Evaluation of two-stage bioequivalence design

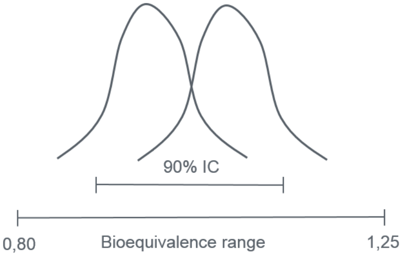
Mélanie Bovo

École nationale de la statistique et de l'analyse de l'information

28/11/2016

Table of contents

BE studies



Studies are assumed to be

- Comparing two treatments
- Using randomized cross-over designs
- Balanced

Test are assumed to be

Performed for average BE

Two-stage BE designs

Interest for case with

- Lack of information on some parameters
  - Geometric Mean Ratio between treatments (GMR)
  - Within-subject variability

  $\rightarrow$ Two-stage design to fine-tune non fully reliable information
- High within-subject variability $\rightarrow$ big sample size

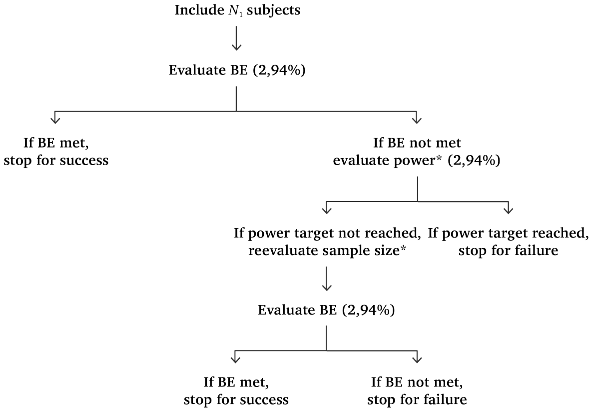
  $\rightarrow$ Two-stage design to be able to stop early

Main issue

Maintaining a global type I error at a 5% level
$\rightarrow$ Creation of adapted designs

Table of contents

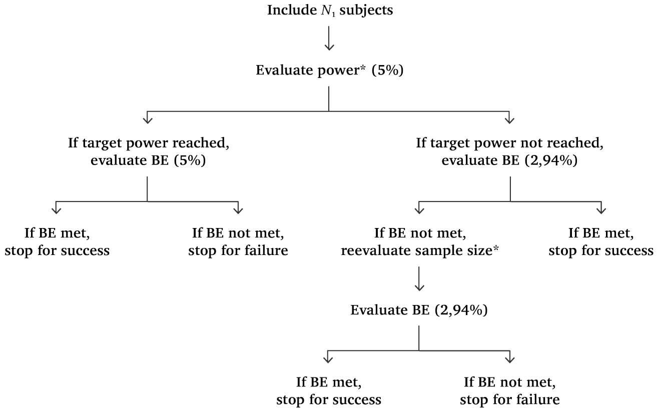
Potvin's method B



$*$ based on estimated within-subject variability and assumed GMR ▼ ▼

- Possibility of adding a futility rule: maximal sample size N_{max}
- Alternative: Zheng's method MSDBE
  - $\rightarrow$ Same algorithm, test levels: α_1=1% and α_2=4%

Potvin's method C



$*$ based on estimated within-subject variability and assumed GMR

- Possibility of adding a futility rule: maximal sample size N_{max}
- Alternative: Potvin's method D
  - $\rightarrow$ Same algorithm, test levels: α_1=2.8% and α_2=2.8%

Literature evaluation of designs

Notations

N: Total sample size of the two-stage design

N_1: First-stage sample size

N_0: Classical BE one-stage design sample size

Incomplete evaluation

Absolute N_1: 12, 24, 36, 48

Absolute N_{max}: no N_{max}, 150

Issue: N_1 and N_{max} not adapted to N_0

Table of contents

Lan & DeMets spending functions

Test levels issue with literature designs

- Each method has different test levels
- Type I error inflations with all literature test levels

Pocock spending function

$\alpha_1(t^*) = \alpha \log(1 + (e - 1)t^*)$
$\rightarrow$ t^*: information fraction at first stage
α_2 determine from α_1 and estimated t^*, assuming correlated bivariate standard normal distribution

Issue: N unknown $\rightarrow$ estimation of t^*

- Naive estimation: $\frac{N_1}{N_0}$ $\rightarrow$ Some type I error inflations ▼
- Better estimation: $\frac{N_1}{1.2N_0}$ $\rightarrow$ Better properties

Evaluated designs

Evaluated designs

- Method B
- Method C
- Method D
- Method MSDBE
- Lan & DeMets - Pocock adapted method B with information fraction estimated at $\frac{N_1}{1.2N_0}$ (denoted BP)
- Lan & DeMets - Pocock adapted method C with information fraction estimated at $\frac{N_1}{1.2N_0}$ (denoted CP)

Implemented scenarios

Studies characteristics

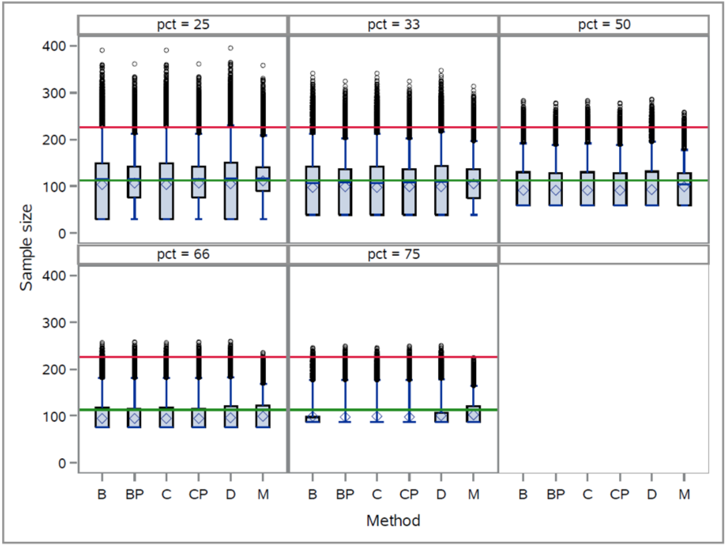
- GMR: 0.80 (type I error), 0.90 and 0.95 (power)
- Target power: 80% and 90%
- SD_w: 0.20, 0.30, 0.40 and 0.50
- N_1: 25%, 33%, 50%, 66% and 75% of N_0
- N_{max}: no limit, $2N_0$ and $1.5N_0$

Simulations: 1 000 000 per scenario

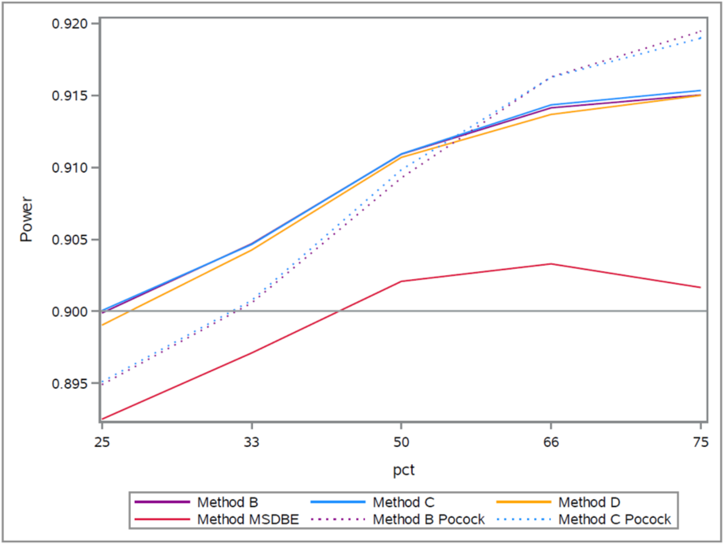
Table of contents

Table of contents

Sample size comparison



Empirical power comparison



Empirical type I error comparison

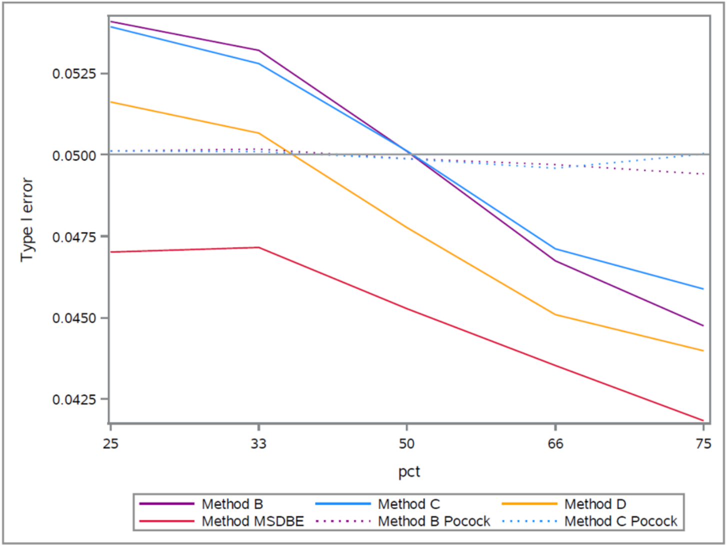
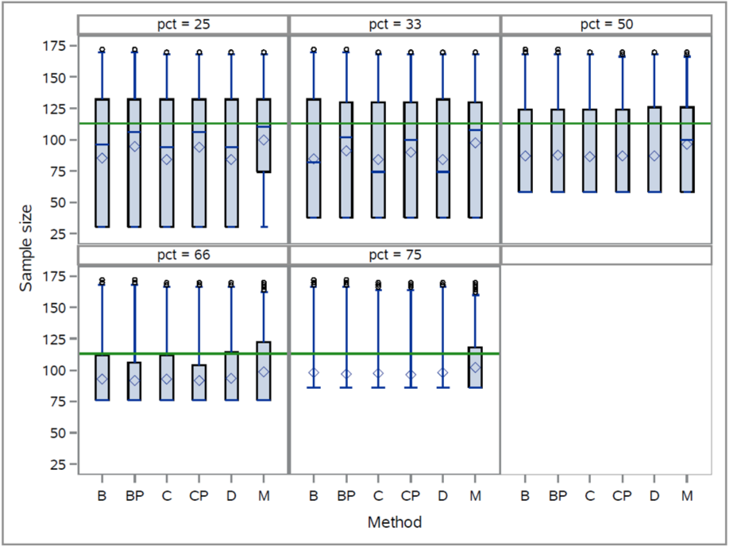
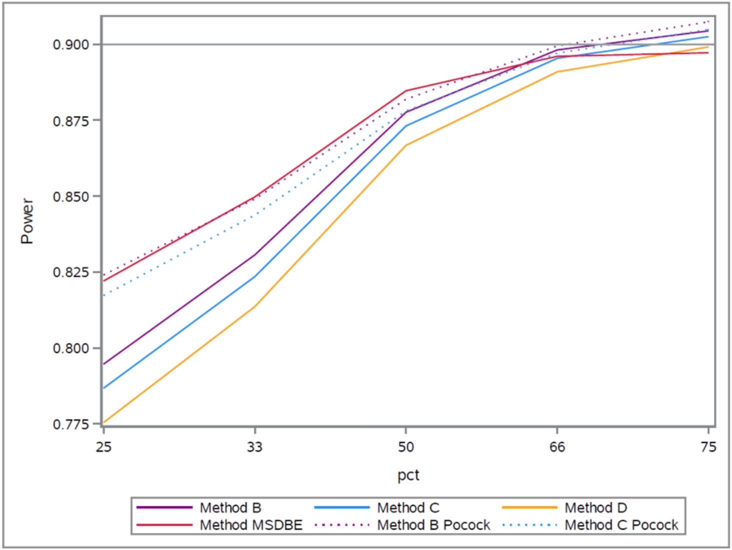


Table of contents

Sample size comparison



Empirical power comparison



Empirical type I error comparison

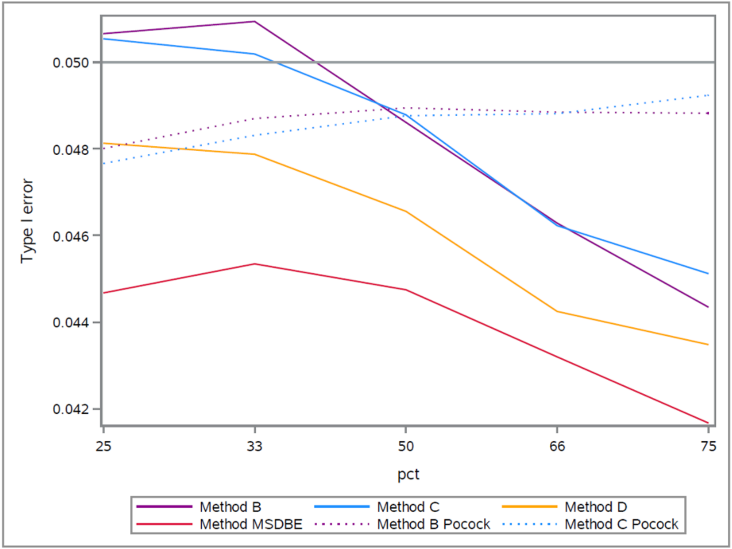


Table of contents

Global patterns

- Similar patterns for other scenarios ▼ ▼
- Target power not reached for:
  - Low $\frac{N_1}{N_0}$ values
  - N_{max}=$1.5N_0$ futility rule
- Small type I errors for scenarios with small sample size (true for any BE designs)
  - Asymmetrical distribution of SD_w and thus SE, more visible for small sample size
  - Visible in the first stage of two-stage BE designs (if small N_1)
- Type I error inflations for scenarios with low $\frac{N_1}{N_0}$ values for BP and CP methods with naive t^* estimation
  - Gap between true $\left(\sqrt{\frac{N_1}{N}}\right)$ and estimated $\left(\sqrt{\frac{N_1}{N_0}}\right)$ correlations

  $\rightarrow$ Over-estimated correlation

Recommendations

Method

Method B with adapted test levels based on Pocock spending function
If not, classical method B (already rather accepted)

Futility rule

Advised

- With a $2N_0$ maximal sample size, impose $\frac{N_1}{N_0}$ values $\geq 50\%$
- With a $1.5N_0$ maximal sample size, impose $\frac{N_1}{N_0}$ values $\geq 66\%$ and a target power of 90%

Limit of evaluated methods

- **Potvin's methods B and C:**
  - No genuine theoretical justification of test levels (information fraction not taken into account)
  - Type I error inflations in some situations (high treatment difference and low $\frac{N_1}{N_0}$)
- **Potvin's method D:**
  - Arbitrarily chosen test levels
  - Too conservative in some situations in terms of power (low treatment difference and high $\frac{N_1}{N_0}$)
- **Zheng's method MSDBE:**
  - Bonferroni correction $\rightarrow$ most conservative method
- **Adapted test levels methods:**
  - Several spending functions explored
  - Only monotonous functions
  - Unjustified information fraction estimation

Further explorations

To continue on evaluated designs

- Optimized test levels ▼
  - Might be non monotonous against information fraction
  - Simulation-based test levels
- Use of observed GMR
  - Low power $\rightarrow$ use GMR quantile

To explore new designs

- Conditional power
  - Promizing zone approach
- Predictive power

Thank you for your attention!

References

[1] Potvin D. *et al.*, Sequential design approaches for bioequivalence studies with crossover designs, *Pharmaceutical statistics*, 2008, vol. 7, no 4, p. 245-262.

[2] Montague T. H. *et al.*, Additional results for "Sequential design approaches for bioequivalence studies with crossover designs", *Pharmaceutical statistics*, 2012, vol. 11, no 1, p. 8-13.

[3] Zheng C. et al., Modifications of sequential designs in bioequivalence trials, *Pharmaceutical statistics*, 2015, vol. 14, no 3, p. 180-188.

[4] Fuglsang A., Futility Rules in Bioequivalence Trials with Sequential Designs,*The AAPS journal*, 2014, vol. 16, no 1, p. 79-82.

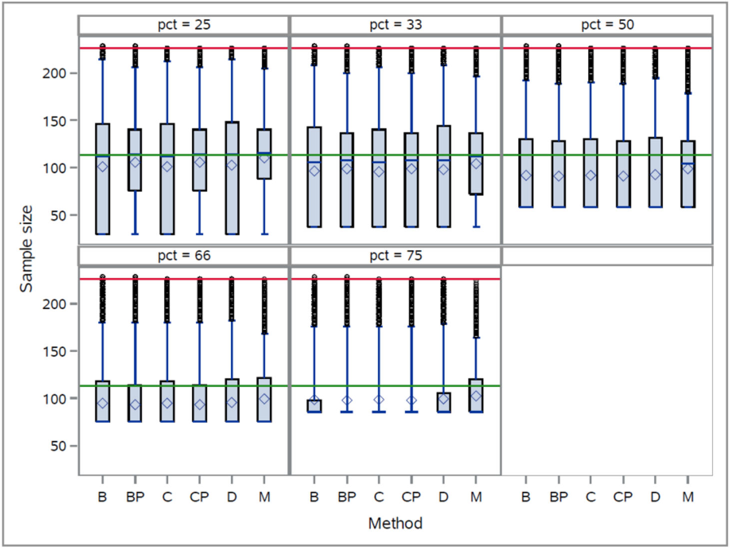
[5] Karalis V., The role of the upper sample size limit in two-stage bioequivalence designs, *International journal of pharmaceutics*, 2013, vol. 456, no 1, p. 87-94.

[6] Pocock S. J., Group sequential methods in the design and analysis of clinical trials, *Biometrika*, 1977, vol. 64, no 2, p. 191-199.

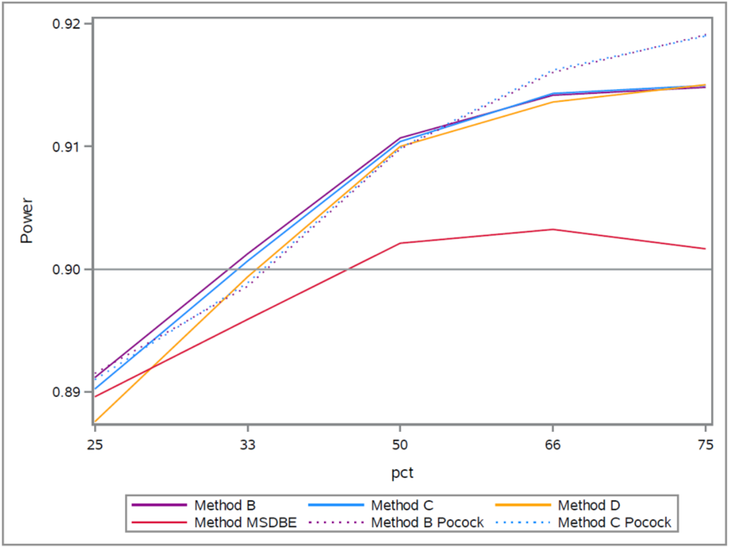
[7] DeMets D.L et Lan K.K.G., Interim analysis: the alpha spending function approach, *Statistics in medicine*, 1994, vol. 13, no 13-14, p. 1341-1352.

[8] Kieser M., *et al.*, Two-stage designs for cross-over bioequivalence trials, *Statistics in medicine*, 2015, vol. 34, no. 16, p. 2403-2416.

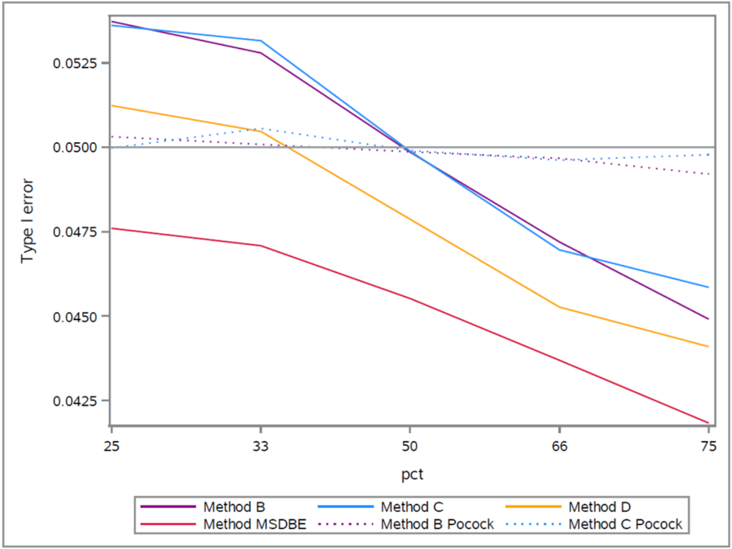
Sample size comparison - $N_{max} = 2N_0$



Empirical power comparison - $N_{max} = 2N_0$



Empirical type I error comparison - $N_{max} = 2N_0$



Other results

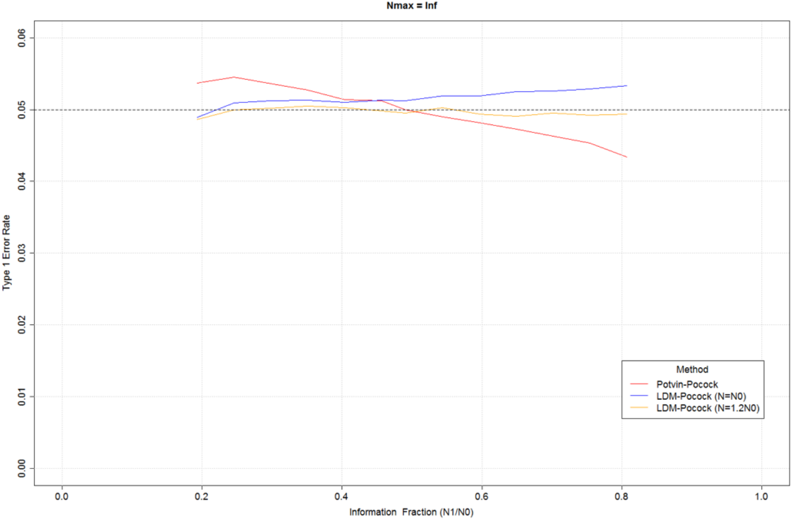
- Results for literature designs, GMR of 0.90
- Results for literature designs, GMR of 0.95

- Results for Lan-DeMets adapted designs, GMR of 0.90 and $t^* = \frac{N_1}{N_0}$
- Results for Lan-DeMets adapted designs, GMR of 0.95 and $t^* = \frac{N_1}{N_0}$

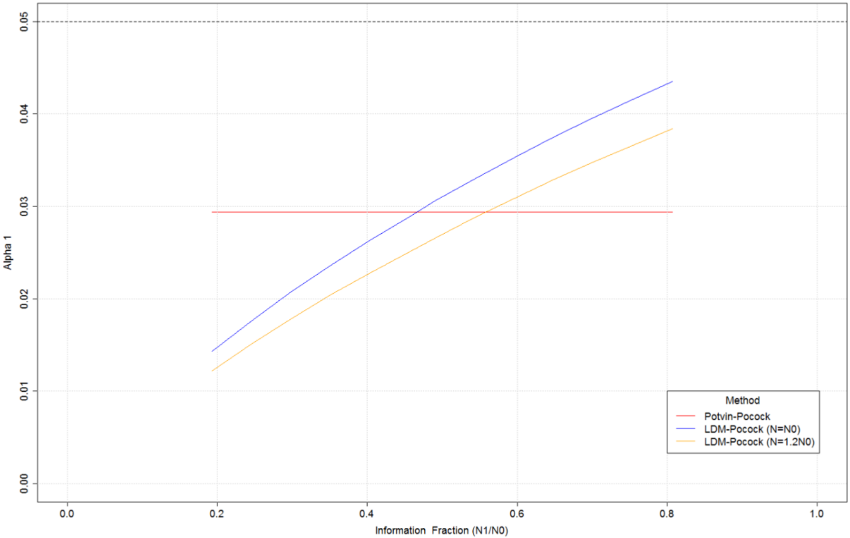
- Results for Lan-DeMets adapted designs, GMR of 0.90 and $t^* = \frac{N_1}{1.2N_0}$
- Results for Lan-DeMets adapted designs, GMR of 0.90 and $t^* = \frac{N_1}{1.2N_0}$

Back

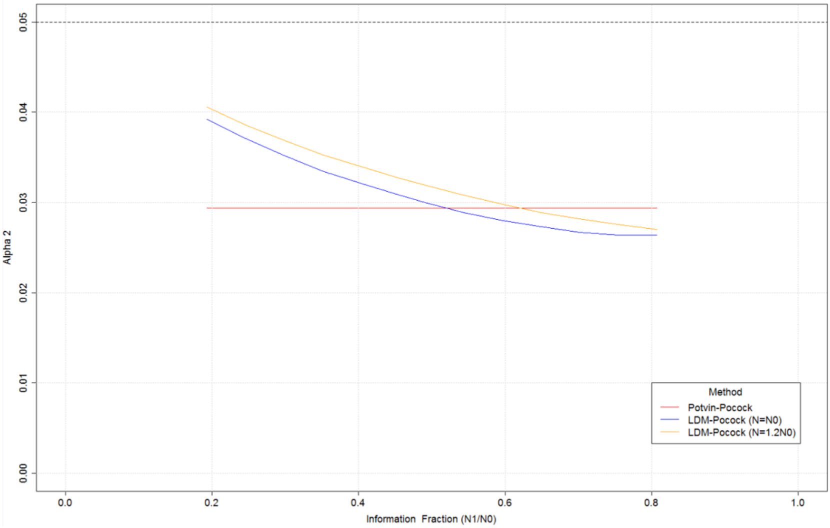
Lan & DeMets type I errors



Lan & DeMets α_1 test levels

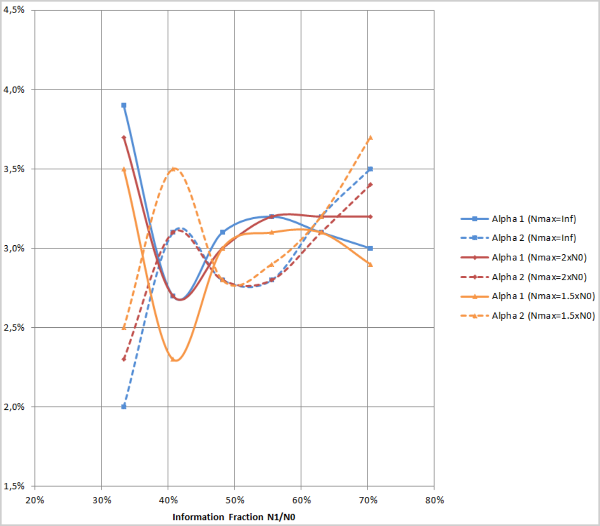


Lan & DeMets α_2 test levels



Back

Optimized test levels



Back

Power computation

Let $\hat{\theta} = \bar{X}_T - \bar{X}_R \sim \mathcal{N}\left(\theta, \frac{2\sigma_w^2}{n}\right)$

Therefore, $\mu\frac{\hat{\sigma}_w^2}{\sigma_w^2} \sim \chi_\mu^2$ and $\frac{\hat{\theta}-\theta}{\hat{\sigma}_w\sqrt{\frac{2}{n}}} \sim \tau_\mu$ with μ: dl

Test: H_0: $H_0^L \cup H_0^U$ vs H_1: $H_1^L \cup H_1^U$

$$H_0^L : \theta < \theta_L \text{ vs } H_1^L : \theta \geq \theta_L \text{ and } H_0^U : \theta > \theta_U \text{ vs } H_1^U : \theta \leq \theta_U$$

Test statistics: $T_L = \frac{\hat{\theta}-\theta_L}{\hat{\sigma}_w\sqrt{\frac{2}{n}}}$ and $T_U = \frac{\hat{\theta}-\theta_U}{\hat{\sigma}_w\sqrt{\frac{2}{n}}}$

H_0 is rejected if $T_L \geq t_{\mu,1-\alpha}$ and $T_U \leq t_{\mu,\alpha}$
Hence power can be expressed:

$$P(\theta, \sigma_w, \mu) = \mathbb{P}[T_L \geq t_{\mu,1-\alpha} \text{ and } T_U \leq -t_{\mu,1-\alpha} | \theta, \sigma_w, \mu]$$

Power computation

Yet, $T_L = \frac{\hat{\theta}-\theta_L}{\hat{\sigma}_w\sqrt{\frac{2}{n}}} = \frac{\hat{\theta}-\theta+\theta-\theta_L}{\sigma_w\sqrt{\frac{2}{n}}} * \frac{1}{\sqrt{\frac{\hat{\sigma}_w^2}{\sigma_w^2}\frac{\mu}{\mu}}}$

As $\hat{\theta} \sim \mathcal{N}\left(\theta, \frac{2\sigma_w^2}{n}\right)$, $\frac{\hat{\theta}-\theta+\theta-\theta_L}{\sigma_w\frac{2}{n}} \sim \mathcal{N}(\theta-\theta_L, 1)$

And as $\frac{\hat{\sigma}_w^2}{\sigma_w^2}\mu \sim \chi^2_\mu$, $T_L \sim \tau_\mu\left(\frac{\theta-\theta_L}{\sigma_w\sqrt{\frac{2}{n}}}\right)$ (Decentered Student) Hence,

$$P(\theta, \sigma_w, \mu) = \mathbb{P}\left[\tau_\mu\left(\frac{\theta-\theta_L}{\sigma_w\sqrt{\frac{2}{n}}}\right) \geq t_{\mu,1-\alpha}\right.$$
$$\left.\text{and } \tau_\mu\left(\frac{\theta-\theta_U}{\sigma_w\sqrt{\frac{2}{n}}}\right) \leq -t_{\mu,1-\alpha} \middle| \theta, \sigma_w, \mu\right]$$

Back

Within-subject variability estimation

$ln(T) - ln(R) \sim \mathcal{N}(ln(\theta), 2\sigma^2)$
With θ: true GMR and σ^2: true intra-subject variance

$$Power = F_t\left(\frac{\ln(\frac{1.25}{\theta})}{SD\sqrt{\frac{2}{n}}} + t_{1-\alpha,DF}, DF\right) - F_t\left(-\frac{\ln(1.25\theta)}{SD\sqrt{\frac{2}{n}}} + t_{1-\alpha,DF}, DF\right)$$

With

- SD: *sigma* estimator
- DF: degrees of freedom
- $F_t(x, DF)$: cumulative distribution function of Student's t density function with DF degrees of freedom
- $t_{1-\alpha,DF}$: $(1-\alpha)$th percentile of a Student's t density function

Within-subject variability estimation

First stage SD calculation

Let X_{ijk} (normally distributed) represent $ln(T) - ln(R)$ for the kth subkect in sequence j of stage i, the error sum of square SS1 is:

$$SS1 = \frac{1}{2}\sum_{j=1}^{2}\sum_{k=1}^{\frac{n_1}{2}}(X_{1ij} - \bar{X}_{1j.})^2 = \frac{1}{2}\sum_{j=1}^{2}\left[\left(\sum_{k=1}^{\frac{n_1}{2}} X_{1jk}^2\right) - \frac{\left(\sum_{k=1}^{\frac{n_1}{2}} X_{1jk}\right)^2}{\frac{n_1}{2}}\right]$$

With $\bar{X}_{1j.} = \frac{2}{n_1}\sum_{k=1}^{\frac{n_1}{2}} X_{1jk}$. Hence

$$SD_1^2 = \frac{SS1}{n_1 - 2}$$

Within-subject variability estimation

Second stage SD calculation

If $n_2 = 0,\ SD_2^2 = \frac{SS1}{n_1-2}$

If $n_2 = 2,\ SD_2^2 = \frac{SS1+SSmean}{n-3}$

If $n_2 > 2,\ SD_2^2 = \frac{SS1+SSmean+SS2}{n-3}$

With

- $SS2 = \frac{1}{2}\sum_{j=1}^{2}\sum_{k=1}^{\frac{n_1}{2}}(X_{2jk} - \bar{X}_{2j.})^2 = \frac{1}{2}\sum_{j=1}^{2}\left[\left(\sum_{k=1}^{\frac{n_1}{2}} X_{2jk}^2\right) - \frac{\left(\sum_{k=1}^{\frac{n_1}{2}} X_{2jk}\right)^2}{\frac{n_2}{2}}\right]$
- $\bar{X}_{2j.} = \frac{2}{n_2}\sum_{k=1}^{\frac{n_1}{2}} X_{2jk}$
- $SSmean = \frac{(\bar{X}_{1..} - \bar{X}2..)^2}{\frac{2}{n_1}+\frac{2}{n_2}}$

Back

Confidence intervals computation

First stage:

$$\bar{X}_{1..} + -t_{1-\alpha(n_1-2)}\sqrt{s_1^2\frac{2}{n_1}}$$

Second stage:

$$\bar{X}_{...} + -t_{1-\alpha(n-3)}\sqrt{s_2^2\frac{2}{n}}$$

Back