Blinded sample size re-estimation in three-arm trials with 'gold standard' design

Tobias Mütze [1] Tim Friede [1,2]

[1] Department of Medical Statistics, University Medical Center Göttingen, Germany

[2] DZHK (German Centre for Cardiovascular Research), partner site Göttingen, Göttingen, Germany

10th International Conference on Multiple Comparison Procedures

Riverside, June 20 – 23 2017

'Gold standard' design

- Three-arm trial design with experimental, reference, and placebo arm
- Assessment of both
  - assay sensitivity
  - non-inferiority of experimental treatment compared to reference treatment
- Recommended when
  - the reference only offers limited benefits compared to placebo
  - the effect of the reference compared to placebo is volatile (Koch & Röhmel, 2004)
- Recommended design in conditions such as asthma, schizophrenia, and migraine (CHMP guidelines)

Nuisance parameter based sample size re-estimation

- Initially planned sample size is adjusted mid-trial based on nuisance parameter estimates from an internal pilot study
- Limits the negative effect of nuisance parameter misspecification on statistical power
- Maintain blinding during re-estimation reduces organizational effort: no independent DMC required
- Blinded sample size re-estimation well-accepted from a regulatory perspective (FDA draft guidance, 2010)

Statistical model and testing non-inferiority

- Independent random variables

$$X_{k,i} \sim \mathcal{N}(\mu_k, \sigma^2) \quad i = 1, \ldots, n_k, \quad k = E, R, P$$

- Smaller means μ_k are better
- Non-inferiority of the experimental treatment compared to the reference

$$H_0^{ER} : \mu_E \geq \mu_R + \delta_{ER} \quad \text{vs.} \quad H_1^{ER} : \mu_E < \mu_R + \delta_{ER}$$

- Non-inferiority margin $\delta_{ER} > 0$

Assay sensitivity and global hypothesis

- Two hypotheses for testing assay sensitivity commonly used

$$H_0^{EP} : \mu_E \geq \mu_P \quad \text{vs.} \quad H_1^{EP} : \mu_E < \mu_P$$
$$H_0^{RP} : \mu_R \geq \mu_P \quad \text{vs.} \quad H_1^{RP} : \mu_R < \mu_P$$

- In this presentation: global hypothesis is the union of the single hypotheses

$$H_0 : H_0^{ER} \cup H_0^{EP} \cup H_0^{RP} \quad \text{vs.} \quad H_1 : H_1^{ER} \cap H_1^{EP} \cap H_1^{RP}$$

- Intersection-union test method: Global hypothesis can be tested with level α by testing the local hypotheses with level α

Hypothesis testing and power

- Local hypotheses are each tested with a one-sided Student's t-test with significance level α
- Approximate power of the test procedure for the global hypothesis by multivariate normal distribution (Stucke & Kieser, 2012)

$$B(n) = \Phi\left(t_{\alpha,\nu_{ER}} - \frac{\delta^*_{ER} - \delta_{ER}}{\sigma\sqrt{\frac{1}{n_E} + \frac{1}{n_R}}}, t_{\alpha,\nu_{EP}} - \frac{\delta^*_{EP}}{\sigma\sqrt{\frac{1}{n_E} + \frac{1}{n_P}}}, t_{\alpha,\nu_{RP}} - \frac{\delta^*_{RP}}{\sigma\sqrt{\frac{1}{n_R} + \frac{1}{n_P}}} \right)$$

- Cumulative distribution function of the multivariate normal distribution $\mathcal{N}_3(0, \Sigma)$: $\Phi(\cdot)$
- Mean differences in the alternative: $\delta^*_{ER}, \delta^*_{EP}, \delta^*_{RP}$
- Degrees of freedom: $\nu_{ij} = n_i + n_j - 2$

Sample size planning

- Total sample size required to obtain power $1-\beta$ for fixed allocation $w_k = n_k/n$

$$n = \min \left\{ \tilde{n} = \sum_{k=E,R,P} \tilde{n}_k : \tilde{n}_k \in \mathbb{N},\ \tilde{n}_k/\tilde{n} = w_k,\ B(\tilde{n}) \geq 1-\beta \right\}$$

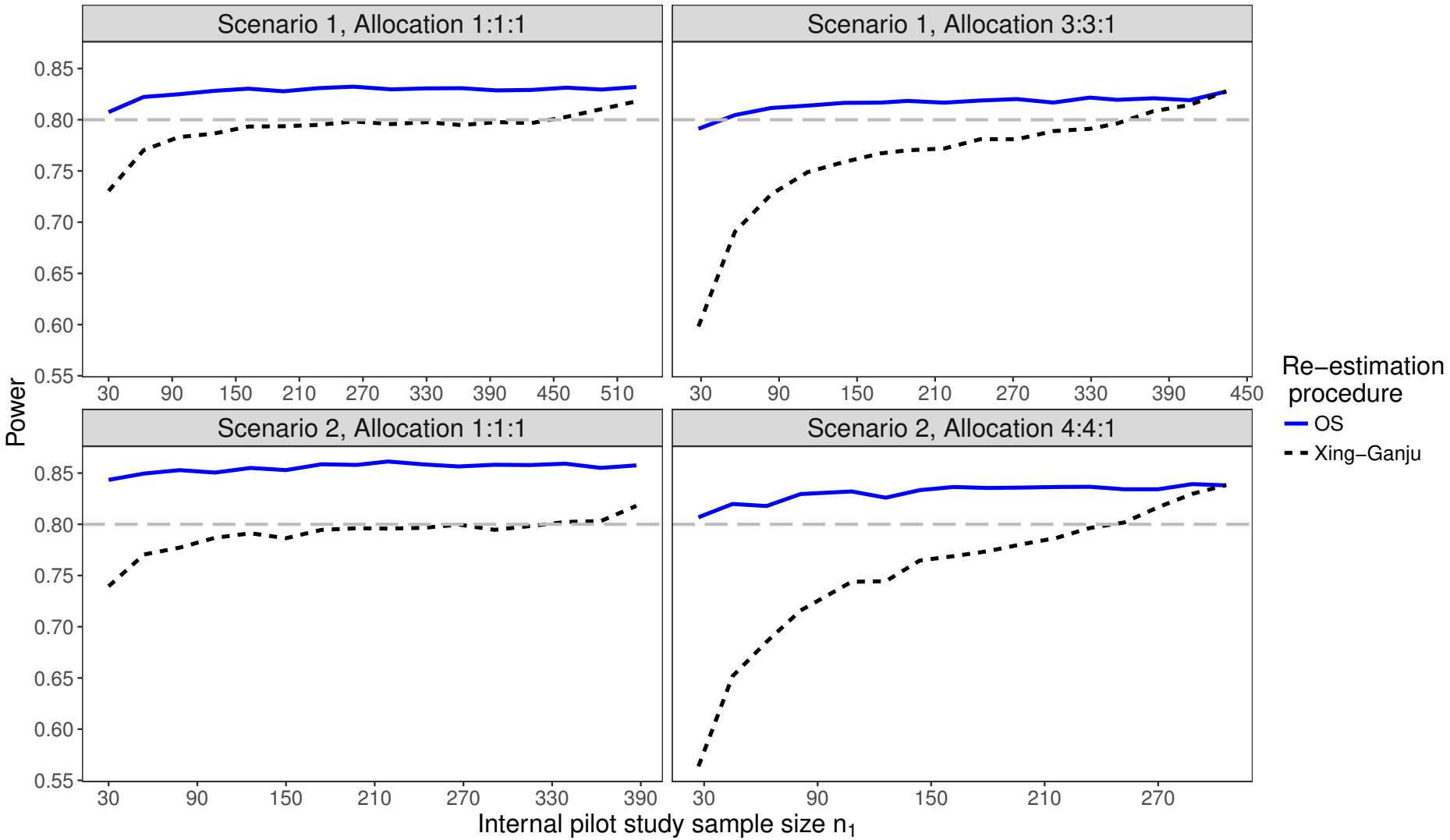
Nuisance parameter estimation I

- Blinded observations from internal pilot study: $Y_1, \ldots, Y_{n_{IPS}}$
- Blinded one-sample variance estimator
  - $\hat{\sigma}^2_{OS} = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} \left(Y_i - \bar{Y}\right)^2$
  - Unbiased estimator when group means are equal, otherwise overestimates the variance σ^2
  - Recommended approach in two-arm trials
- Blinded variance estimator by Xing and Ganju (Xing & Ganju, 2005)
  - Available in a randomized block design
  - Balanced blocks of length m and T_k the sum of the observations in block k
  - $\hat{\sigma}^2_{XG} = \frac{1}{n_1 - m} \sum_k \left(T_k - \bar{T}\right)^2$
  - Unbiased estimator for σ^2

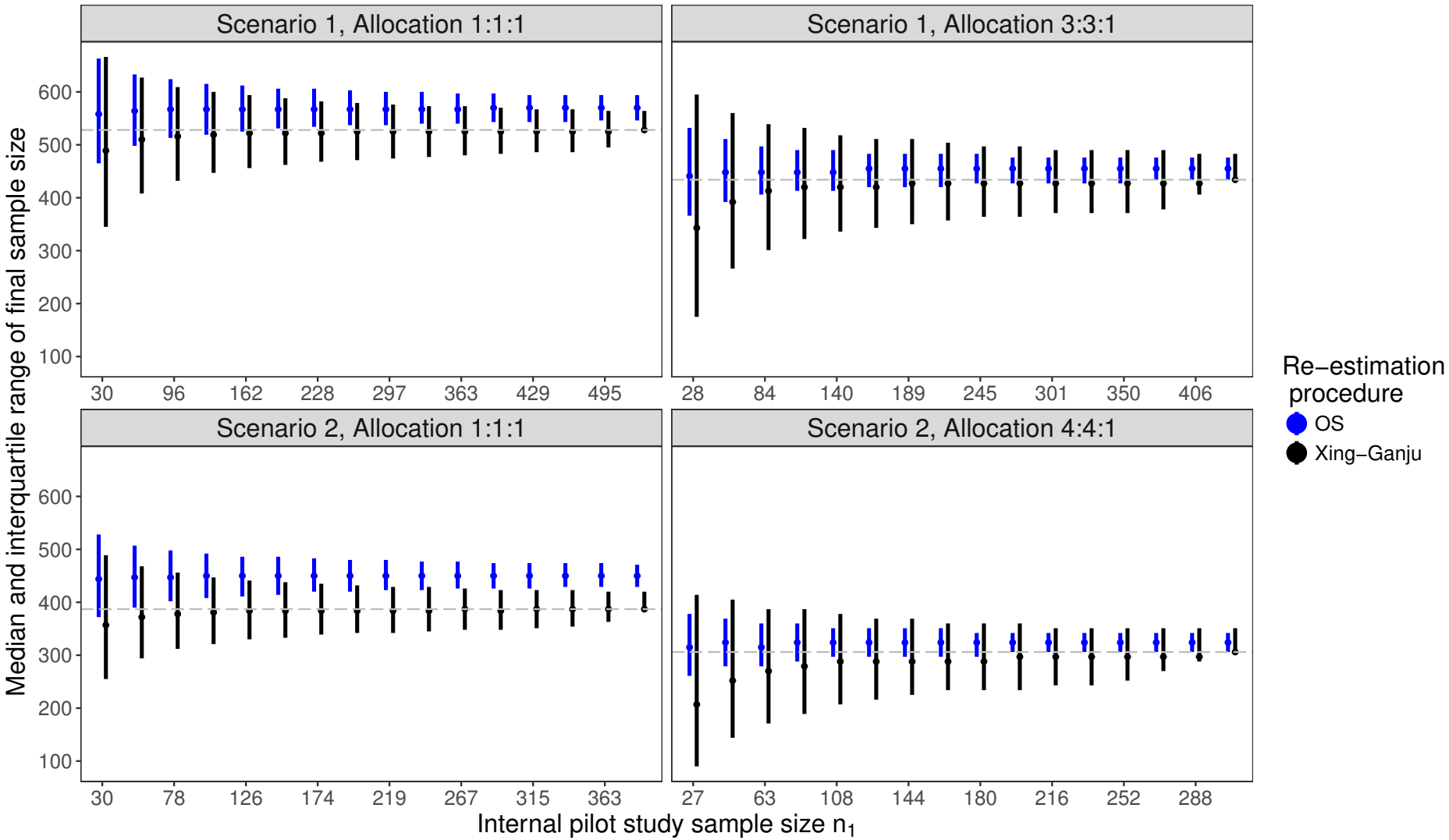
Simulation scenarios

Parameter	**Scenario 1**	**Scenario 2**
One-sided significance level α	0.025	0.025
Target power $1-\beta$	0.8	0.8
Non-inferiority margin δ_{ER}	$\delta_{ER}=0.3$	$\delta_{ER}=0.45$
Mean μ_R under H_1	$\mu_R=0$	$\mu_R=0$
Mean μ_E under H_1	$\mu_E=0$	$\mu_E=0.1$
Mean μ_P under H_1	$\mu_P=0.6$	$\mu_P=0.9$
Standard deviation σ under H_1	$\sigma=1$	$\sigma=1$
Sample size allocation $n_E:n_R:n_P$	1:1:1, 3:3:1	1:1:1, 4:4:1

Results - Power



Final sample size distribution



Power approximation

- Goal: modify the Xing-Ganju sample size re-estimation procedure such that it meets the target power in the 'gold standard' design
- Power approximation of nuisance parameter based sample size re-estimation procedure (Zucker *et al.*, 1999)

$$\text{Power} \approx \int_0^\infty B\left(\tilde{n}\left(x\right)\right) f_{\hat{\sigma}^2}(x)\mathrm{d}x$$

- $\tilde{n}(\cdot)$ denotes the re-estimated sample size as a function of the variance, $f_{\hat{\sigma}^2}(\cdot)$ denotes the density of the Xing-Ganju variance estimator from the internal pilot study

Sample size inflation factor

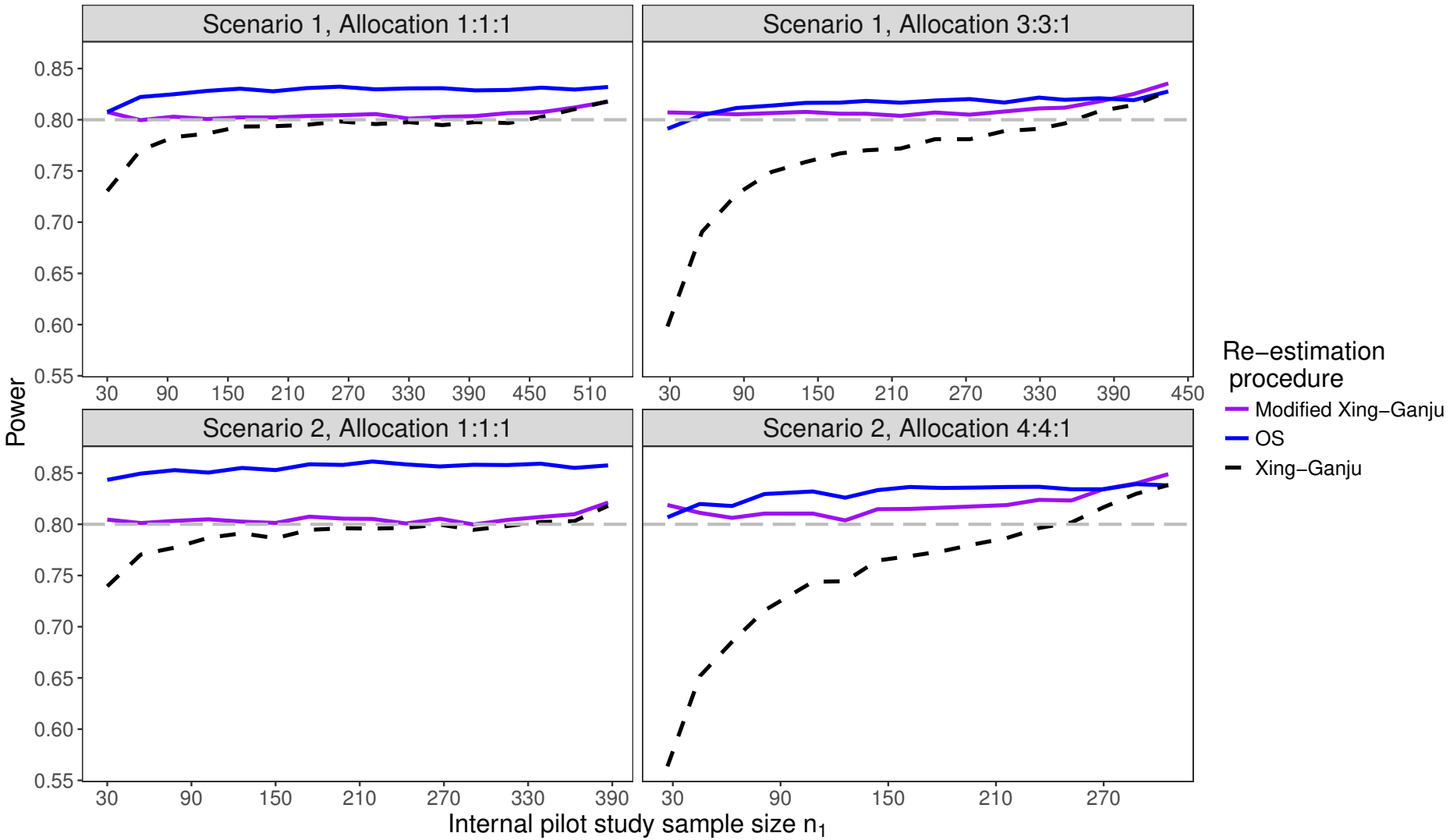
- Idea: inflate the re-estimated sample size with a factor ζ such that the sample size re-estimation procedure meets the target power (Zucker *et al.*, 1999)

$$\int_0^\infty B\left(\zeta\tilde{n}\left(x\right)\right) f_{\hat{\sigma}^2}(x)\mathrm{d}x = 1-\beta$$

- Inflation factor will affect variability of the final sample size

- Inflation factor for Xing-Ganju variance estimator is (mostly) constant in unknown variance σ^2

Results - Power of modified Xing-Ganju procedure



Estimation of treatment effect following sample size re-estimation

- Nuisance parameter based sample size re-estimation can result in biased effect and variance estimates at the final analysis
- Unbiased effect estimates when sample size is adjusted based on (modified) Xing-Ganju approach
- Bias of effect estimate is given by

$$\mathbb{E}\left[\hat{\delta}_{ij} - \delta^*_{ij}\right] = \mathrm{Cov}\left(n_1/\hat{n}_{final}, \hat{\delta}^{(1)}_{ij}\right)$$

- $\hat{\delta}^{(1)}_{ij}$ is the effect estimate based on the data of the internal pilot study
- Covariance can be shown to be zero

Discussion

- Sample size re-estimation based on the blinded one-sample variance estimator results in overpowered clinical trials in the 'gold standard' design
- Performance of the sample size re-estimation procedure based on the blinded one-sample variance estimator differs between three-arm trials and two-arm trials
- Quantitatively similar performances are observed when the non-inferiority is tested through the retention-of-effect hypothesis

$$H_0 : \frac{\mu_P - \mu_E}{\mu_P - \mu_R} \leq \Delta \quad \text{vs.} \quad H_1 : \frac{\mu_P - \mu_E}{\mu_P - \mu_R} > \Delta$$

Some references

Koch A, Röhmel J. Hypothesis testing in the "gold standard" design for proving the efficacy of an experimental treatment relative to placebo and a reference. *Journal of Biopharmaceutical Statistics* 2004; **14**:315–325.

Committee for Medicinal Products for Human Use (CHMP). Guideline on clinical investigation of medicinal products for the treatment of migraine.
`http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003481.pdf`

Committee for Medicinal Products for Human Use (CHMP). Guideline on the clinical investigation of medicinal products for the treatment of asthma.
`http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2015/12/WC500198877.pdf`

Committee for Medicinal Products for Human Use (CHMP) Guideline on clinical investigation of medicinal products, including depot preparations in the treatment of schizophrenia.
`http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2012/10/WC500133437.pdf`

Friede T, Kieser M. Sample size recalculation in internal pilot study designs: a review. *Biometrical Journal* 2006; **48**:537–555.

Stucke K, Kieser M. A general approach for sample size calculation for the three-arm 'gold standard' non-inferiority design. *Statistics in Medicine* 2012; **31**:3579–3596.

Zucker DM, Wittes JT, Schabenberger O, Brittain E. Internal pilot studies II: comparison of various procedures. *Statistics in Medicine* 1999; **18**:3493–3509.

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

Xing B, Ganju J. A method to estimate the variance of an endpoint from an on-going blinded trial. *Statistics in Medicine* 2005; **24**:1807–1814.

Student's t-test statistics

- Test statistic for H_0^{ER}

$$T_{ER} = \sqrt{\frac{n_E + n_R}{n_E n_R}} \frac{\bar{X}_E - \bar{X}_R - \delta_{ER}}{\hat{\sigma}_{ER}}.$$

- $\hat{\sigma}^2_{ER}$ denotes the pooled variance of the experimental treatment group and the reference group

$$\hat{\sigma}^2_{ER} = \frac{(n_E - 1)\hat{\sigma}^2_E + (n_R - 1)\hat{\sigma}^2_R}{n_E + n_R - 2}.$$

- $\hat{\sigma}^2_k$ the sample variance of group $k = E, R$.

Covariance matrix for power calculation

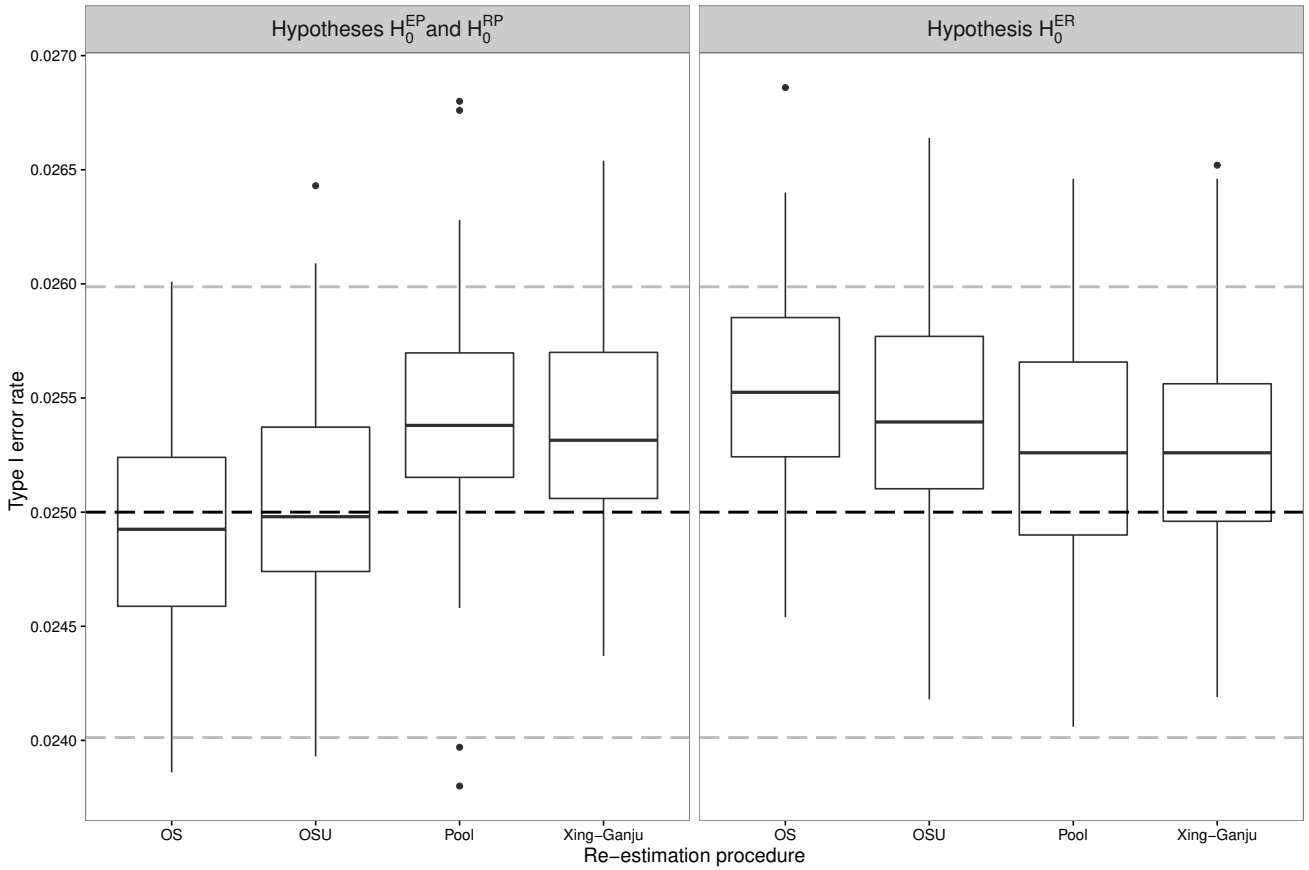
$$\Sigma = \begin{pmatrix} 1 & \frac{1}{\sqrt{\left(1+\frac{n_E}{n_R}\right)\left(1+\frac{n_E}{n_P}\right)}} & -\frac{1}{\sqrt{\left(1+\frac{n_R}{n_E}\right)\left(1+\frac{n_R}{n_P}\right)}} \\ \frac{1}{\sqrt{\left(1+\frac{n_E}{n_R}\right)\left(1+\frac{n_E}{n_P}\right)}} & 1 & \frac{1}{\sqrt{\left(1+\frac{n_P}{n_R}\right)\left(1+\frac{n_P}{n_E}\right)}} \\ -\frac{1}{\sqrt{\left(1+\frac{n_R}{n_E}\right)\left(1+\frac{n_R}{n_P}\right)}} & \frac{1}{\sqrt{\left(1+\frac{n_P}{n_R}\right)\left(1+\frac{n_P}{n_E}\right)}} & 1 \end{pmatrix}$$

Nuisance parameter estimation II

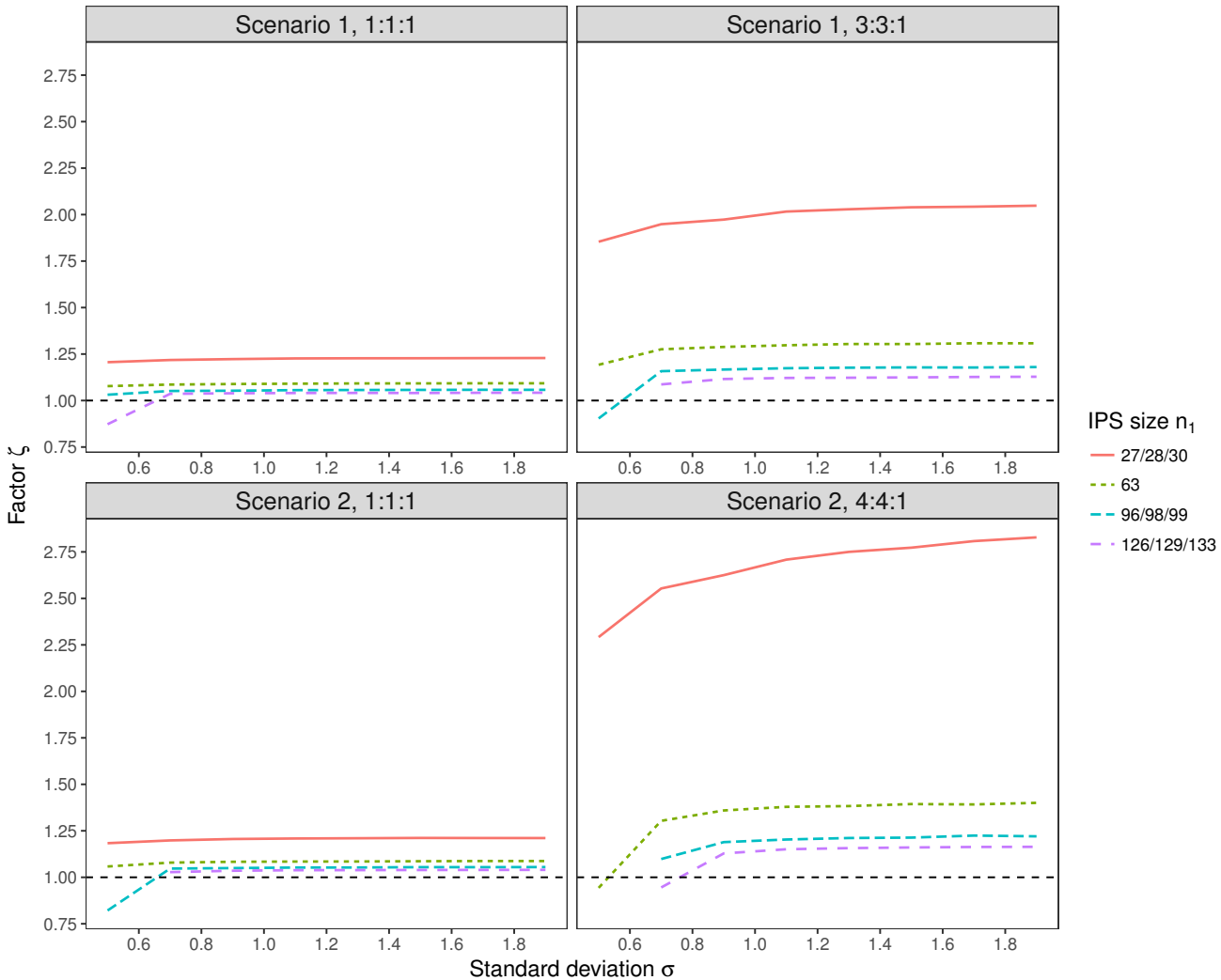
- Blinded adjusted one-sample variance
  - Idea: Shift the one-sample variance estimator to be unbiased under the planning alternative
  - $\hat{\sigma}^2_{OSU} = \hat{\sigma}^2_{OS} - c$
  - Bias of $\hat{\sigma}^2_{OS}$ under planning alternative: c

- Unblinded pooled variance estimator
  - $\hat{\sigma}^2_{Pool} = \frac{(n_{1,E}-1)\hat{\sigma}^2_E+(n_{1,R}-1)\hat{\sigma}^2_R+(n_{1,P}-1)\hat{\sigma}^2_P}{n_1-3}$
  - Unbiased estimator for the variance σ^2
  - Does not maintain blinding of internal pilot study: independent DMC required

Type I error inflation



Inflation factor ζ



Sample size distribution of modified Xing-Ganju procedure