Sample Size Reestimation in Non-Inferiority Trials

Tim Friede[1] and Meinhard Kieser[2]

[1] Warwick Medical School, The University of Warwick, UK
[2] Institute of Medical Biometry and Informatics, University of Heidelberg, Germany

This presentation involves joint work with
Charles Mitchell (ETH Zürich) and Günther Müller-Velten (Novartis)

Example: Anti-hypertensive Trial

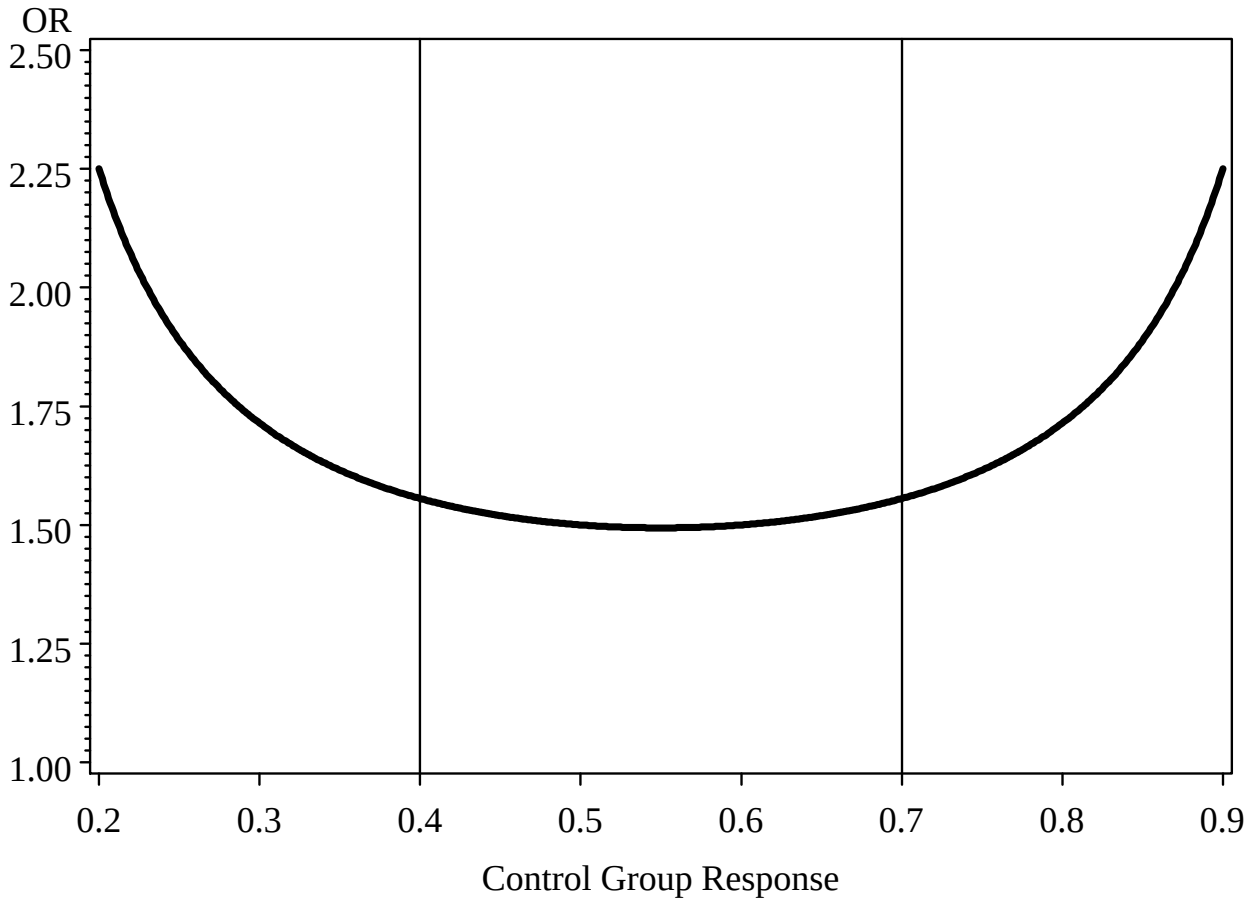
- **design**: randomized, double-blind, parallel group, active-controlled
- **patients** with hypertension and non-insulin dependent diabetes
- **primary endpoint**: proportion of patients who . . .
  - completed study on treatment (tolerability, safety)
  - with mean 24h blood pressure $<$ 130/80 mmHg (syst./diast.) (efficacy)
- **non-inferiority margin**
  - defined in terms of risk differences: 10 percentage points

Example: Anti-hypertensive Trial (cont.)

- **sample size**
  - assumed overall response rate 70%
  - target power 80% $\Rightarrow$ 330 patients per group
- **results**: overall response 42%
  - experimental treatment 133/327
  - control treatment 141/326
- **problem**: power 75% rather than 80%

Example: Anti-hypertensive Trial

Non-inferiority Margin in Terms of Odds Ratio



Designs with Sample Size Re-estimation

- **interim analysis**
  - estimation of **treatment effect**
  - hypothesis test (offers opportunity for early stopping)
  - basically two types
    - classical group sequential designs (e.g. Jennison & Turnbull 1999)
    - designs based on combination of p-values (e.g. Bauer & Köhne 1994)
  - sample size re-estimation could be based on observed treatment effect
- **sample size review**
  - estimation of **nuisance parameters** (e.g. variance), no hypothesis test
  - design with internal pilot study (e.g. Wittes & Brittain 1990)

Internal Pilot Study Design (Wittes & Brittain 1990)

- **initial sample size estimation** $n_0 = n(\alpha, 1-\beta, \Delta^\star, \widehat{\sigma}_0^2)$
  - significance level α, desired power $1-\beta$, clinically relevant effect $\Delta^\star$
  - initial estimate $\widehat{\sigma}_0^2$ of the nuisance parameter σ^2 (from other studies)
- **sample size review**:
  - after recruitment of $n_1 = \pi n_0$ patients (e.g., $\pi = 1/2$)
  - estimation of nuisance parameter $\rightarrow \widehat{\sigma}^2$
  - sample size *re-estimation* $\widehat{N} = n(\alpha, 1-\beta, \Delta^\star, \widehat{\sigma}^2)$
    - "restricted": $n_2 = \max(n_0, \widehat{N}) - n_1$
    - "unrestricted": $n_2 = \max(n_1, \widehat{N}) - n_1$ (Birkett & Day 1994)
- **final analysis**
  - estimation of treatment effect and hypothesis test
  - with all $n_1 + n_2$ patients

Sample Size Re-estimation and International Guidelines

- **ICH Guideline E9 (1998)**, Section 4.4 Sample size adjustment:

  The steps taken to preserve blindness and consequences, if any, for the type I error [. . .] should be explained.

- **CHMP Reflection Paper on Adaptive Designs (2007)**, Section 4.2.2 Sample size reassessment:

  Whenever possible, methods for blinded sample size reassessment [. . .] that properly control the type I error should be used.

- requirements: **blinding** and **control of type I error rate**

Continuous Data

- **data**: normally distributed with equal within-group variances σ^2

- **hypotheses:** $H_0 : \mu_T - \mu_C \leq D$ vs. $H_1 : \mu_T - \mu_C \leq D$ with non-inferiority margin $D < 0$

- **approximate sample size** (Kieser & Hauschke 1999)

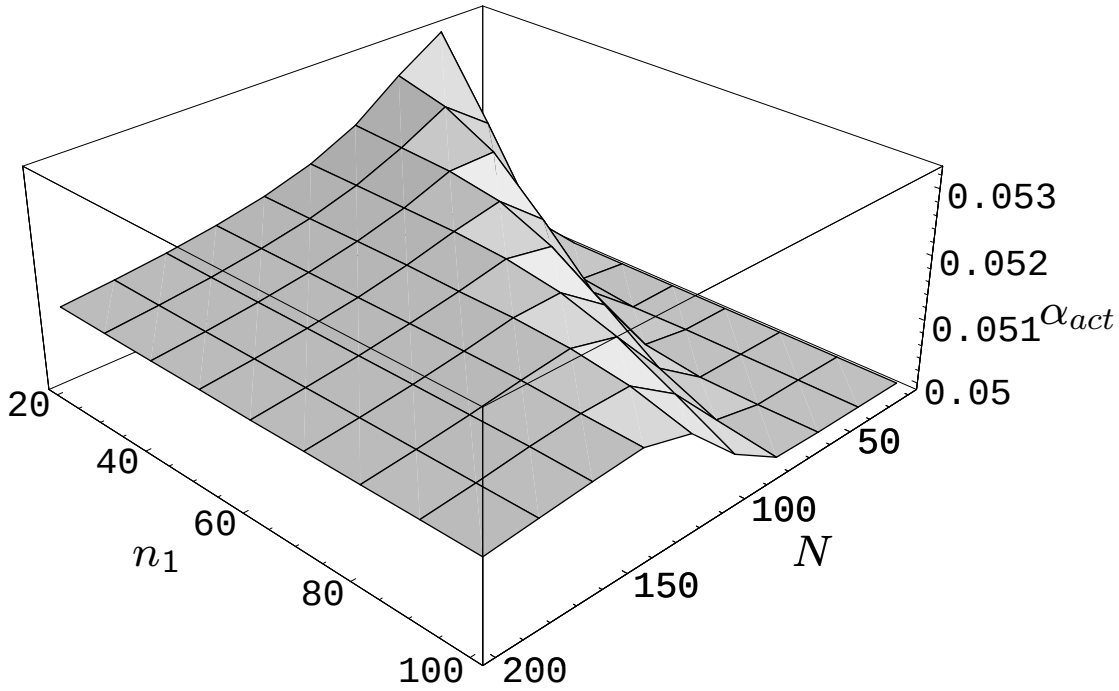
$$n \approx 4 \frac{(\Phi^{-1}(\alpha) + \Phi^{-1}(\beta))^2}{(\mu_T - \mu_C - D)^2} \sigma^2$$

- **sample size adjustment**: re-estimating σ^2

Variance Estimators for Blinded Sample Size Reviews

- **idea**: total variance = within-group + between-group variance
- **one-sample variance** $S^2_{OS} = \frac{1}{n_1-1}\sum_{i,j}(X_{ij} - \bar{X})^2$
  - typically between-group variance relatively small compared to within-group variance; in particular in non-inferiority trials
- **adjusted one-sample variance** (Zucker et al. 1999)
  - idea: S^2_{adj} unbiased under alternative $\Delta = \Delta^\star$
  - $S^2_{adj} = S^2_{OS} - \frac{1}{4}\frac{n_1}{n_1 - 1}\Delta^{\star 2}$

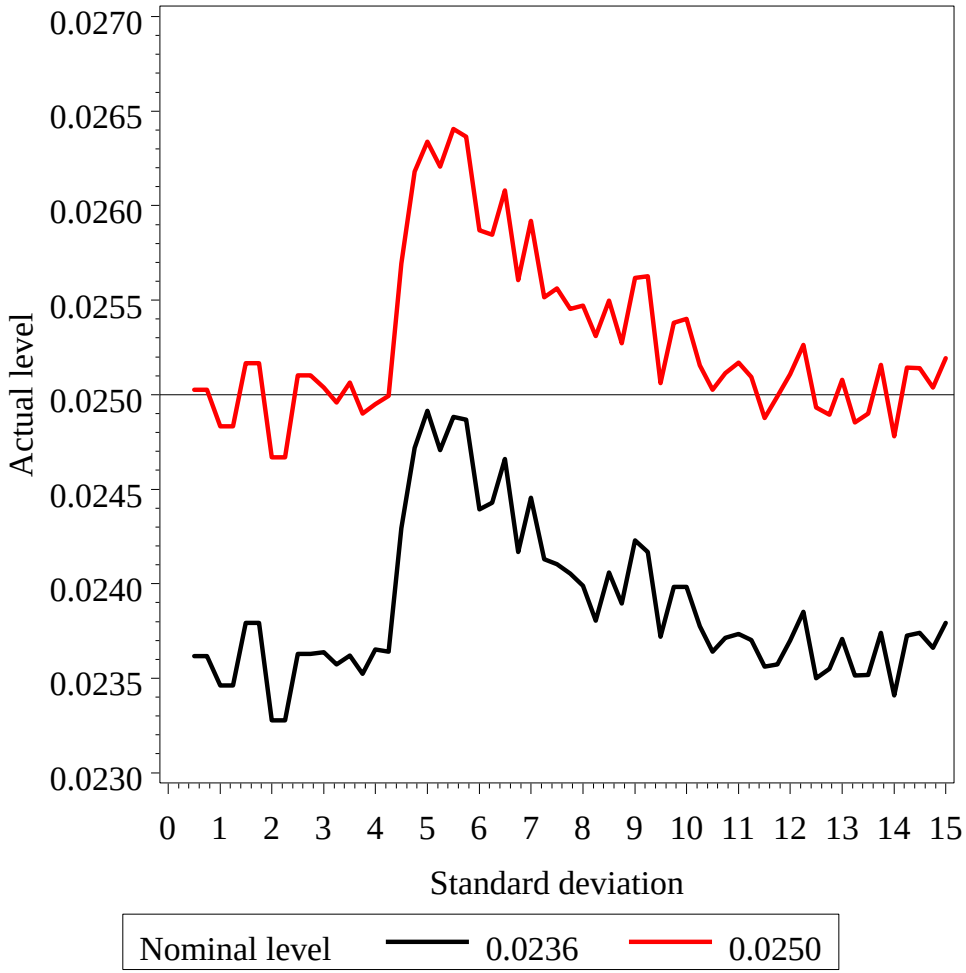
Blinded Sample Size Reviews: Acutal Type I Error Rate



Std. non-inf. margin $D/\sigma = -0.3$

- actual type I error rate α_{act}
- sample size n_1 of internal pilot study
- unknown, required total sample size N
- small type I error rate inflation
- reference: Friede & Kieser (2003)

Control of Type I Error Rate: Example in Depression



- study in severe depression
- endpoint: change in HAMD
- non-inf. margin $D = -2.5$
- internal pilot size $n_1 = 100$
- maximum sample size $N_{max} = 500$
- simulated actual level (1,000,000 replications)

Binary Data: Non-inferiority Tests

- **hypotheses**: $H_0 : \pi_1 \leq \pi_2 - \delta$ vs. $H_1 : \pi_1 > \pi_2 - \delta$
  - non-inferiority margin $\delta > 0$
- **test statistic**: $z = \dfrac{\widehat{\pi}_1 - \widehat{\pi}_2 + \delta}{\text{SE}}$
  - $\widehat{\pi}_i$ MLE of π_i
  - unrestricted (Blackwelder) or restricted (Farrington & Manning) MLE of π_i to compute SE
- **reference distribution**: standard normal distribution

Sample Size Formulae

- **notation**: $\delta_1 = \pi_2 - \pi_1$, $n_2 = \theta n_1$

- **Blackwelder** (1982, Table 1)

$$n_1 = \frac{(\Phi^{-1}(\alpha) + \Phi^{-1}(\beta))^2}{(\delta_1 - \delta)^2} \cdot (\pi_1(1-\pi_1) + \pi_2(1-\pi_2)/\theta)$$

- **Farrington & Manning** (1990, Formula 4)

$$n_1 = \frac{(\Phi^{-1}(\alpha)\sqrt{V_0} + \Phi^{-1}(\beta)\sqrt{V_1})^2}{(\delta_1 - \delta)^2}$$

  where V_0 and V_1 variances under H_0 and H_1 respectively

Blinded Sample Size Reestimation

- **overall response rate** $\widehat{\pi}_{1\bullet}$ estimated from internal pilot

- **response rates** estimated in a blinded fashion by

$$\widehat{\pi}_{11} = \widehat{\pi}_{1\bullet} - \frac{\theta}{1+\theta}\,\delta_1^a \quad \text{and} \quad \widehat{\pi}_{12} = \widehat{\pi}_{1\bullet} + \frac{1}{1+\theta}\,\delta_1^a$$

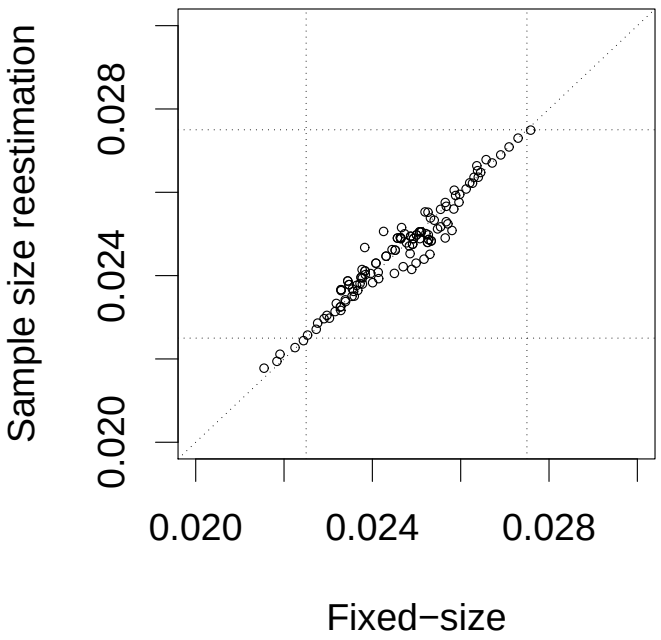
  where δ_1^a assumed treatment difference (often $\delta_1^a = 0$)

- substituting parameters by their estimates in the sample size formulae above gives new **sample size estimate** $\widehat{N}$

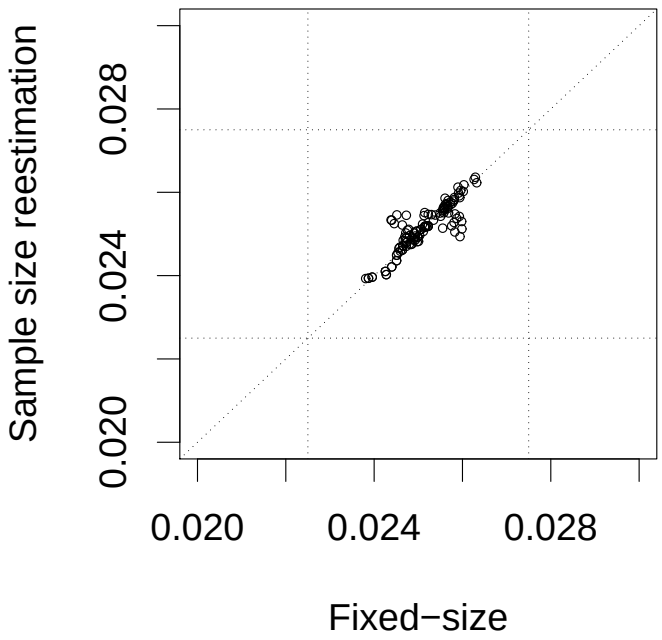
- **sample size adaptation**: $\widehat{N}$ if larger than IPS, otherwise stop recruitment

Type I Error Rate for Blinded Sample Size Reestimation

Blackwelder

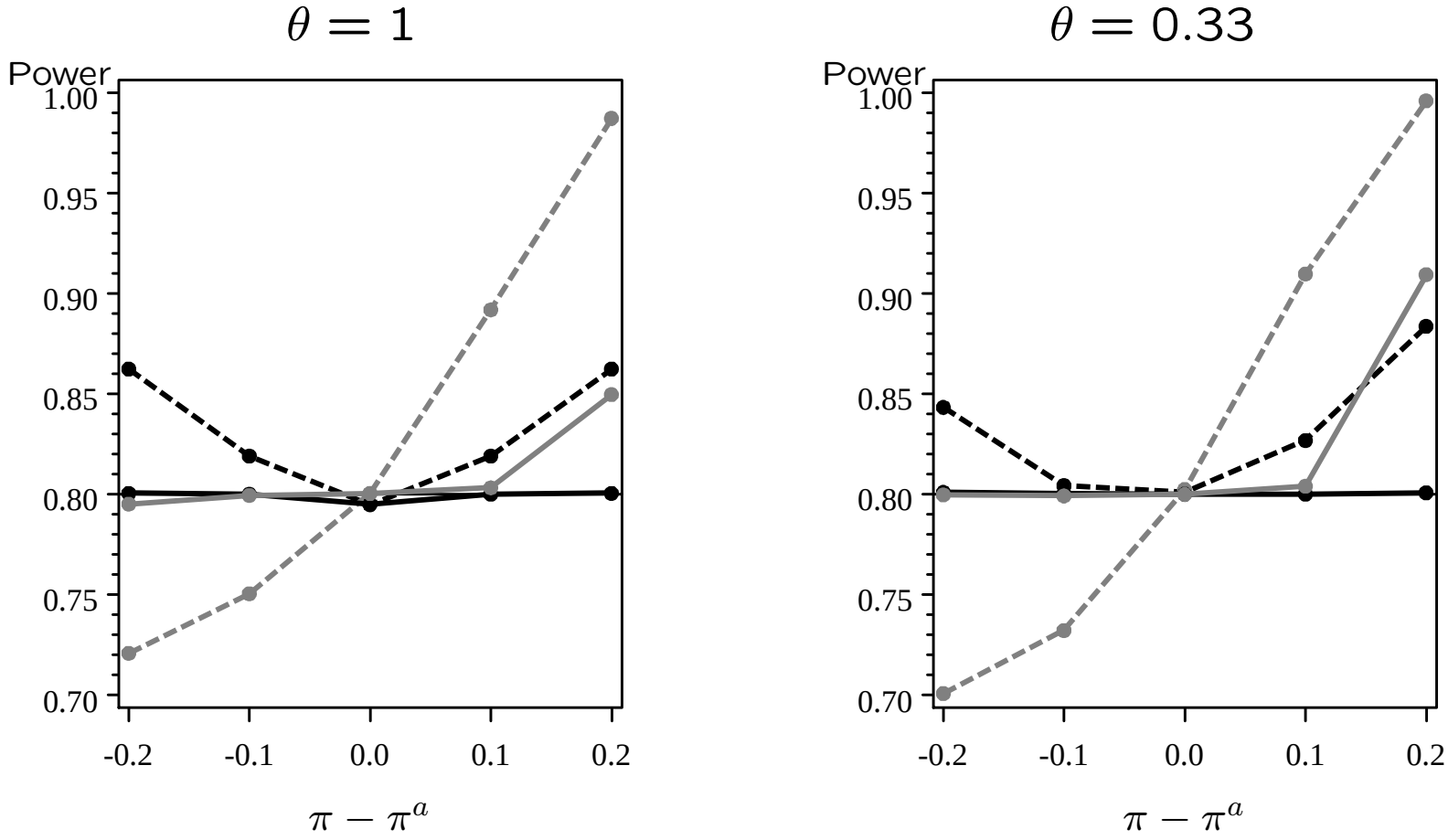


F&M



$\alpha = 0.025,\ 1-\beta = 0.80,\ \theta = 1/3, 1/2, 1,\ \delta = 0.1,\ \delta_1^a = 0,\ \pi = 0.30, 0.31, \ldots, 0.7$

Power of F&M Test: Misspecification of Overall Response



fixed (dashed) and reest. (solid) for $\pi^a = 0.5$ (black), 0.7 (grey) with $\delta_1 = \delta_1^a = 0$

Power and Sample Size for Farrington & Manning Test

$\pi = \pi^a$ and $\delta_1 = \delta_1^a$

		Power		Total sample size			
		Fixed-size	Reest.	Fixed-size	Reestimation		
π^a	θ				5%	Mean	95%
0.5	1/3	80.1%	80.0%	1035	1018	1031.8	1039
	1/2	79.9%	80.0%	876	863	872.9	877
	1	79.5%	79.5%	780	772	776.9	778
0.7	1/3	80.2%	80.0%	815	736	811.7	881
	1/2	80.1%	80.0%	707	641	703.9	761
	1	80.1%	80.0%	658	602	655.8	704

Example: Anti-hypertensive Trial (continued)

- $\alpha = 0.025$, $1 - \beta = 0.80$, $\delta = 0.10$, $\delta_1 = \delta_1^a$, $\theta = 1$
- assumed overall response 70%, actual overall response rate 42%
- **fixed design**
  - total sample size 660 $\Rightarrow$ power 74.5% (B), 74.9% (F&M)
- **blinded sample size reestimation**
  - exp. sample size 759 (B), 754 (F&M); power 79.7% (B, F&M)

Key References

- Birkett MA, Day SJ (1994). Internal pilot studies for estimating sample size. *Statistics in Medicine* **13**: 2455–2463.

- Blackwelder WC (1982). "Proving the Null Hypothesis" in Clinical Trials. *Controlled Clinical Trials* **3**, 345–353.

- Farrington CP, Manning G (1990). Test statistics and sample size formulae for comparative binomial trials with null hypothesis of non-zero risk difference or non-unity relative risk. *Statistics in Medicine* **9**, 1447–1454.

- Friede T, Kieser M (2003). Blind sample size reassessment in non-inferiority and equivalence trials. *Statistics in Medicine* **22**, 995–1007.

- Friede T, Mitchell C, Müller-Velten G (2007). Blinded sample size reestimation in non-inferiority trials with binary endpoints. *Biometrical Journal* (in press).

- Kieser M, Hauschke D (1999). Approximate sample sizes for testing hypotheses about the ratio and difference of two means. *Journal of Biopharmaceutical Statistics* **9**: 641–650.

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

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