Variance estimation in clinical studies with interim sample size re-estimation

Frank Miller

AstraZeneca, Clinical Science, S-15185 Södertälje, Sweden

email: frank.miller@astrazeneca.com

The definitive version is available at www.blackwell-synergy.com.

This is a preprint of an Article published 2005 in *Biometrics*, 61, 355-361.

SUMMARY. We consider clinical studies with a sample size re-estimation based on the unblinded variance estimation at some interim point of the study. Since the sample size is determined in that flexible way, the usual variance estimator at the end of the trial is biased. We derive sharp bounds for this bias. These bounds have a quite simple form and can help for the decision if this bias is negligible for the actual study or if a correction should be done. An exact formula for the bias is also provided. We discuss possibilities to get rid of this bias or at least to reduce the bias substantially. For this purpose, we propose a certain additive correction of the bias. We see in an example that the significance level of the test can be controlled when this additive correction is used.

KEY WORDS: Adaptive design; Bias of variance estimation; Bounds for bias; Clinical studies; Interim sample size re-estimation; Variance estimation.

1 Introduction

Since it is important that clinical studies are adequately powered, the sample size of a study is usually determined with some sample size formula before the study starts. However, sample size formulas depend on parameters, such as assumed treatment difference and variance, and one has to have information on these parameters before study start. But often is the information on these parameters not very good before the study. Hence, flexible designs where the design can be adapted using information of the first part of the study became recently very popular.

That adaptive designs have great advantages in certain situations is in the meantime widely accepted both by the pharmaceutical industry and the regulatory authorities. However, from a statistical perspective, an investigation of such designs is not at all trivial. After a careful inspection it becomes often unclear what happens to properties of estimators or to the type I error and power of tests if a certain design-adaptation rule is used during the study.

Let us consider a quite simple situation: a parallel design for two treatment groups with normally distributed outcomes and equal variances for both groups. Let us assume that we have only uncertain or no information about the true variance. Therefore we estimate after a certain number n_1 of patients in each group the variance with the usual two-sample variance estimator (which requires that we know to which treatment the patients were assigned to, that is we unblind the treatment codes from the first $2n_1$ patients). With this estimation, we calculate a sample size for the whole study which should provide a certain power based on the information we got from the first $2n_1$ patients. We call the part of the study with the first $2n_1$ patients first stage of the study. If the variance estimator has a low value a small study should be sufficient to

achieve a certain power, if the variance estimator has a high value we need certainly a larger study. Then, at the end of the study we estimate the variance with the usual variance estimator S^2 for two groups. A nice property of this variance estimator S^2 is common to us: in fixed designs without sample size adaptation it is an unbiased estimator of the variance. But the situation changes for the here described adaptive design with sample size re-estimation: this estimator is then negatively biased, it underestimates in average the true variance! This fact is well known in the literature. Unfortunately, it is very little known about the amount of bias. Some calculations were done in Wittes *et. al.* (1999), but their bound for the bias is not very sharp. According to Wittes *et. al.* (1999), we call this usual variance estimator S^2 the naive variance estimator when used in uncorrected form in the here described sample size re-estimation procedure.

For some studies, for example studies of more explorative nature in the early phase of drug development, it may be no problem to accept a biased estimation as long as the bias is not too big. We present in Section 4 a sharp bound for this bias, derived by algebraic calculations. The bound has surprisingly a very simple form. With this bound, we get help to judge if we can live with this bias or not. We develop also an exact formula for the bias. However the shape of the exact bias is not that nice as the bound developed before.

In some studies, for example studies in phase III for registration purpose of a drug, it may be important to adjust for any bias induced by the adaptation in the design. We discuss possibilities to get rid of this bias or at least to reduce the bias substantially in Section 5. For this purpose, we propose a new method, the so-called variance estimator with additive correction of the bias.

Why is the variance estimator S^2 so important for us? One thing is that in the

t-test statistic usually used in this model for the treatment comparison, it is divided by the square root of the variance estimator. Due to the negative bias of S^2 in the above described adaptive design, the type I error increases over the nominal size α if no adjustment of the critical value is done. An early paper dealing with this adaptive test problem is Stein (1945). Stein uses the unbiased S_1^2, the variance estimator of the first stage, in the final test statistic instead of S^2. With this choice, his test strongly controls both the significance level α and the power. Proschan and Wittes (2000) describe an improvement of Steins procedure.

Even if we do not have the test in mind, the variance estimator is important itself: Clinical projects for drug development consist usually of a lot of single studies. Hence, it is desirable to have a good variance estimator which can be used as prior information for sample size calculation for comparable studies in the future. In this paper, we focus on the variance estimator itself. However, we see in Section 6 for an example that the significance level of the final test can be controlled if we use the variance estimator with additive correction instead of the naive variance estimator in the test statistic. In the discussion we give some recommendations. Technical proofs are postponed to the appendix.

In the next section, we discus differences between unblinded and blinded sample size re-estimation before we describe the model more formally in Section 3.

2 Unblinded or blinded sample size re-estimation

In this paper, we focus on sample size re-estimation based on the usual two-sample variance estimator S_1^2 after the first stage, i.e.

$$S_1^2 = \frac{1}{2n_1 - 2} \sum_{i=1}^{2} \sum_{j=1}^{n_1} \left(Y_{ij} - \bar{Y}_{i\cdot}^{(1)} \right)^2,$$

where Y_{ij} is the outcome of patient number j in treatment group i, $i \in \{1, 2\}$ and $\bar{Y}_{i\cdot}^{(1)}$ is the mean of the n_1 observations for treatment i in the first stage. For this estimator it is necessary to know the treatments to which the patients of the first stage were assigned. This means that it is necessary for blinded studies to unblind them.

An alternative to the unblinded sample size re-estimation procedure considered in this paper is to use blinded sample size re-estimation procedures. Among several possibilities, the investigation of Friede and Kieser (2002) suggests that the best blinded approach is the sample size re-estimation with the one-sample variance proposed by Gould and Shih (1992) or with an adjusted version of the one-sample variance proposed by Zucker *et. al.* (1999). Using numerical calculation, Kieser and Friede (2003) showed that the type I error is approximately controlled with these procedures. Zucker *et. al.* (1999) report power comparisons with some unblinded procedures.

The use of the one-sample variance has the disadvantage that it is increasing with an increasing treatment difference. This implies unnecessary big sample sizes if the true treatment difference is bigger than the pre-specified difference in the sample size formula. However, Friede and Kieser (2001) argue that the inflation in sample size for the one-sample variance procedure is not very big as long as the true treatment difference is only slightly larger than the pre-specified difference.

On the other hand, with an unblinded procedure there may be a risk that information about the observed treatment difference at the interim analysis reaches the investigators. Knowing the observed difference, the investigators could be influenced in their choice of future patients or in the way of taking care of their patients. Therefore, one should try to minimize this risk. For example, only an independent statistician should be unblinded to perform the interim analysis. Further, since the within treatment variability can be calculated with the within group variability and the total

variability, it should be avoided if possible that somebody has access to data which allows the computation of the total variance.

The decision for a blinded or an unblinded procedure should be done on a case-to-case basis. In the following example, we prefer strongly an unblinded procedure.

Example 2.1 *Let us assume that we have a study with three treatment arms: a new treatment developed by the sponsor company of the study, an active comparator and placebo. The primary objective is to show superiority of the new treatment over the active comparator using a two-sample t-test. The additional placebo arm not needed for primary purpose is useful to judge the sensitivity of the study to identify treatment differences. In historical studies it was shown that the active comparator is better than placebo. If the actual study is not even able to re-prove this, the possible acceptance of the primary null hypothesis could question the study itself and not necessarily the performance of the new treatment.*

In this example, the final two-sample t-test uses no information from the placebo arm, since the variance in this arm might differ from the variance in the other arms. We want also the interim variance estimation not to be influenced from the placebo arm to get a proper sample size re-estimation. Hence, it is necessary that an independent statistician is unblinded to identify placebo patients. Further, due to the placebo arm, it is not possible to calculate the primary within treatment variability with the interim estimation and the total variability. In such a situation, sample size re-estimation with two-sample variance estimation (unblinded) is preferable compared to one-sample variance estimation. This design has to be analysed as two-sample problem and hence the theory of this paper applies directly to this example.

3 Model and notation

We consider in this paper a parallel two-group trial with normally distributed outcomes, i.e. $Y_{ij} \sim N(\mu_i, \sigma^2)$. The mean values μ_1 and μ_2 and the variance σ^2 are unknown. The main aim of the study is the test of $H_0 : \mu_1 = \mu_2$ against the alternative $H_1 : \mu_1 \neq \mu_2$ with a test of size α and, for $|\mu_1 - \mu_2| = \Delta$, power $1 - \beta$. The values α (usually 0.05), β (usually 0.1 or 0.2) and Δ (expected treatment difference or minimal clinical relevant difference) are assumed to be specified in advance.

We observe $n_1 \geq 3$ patients per treatment. Then, the variance is estimated by the usual two-sample variance estimator S_1^2. This estimator is used to compute the final sample size of n patients per treatment and additional $n_2 = n - n_1$ patients per treatment will be observed. A popular sample size formula is the formula for known variance, replacing the true variance with the estimator S_1^2, namely

$$n = vS_1^2$$

with

$$v = \frac{2}{\Delta^2} \left\{ \Phi^{-1}(1 - \alpha/2) + \Phi^{-1}(1 - \beta) \right\}^2 . \tag{3.1}$$

We modify here the formula $n = vS_1^2$ slightly to $n = vS_1^2 + 1$ since the later calculations show that it is appropriate to add 1. However, we ignore here the fact that we have to round in practice n to an integer value.

Generally, if the above calculation yield to a n lower than n_1, we would stop the study immediately but analyze all patients in the study. Moreover, it is a quite common situation that the treatment time in the study is some weeks or months. Hence, more than n_1 patients per treatment may be already in our study when the results of n_1 are available. Further, performing the interim sample size re-estimation may require some days or weeks, which also increases the number of patients already in the study.

The results of these patients should be included in the final analysis independent of the result of the sample size formula. To include this phenomenon in our sample size formula, we generalize the formula to

$$n = \max(v \cdot S_1^2 + 1, n_1 + n_{2\,\min}) \tag{3.2}$$

where $n_{2\,\min} \geq 0$ is arbitrary. In the sequel, we consider the sample size formula (3.2) where $n_1, n_{2\,\min}, v$ are specified in advance. The factor v may be calculated by (3.1), but our considerations are not restricted to this formula for v.

Note here that the so called restricted sample size re-estimation investigated by Wittes and Brittain (1990) can be seen as special case of (3.2). They recommend to calculate an initial sample size based on a vague guess of the variance, σ_0^2, say, and then in the interim re-estimation only to update the sample size upwards if necessary. In our notation, they use $n_{2\,\min} = v\sigma_0^2 - n_1$.

At the end of the trial (after the observation of n patients for every treatment), we compute the naive estimator for the variance in the whole study

$$S^2 = \frac{1}{2n-2} \sum_{i=1}^{2} \sum_{j=1}^{n} (Y_{ij} - \bar{Y}_{i\cdot})^2$$

where $\bar{Y}_{i\cdot}$ is the mean of all n observations for treatment i. In this paper, we investigate properties of S^2, especially the bias of this estimator.

4 Bounds and exact formula for the bias of the variance estimator

The fact that the naive variance estimator S^2 is negatively biased can be explained in the following way: If we underestimate the true variance after n_1 patients in each

group (small S_1^2), then n_2 will be small and hence it will be hard to correct the underestimated variance within the first $2n_1$ patients by the last $2n_2$ patients. If we overestimate the true variance after n_1 patients in each group (large S_1^2), then n_2 will be large and hence it will be easier to correct the overestimated variance within the first $2n_1$ patients by the last $2n_2$ patients. Bounds for the bias of S^2 are provided in the following theorem.

Theorem 4.1 *The bias of the naive variance estimator has the following bounds:*

$$-\frac{n_1-1}{n_1-2}\cdot\frac{1}{v} \leq ES^2-\sigma^2 \leq 0.$$

The proof is contained in the appendix. Note that the proof of Theorem 4.1 shows that we would have exactly $ES^2-\sigma^2 = -\frac{n_1-1}{n_1-2}\cdot\frac{1}{v}$, if we would use $n = vS_1^2+1$ instead of (3.2), $n = \max(v\cdot S_1^2+1, n_1+n_{2\,\min})$. Hence, we expect that this lower bound will be nearly attained if the probability that $vS_1^2+1 \geq n_1+n_{2\,\min}$ is high. This is the case for large σ^2.

Example 4.2 *Let us assume that we plan a study where a minimal clinical relevant difference of* $\Delta = 1$ *between the two treatments has to be detected with a power of* $1-\beta = 0.9$. *The two-sided test has to have a significance level of* $\alpha = 0.05$. *For the variance, we have only a very vague guess: it may be about 16 (standard deviation=4). A fixed sample size trial would be planned with the sample size of about* $v\cdot 16$ *with* $v = \frac{2}{\Delta^2}\left\{\Phi^{-1}(1-\alpha/2)+\Phi^{-1}(1-\beta)\right\}^2 \approx 21.016$. *Hence, about 336 patients per treatment would be necessary. Since the variance guess was very vague, we choose an adaptive design with interim sample size re-estimation and update the sample size after* $n_1 = 168$ *patients per treatment. For the bias of the naive variance estimator, we have the lower bound* $-(n_1-1)/\{(n_1-2)v\} \approx -0.0479$. *If the true variance is in reality not very*

much lower than our vague guess of 16, we may conclude that the bias of the naive variance estimator is negligible in this situation.

We state in the following theorem a formula for the exact bias.

Theorem 4.3 *Let* $d = (2n_1 - 2)(n_1 + n_{2\,\min} - 1)/(v\sigma^2)$ *and* F_ν *be the distribution function of the chi-square distribution with* ν *degrees of freedom. Then the bias* $ES^2 - \sigma^2$ *of the naive variance estimator is*

$$\frac{2(n_1-1)^2}{vd}\left\{F_{2n_1}(d) - F_{2n_1-2}(d)\right\} + \frac{n_1-1}{v}\left\{1 - F_{2n_1-2}(d)\right\} - \frac{(n_1-1)^2}{v(n_1-2)}\left\{1 - F_{2n_1-4}(d)\right\}.$$

The proof can be found in the appendix.

Wittes *et. al.* (1999) have computed the following bounds for the bias of S^2:

$$-\sigma^2/\sqrt{n_1 - 1} \leq ES^2 - \sigma^2 \leq 0.$$

Their lower bound applies for arbitrary sample size formulas with $n \geq n_1$, not only for (3.2). However, it is usually a very rough bound, which can be seen in the following example.

Example 4.4 *Let* $n_1 = 20, n_{2\,\min} = 10$ *and* $v = 4.3421$. *The value of* v *is the result of formula (3.1) with* $\alpha = 0.05, 1 - \beta = 0.9$ *and* $\Delta = 2.2$. *Figure 1 shows the bias of* S^2, *the bounds of Theorem 4.1 and the lower bound of Wittes et. al. (1999) as a function of the true* σ^2.

5 An 'almost' unbiased variance estimator

The proof of Theorem 4.1 shows that the lower bound $-\frac{n_1-1}{n_1-2} \cdot \frac{1}{v}$ would be the exact bias, if $P(v \cdot S_1^2 + 1 > n_1 + n_{2\,\min}) = 1$. On the other hand it is clear that we would have

no bias if $P(v \cdot S_1^2 + 1 \leq n_1 + n_{2\,\min}) = 1$. Even if this is only a theoretic consideration since these probabilities will be never exact 1, it suggests the following correction of the variance estimator:

$$S^2_{\text{ac}} = \begin{cases} S^2 + \frac{n_1-1}{n_1-2} \cdot \frac{1}{v}, & \text{if } n > n_1 + n_{2min}, \\ S^2, & \text{if } n = n_1 + n_{2min}. \end{cases}$$

We call this estimator 'Variance estimator with additive correction'. Using Theorem 4.3, we are able to calculate the bias of S^2 and S^2_{ac} for given values of $n_1, n_{2\,\min}, \sigma^2$ and v.

Example 5.1 *Let us consider again* $n_1 = 20, n_{2\,\min} = 10$ *and* $v = 4.3421$. *Figure 2 shows the bias of* S^2 *and* S^2_{ac} *as a function of the true* σ^2.

The figure shows that we can reduce the bias substantially with the additive correction. Furthermore, the remaining bias is a bias in the other direction. An overestimation of the variance is often preferable to an underestimation since it results usually in a more conservative interpretation of study results.

Of course, the question of existence or amount of bias is not the only criterion if we look at properties of estimators. Therefore, we investigate the variability of the estimators and compare this variability with other estimators. More precisely, we compare the standard deviation of different variance estimators by simulations. We compare the naive variance estimator S^2 and the variance estimator with additive correction S^2_{ac} with the variance estimator proposed by Proschan and Wittes (2000). This estimator can be defined as weighted sum of the variance estimator of the variance estimator in the first stage S_1^2 and the 'rest' S_*^2 with

$$S_*^2 = \frac{n-1}{n-n_1} S^2 - \frac{n_1-1}{n-n_1} S_1^2$$

(S_*^2 is the 'rest' since S^2 can be decomposed into $(n_1-1)/(n-1)S_1^2+(n-n_1)/(n-1)S_*^2$). The variance estimator of Proschan and Wittes is then

$$S_{\mathrm{PW}}^2 = \frac{n_1 - 1}{n_1 + n_{2\,\min} - 1} S_1^2 + \frac{n_{2\,\min}}{n_1 + n_{2\,\min} - 1} S_*^2.$$

This estimator is an unbiased estimator for the variance and it is the best estimator in the class of estimators $\lambda S_1^2 + (1-\lambda)S_*^2$ with $\lambda \in [0,1]$ fixed. It is worth mentioning that $S_{\mathrm{PW}}^2 = S_1^2$ if $n_{2\,\min} = 0$.

We include in our simulations of the standard deviation two other unbiased variance estimators: the variance estimator of the first stage S_1^2 (which was used by Stein (1945) in the test statistic at the end of the trial) and the variance estimator of the second stage S_2^2. Note that S_2^2 is well defined if $n_2 \geq 2$ which is fulfilled if $n_{2\,\min} \geq 2$.

Example 5.2 *Let us consider again $n_1 = 20, n_{2\,\min} = 10$ and $v = 4.3421$. We have done 20 000 simulations for each of $\sigma^2 \in \{2, 4, 6, \ldots, 24\}$ to compute the standard deviation of the different variance estimators.*

We see in Figure 3 that the naive variance estimator S^2 has the lowest standard deviation. The additive correction in the estimator S_{ac}^2 has almost no influence on the standard deviation, it is only slightly higher than the standard deviation of S^2 in a small region of σ^2 values. The estimator of Proschan and Wittes has a higher standard deviation than S^2 and S_{ac}^2, especially for high σ^2 when n_2 usually is much bigger than $n_{2\,\min}$. The variance estimator which uses only the information of the first stage, S_1^2, has a much higher standard deviation than S^2, S_{ac}^2 and S_{PW}^2 for all σ^2 values. The variance estimator S_2^2 has a relatively high standard deviation for small values of σ^2. For large values of σ^2 it becomes better than S_{PW}^2 but remains worse than S^2 and S_{ac}^2 due to the fact that the first stage information is not used.

6 The test with the corrected variance estimator

Usually, at the end of the study a test is performed for the null hypothesis that the two treatments have equal effect $H_0 : \mu_1 = \mu_2$ against the alternative that the effect is different $H_1 : \mu_1 \neq \mu_2$. This is in fixed designs usually done with the t-test

$$t = \frac{\bar{Y}_{1\cdot} - \bar{Y}_{2\cdot}}{\sqrt{2S^2/n}}, \qquad \text{reject } H_0 \text{ if and only if } |t| > t_{2n-2,1-\alpha/2}, \tag{6.1}$$

where $t_{\nu,\gamma}$ is the γ-quantile of the t-distribution with ν degrees of freedom. In the adaptive design with interim sample size re-estimation considered in this paper, the naive variance estimator S^2 is negatively biased and therefore the type I error of the 'naive' t-test (6.1) increases over the nominal size α. This can be corrected using an adjusted critical value $t_{2n-2,1-\alpha_{adj}/2}$ instead of $t_{2n-2,1-\alpha/2}$, see Kieser and Friede (2000).

Another possibility to adjust for the influence of the interim sample size re-estimation is to use the variance estimator with additive correction S^2_{ac} instead of S^2 in the t-test statistic:

$$t_{ac} = \frac{\bar{Y}_{1\cdot} - \bar{Y}_{2\cdot}}{\sqrt{2S^2_{ac}/n}}, \qquad \text{reject } H_0 \text{ if and only if } |t_{ac}| > t_{2n-2,1-\alpha/2}.$$

We see in the following example by simulations, that the additive correction for the bias in the variance estimator is enough to bring the type I error of the test back to the nominal level α. Even if an investigation of the test statistic is not the main scope of this paper, the example underlines the usefulness of the variance estimator with additive correction.

Example 6.1 *Let us consider again* $n_1 = 20, n_{2\,\min} = 10$ *and* $v = 4.3421$ *(appropriate for* $\alpha = 0.05, 1 - \beta = 0.9, \Delta = 2.2$*). We have done 4 000 000 simulations for each of* $\sigma^2 \in \{2, 4, 6, \ldots, 24\}$ *to compute the type I error of the t-test and the* t_{ac}*-test. The*

type I errors are shown in Figure 4. The t-test has a maximal type I error of 0.0526 *for* $\sigma^2 = 10$. *Using the* t_{ac}*-test, i.e. using simply the additive correction in the variance estimator of the t-test, the maximal type I error is reduced to (or at least very near to) the nominal level of* α. *Hence for this example, the use of the* t_{ac}*-test instead of the t-test should be sufficient to fulfill the requirement of the ICH guideline E9 (1998) for clinical studies: "... the consequences, if any, for the type I error ... should be explained". Moreover, the type I error of the* t_{ac}*-test as function of the true variance is a quite flat curve which implies that the test is not conservative for some values of the unknown variance* σ^2.

7 Discussion

The bounds for the bias of the naive variance estimator in designs with interim re-estimation of the sample size derived in this paper are quite sharp and simple. Hence, they are useful for the judgment if the bias is negligible in a certain situation or not. If an adjustment for the effects of the interim re-estimation is required, we recommend doing an additive correction on the variance estimator, which reduces the bias with minor impact on the standard deviation of the variance estimator. In the example considered in this paper, it was satisfactory (i.e. the significance level was controlled) to use the variance estimator with additive correction instead of the naive variance estimator in the test statistic.

Acknowledgement

The author likes to thank an unknown associate editor for comments improving the presentation of the paper.

References

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

Friede, T. and Kieser, M. (2002). On the inappropriateness of an EM algorithm based procedure for blinded sample size re-estimation. *Statistics in Medicine* **21**, 165-176.

Gould, A. and Shih, W. (1992). Sample size re-estimation without unblinding for normally distributed outcomes with unknown variance. *Communication in Statistics – Theory and Methods* **21**, 2833-2853.

ICH guideline E9 (1998). International conference on harmonisation of technical requirements for registration of pharmaceuticals for human use. ICH topic E9: statistical principles for clinical trials. ICH technical coordination, EMEA, London.

Kieser, M. and Friede, T. (2000). Re-calculating the sample size in internal pilot study designs with control of the type I error rate. *Statistics in Medicine* **19**, 901-911.

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

Johnson, N.L., Kotz, S. and Balakrishnan, N. (1994). *Continuous univariate distributions,* Volume 1, Second ed. Wiley, New York.

Proschan, M. A. and Wittes, J. (2000). An improved double sampling procedure based on the variance. *Biometrics* **56**, 1183-1187.

Stein, C. (1945). A two-sample test for a linear hypothesis whose power is independent of the variance. *Annals of Mathematical Statistics* **16**, 243-258.

Wittes, J. and Brittain, E. (1990). The role of internal pilot studies in increasing the efficiency of clinical trials. *Statistics in Medicine* **9**, 65-72.

Wittes, J., Schabenberger, O., Zucker, D., Brittain, E., Proschan, M. (1999). Internal pilot studies I: type I error rate of the naive t-test. *Statistics in Medicine* **18**, 3481-3491.

Zucker, D., Wittes, J., Schabenberger, O. and Brittain, E. (1999). Internal pilot studies II: comparison of various procedures. *Statistics in Medicine* **18**, 3493-3509.

A Proofs

Lemma A.1 *Let $c > 0$ be arbitrary. Then we have for the variance estimator S_1^2 of the first stage the inequality*

$$Cov\left\{S_1^2, \frac{1}{\max(S_1^2, c)}\right\} > Cov\left(S_1^2, \frac{1}{S_1^2}\right)$$

Proof: Let g be the Lebesgue density of S_1^2. We have

$$0 = \frac{1}{c}\int_0^\infty (x-\sigma^2)g(x)dx > \frac{1}{c}\int_0^c (x-\sigma^2)g(x)dx > \int_0^c \frac{1}{x}(x-\sigma^2)g(x)dx \qquad \text{(A.1)}$$

where the first equation follows since S_1^2 is unbiased for σ^2 and the last inequality is clear in the case $c \leq \sigma^2$. In the case $\sigma^2 < c$, the last inequality follows by

$$\frac{1}{c}\underbrace{\int_0^c (x-\sigma^2)g(x)dx}_{<0} > \frac{1}{\sigma^2}\int_0^c (x-\sigma^2)g(x)dx > \int_0^c \frac{1}{x}(x-\sigma^2)g(x)dx.$$

The expression (A.1) implies

$$\int_0^\infty (x-\sigma^2)\frac{1}{\max(x,c)}g(x)dx > \int_0^\infty (x-\sigma^2)\frac{1}{x}g(x)dx$$

which is equivalent to the assertion of the lemma. □

Lemma A.2 *(Wittes et. al. (1999)) The bias of the variance estimator S^2 is given by*

$$E(S^2-\sigma^2) = (n_1-1)\,Cov\left(S_1^2, \frac{1}{n-1}\right)$$

Proof: Wittes *et. al.* (1999). □

Proof of Theorem 4.1: The upper bound, 0, is well known, see for example Wittes *et. al.* (1999). For the lower bound, we use the preceding lemmas:

$$\begin{aligned} E(S^2-\sigma^2) &= (n_1-1)\mathrm{Cov}\left(S_1^2, \frac{1}{n-1}\right) \\ &= (n_1-1)\mathrm{Cov}\left\{S_1^2, \frac{1}{\max(v\cdot S_1^2, n_1+n_{2\min}-1)}\right\} \\ &> (n_1-1)\mathrm{Cov}\left(S_1^2, \frac{1}{v\cdot S_1^2}\right) \\ &= \frac{n_1-1}{v}\left\{1-E\left(\frac{\sigma^2}{S_1^2}\right)\right\} = -\frac{n_1-1}{n_1-2}\cdot\frac{1}{v}. \end{aligned}$$

For the last equality, we have used $E\left(\sigma^2/S_1^2\right) = E\left\{(2n_1-2)/X\right\} = 2(n_1-1)EX^{-1} = (n_1-1)/(n_1-2)$ where X is a chi-squared distributed random variable with $2n_1-2$ degrees of freedom. The expected value of X^{-1} can be derived, for example, from Johnson, Kotz and Balakrishnan (1994), p.421. □

Proof of Theorem 4.3: Using Lemma A.2, we have

$$\begin{aligned} E(S^2-\sigma^2) &= (n_1-1)\mathrm{Cov}\left(S_1^2, \frac{1}{n-1}\right) \\ &= (n_1-1)E\left(\frac{S_1^2-\sigma^2}{n_1+n_2-1}\right) \\ &= (n_1-1)E\left\{\frac{S_1^2-\sigma^2}{n_1+n_{2\min}-1}\mathbf{1}\left(S_1^2 \le \frac{n_1+n_{2\min}-1}{v}\right)\right\} \\ &\quad +(n_1-1)E\left\{\frac{S_1^2-\sigma^2}{vS_1^2}\mathbf{1}\left(S_1^2 > \frac{n_1+n_{2\min}-1}{v}\right)\right\} \\ &= \frac{n_1-1}{vd}E\left\{X\mathbf{1}(X\le d)\right\} - \frac{2(n_1-1)^2}{vd}P(X\le d) \\ &\quad +\frac{n_1-1}{v}P(X>d) - \frac{2(n_1-1)^2}{v}E\left\{\frac{1}{X}\mathbf{1}(X>d)\right\}. \end{aligned} \tag{A.2}$$

Here, $\mathbf{1}$ is the indicator function and X is again a chi-squared distributed random variable with $2n_1-2$ degrees of freedom. Let f_ν be the density of the chi-square

distribution with ν degrees of freedom and $c > 0$ arbitrary. It can be shown that

$$\begin{aligned}\int_0^c x f_{2n_1-2}(x)dx &= (2n_1-2)F_{2n_1}(c),\\ \int_c^\infty \frac{1}{x} f_{2n_1-2}(x)dx &= \frac{1}{2n_1-4}\left\{1-F_{2n_1-4}(c)\right\}.\end{aligned}$$

Therefore, we see that (A.2) is equal to the formula in Theorem 4.3 and hence the assertion of the theorem is shown. □

List of Figures

Figures

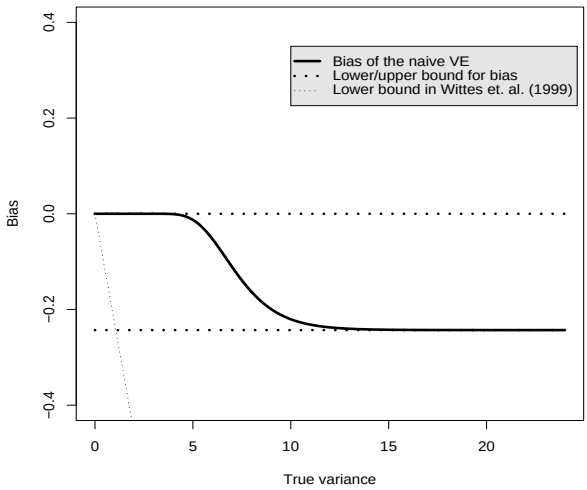

Figure 1: Bias of the naive variance estimator and bounds ($n_1 = 20$, $n_{2\,\min} = 10$, $v = 4.3421$).

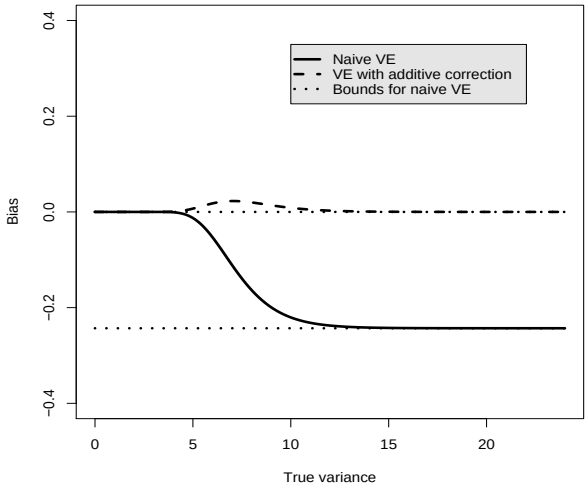

Figure 2: Bias of the naive variance estimator and the variance estimator with additive correction ($n_1 = 20$, $n_{2\,\min} = 10$, $v = 4.3421$).

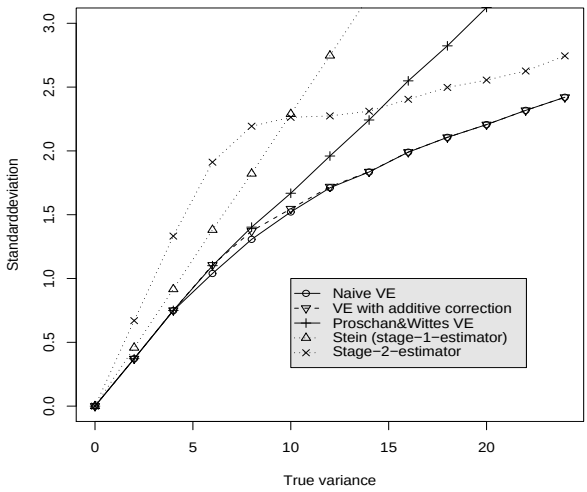

Figure 3: Standard deviation of different variance estimators ($n_1 = 20$, $n_{2\,\min} = 10$, $v = 4.3421$).

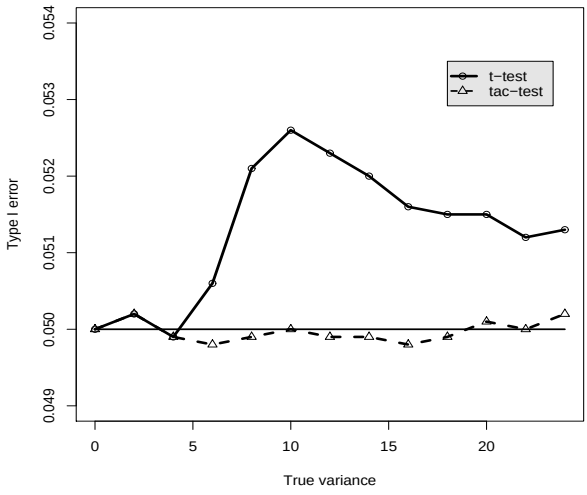

Figure 4: Type I error of t-test and t_{ac}-test using the variance estimator with additive correction ($n_1 = 20$, $n_{2\,\min} = 10$, $\alpha = 0.05$, $\beta = 0.9$, $\Delta = 2.2$, $v = 4.3421$, 4 000 000 simulations).