Sample size Re-estimation and Bias

Medical University of Vienna, January 24, 2013

Frank Miller, Stockholm University, Sweden

Partly joint work with
Tim Friede, Dept. of Medical Statistics,
University Medical Center Göttingen, Germany

Content

- Sample size re-estimation, bias and correction
- Continuous variance monitoring
- Blinded continuous variance monitoring
- More insight into the bias and it's consequences

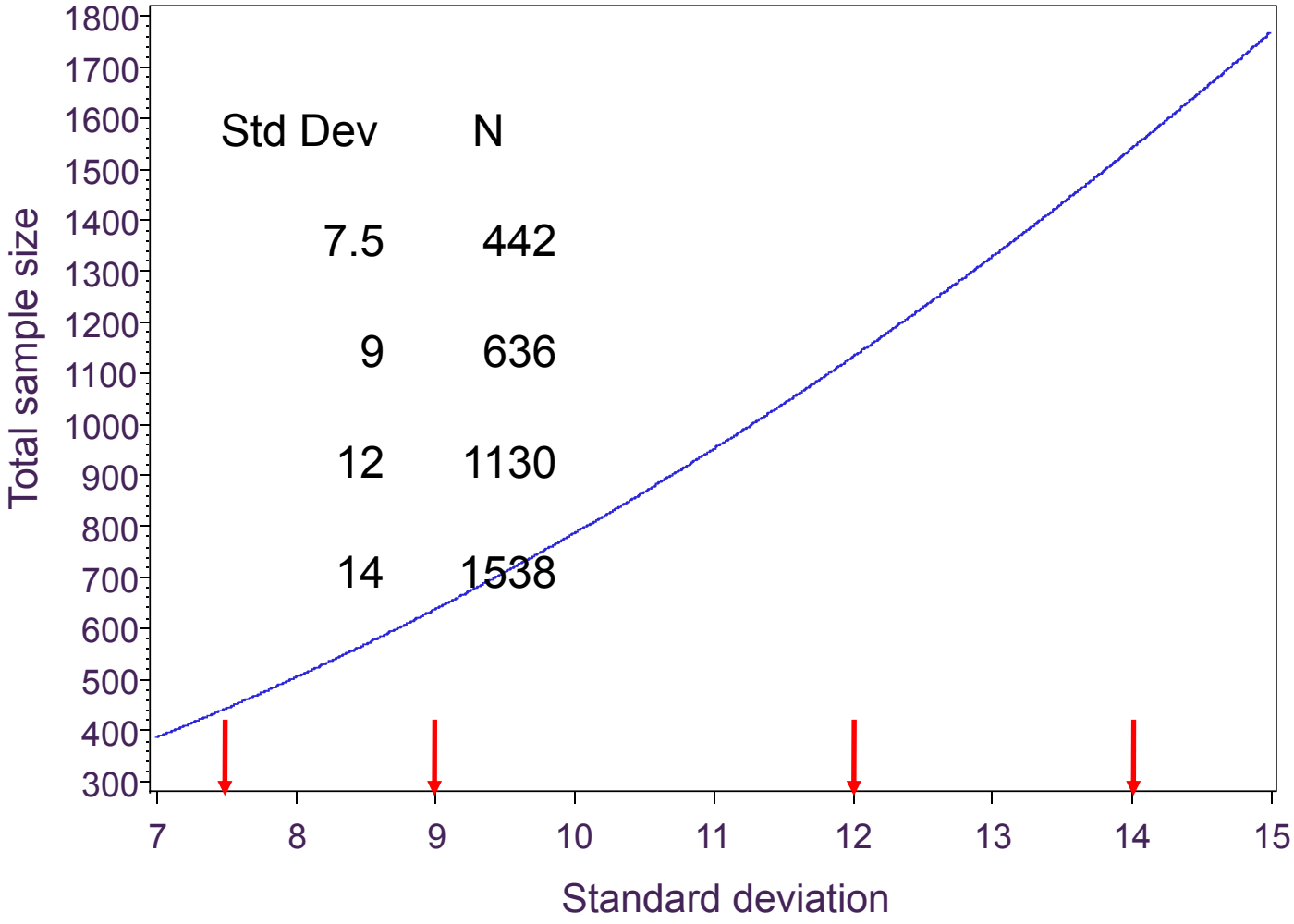
Clinical study and sample size calculation

- Randomized study comparing two treatments (e.g. new treatment versus control)
- A variable is observed which is assumed to be independently **normally distributed** with common unknown variance σ^2 and treatment difference δ
- Significance test with level α and power $1-\beta$ for $\delta = \delta_a$,
  $H_0 : \delta = 0$ (new treatment equal to control) versus
  $H_a : \delta > 0$ (new treatment better than control)
- Sample size per treatment $n = v\sigma^2,\ v = 2\left(z_{1-\alpha} + z_{1-\beta}\right)^2 / \delta_a^2$
- Since σ^2 is unknown, an initial guess required

Example: MacDonald et al. (2008)

- **Objective:** Assessment Lumiracoxib's effect on blood pressure in Osteoarthritis patients with hypertension
- **Treatments:** Lumiracoxib or Ibuprofen
- **Primary endpoint:** change from baseline at week 4 in average 24 h systolic blood pressure
- Significance level α=0.025 (1-sided), power 80% for $\delta_a = 2$ mmHg
- **Standard deviation σ= ??? mmHg**
  - White et al (2002): **9** mmHg observed (but different population)
  - Sowers et al (2005): assumed **7.5**, observed **12** mmHg (6 week follow-up)
  - Other studies in non-OA population: up to **14** mmHg

Uncertainty in the planning phase



Uncertainty about variance

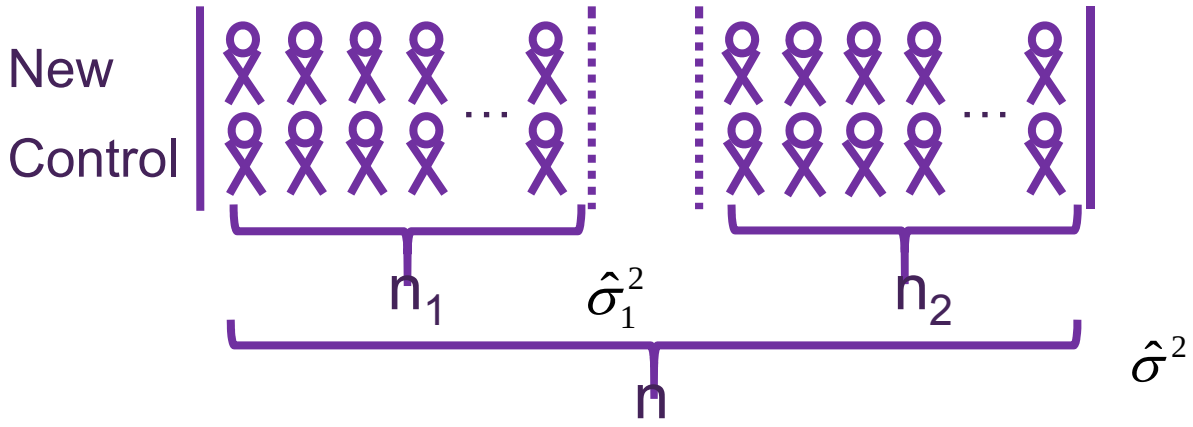
- Problem: Often considerable uncertainty regarding the variance
- Solutions include:
  - **One interim look**: estimate σ^2 from accruing data in an interim sample size review and use this estimate to adjust sample size
  - **Several interim looks** with update of variance estimate and sample size
  - **Continuous monitoring** of the variance

Sample size based on one interim analysis

Observe n_1 patients ("Stage 1") & estimate variance: $\hat{\sigma}_1^2$

Final sample size $n = v\hat{\sigma}_1^2$

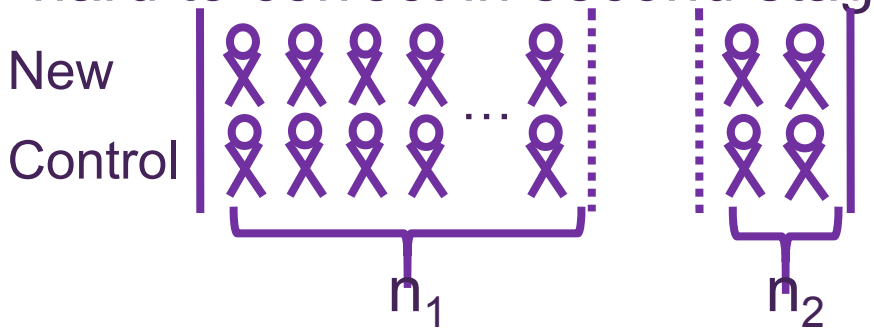
Observe additional n_2=n-n_1 patients ("Stage 2")



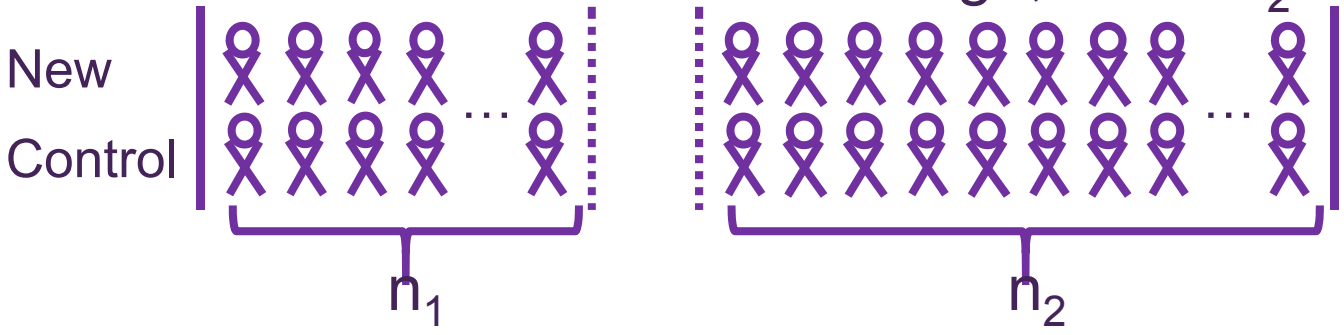
- The variance estimator $\hat{\sigma}^2$ at the end of the trial is biased. It underestimates the true variance!

Why is there a bias?

$\hat{\sigma}_1^2$ lower than true σ^2
$\rightarrow$ hard to correct in second stage, since n_2 low



- $\hat{\sigma}_1^2$ larger than true σ^2
  $\rightarrow$ easier to correct in second stage, since n_2 large



➢ $\hat{\sigma}^2$ underestimates the variance

Consequence of the bias in the final variance estimator

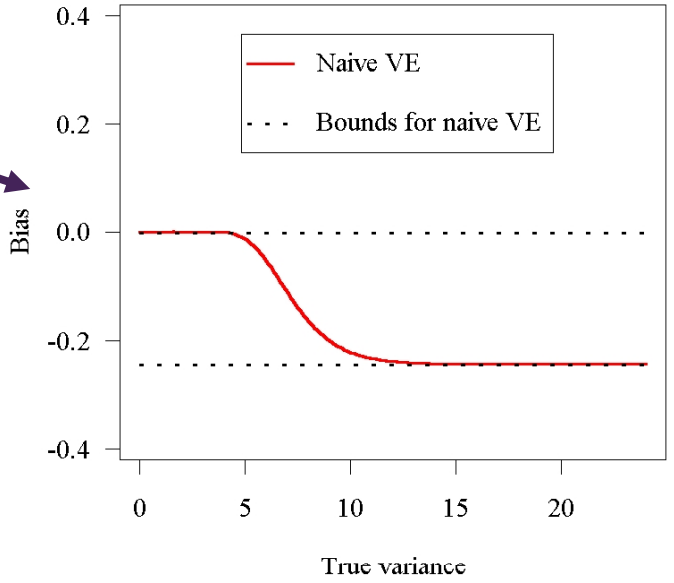
- The t-test Reject $H_0 \Leftrightarrow t > t_{2n-2,1-\alpha}$ with

  $t = \hat{\delta} \Big/ \sqrt{2\hat{\sigma}^2 / n}$ does not control the alpha level

How large is the bias?

- Bias of the ("naïve") variance estimator can be computed
- Bounds for bias (Miller, 2005)

$$-\frac{n_1-1}{n_1-2}\cdot\frac{1}{v}\le E\hat{\sigma}^2-\sigma^2\le 0$$



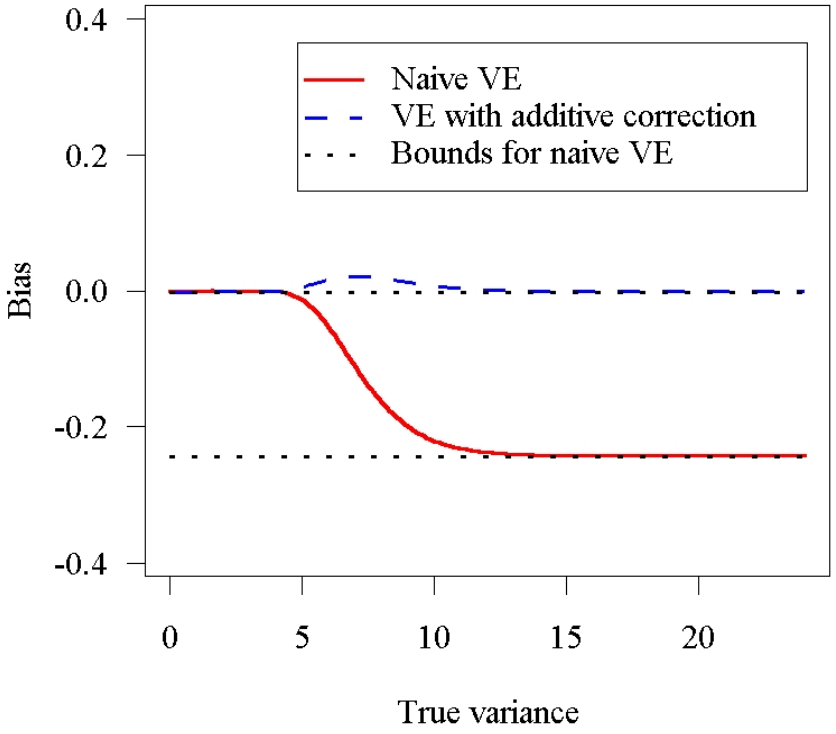
recall: $n=v\hat{\sigma}_1^2,\ v=2\left(z_{1-\alpha}+z_{1-\beta}\right)^2/\delta_a^2$

- Additive correction of final variance estimate:

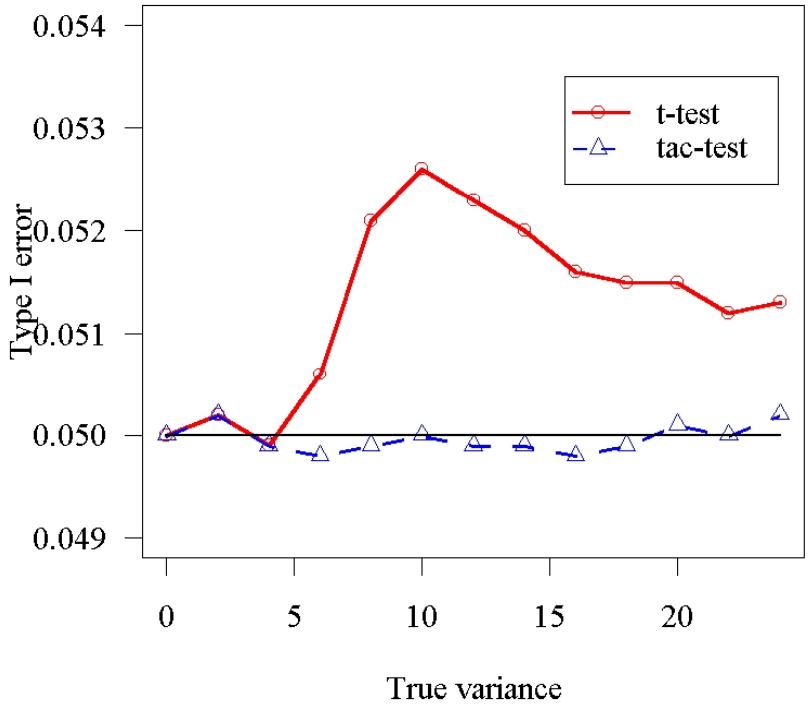
$$\hat{\sigma}_{ac}^2=\hat{\sigma}^2+\begin{cases}0, & \text{if stopped directly after interim}\\ \dfrac{n_1-1}{n_1-2}\cdot\dfrac{1}{v}, & \text{otherwise}\end{cases}$$

Correction of bias

Bias of variance estimator



Type I error of t-test and t-test with additive correction



Continuous variance monitoring

- **Continuous variance monitoring procedure**:
  - Monitor variance after each (pair of) patients: $\hat{\sigma}_n^2$ starting after n_1 patients ($n_1 \geq 2$)
  - Stop study as soon as “sample size sufficient” according to this estimate ($n \geq v\hat{\sigma}_n^2$)
- This is a stochastic process with stop-time

$$N = \min\left\{ n = n_1, n_1 + 1, \ldots \,\middle|\, \hat{\sigma}_n^2 \leq n / v \right\}$$

- Discussed in the context of clinical studies e.g. by Mehta & Tsiatis (2001) and Jennison & Turnbull (2007)
- Investigated by Friede & Miller (2012)

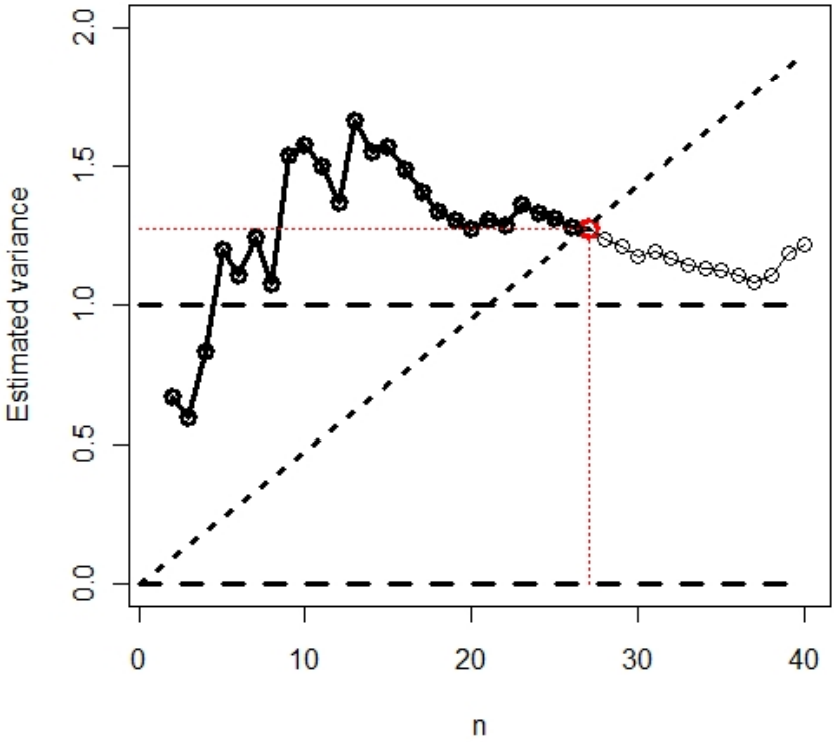
Continuous variance monitoring as stochastic process

- Example study with true variance=1, v=21

$$v = 2\left(z_{1-\alpha} + z_{1-\beta}\right)^2 \Big/ \delta_a^2$$

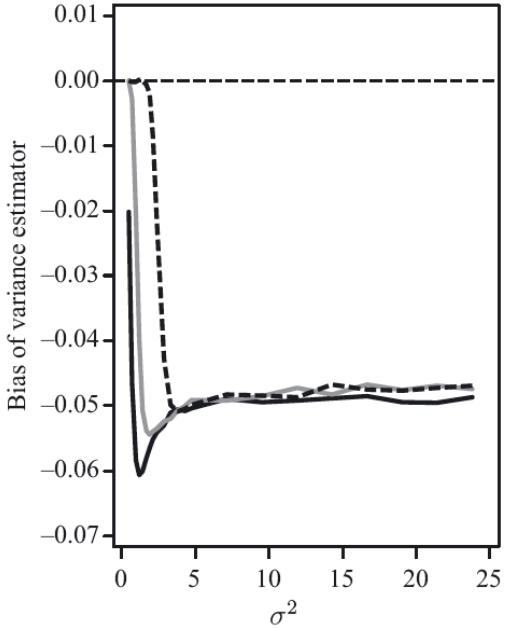
$$v = 21.0 \text{ for } \alpha = 0.025, \beta = 0.1, \delta_a = 1$$

- First time under linear boundary stops study
- $E(\hat{\sigma}_N^2) < \sigma^2 = 1$ (negative bias of variance estimator)

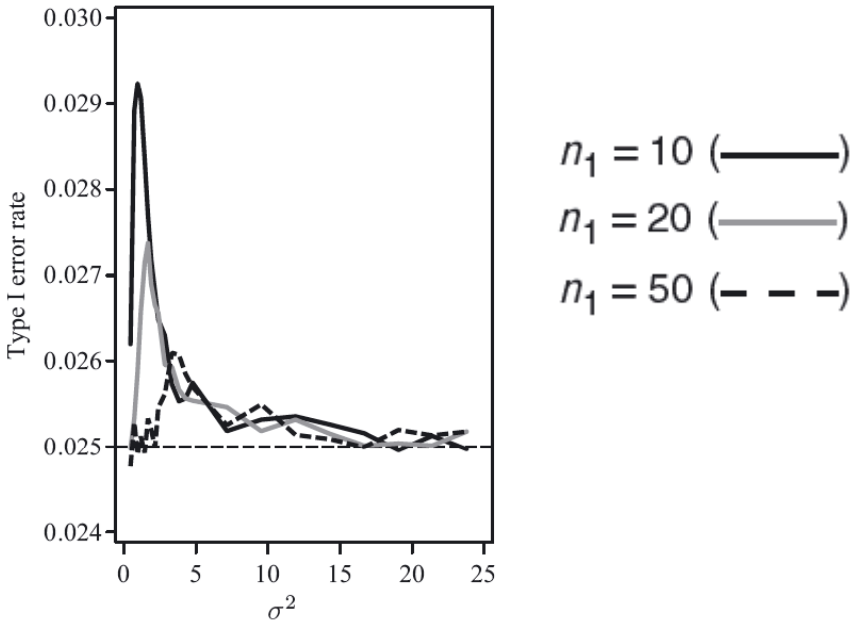


Bias of variance estimator and test

Bias of variance estimator



Bias of test size



Variance estimator $\hat{\sigma}_N^2$ at stop-time *N* negatively biased

Therefore, the t-test Reject $H_0 \Leftrightarrow t > t_{2N-2,1-\alpha}$ with $t = \hat{\delta} \Big/ \sqrt{2\hat{\sigma}_N^2 / N}$ does not control the alpha level

Blinding in clinical trials

- **Randomized clinical trials** for drug development are usually **blinded**
- Database of ongoing study has no treatment information
- Separate file with treatment info kept secretly

Database

Patient	Treatment	Result1	Result2	...
1	*	1	7.7	...
2	*	0	6.8	...
3	*	1	8.0	...
4	*	0	6.5	...
5	*	0	8.9	...

Treatments

Patient	Treatment
1	Placebo
2	Active
3	Active
4	Placebo
5	Active

Blinding and sample size re-estimation

- To perform the sample size re-estimation shown before, the **treatment of all patients needs to be known for the computation of the variance (unblinding necessary)**
- Regulatory authorities **prefer methods not requiring unblinding**

$$\hat{\sigma}_n^2 = \frac{1}{2n-2}\left(\sum_{j=1}^{n}(X_{1j}-\bar{X}_{1\bullet})^2+\sum_{j=1}^{n}(X_{2j}-\bar{X}_{2\bullet})^2\right)$$

$$\downarrow$$

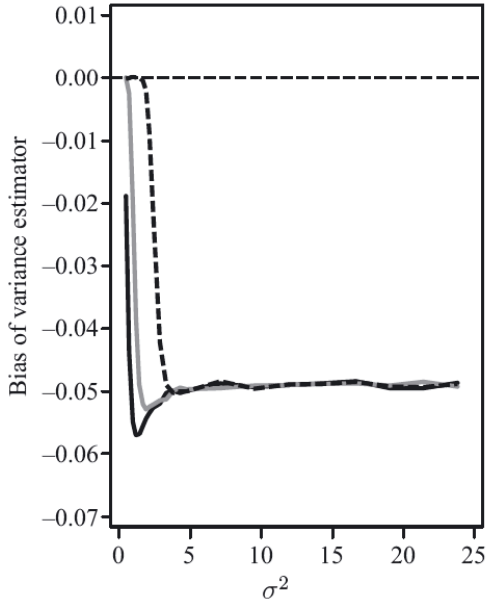
$$\hat{\sigma}_{n,\text{blind}}^2 = \frac{1}{2n-1}\left(\sum_{j=1}^{n}(X_{1j}-\bar{X}_{\bullet\bullet})^2+\sum_{j=1}^{n}(X_{2j}-\bar{X}_{\bullet\bullet})^2\right)$$

Blinded continuous monitoring

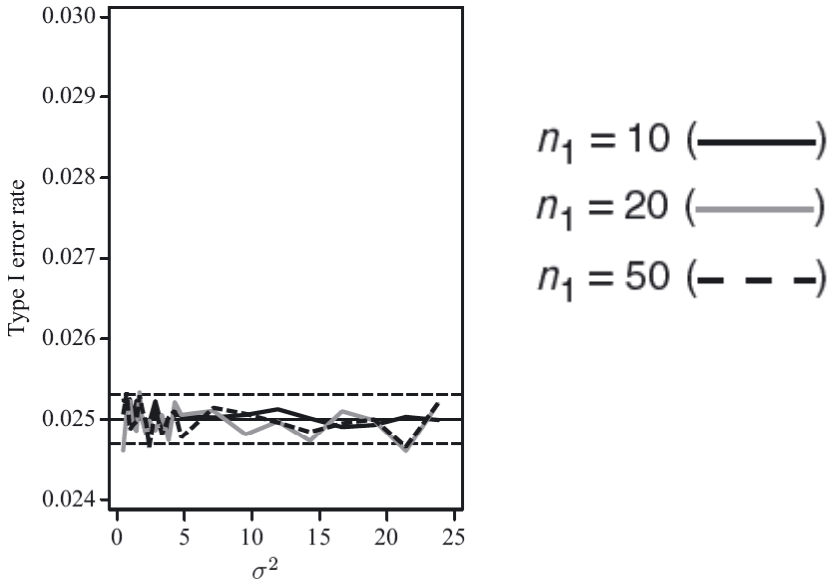
- The different sample size re-estimation procedures can be performed blinded
- Here we show the **blinded version for continuous monitoring**
- Blinded continuous monitoring procedure:
  - Monitor **overall (one sample) variance** $\hat{\sigma}^2_{n,\text{blind}}$ after each (pair of) patients, ignoring the different treatments
  - Stop study as soon as sample size sufficient according to this blinded estimate ($n \Big/ \hat{\sigma}^2_{n,\text{blind}} \geq v$)
  - Estimate final variance unblinded

Bias of variance estimator and test – blinded procedure

Bias of variance estimator

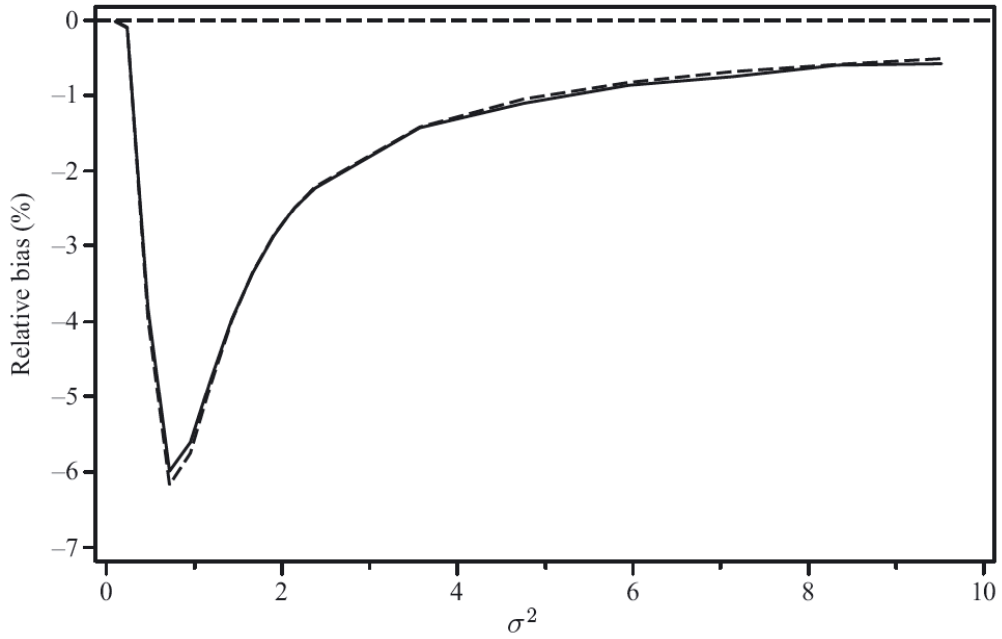


Bias of test size



Still bias for variance estimation but essentially no bias for test.

Blinded procedure: Why is the test size (almost) unbiased?



Relative bias of the numerator (———) and denominator (– – –) of the *F*-test

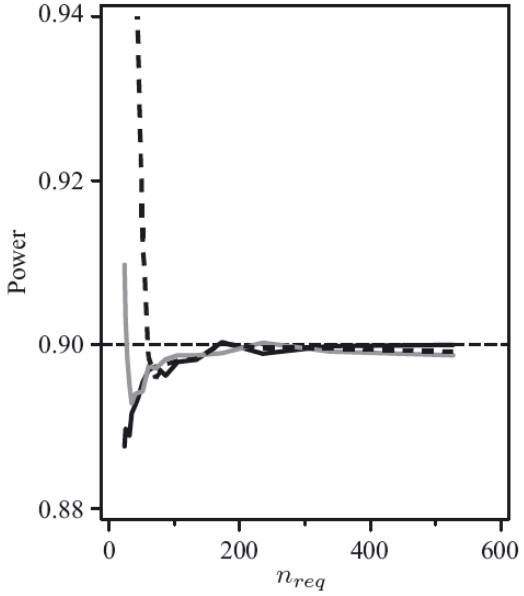
$$t = \hat{\delta}_N \Big/ \sqrt{2\hat{\sigma}_N^2 / N}$$

$$F = \frac{\frac{N}{2}\hat{\delta}_N^2}{\hat{\sigma}_N^2}$$

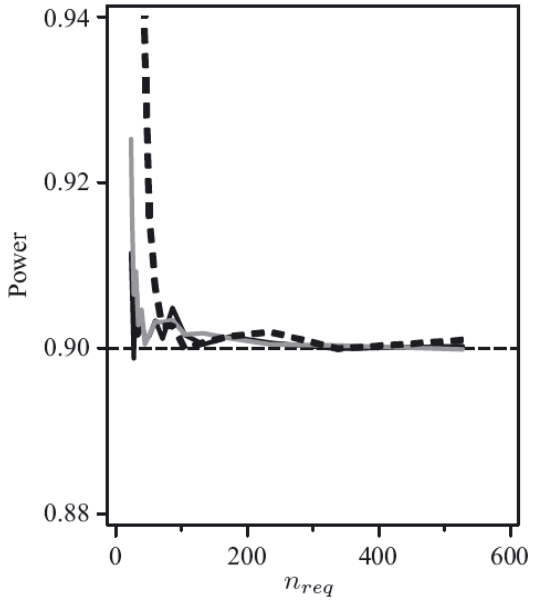
If the (two-sided) test is written as F-test, the **numerator and the denominator have the same bias** after blinded continuous monitoring

Power of test after continuous monitoring

Unblinded continuous monitoring



Blinded continuous monitoring

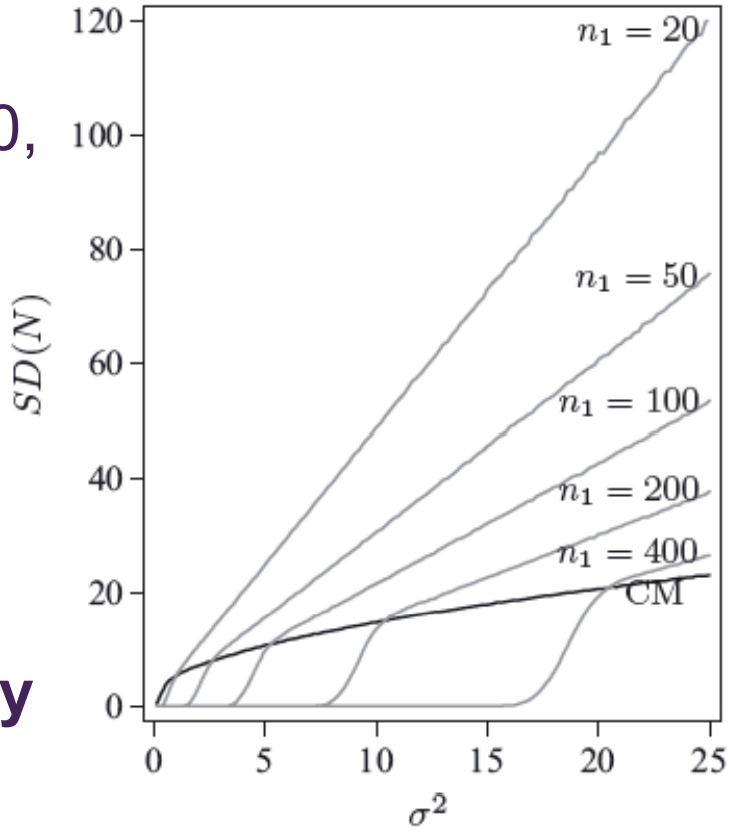


$n_1 = 10$ (——)
$n_1 = 20$ (——)
$n_1 = 50$ (– – –)

Desired power approximately maintained; better so for the blinded procedure.

Blinded procedures: Continuous monitoring versus sample size re-estimation with one interim look

- $\alpha = 0.025, 1\text{-}\beta = 0.90, \delta = \delta_a = 1$
- One look after n_1=20, 50, 100, 200, or 400 patients per treatment to estimate variance
- Compare with continuous monitoring (CM)
- If n_1 chosen too large: risk to overshoot necessary sample size
- If n_1 chosen too small: **SD for sample size can be considerably higher than for CM**



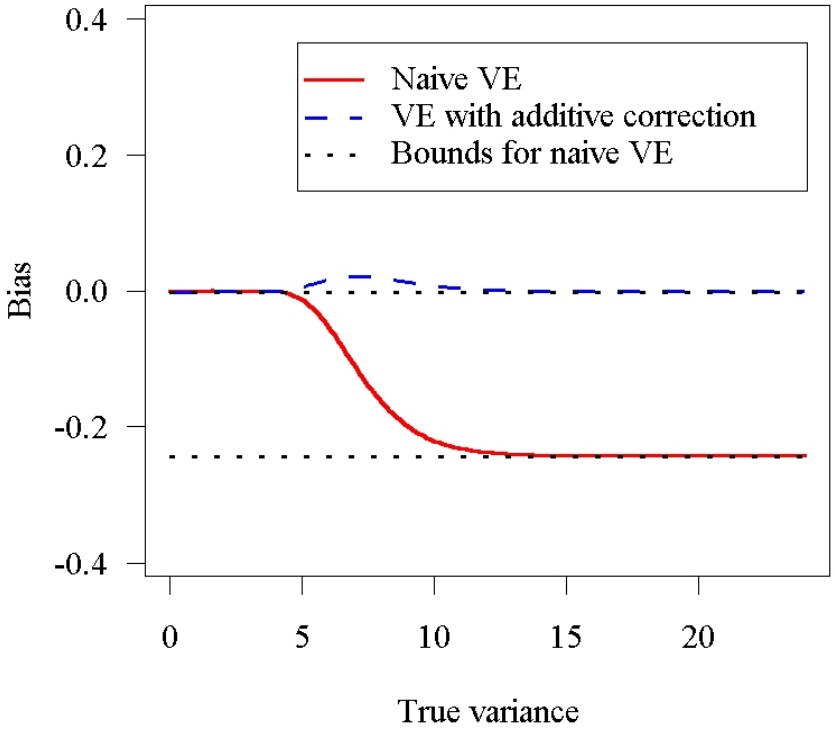
Continuous monitoring – is it logistically feasible?

- Increased use of electronic data capture techniques in clinical studies
- E.g. in chronic pain studies
  - Patients are provided with electronic diaries for their daily pain recording
  - SMS reporting has successfully been applied (Axén, Bodin, Bergström et al, 2012)
  - Pain intensity ratings can immediately be transferred to central database of sponsor

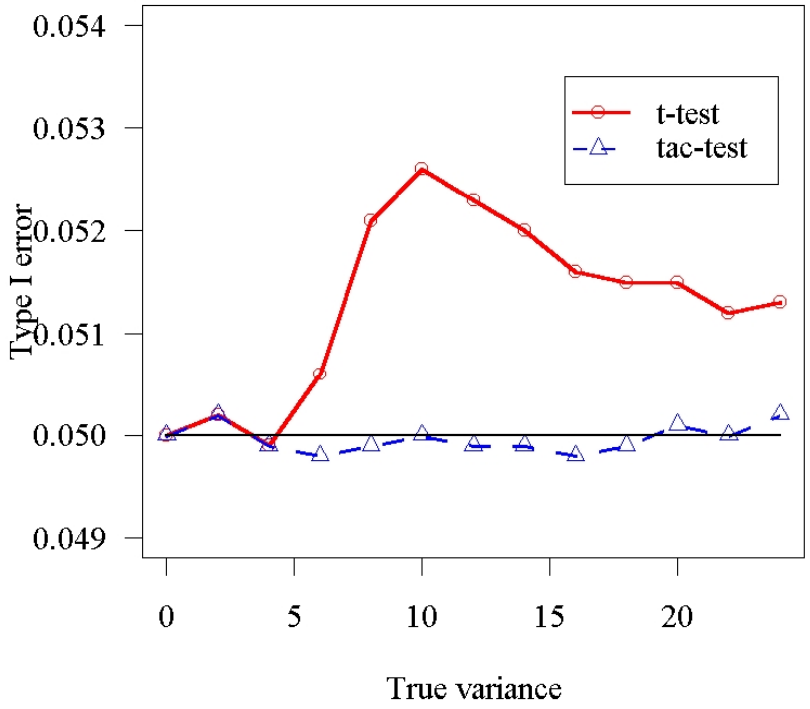
➢ Continuous monitoring feasible in some (but not all) clinical study situations

Back to the bias for sample size re-estimation with one look

Bias of variance estimator

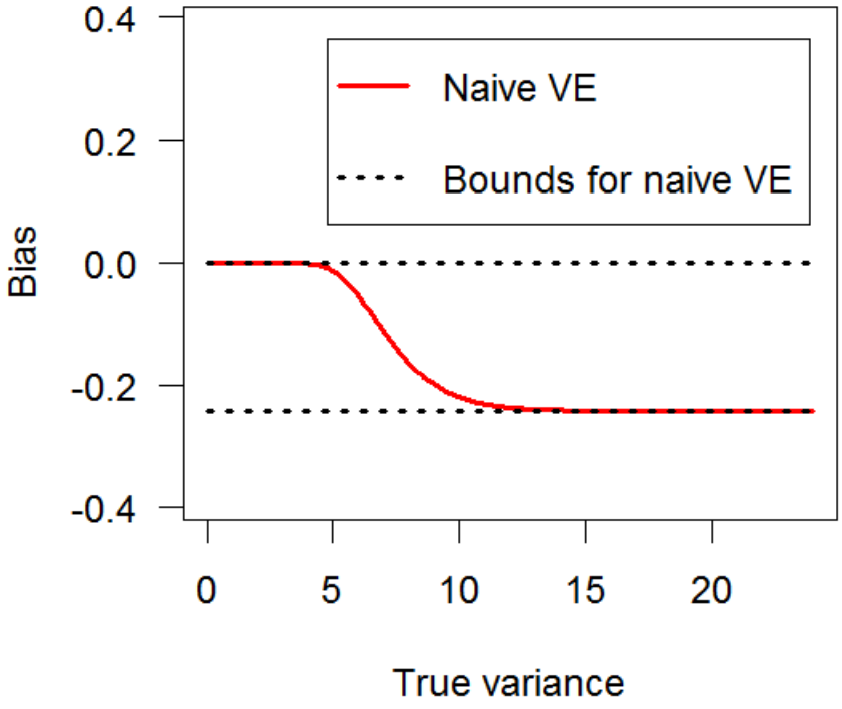


Type I error of t-test and t-test with additive correction



Bias for unblinded case with 1 IA

Bias of variance estimator



$$-\frac{n_1-1}{n_1-2}\cdot\frac{1}{v}\le E\hat{\sigma}^2-\sigma^2\le 0$$

Exact bias:

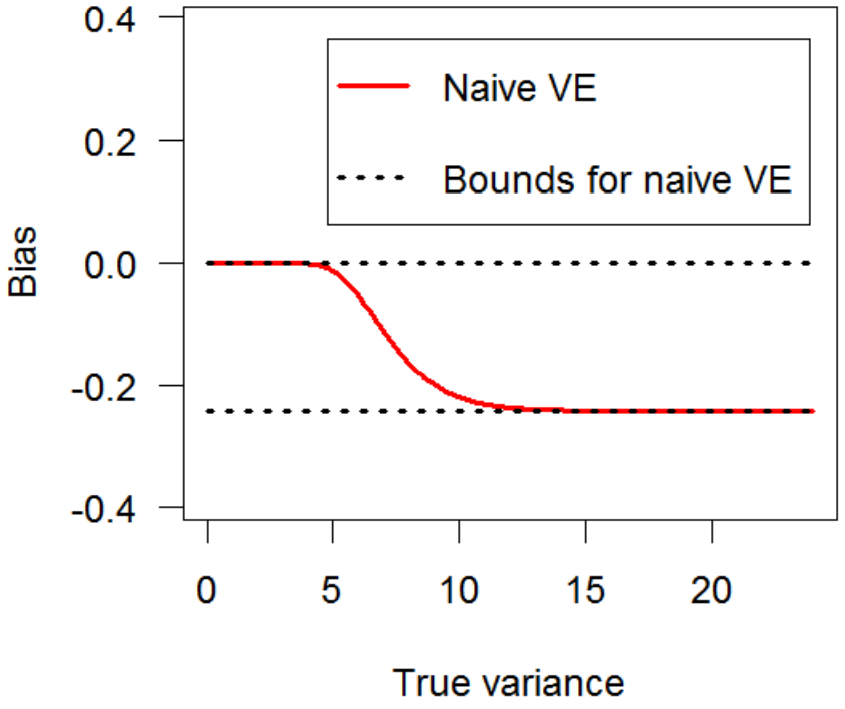
$$\frac{2(n_1-1)^2}{vd}\{F_{2n_1}(d)-F_{2n_1-2}(d)\}+\frac{n_1-1}{v}\{1-F_{2n_1-2}(d)\}$$
$$-\frac{(n_1-1)^2}{v(n_1-2)}\{1-F_{2n_1-4}(d)\}.$$

where F is the chi-square distribution function and

$$d=(2n_1-2)(n_1+n_{2\min}-1)/(v\sigma^2)$$

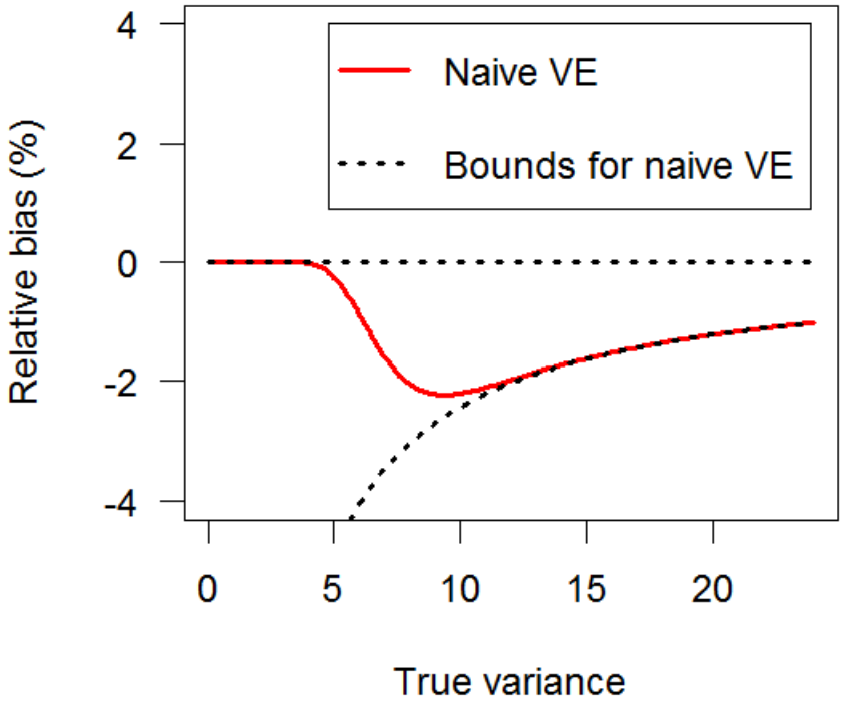
Bias for unblinded case with 1 IA

Absolute bias of variance estimator



$$-\frac{n_1-1}{n_1-2}\cdot\frac{1}{v}\le E\hat{\sigma}^2-\sigma^2\le 0$$

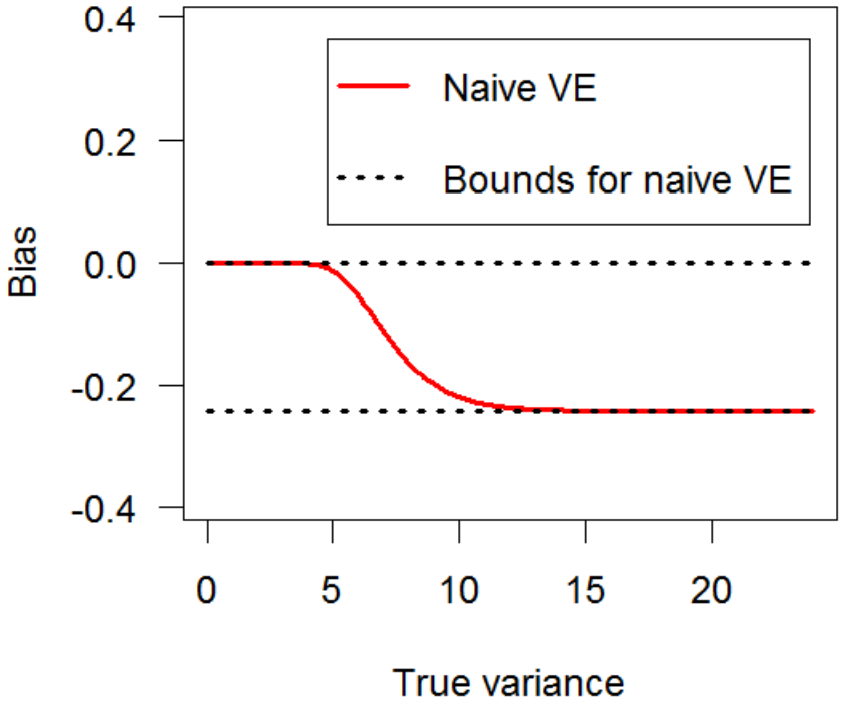
Relative bias of variance estimator



$$-\frac{n_1-1}{n_1-2}\cdot\frac{1}{v\sigma^2}\le-\frac{E\hat{\sigma}^2-\sigma^2}{\sigma^2}\le 0$$

Bias for unblinded case with 1 IA

Absolute bias of variance estimator



$$-\frac{n_1-1}{n_1-2}\cdot\frac{1}{v}\le E\hat{\sigma}^2-\sigma^2\le 0$$

Assumptions for this graph

- Sample size in first stage n_1=20 per group
- Minimum sample size in second stage: n_{2min}=10
- α=0.025, power=90%, δ_a=2.2

$$v=2\left(z_{1-\alpha}+z_{1-\beta}\right)^2\Big/\delta_a^2$$

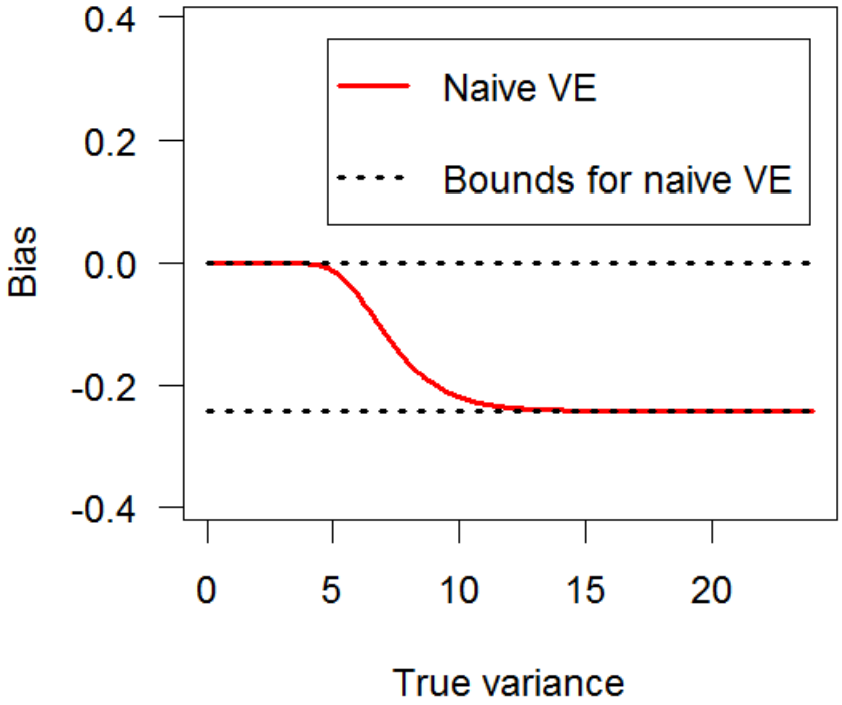
- Sample size formula for interim

$$N=\max\left\{v\hat{\sigma}_1^2+1,n_1+n_{2\min}\right\}$$

- How does the bias depend on n_{2min}, n_1, δ_a? ⊤

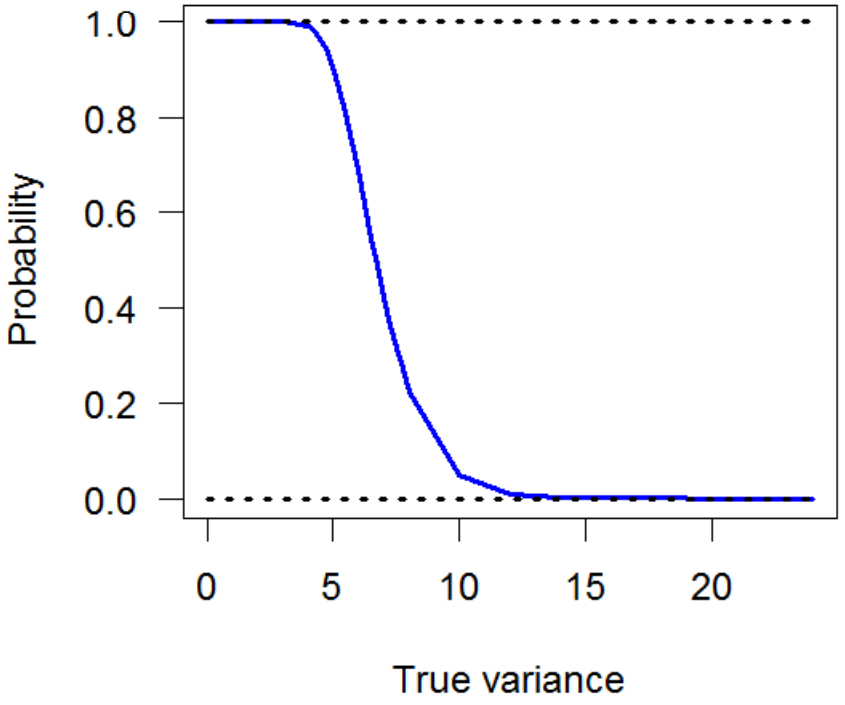
Bias for unblinded case with 1 IA

Absolute bias of variance estimator



$$E\hat{\sigma}^2 - \sigma^2$$

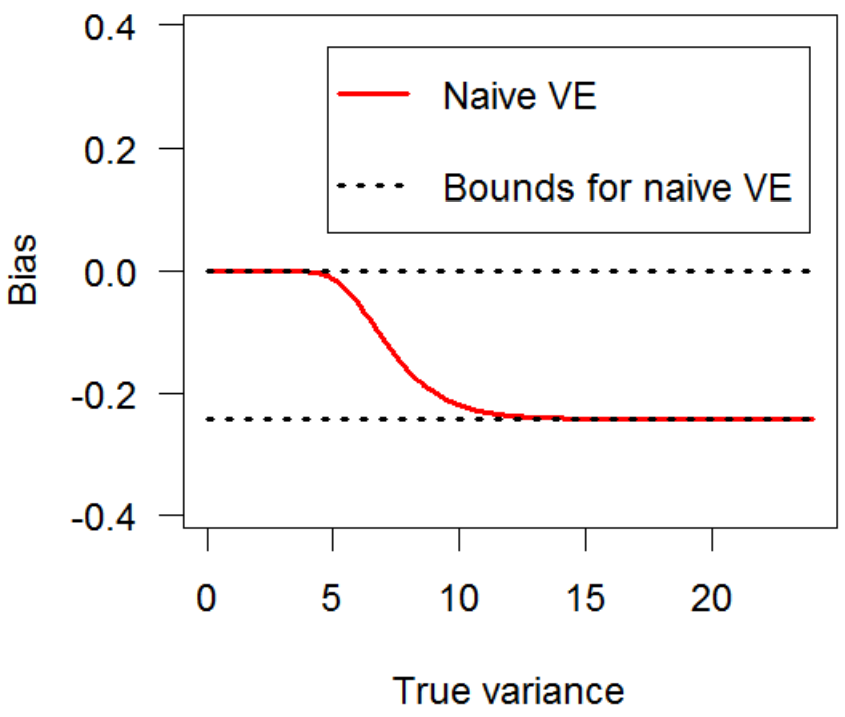
Probability for stopping the study with minimal sample size n_1+n_{2min}



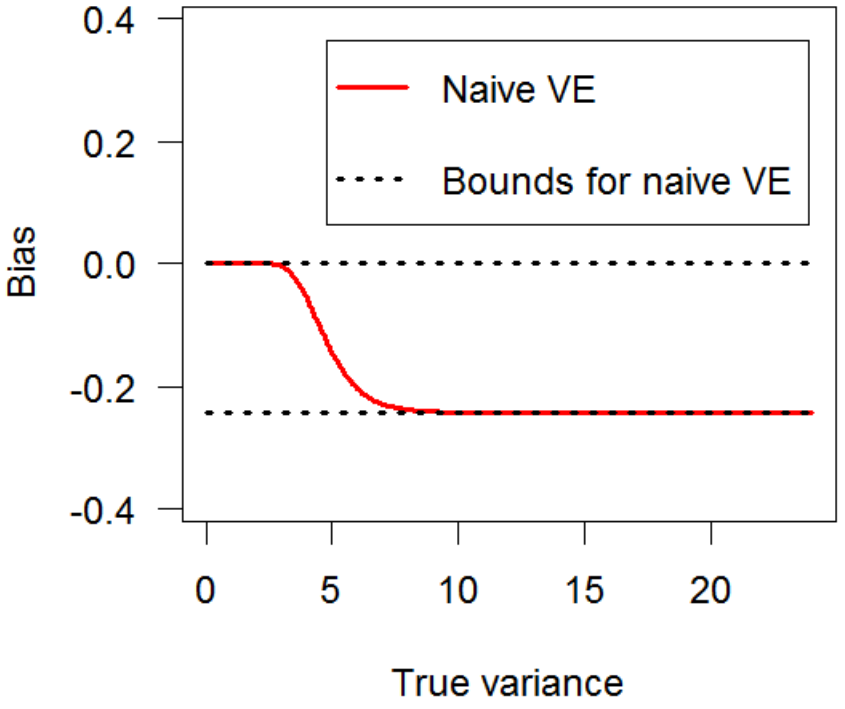
$$P(N = n_1 + n_{2\min})$$

Bias for unblinded case with 1 IA

Bias of variance estimator, n_{2min}=10

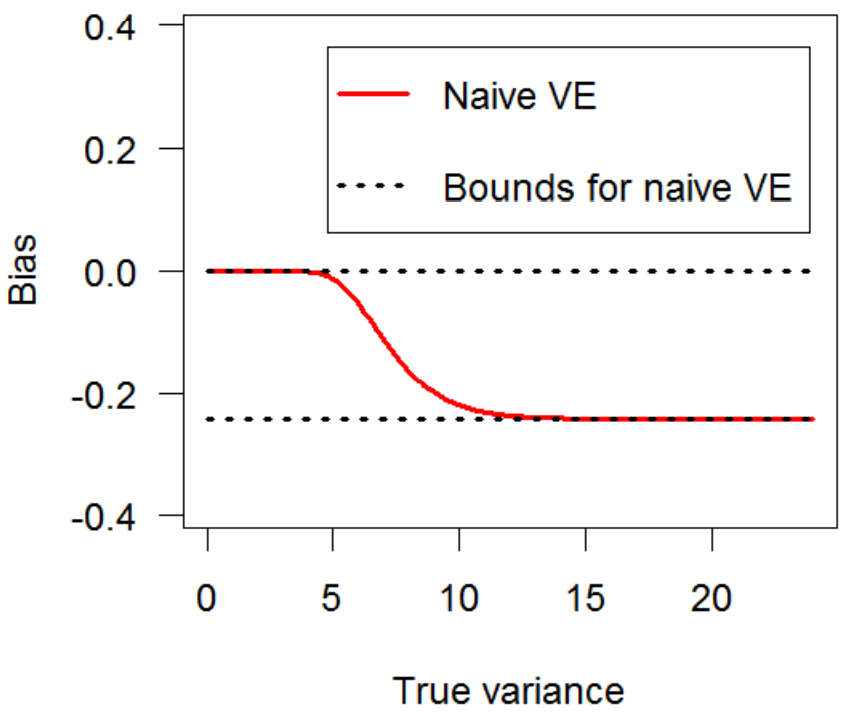


Bias of variance estimator, n_{2min}=0

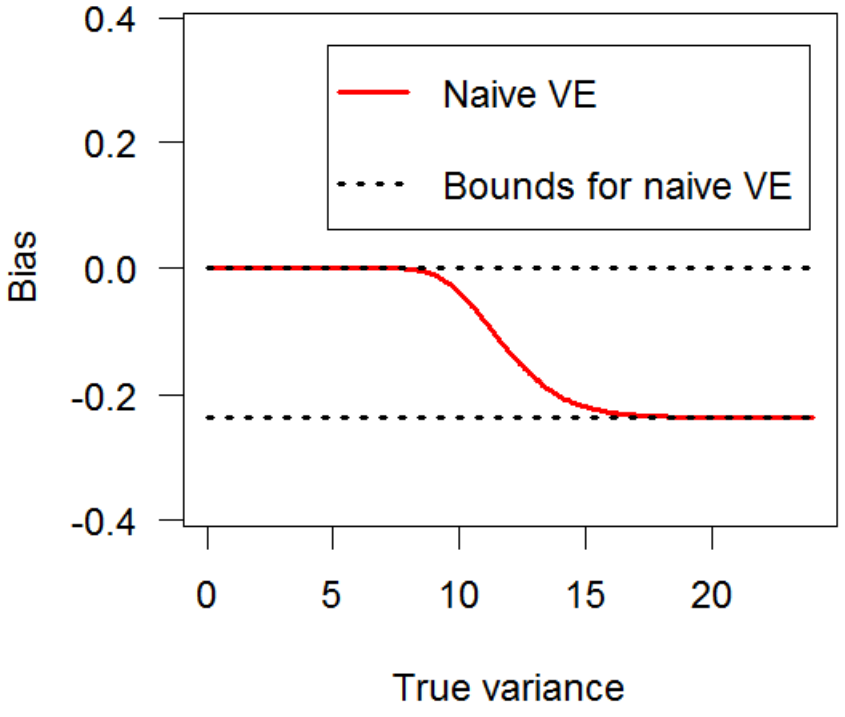


Bias for unblinded case with 1 IA

Bias of variance estimator, n_1=20

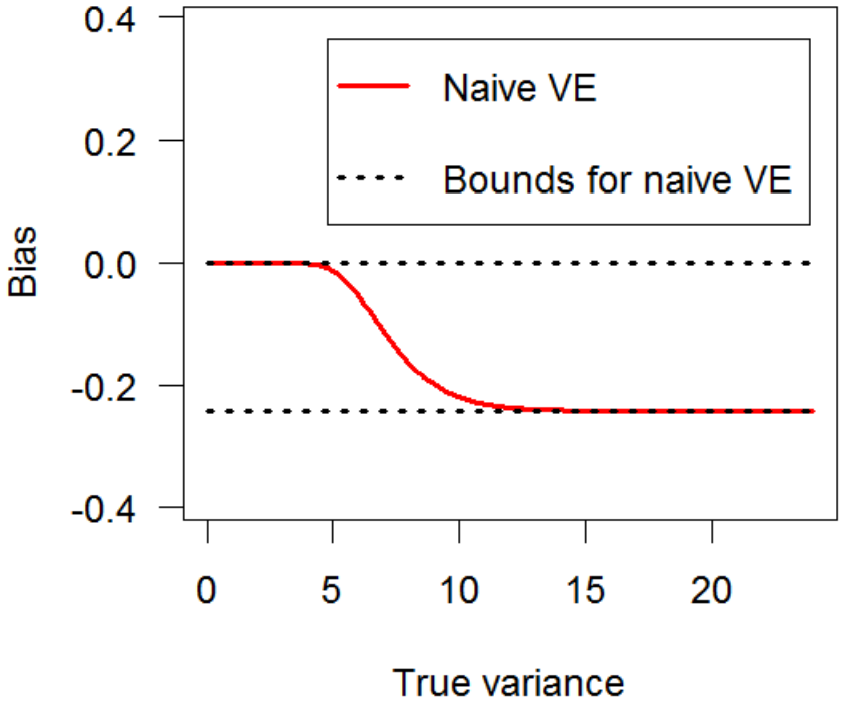


Bias of variance estimator, n_1=40

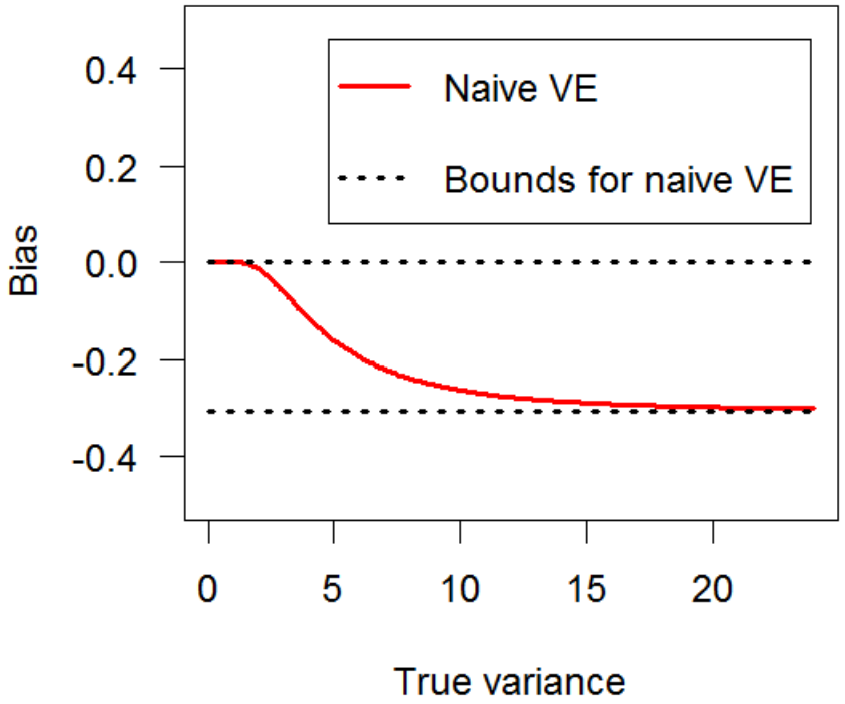


Bias for unblinded case with 1 IA

Bias of variance estimator, n_1=20

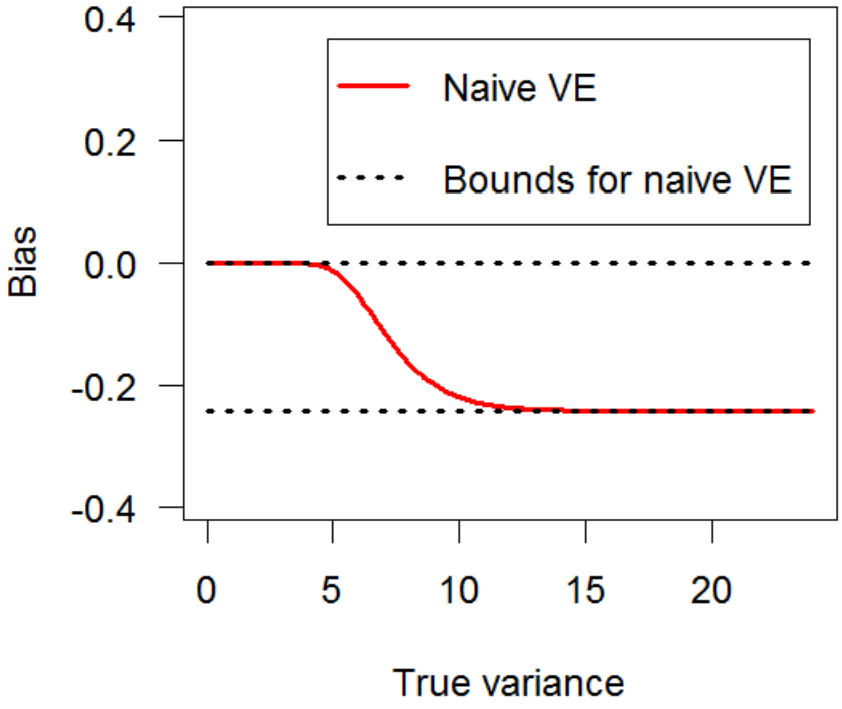


Bias of variance estimator, n_1=5

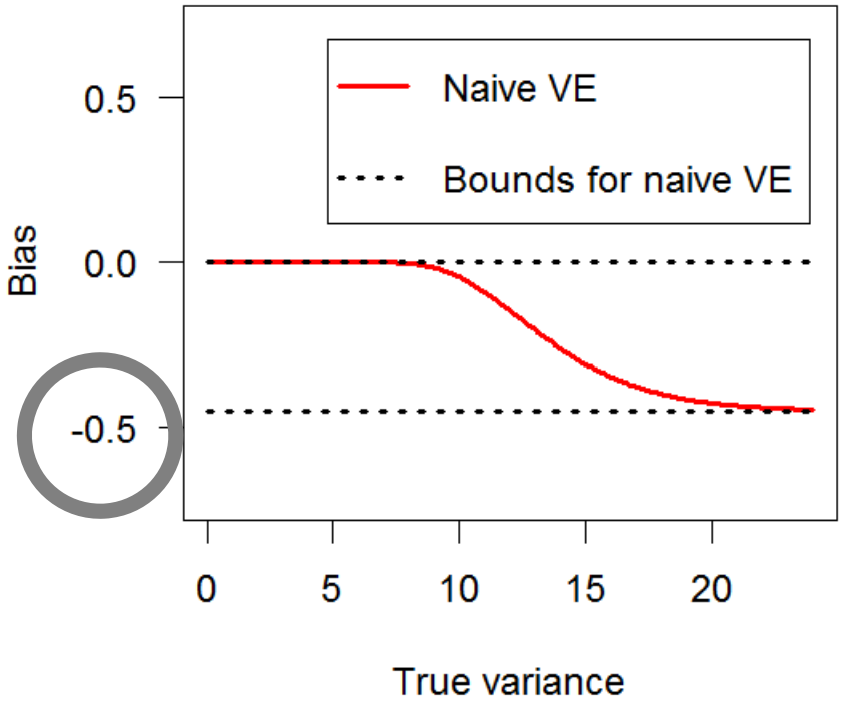


Bias for unblinded case with 1 IA

Bias of variance estimator, v=4.34
(α=0.025, β=0.1, δ_a=2.2)



Bias of variance estimator, v=2.34
(α=0.025, β=0.1, δ_a=3)



Maximal relative bias for both cases ~2.5%

Dependence of bias on parameters

- The **absolute bias** for large variances depends almost not on n_1 and not on the minimum number of patients in Stage 2
- The **relative bias** is largest if there is no minimum number of patients for Stage 2 and n_1 is small; however, decreasing n_1 to very small values (e.g. from 10 to 5) does not increase the relative bias much
- The relative bias in the considered scenarios was **at most 4%**

Summary

- Sample size re-estimation **can ensure appropriate power** (neither over- or under-powered) even under uncertainty of nuisance parameters
- The **effects on estimates and tests are usually small** and often, they might be totally acceptable
- In specific situations (small studies) it might be worth to investigate the bias further

References

- **Axén, Bodin, Bergström et al (2012).** The use of weekly text messaging over 6 months was a feasible method for monitoring the clinical course of low back pain in patients seeking chiropractic care. J. Clin. Epidem. 65: 454-361.
- **Friede & Kieser (2003).** Blinded sample size reassessment in non-inferiority and equivalence trials. Statist. Med., 22, 995–1007.
- **Friede & Miller (2012).** Blinded continuous monitoring of the nuisance parameter in clinical trials. J. Royal Stat. Soc. – Series C, 61, 601-618.
- **Jennison & Turnbull (2007).** Adaptive seamless designs with control of the type I error rate. J. Biopharm. Statist. 17:1135-1161.
- **MacDonald et al (2008).** Effect on blood pressure of lumiracoxoib versus ibuprofen in patients with osteoarthritis and controlled hypertension: a randomized trial. Journal of Hypertension 26: 1695-1704.
- **Mehta & Tsiatis (2001).** Flexible sample size considerations using information-based monitoring. Drug Inf. J. 35:1095-1112.
- **Miller (2005).** Variance estimation in clinical studies with interim sample size re-estimation. Biometrics, 61, 355-361.
- **Siegmund (1985).** Sequential analysis. Springer.