. For large positive and negative effect sizes, the bias approaches zero, again, because the overall estimate becomes essentially equal to the (unbiased) second stage estimates. In general, we observe that using
the adjusted interim variance estimate for sample size reassessment results in a larger absolute bias of mean and variance estimates.

The difference of the coverage probabilities of the confidence intervals to the nominal coverage probabilities (i.e., 0.975 for the one-sided, 0.95 for the two-sided confidence interval) is shown in Figure 3 for the unadjusted sample size reassessment rule (see Figure 1 of the Supplementary Material for corresponding results for the adjusted sample size reassessment rule). For positive $\delta$, the lower confidence bound is conservative (with coverage probability larger than 0.975) while the upper bound is anti-conservative, and vice-versa for negative $\delta$. In fact, due to symmetry, the coverage probability for the upper confidence bound at a certain true effect size $\delta$ is the same as the coverage probability of the lower confidence bound at $-\delta$. If the true $\sigma^2$ and the true $|\delta|$ are small, the trial will stop with minimal sample size almost with probability 1. Then, sample size is essentially fixed, no bias occurs and coverage is as desired. This is, e.g., the case for $\sigma = 1, n_1 = 32$ and $|\delta| < 1.3$ and explains the exact coverage there. Similar cases occur for small $|\delta|$ in the three pictures for $\sigma = 0.5$ as well.

The two-sided coverage probability (which is given by one minus the sum of the non-coverage probabilities of the lower and upper bounds) is not controlled over a large range of $\delta$. However, for $\delta = 0$ (i.e., under the null hypothesis), the inflation of the non-coverage probability, which corresponds to the Type I error rate of the corresponding test, is minor for small sample sizes (0.5 to 0.01 percentage points $n_1 < 10$) and essentially controlled for larger samples sizes. (See Figure 15 of the Supplementary Material for a simulation study that aims to identify the $\sigma$ where the Type I error rate of the original hypothesis test is maximized.) The good coverage under the null hypothesis is at first sight surprising, given that the variance estimate is negatively biased and the mean estimate is unbiased. We therefore computed the actual variance $\sigma_*^2$ of the mean estimate in the simulation study. To compare the actual variance of the mean to the estimated variance, we computed for each simulated trial the variance estimate $\hat{S}_v^2 = \frac{2N}{n_1 + n_2}$ as well as the actual variance $\sigma_*^2 = \frac{2\sigma^2}{n_1 + n_2}$ of a design with fixed sample sizes $n_1, n_2$.
Thus, ignoring that \( n_2 \) is dependent on the first stage sample. Both quantities were then averaged over the Monte-Carlo samples. The results show that the true variance of the mean estimate of an adaptive design with blinded sample size reassessment is smaller than the average variance of a fixed sample design with the same sample sizes. The bias of the estimate of the variance of the mean estimate is close to zero around the null hypothesis. This holds both for designs using the adjusted and unadjusted interim variance estimates to reassess the sample size and gives some intuition why there is no relevant inflation of the coverage probability of confidence intervals under the null hypotheses, even though the variance estimate has a considerable negative bias (see Figures 5 and 6 of the Supplementary Material for results that show the variance of the mean estimate for designs that use the unadjusted and adjusted sample size rule).

While Figures 1 to 3 demonstrate the dependence of the bias and coverage probabilities on the true mean and standard deviation, in an actual trial these parameters are typically unknown. Therefore, we computed the maximum absolute bias of the mean, variance and the maximum difference between nominal and actual coverage probabilities for fixed first stage sample sizes over a range of values for the true mean and variance (see Figure 4). The optimization was performed in several steps based on simulations across an increasingly finer grid of true mean differences \( \delta \in [0, 4] \) and standard deviations \( \sigma \in [0.5, 4] \) and using increasingly larger numbers of up to \( 10^8 \) simulation runs (see Section 2 of the Supplementary Material for computational details of this simulation).

Overall, the maximum bias of the mean and variance estimates as well as the maximum difference between nominal and actual coverage probabilities using the adjusted sample size reassessment rule exceed that of the unadjusted rule. The maximum absolute bias of the mean bias (Columns 1 and 2 of Figure 4) drops from 0.18 (0.08) for trials with a per group first stage sample size of \( n_1 = 2 \) to 0.02 (0.02) for trials with a per group first stage sample size of \( n_1 = 50 \) if the adjusted (unadjusted) interim variance estimate is used to reassess the sample size. For increasing values of the first stage sample size, the maximum of the absolute mean bias is found for larger values of the true mean and variance.
For the unadjusted rule, the simulations suggest that the absolute variance bias is maximized for $C14 = 0$ and $C27 \neq 1$ (see Figure 2 and the additional simulations in Section 2 of the Supplementary Material). Thus, Column 3 of Figure 4 shows the absolute bias for $C14 = 0$ and $C27 = 4$ (the largest value of $C27$ considered in the parameter grid).

For the adjusted rule, Column 4 of Figure 4 shows the maximum absolute bias for $C14 = 0$ and $C27$ maximized over the grid. For the adjusted rule, the maximum bias is attained for a finite $C27$. The maximum absolute bias of the variance estimate drops from 0.27 (0.19) for trials with a per group first stage sample size of $n1 = 2$ to 0.07 (0.06) for trials with a per group first stage sample size of $n1 = 50$ if the adjusted (unadjusted) interim variance estimate is used to reassess the sample size. In the simulation, the maxima of the absolute variance bias, using the unadjusted reassessment rule, is practically identical to the theoretical boundary derived in Theorem 4 (see the gray line in the first graph of Column 3 of Figure 4). If the adjusted rule is used, the maximum is attained for values of $C27$ ranging from 1 to 2, increasing with the first stage sample size.

The maximum negative difference between actual and nominal coverage probabilities ("AC-NC," in Columns 5 and 6 of Figure 4) of the one-sided confidence intervals ranges from about 2.8 (2.4) to 0.4 (0.3) percentage points if the adjusted (unadjusted) variance estimate is used to reassess the sample size. For the two-sided confidence intervals, differences between actual and nominal coverage are slightly lower ranging from 2.4 (2.0) to 0.2 (0.2) percentage points, respectively. For increasing first stage sample sizes, the maximum is attained for increasing values of both $\sigma$ and $\delta$ where the former ranges between 1 and 3 and the latter between 0.5 and 1.5.

In Section 4 of the Supplementary Material, the corresponding results for restricted sample size rules that limit the maximum second stage sample size to twice the preplanned second stage sample size, are given. We observe that for scenarios where $\sigma_0 < \sigma$, limiting the second stage sample size results in a reduction of the (absolute) bias and higher coverage probabilities. This, however, comes at the price of limited control in terms of power. When $\sigma_0 > \sigma$ we observe that the (absolute) bias of the mean estimate (in line with the theoretical results in Theorem 2) is reduced by a factor $n_{2max}/(n_1 + n_{2max})$. The bias of the variance estimate as well as the difference between actual
and nominal coverage probabilities are less affected by restricting the second stage sample size. When \( \sigma_0 \) bias and coverage probabilities under the restricted and unrestricted sample size reassessment rules are nearly identical.

5 Case study

We illustrate the estimation of treatment effects with a randomized placebo controlled trial to demonstrate the efficacy of the kava-kava special extract “WS\(^8\) 1490” for the treatment of anxiety\(^{35}\) that was used to illustrate procedures for blinded sample size adjustment.\(^{17}\) The primary endpoint of the study was the change in the Hamilton Anxiety Scale (HAMA) between baseline and end of treatment. Assuming a mean difference between treatment and control of \( \delta_0 = 5.5 \) and a standard deviation of \( \sigma_0 = 8 \) (i.e., a variance of \( \sigma_0^2 = 64 \)) results in a sample size of 34 patients per group to provide a power of 80%.

As in the previous case study,\(^{17}\) we assume that sample size reassessment based on blinded data is performed after 15 patients per group, that the sample size reassessment rules were applied with \( n_{2 \min} = 0 \) and \( n_{2 \max} = \infty \) and consider that the interim estimate of the standard deviation is \( S_{1,OS} = 6 \). Consequently, we get \( n_{2}^* = 4.7 \) and \( n_{1}^* = 0.6 \), such that in the second stage five patients per group would have been recruited based on the unadjusted sample size reassessment rule \( n_{2}^* \), and only one patient per group based the adjusted rule \( n_{1}^* \).

To estimate the bias of the mean and variance estimate as well as the coverage probabilities of confidence intervals, we performed a simulation study for true effect sizes \( \delta \) ranging from \(-11 (= -2\delta_0) \) to 11 (= 2\( \delta_0 \)) in steps of 0.05, and \( \sigma \) from 1 to 20 in steps of 1 (see Figure 5).

5.1 Unadjusted sample size rule

For fixed \( \delta \), the absolute bias of the variance estimate is increasing in \( \sigma \). For \( \sigma = 20 \), the bias of the variance becomes \(-2.06 \) which is within simulation error of the theoretical lower bound. The absolute mean bias takes its maximum (minimum) 0.2 (−0.2) for the effect sizes \( \delta = -7.98 \) (7.98), respectively, and a standard deviation of \( \sigma = 5 \). The maximum inflation of the non-coverage probabilities of one-sided 97.5% and two-sided 95% confidence intervals is 0.7 and 0.5 percentage points, respectively. The actual coverage probabilities are smallest for large absolute values of the true mean difference and standard deviations of around 5.

![Figure 4](image-url)
5.2 Adjusted sample size rule

If the sample size reassessment is based on the adjusted variance estimate, the absolute bias of the variance and mean estimate will be even larger taking values up to 2.49 for the variance and up to 0.25 for the mean, respectively. The inflation of the non-coverage probabilities goes up to 0.9 percentage points for the one-sided confidence intervals and 0.6 percentage points for the two-sided intervals.

Figure 5. Coverage probabilities and bias of mean and variance estimates for the case study. The first row of panels shows actual-nominal coverage probabilities (AC-NC) for the 97.5% upper (dashed line), lower (dotted line) and the 95% two-sided confidence intervals, for the unadjusted reassessment rule $n^2_f$. The second row shows the bias of the mean estimate if $n^2_f$ (solid line) or $n^2_a$ (dashed line) is used. The dotted line shows upper and lower bounds for the bias for general blinded sample size reassessment rules based on $S^2_{1D5}$. The dashed gray line shows the bounds for the bias for a general unblinded sample size reassessment rule. The third panel shows the bias of the variance estimate if either $n^2_f$ (solid line) or $n^2_a$ (dashed line) is used. The red line shows the theoretical boundary for the bias give in Theorem 4.

6 Discussion

We investigated the properties of point estimates and confidence intervals in adaptive two-stage clinical trials where the sample size is reassessed based on blinded interim variance estimates. Such adaptive designs are in accordance with current regulatory guidance that proposes blinded sample size reassessment procedures based on aggregated interim data.9-11 We showed that such blinded sample size reassessment may lead to biased effect size estimates, biased variance estimates and may have an impact on the coverage probability of confidence intervals. The extent of the biases depends on the specific sample size reassessment rule, the first stage sample size, the true effect size and the variance.

We showed that for the unadjusted and the adjusted sample size reassessment rules that aim to control the power at the pre-specified level, the bias of confidence intervals may be large for very small first stage sample sizes but is small otherwise. Under the null hypothesis, even for first stage sample sizes as low as 16, the confidence intervals do not exhibit a relevant inflation of the non-coverage probability. For positive effect sizes, inflations (even though minor) are observed also for somewhat larger sample sizes. This corresponds to previous findings.
that the type I error rate of superiority as well as non-inferiority tests is hardly affected by blinded sample size reassessment.\textsuperscript{17,28} In addition, we show that for positive treatment effects the lower confidence interval (which is often the most relevant because it gives a lower bound of the treatment effect) is strictly conservative in all considered simulation scenarios. The upper bound in contrast shows an inflation of the non-coverage probability, which is however only of relevant size if sample sizes are small.

An approach to obtain conservative confidence intervals even for trials with small first stage sample sizes is to apply methods proposed for flexible designs with unblinded interim analyses\textsuperscript{32,33,36–39} which are based on combination tests or the conditional error rate principle. Such approaches allow to construct confidence intervals which control the coverage probabilities. However, they are based on test statistics that do not equally weigh outcomes from patients recruited in the first and second stage. A hybrid approach that strictly controls the coverage probability could be based on confidence intervals that are the union of the fixed sample confidence interval and the confidence interval obtained from the flexible design methodology. Similar approaches have been proposed for hypothesis testing in adaptive designs.\textsuperscript{40,41}

We demonstrated that the treatment effect estimates are negatively (positively) biased under the adjusted and unadjusted sample size reassessment rules for all positive (negative) effect sizes. The bias decreases with the first stage sample size but for worst case parameter constellations it remains noticeable also for larger sample sizes. In addition, for the worst case sample size reassessment rule based on blinded variance estimates, the bias may be substantial even for large first stage sample sizes: As the effect size increases the maximum bias of the blinded sample size reassessment rule approaches the bias of unblinded sample size reassessment. As a consequence to maintain the integrity of confirmatory clinical trials with blinded sample size reassessment, binding sample size adaptation rules must be pre-specified and it is important to verify on a case-by-case basis that the reassessment rules used do not have a substantial impact on the properties of estimators should be used.

Declaration of conflicting interests
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Martin Posch and Franz Koenig were supported by Austrian Science Fund Grant P23167. Florian Klinglmueller was supported by Austrian Science Fund Grants Number P23167 and J3721.

Supplementary material
Supplementary figures and additional details on the simulation study are presented in an additional document, which is distributed with the paper as Supplementary Material.

The R code used to compute the results of the simulation, and the case study is available as an R-package named blindConfidence which is available on github or on request from the authors.

References
1. Stein C. A two-sample test for a linear hypothesis whose power is independent of the variance. Ann Math Stat 1945; 16: 243–258.
2. Wittes J and Brittain E. The role of internal pilot studies in increasing the efficiency of clinical trials. Stat Med 1990; 9: 65–72.
3. Betensky RA and Tierney C. An examination of methods for sample size recalculation during an experiment. Stat Med 1997; 16: 2587–2598.
4. Wittes J, Schabenberger O, Zucker D, et al. Internal pilot studies I: type I error rate of the naive t-test. Stat Med 1999; 18: 3481–3491.
5. Zucker DM, Wittes JT, Schabenberger O, et al. Internal pilot studies II: comparison of various procedures. Stat Med 1999; 18: 3493–3509.
6. Denne JS and Jennison C. Estimating the sample size for a t-test using an internal pilot. Stat Med 1999; 18: 1575–1585.
7. Proschan MA and Wittes J. An improved double sampling procedure based on the variance. Biometrics 2000; 56: 1183–1187.
8. Miller F. Variance estimation in clinical studies with interim sample size reestimation. Biometrics 2005; 61: 355–361.
9. EMEA. Reflection paper on methodological issues in confirmatory clinical trials planned with an adaptive design, London, www.ema.europa.eu/ema/pages/includes/document/open_document.jsp?webContentId=WC500003616 (2007, accessed 8 September 2016).

10. FDA, CDER. Guidance for industry: adaptive design clinical trials for drugs and biologics (Draft Guidance). Rockville: Food and Drug Administration, 2010.

11. FDA, CDRH. Draft guidance: adaptive designs for medical device clinical studies. Rockville: Food and Drug Administration, 2015.

12. Elsäßer A, Regnstrom J, Vetter T, et al. Adaptive clinical trial designs for European marketing authorization: a survey of scientific advice letters from the European Medicines Agency. Trials 2014; 15: 383.

13. Pritchett YL, Menon S, Marchenko O, et al. Sample size re-estimation designs in confirmatory clinical trials: current state, statistical considerations, and practical guidance. Stat Biopharm Res 2015; 7: 309–321.

14. Gould LA and Shih WJ. Sample size re-estimation without unblinding for normally distributed outcomes with unknown variance. Commun Stat Theor Meth 1992; 21: 2833–2853.

15. Govindarajulu Z. Robustness of sample size re-estimation procedure in clinical trials (arbitrary populations). Stat Med 2003; 22: 1819–1828.

16. Friede T and Kieser M. Blinded sample size re-estimation in superiority and noninferiority trials: bias versus variance in variance estimation. Pharmaceut Stat 2013; 12: 141–146.

17. Kieser M and Friede T. Simple procedures for blinded sample size adjustment that do not affect the type I error rate. Stat Med 2003; 22: 3571–3581.

18. Proschan MA. Two-stage sample size re-estimation based on a nuisance parameter: a review. J Biopharm Stat 2005; 15: 559–574.

19. Proschan MA, Wittes JT and Lan K. Statistical monitoring of clinical trials. New York: Springer, 2006.

20. Posch M and Proschan MA. Unplanned adaptations before breaking the blind. Stat Med 2012; 31: 4146–4153.

21. Posch M, Glimm E and Posch M. Connections between permutation and t-tests: relevance to adaptive methods. Stat Med 2014; 33: 4734–4742.

22. Friede T and Kieser M. Sample size recalculation for binary data in internal pilot study designs. Pharmaceut Stat 2004; 3: 269–279.

23. Friede T and Schmidli H. Blinded sample size reestimation with count data: methods and applications in multiple sclerosis. Stat Med 2010; 29: 1145–1156.

24. Schneider S, Schmidli H and Friede T. Blinded and unblinded internal pilot study designs for clinical trials with count data. Biometrical J 2013; 55: 617–633.

25. Stucke K and Kieser M. Sample size calculations for noninferiority trials with Poisson distributed count data. Biometrical J 2013; 55: 203–216.

26. Wachtlin D and Kieser M. Blinded sample size recalculation in longitudinal clinical trials using generalized estimating equations. Ther Innov Regul Sci 2013; 47: 460–467.

27. Friede T and Miller F. Blinded continuous monitoring of nuisance parameters in clinical trials. J R Stat Soc Series C (Appl Stat) 2012; 61: 601–618.

28. Friede T and Kieser M. Blinded sample size reassessment in non-inferiority and equivalence trials. Stat Med 2003; 22: 995–1007.

29. Lu K. Distribution of the two-sample t-test statistic following blinded sample size re-estimation. Pharm Stat 2016; 15: 208–215.

30. Zebrowska M, Posch M and Magirr D. Maximum type I error rate inflation from sample size reassessment when investigators are blind to treatment labels. Stat Med 2015; 35: 1972–1984.

31. Bretz F, Koenig F, Brannath W, et al. Adaptive designs for confirmatory clinical trials. Stat Med 2009; 28: 1181–1217.

32. Brannath W, Mehta CR and Posch M. Exact confidence bounds following adaptive group sequential tests. Biometrics 2009; 65: 539–546.

33. Brannath W, König F and Bauer P. Estimation in flexible two stage designs. Stat Med 2006; 25: 3366–3381.

34. R Core Team. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing, www.R-project.org (2014, accessed 8 September 2016).

35. Malsch U and Kieser M. Efficacy of kava-kava in the treatment of non-psychotic anxiety, following pretreatment with benzodiazepines. Psychopharmacology 2001; 157: 277–283.

36. Brannath W, Posch M and Bauer P. Recursive combination tests. Stat Med 2005; 24: 3697–3714.

37. Lawrence J and Hung H. Estimation and confidence intervals after adjusting the maximum information. Biometrical J 2003; 45: 143–152.

38. Mehta CR, Bauer P, Posch M, et al. Repeated confidence intervals for adaptive group sequential trials. Stat Med 2007; 26: 5422–5433.

39. Posch M, Koenig F, Branson M, et al. Testing and estimation in flexible group sequential designs with adaptive treatment selection. Stat Med 2005; 24: 3697–3714.

40. Posch M, Bauer P and Brannath W. Issues in designing flexible trials. Stat Med 2003; 22: 953–969.

41. Burman CF and Sonesson C. Are flexible designs sound? Biometrics 2006; 62: 664–669.
Appendix 1: Proofs of the Theorems

Note that the blinded first stage variance estimate can be written as \[ S_{1,OS}^2 = \frac{2(n_1 - 1)}{2n_1 - 1} S_1^2 + \frac{n_1}{2(2n_1 - 1)} \Delta_1^2 \] (6)

where \( \Delta_1 \) denotes the unblinded first stage estimates of the mean treatment effect and \( S_1^2 \) the (unblinded) pooled variance estimate. Thus, every second stage sample size rule \( n_2(S_{1,OS}^2) \) can be written as function \( n_2(S_1^2, \Delta_1) \) of \( S_1^2 \) and \( \Delta_1 \) which is symmetric in \( \Delta_1 \) around 0. We can then write \( r(x, y) = n_2(x, y)/n_1 \). Below we switch between both parameterizations of \( n_2 \).

Proof of Theorem 1

Proof. (a) Let the true effect size \( \delta \) be fixed. The bias of \( \hat{\Delta}_1 \) is given by \( Cov\left(\frac{1}{1 + r(S_1^2, \Delta_1)}, \hat{\Delta}_1\right) \), see Brannath et al.,\(^3\) which can be computed by

\[
E\left(\frac{\hat{\Delta}_1}{1 + r(S_1^2, \Delta_1)}\right) - E\left(\frac{1}{1 + r(S_1^2, \Delta_1)}\right)E(\hat{\Delta}_1) = E\left(\frac{\hat{\Delta}_1 - \delta}{1 + r(S_1^2, \Delta_1)}\right)
\]

\[
= \int_0^\infty \int_{-\infty}^\infty \frac{t}{\left(1 + r(q, \delta + t)\right)} \varphi_0(\sqrt{2\sigma^2}t) \gamma dq dt
\]

(7)

where \( \gamma = (2n_1 - 2)/\sigma^2 \).

(b) and (c) The sign of the integral is determined by the sign of \( n_2(q, t + \delta) - n_2(q, -t + \delta) \). Note that \( r(q, y) \) is symmetric in \( y \) around 0 and increasing for \( y > 0 \) since this is true for \( n_2(q, y) \). Thus, the integral is zero for \( \delta = 0 \), negative for positive \( \delta \) and positive for negative \( \delta \). Also the symmetry of the absolute bias follows by the symmetry of \( n_2(q, y) \) in \( y \) around 0. (d) The first factor in equation (7) tends to 0, and the proof follows with the dominated convergence theorem. \( \square \)

Proof of Theorem 2

Proof. (a) By equation (6), the joint density of \( \hat{\Delta}_1 \) and \( S_{1,OS}^2 \) is given by

\[
h(x, v) = \varphi_b(\sqrt{2\sigma^2}n_1)(x^2 - 2n_1 - 2) [(2n_1 - 1)v - x^2 n_1 / 2] / (2n_1 - 1) \}
\]

for \( v \geq 0 \) and \( |x| \leq b \) and \( h(x, v) = 0 \) otherwise. Therefore, the conditional distribution of \( \hat{\Delta}_1 \) given \( S_{1,OS}^2 = v \) is \( h(x, v)/ \int_{-b}^b h(x, v) dx \) and (a) follows.

(b) Note that the bias is given by

\[
E(\hat{\Delta} - \delta) = E\left[\frac{\hat{\Delta}_1 - \delta}{n_1 + n_2(S_{1,OS}^2)}\right] = E\left[\frac{n_1}{n_1 + n_2(S_{1,OS}^2)} | S_{1,OS}^2 \right] (8)
\]

\[
= E\left[\frac{E((\hat{\Delta}_1 - \delta) | S_{1,OS}^2)}{n_1 + n_2(S_{1,OS}^2)} | S_{1,OS}^2 \right] (9)
\]
For fixed $S^2_{1,OS}$, the term within the outer conditional expectation in equation (8) is deterministic and decreases in $n_2$ if $E(\Delta_1 | S^2_{1,OS}) > \delta$ and increases if $E(\Delta_1 | S^2_{1,OS}) < \delta$. Therefore, the sample size rule equation (4) maximizes the bias.

(c) We plug-in the worst case sample size rule equation (4) into equation (9) and obtain

$$E(\tilde{\Delta} - \delta) = E\left[ \frac{n_1}{n_1 + n_{2,\min}} E(\Delta_1 - \delta | S^2_{1,OS}) I_{1} | E(\Delta_1 | S^2_{1,OS}) > \delta \right]$$

$$+ E\left[ \frac{n_1}{n_1 + n_{2,\max}} E(\Delta_1 - \delta | S^2_{1,OS}) I_{2} | E(\Delta_1 | S^2_{1,OS}) \leq \delta \right]$$

Because $\tilde{\Delta}_1$ is unbiased such that $E(\tilde{\Delta}_1 - \delta | S^2_{1,OS}) I_{1} = \delta$ we get equation (5).

(d) Let $N$ denotes a set of sample size reassessment rules $n_{2,\delta}(S^2_{1,OS}, \tilde{\Delta}_1)$ that may depend on $\delta, S^2_{1,OS}, \tilde{\Delta}_1$. Given a sample size rule $n_{2,\delta}() \in N$, the bias can be written as

$$E_3(\tilde{\Delta} - \delta) = E_3\left( (\tilde{\Delta}_1 - \delta | n_1 + n_{2,\delta}(S^2_{1,OS}, \tilde{\Delta}_1)) \right)$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} n_1 \int_{0}^{n_{2,\min}} \varphi_{0,\sigma/\sqrt{n_1}}(y)$$

$$\chi^2_{2m} - \frac{1}{2} [(2n_1 - 1)\nu - (\nu + \delta)^2n_1/\sigma^2] \, dy$$

The set of integrands of the outer integral in equation (10) for second stage sample sizes in $N$ is uniformly integrable because the fraction $n_1/[n_1 + n_{2,\delta}(y + \delta, \nu)] \leq 1$ and the $\chi^2$-density is bounded (since $n_1 \geq 2$).

The unblinded sample size reassessment rule that maximizes the bias for given $\delta$ is

$$n_{2,\delta}(\tilde{\Delta}) = \begin{cases} n_{2,\min} & \text{if } \tilde{\Delta}_1 > \delta \\ n_{2,\max} & \text{otherwise} \end{cases}$$

Consider the blinded sample size rules

$$\tilde{n}_{2,\delta}(S^2_{1,OS}) = \begin{cases} n_{2,\min} & \text{if } S^2_{1,OS} > \delta^2 \frac{n_1}{2(2n_1 - 1)} \\ n_{2,\max} & \text{otherwise} \end{cases}$$

We show that for fixed $\sigma$ and $\delta \to \infty$ the second stage sample sizes of the blinded sample size rule $\tilde{n}_{2}(S^2_{1,OS}) \to n_{2}(\tilde{\Delta}_1)$, almost surely. Because, as shown above, the corresponding integrands in equation (10) are uniformly integrable, and it follows that also the respective biases converge and (d) follows.

Let $\epsilon > 0$. Then, there exists an $\epsilon_1, 1/3 > \epsilon_1 > 0$, such that for all $\delta$ we have $P_3(|\Delta_1 - \delta| > 3\epsilon_1) > 1 - \epsilon$. Furthermore, there exists a $\delta_0 > 2$ such that for all $\delta > \delta_0$ we have $P_3(S^2_{1,\delta} X < \epsilon) > 1 - \epsilon$ and $P_3(\Delta_1 > 0) > 1 - \epsilon$. Thus, for the event $A = \{\Delta_1 > 0\} \cap \{S^2_{1,\delta} X < \epsilon\} \cap \{|\Delta_1 - \delta| > 3\epsilon_1\}$ we have $P_3(A) \geq 1 - 3\epsilon$. However, on the set $A$ we have $n_{2,\delta}(\Delta_1) = \tilde{n}_{2,\delta}(S^2_{1,OS})$: If $n_{2,\delta}(\Delta_1) = n_{2,\min}$ then $\Delta_1 > \delta$ and on $A$ also $\tilde{\Delta}_1 > \delta + 3\epsilon_1$. Therefore, by equation (6)

$$S^2_{1,OS} > \delta^2(1 + 6\epsilon_1/\delta + 9\epsilon_1^2/\delta^2) \frac{n_1}{2(2n_1 - 1)} > \delta^2 \frac{n_1}{2(2n_1 - 1)}$$

and $\tilde{n}_{2,\delta}(S^2_{1,OS}) = n_{2,\min}$. On the other hand, if $n_{2,\delta}(\tilde{\Delta}_1) = n_{2,\max}$ then $\tilde{\Delta}_1 < \delta$ and on $A$ also $0 \leq \tilde{\Delta}_1 < \delta - 3\epsilon_1$. Therefore, by equation (6), because $2(n_1 - 1)/(2n_1 - 1) \leq 1$ and $1/3 \geq n_1/2(2n_1 - 1) \geq 1/4$ and on $A$ we have $S^2_{1,\delta} \leq \epsilon_1 \delta$

$$S^2_{1,OS} < \epsilon_1 \delta + \frac{n_1}{2(2n_1 - 1)}(\delta^2 - 6\epsilon_1 \delta + 9\epsilon_1^2)$$

$$\leq \delta^2 \frac{n_1}{2(2n_1 - 1)} + \epsilon_1 \delta - \frac{3}{2} \epsilon_1 \delta + 3\epsilon_1^2$$

$$\leq \delta^2 \frac{n_1}{2(2n_1 - 1)} - \epsilon_1 \frac{n_1}{2} + 3\epsilon_1^2 \leq \delta^2 \frac{n_1}{2(2n_1 - 1)}$$

where the last inequality follows because $\delta > 2$ and $\epsilon_1 < 1/3$. \[\square\]
Proof of Theorem 3

Proof. Let \( \bar{\Delta}_j = \bar{X}_{aj} - \bar{X}_{bj} \) be the observed mean difference between the treatments in stage \( j = 1 \) and (if \( n_2 > 0 \)) in stage \( j = 2 \).

We consider first the case that \( n_2 > 0 \); the following considerations are under the condition of \( S_2^2 \) and \( \bar{\Delta}_1 \) such that they lead to \( n_2 > 0 \). Let \( \Omega_i = \bar{X}_{ai} - \bar{X}_{bi} \) be the observed mean difference between the stages for treatment \( i = a, b \).

Let further \( S_1^2 \) be the (unblinded) pooled variance estimate of the second stage (for \( n_2 = 1 \), we define \( S_1^2 = 0 \), \( \Omega_+ = \Omega_a + \Omega_b = 2(\bar{X}_1 - \bar{X}_2) \) be twice of the overall difference between the stages and \( \Omega_- = \Omega_a - \Omega_b = \bar{\Delta}_1 - \bar{\Delta}_2 \).

With this notation, we can split the total variance estimate \( S_2^2 \) into parts

\[
S_2^2 = \frac{2n_1 - 2}{2n} S_1^2 + \frac{2n_2 - 2}{2n} S_2^2 + \frac{n_1 n_2}{2n(2n - 2)} \Omega_+^2 + \frac{n_1 n_2}{2n(2n - 2)} \Omega_-^2 
\]  

(11)

Obviously, \( E(S_2^2|n) = E(S_1^2) = \sigma^2 \) for all \( n \). As \( \bar{X}_1, S_1^2 \) and \( \bar{\Delta}_1 \) are mutually independent (and consequently \( n \) and \( \bar{X}_1 \)), and as \( \bar{X}_1 \) and \( \bar{X}_2 \) are independent under the condition of \( n \), \( \Omega_i \) is conditionally on \( n \) distributed as \( N(0, 2\sigma^2(1/n_1 + 1/n_2)) \).

Under the condition of \( n \), \( \bar{\Delta}_2 \) is distributed as \( N(\bar{\Delta}_1 - \bar{\Delta}_2, 2\sigma^2/n_2) \). Further, under the condition of \( S_1^2 \) and \( \bar{\Delta}_1 \), \( \frac{n_2}{2\sigma^2}(\Delta_1 - \bar{\Delta}_2)^2 = \frac{n_2}{2\sigma^2} \Delta_2^2 \) has a non-central \( \chi^2 \) distribution with 1 degree of freedom and non-centrality parameter \( \frac{n_2(\bar{\Delta}_1 - \bar{\Delta}_2)^2}{2\sigma^2} \).

Therefore, it follows

\[
E(S_2^2 - \sigma^2 | S_1^2, \bar{\Delta}_1) = E \left[ \frac{2n_1 - 2}{2n} (S_1^2 - \sigma^2) + \frac{2n_2 - 2}{2n} (S_2^2 - \sigma^2) + \frac{\sigma^2}{2n - 2} \left( \frac{n_1 n_2}{2n_2^2} \Omega_+^2 - 1 \right) \right] 
\] 

\[
+ \frac{\sigma^2}{2n - 2} \left( \frac{n_1 n_2}{2n_2^2} \Omega_-^2 - 1 \right) \bigg] | S_1^2, \bar{\Delta}_1 \bigg] 
\] 

(12)

For cases \( S_1^2 \) and \( \bar{\Delta}_1 \) leading to \( n_2 = 0 \), we can go directly from the first to the last row in the above derivation. Using \( n = (1 + r(S_1^2, \bar{\Delta}_1)) n_1 \) and \( n_2/n = (1 + r(S_1^2, \bar{\Delta}_1))^{-1} \), the bias \( E(S_2^2 - \sigma^2) \) can therefore be written as

\[
\int_0^\infty \int_{-\infty}^\infty \left( n_1 - 1 + n_1 r(q, t, \delta) \right) \left( (n_1 - 1)(q - \delta^2) + \frac{r^2 n_1/2 - \sigma^2}{2 + 2r(q, t, \delta)} \right) \varphi_0 \sqrt{\frac{n_2 m}{m_2}} (t) \chi^2_{n_2 - (q'y)} y \, dq \, d\delta 
\]

(a) The integral \( \int_{-\infty}^\infty \ldots \, dt \) can again as in Theorem 1 be divided into \( \int_{-\infty}^0 \ldots \, dt \) and \( \int_0^\infty \ldots \, dt \). By replacing \( t \) with \( -t \) in the first integral, it can be written as one integral over \([0, \infty]\) in the same way as in Theorem 1.

(b) One can conclude that the bias is symmetric around 0 for \( \delta \) since \( r(q, y) = r(q, -y) \).

(c) The asymptotic unbiasedness for \( \delta \to \pm \infty \) follows from the integral expression with the dominated convergence theorem. \( \square \)

Proof of Theorem 4

For the proof of Theorem 4, we show the following lemmas:

Lemma 1. Under the null hypothesis \( \delta = 0 \)

\[
E(S_1^2 | S_{1,OS}) = E \left[ \frac{n_1}{2} \bar{\Delta}_1^2 | S_{1,OS} \right] = S_{1,OS}^2 
\]
Lemma 2. Let \( \alpha \) be arbitrary with \( 0 < \alpha < \beta \leq \infty \). For \( S_{1,OS}^2 \), the following inequality is fulfilled

\[
\text{Cov}\left\{ \frac{S_{1,OS}^2}{\text{min}(\text{max}(S_{1,OS}^2, c), d)} \right\} > \text{Cov}\left( S_{1,OS}^2, \frac{1}{S_{1,OS}^2} \right)
\]

Proof. Note that a proof for \( \beta = \infty \) and for \( S_1^2 \) instead of \( S_{1,OS}^2 \) is in Miller.\(^8\) To include the possibility of \( \beta < \infty \), we show

\[
\text{Cov}\left\{ \frac{S_{1,OS}^2}{\text{min}(\text{max}(S_{1,OS}^2, c), d)} \right\} = \text{Cov}\left( S_{1,OS}^2, \text{max}(1/S_{1,OS}^2, 1/c), 1/d \right) = \text{Cov}\left( S_{1,OS}^2, \frac{1}{S_{1,OS}^2} \right) - \text{Cov}\left( S_{1,OS}^2, (1/S_{1,OS}^2 - 1/c) \cdot 1_{|S_{1,OS}^2| < \alpha} \right) - \text{Cov}\left( S_{1,OS}^2, (1/S_{1,OS}^2 - 1/d) \cdot 1_{|S_{1,OS}^2| > \alpha} \right)
\]

As \((1/x - 1/c) \cdot 1_{x<\alpha} \) and \((1/x - 1/d) \cdot 1_{x>\alpha} \) are monotonically decreasing functions of \( x \), the second and third covariances above are \( \leq 0 \); moreover, as \( c > 0 \), the second covariance is negative.\(^{12}\) This completes the proof. \( \square \)

Proof of Theorem 4. From equation (12) in the proof of Theorem 3 follows with Lemma 1 under the null hypothesis \( \delta = 0 \)

\[
E(S^2 - \sigma^2) = E\left[ E(S^2 - \sigma^2 | S_{1,OS}^2) \right] = E\left[ \frac{n_1 - 1/2 - n_1/(2n)}{n - 1} (S_{1,OS}^2 - \sigma^2) \right]
\]

Since \( E(S_{1,OS}^2) = \sigma^2 \) and \( \frac{n_1 - 1/2 - n_1/(2n)}{n - 1} \) as function of \( S_{1,OS}^2 \) is a monotonically decreasing weight function, the bias of \( S^2 \) is non-negative. Since there is no constant \( c \) such that \( P\left( \frac{n_1 - 1/2 - n_1/(2n)}{n - 1} = c \right) = 1 \), the bias is negative and the upper bound is shown. To show the lower bound, we calculate (writing \( n_{\text{min}} = n_1 + n_{\text{min},2} \) and \( n_{\text{max}} = n_1 + n_{\text{max},2} \))

\[
E(S^2 - \sigma^2) = E\left[ \frac{n_1 - 1/2}{n - 1} (S_{1,OS}^2 - \sigma^2) \right] - E\left[ \frac{n_1}{2n(n - 1)} (S_{1,OS}^2 - \sigma^2) \right]
\]

Proof of Theorem 4. From equation (12) in the proof of Theorem 3 follows with Lemma 1 under the null hypothesis \( \delta = 0 \)

\[
E(S^2 - \sigma^2) = E\left[ E(S^2 - \sigma^2 | S_{1,OS}^2) \right] = E\left[ \frac{n_1 - 1/2 - n_1/(2n)}{n - 1} (S_{1,OS}^2 - \sigma^2) \right]
\]

Since \( E(S_{1,OS}^2) = \sigma^2 \) and \( \frac{n_1 - 1/2 - n_1/(2n)}{n - 1} \) as function of \( S_{1,OS}^2 \) is a monotonically decreasing weight function, the bias of \( S^2 \) is non-negative. Since there is no constant \( c \) such that \( P\left( \frac{n_1 - 1/2 - n_1/(2n)}{n - 1} = c \right) = 1 \), the bias is negative and the upper bound is shown. To show the lower bound, we calculate (writing \( n_{\text{min}} = n_1 + n_{\text{min},2} \) and \( n_{\text{max}} = n_1 + n_{\text{max},2} \))

\[
E(S^2 - \sigma^2) = E\left[ \frac{n_1 - 1/2}{n - 1} (S_{1,OS}^2 - \sigma^2) \right] - E\left[ \frac{n_1}{2n(n - 1)} (S_{1,OS}^2 - \sigma^2) \right]
\]
\[
\begin{align*}
\frac{n_1 - 1/2}{v} & \left[ \frac{S_{1,OS}^2 - \sigma^2}{\min((n_{\text{max}} - 1)/v, \max((n_{\text{min}} - 1)/v, S_{1,OS}^2))} \right] \quad (15) \\
\frac{n_1 - 1/2}{v} & \left( \frac{S_{1,OS}^2 - \sigma^2}{S_{1,OS}^2} \right) \quad (16) \\
\frac{n_1 - 1/2}{v} & \left\{ 1 - E \left( \frac{\sigma^2}{S_{1,OS}^2} \right) \right\} = -\frac{2n_1 - 1}{2n_1 - 3} \quad (17)
\end{align*}
\]

The last expected value in equation (14) is negative since \(1/(n(n - 1))\) as function of \(S_{1,OS}^2\) is monotonically decreasing; therefore, the total expression becomes smaller when this term is removed in equation (15). Further, we used in equation (15) the sample size formula \(n_2 = \min(n_{2,\text{max}}, \max(n_{2,\text{min}}, vS_{1,OS}^2 - n_1 + 1))\). For the inequality in equation (16), we applied Lemma 2. For the equality in equation (17), we have used \(E(\sigma^2/S_{1,OS}^2) = E(2n_1 - 1)/X = (2n_1 - 1)E(X^{-1}) = (2n_1 - 1)/(2n_1 - 3)\) where \(X\) is a chi-squared distributed random variable with \(2n_1 - 1\) degrees of freedom. The expected value of \(X^{-1}\) can be derived, for example, from Johnson et al., \(^{43}\) p. 421.

When \(\sigma^2 \to \infty\), the last expected value in equation (14) converges to 0 which implies that the difference between the expressions in equations (14) and (15) converges to 0. Further, as the probability for \(S_{1,OS}^2 > (n_1 + n_{2,\text{min}})/v\) converges to 1 in the case \(n_{2,\text{max}} = \infty\), also the difference between equations (15) and (16) converges to 0. Therefore, the lower bound becomes sharp when \(\sigma^2 \to \infty\) \(\Box\).

We note that in case of using the adjusted sample size rule \(n_2^*\) instead of \(n_2\), then we cannot go in a similar way from equations (15) to (16). Our simulation results in Figure 2 and in the lower panel of Figure 5 show that the bias for the adjusted sample size rule can be below the lower bound of Theorem 4.