Planning crossover bioequivalence trials accounting for the uncertainty of the assumptions

A. Ring[1,2], B. Lang[3], I. Dan Dije[4] , C. Kazaroho[5], D. Labes[6], R. Schall[1,7], H. Schütz[8]

1 University of the Free State, South Africa
2 medac GmbH, Germany
3 Boehringer Ingelheim, Germany
4 AIMS Senegal, Senegal
5 AIMS Rwanda, Rwanda
6 Consultant, Germany
7 IQVIA Biostatistics, Bloemfontein, South Africa
8 BEBAC, Austria

African Institute of Mathematical Sciences

medac

- 5 African students have worked in this field
  - 2017 – Gilbert Tumusabe – AIMS Rwanda – Superiority trials
  - 2018 – Didier Habima – AIMS Rwanda – Superiority trials
  - 2018 – Claudine Kazaroho – AIMS Rwanda – Bioequivalence trials
  - 2019 – Ibrahim Dan Dije – AIMS Senegal – Bioequivalence trials
  - 2019 – Sinclair Awounvo – University of Bremen – Systematic review of BE trial planning

- *1st paper recently published*
- *2 more manuscripts in preparation*

Received: 10 December 2018 | Revised: 28 May 2019 | Accepted: 3 June 2019

DOI: 10.1111/bcp.14055

ORIGINAL ARTICLE

BJCP BRITISH PHARMACOLOGICAL SOCIETY

Sample size determination in bioequivalence studies using statistical assurance

A. Ring[1,2] | B. Lang[3] | C. Kazaroho[4] | D. Labes[5] | R. Schall[1,6] | H. Schütz[7]

Overview

- H. Schütz presentation – Biosimilars 2017
  - AIMS students
- Setup of bioequivalence trials
  - Power calculation
- Definition of Assurance as Expected Power
  - Expected Coefficient of Variation
  - Expected T/R ratio
- Results
  - Comparison of Power and Assurance
- Conclusion

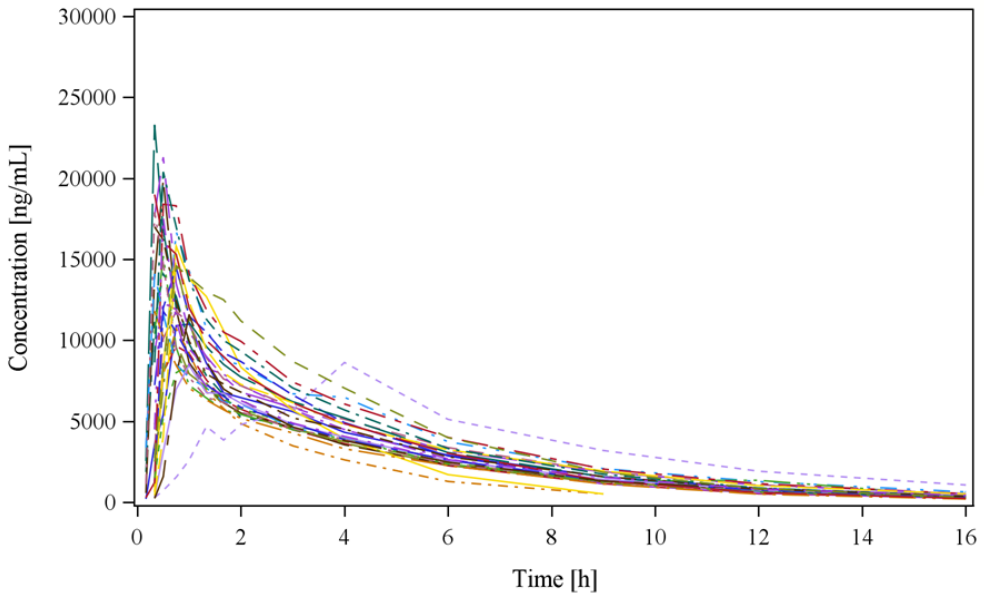
Bioequivalence

- Brand name of a drug on the market (gold standard)
- Patent protection expired
- Other company wants to sell a generic formulation

- WHO Definition:
  Two pharmaceutical products are bioequivalent
  if … their bioavailabilities,
  in terms of **rate** and **extent** of absorption …
  are similar to such a degree that their effects can be expected to be essentially the same

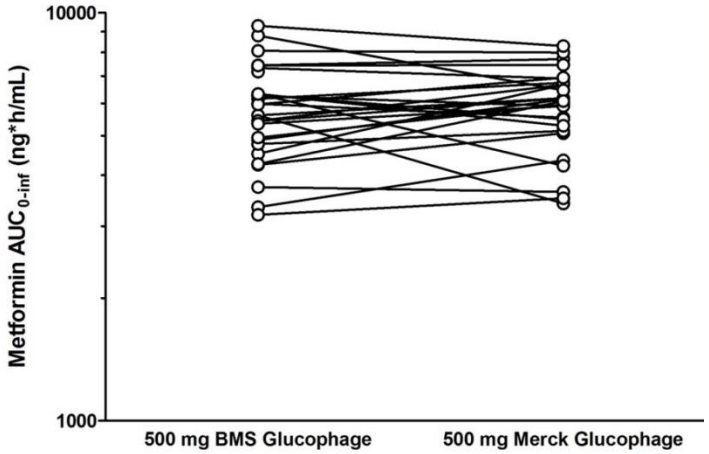
Bioequivalence trials

- New formulation (Test)
- Existing formulation (Reference)
- Comparison of pharmacokinetic profiles in Humans following single p.o. administration
- Typical 2x2 design: 2 treatment, 2 period ***crossover*** trial



Bioequivalence trials

- Primary Endpoints: Pharmacokinetic metrics
  - C_{max}
  - AUC_{tz}, $AUC\infty$
- PK metrics are log-normally distributed
  - Analysis following log-transformation
- To demonstrate that they are „on average similar“

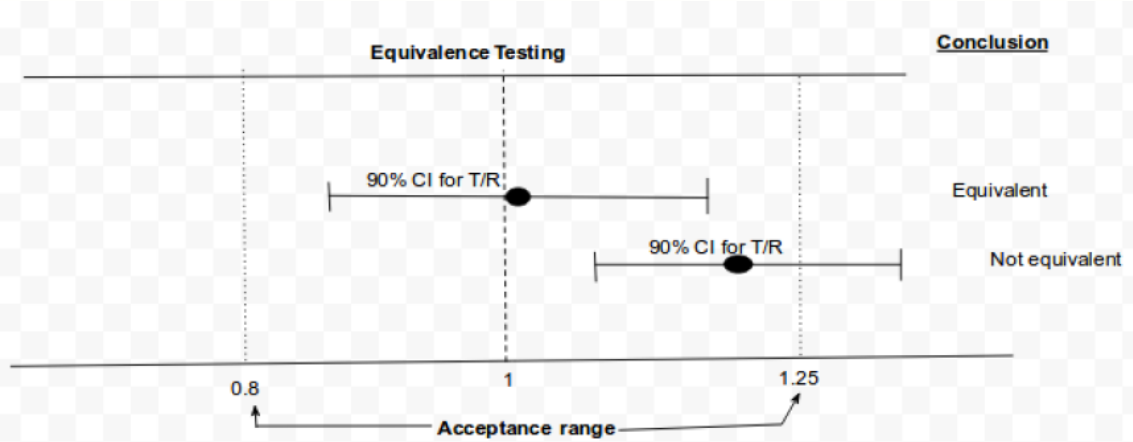


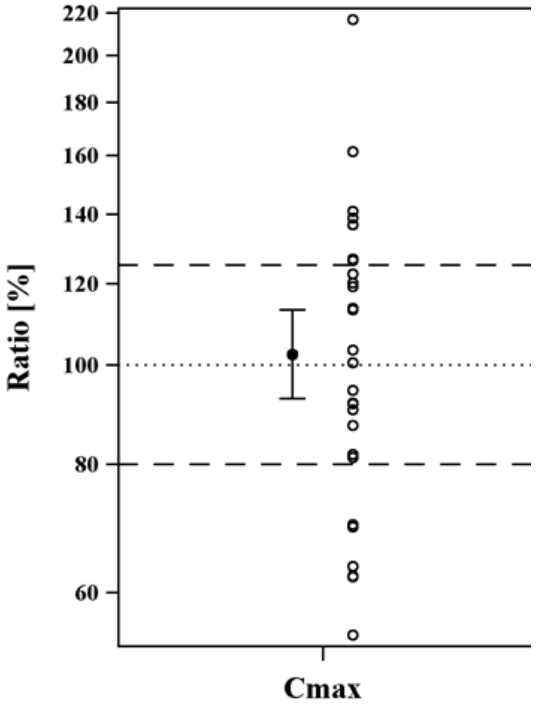
Bioequivalence trials

- Statistical evaluation
  - Ratio of PK endpoints between T and R and its 90%CI

$$\theta = \frac{C_{max}^{T}}{C_{max}^{R}}$$

  - Acceptance range [80.00%,125.00%] (per BE guidelines)



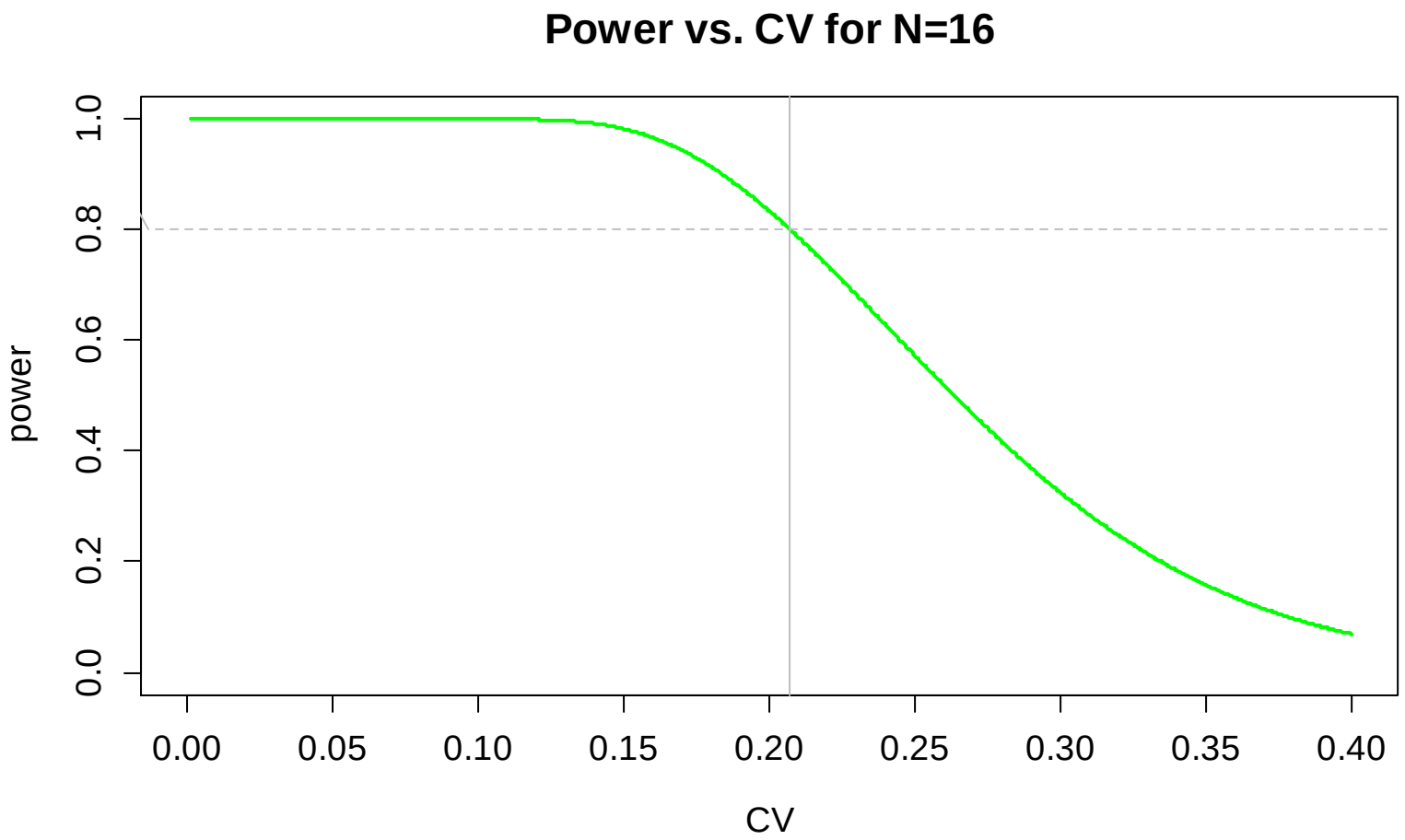


Sample size determination

- Sample size should be sufficiently large to demonstrate BE (with high probability) – but not larger
  - Efficient cost & time

- 6 parameters are combined when considering the power
  - Margin m
  - Type-I error α (consumer risk)
  - Type-II error β ➔ Power $\pi = 1-\beta$
  - Anticipated T/R-ratio θ
  - (intraindividual) Coefficient of variation CV
  - Total sample size N

Power and Coefficient of Variation



Sample size determination

- 6 parameters are combined when considering the power
  - Margin *m=0.8 (1.25)*
  - Type-I error α=0.05 per BE guidelines
  - Type-II error β
    ➔ Power $\pi = 1-\beta$ = 80% or 90% often selected
  - Anticipated T/R-ratio of the metric θ
  - (intraindividual) Coefficient of variation *CV*
    ➔ Total sample size *N*

- *CV*: selected using data of previous PK trials of this drug
- There are generally no data for estimating θ (relative bioavailability is not known for this formulation)

Example

- 6 parameters are combined when considering the power
  - Margin $m=0.8$ (1.25)
  - Type-I error $\alpha=0.05$ per BE guidelines
  - Type-II error β
    ➔ Power $\pi=1-\beta=90\%$
  - Anticipated T/R-ratio of the metric $\theta = 1.00$
  - (intraindividual) Coefficient of variation $CV=0.3$
    ➔ Total sample size $N=30$

- **Can we be sure about the values of CV and θ ?**

How to select the value for CV?

- (intraindividual) Coefficient of variation *CV*
  - Drug property, often determined by the variability of the drug elimination
  - To be obtained from <u>previous</u> pharmacokinetic crossover trials
  - Sample sizes in such PK trials are often rather low
  - **Only an estimate of CV was obtained !!**

How to select the value for CV?

- (intraindividual) Coefficient of variation *CV*
  - **Only an estimate of *CV* was obtained !!**
- Example
  - Previous trial had 24 subjects
  - Estimated *CV* of C_{max}: 20%
  - Estimated *CV* of AUC: 16%
- 90% confidence intervals of *CV* using the χ^2 distribution

$$100(1-\alpha)CI_{CV} = [L_B, U_B] = \left[\sqrt{\exp\left(\frac{(n-1)S^2}{\chi^2_{(n-1),(1-\alpha/2)}}\right)-1}, \sqrt{\exp\left(\frac{(n-1)S^2}{\chi^2_{(n-1),(\alpha/2)}}\right)-1}\right]$$

  - C_{max}: (16.1%, 26.9%)
  - AUC: (12.9%, 21.5%)

➔ **There is considerable uncertainty!**

Statistical Assurance for Clinical Trials

- Power considerations to determine sample size in comparative clinical trials
- Study power depends on fixed values of assumptions
  - No room for uncertainty

- Assurance uses a distribution on one or several of the assumed model parameters
  - Distribution accounts for uncertainty

- Assurance reflects “probability of success” given these distributions

Determining the Statistical Assurance

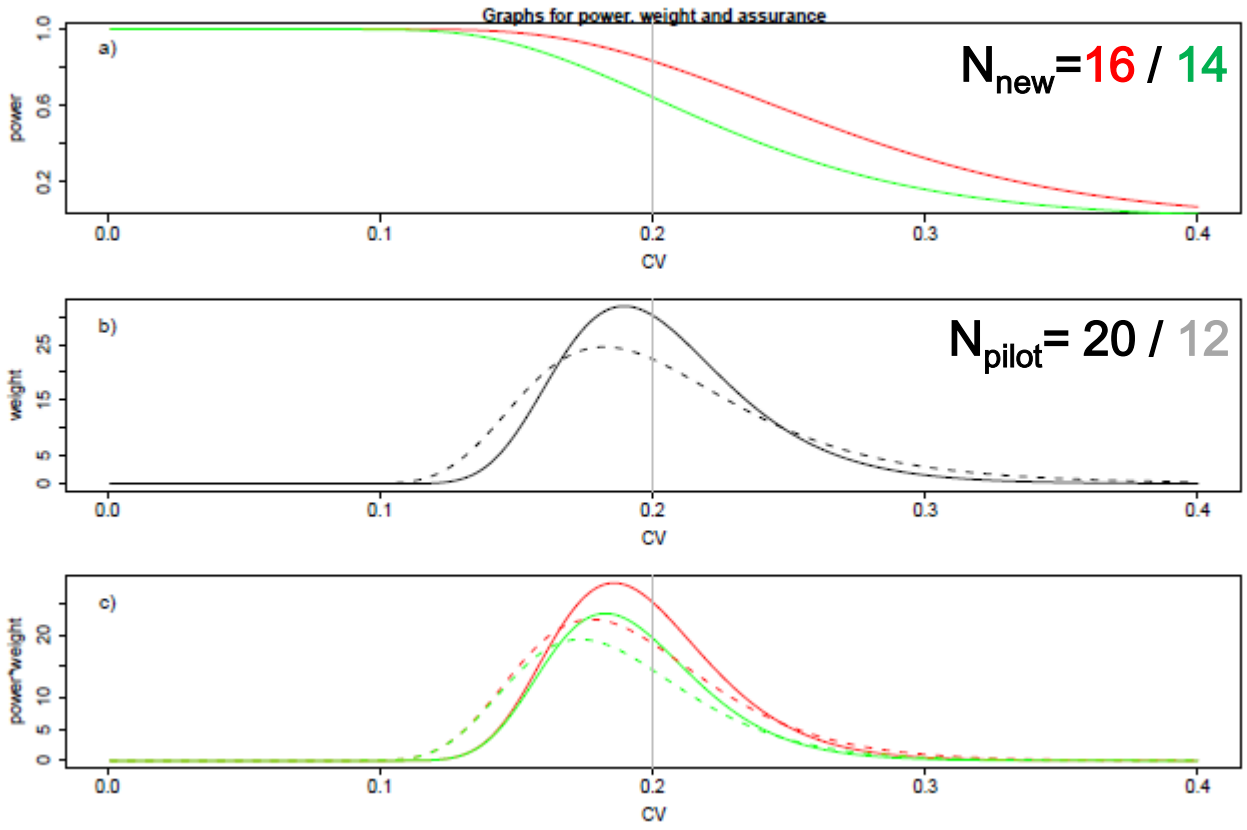
- Instead of taking a single value of CV, consider a distribution

$$\gamma = P(\text{Successful trial}) = \int \pi(\boldsymbol{CV}) W(\boldsymbol{CV}) d\boldsymbol{CV}$$

Power π

Weight W

Product



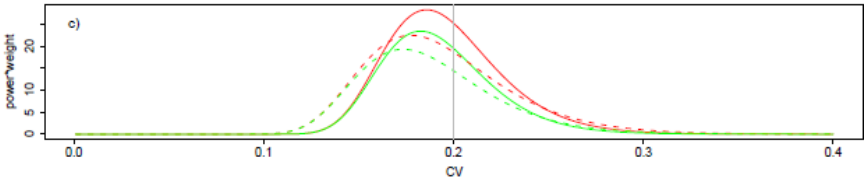
Comparison Power to Assurance

- Instead of taking a single value of CV,
  consider a distribution of potential values of CV

- Power function π for various values of CV
- Weight function W
  (Inverse gamma distribution with CV=0.2 and Pilot-N)

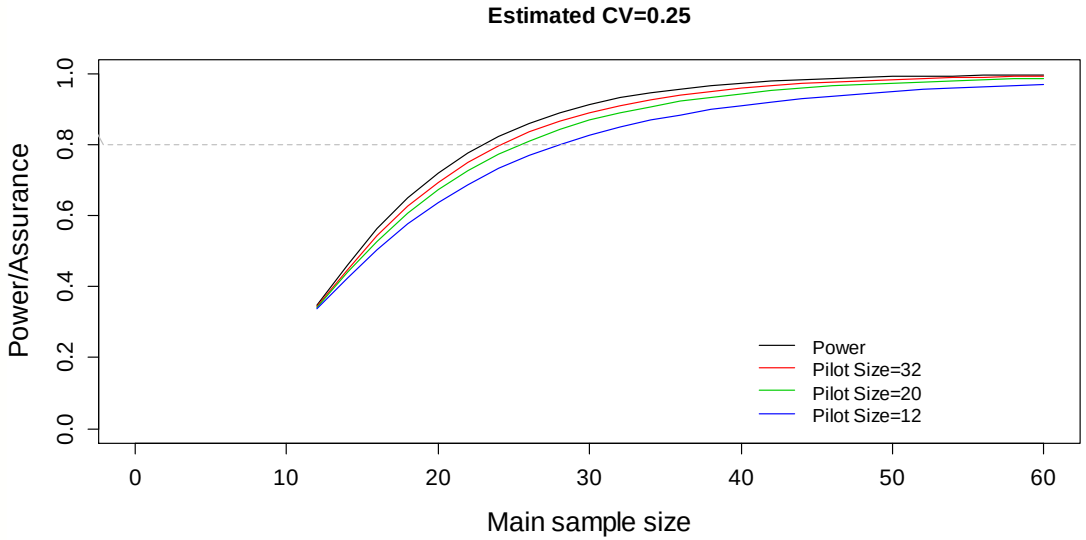
- Assurance of **new** trial (given Pilot N and estimated CV)

	New N=14	New N=16
Pilot N = 12	66%	70%
Pilot N = 20	74%	77%
Power	*76%*	*83%*

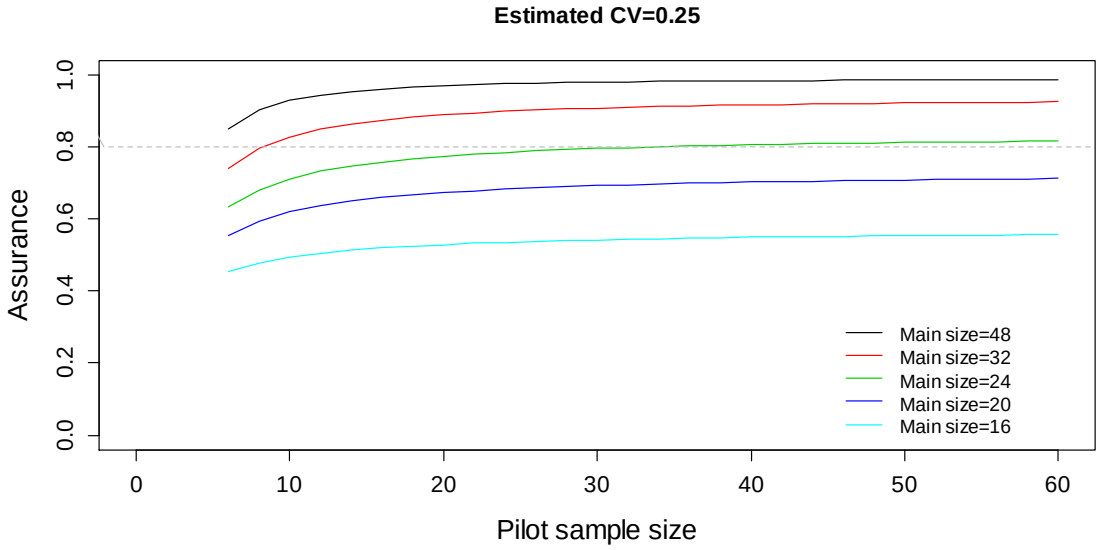


Assurance of CV

Impact of Sample size of New Trial



Impact of Sample size of Pilot Trial(s)

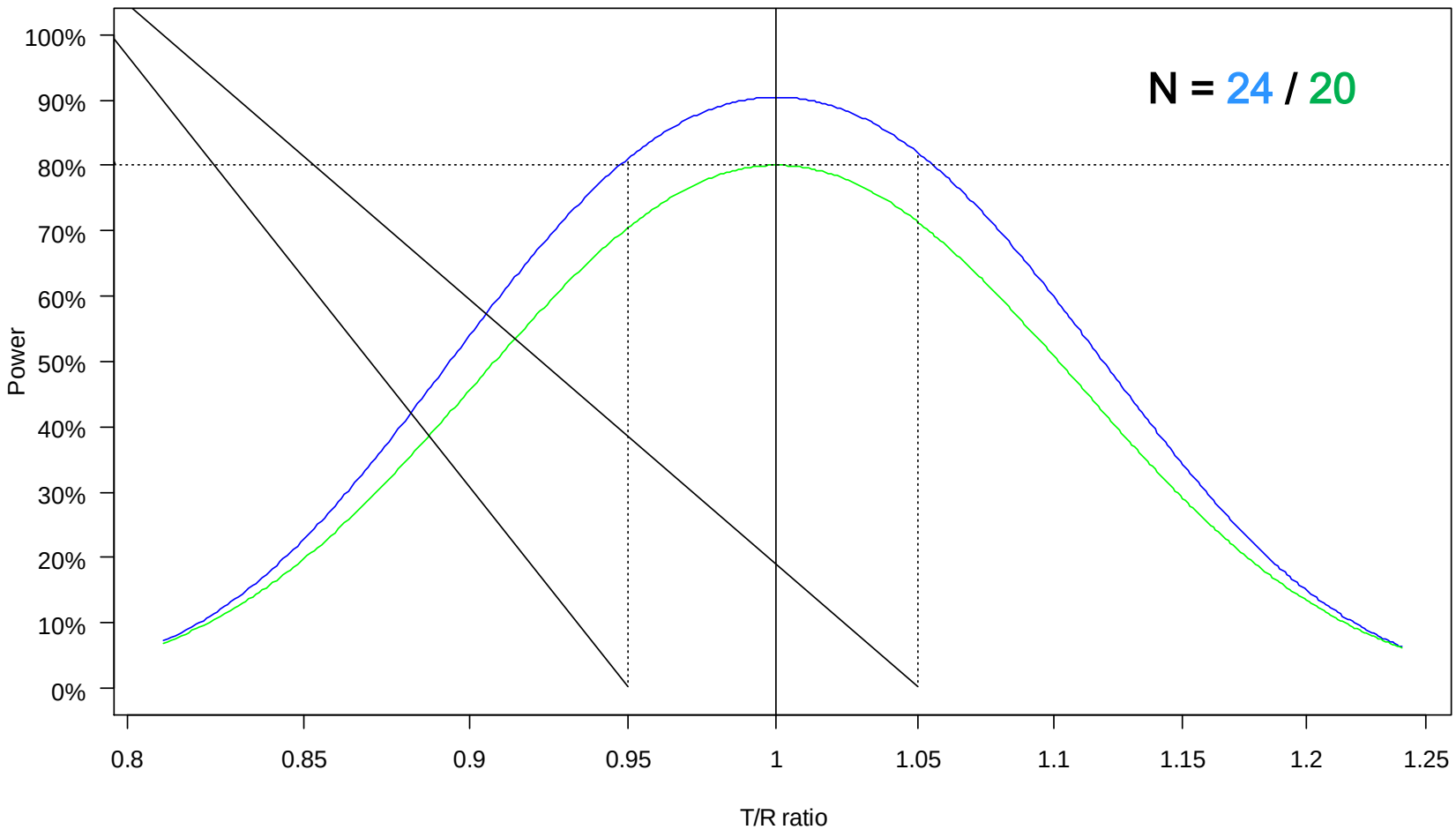


Sample size determination

- 6 parameters are combined when considering the power
  - Margin m
  - Type-I error α
  - Type-II error β ➔ Power $\pi = 1-\beta$
  - Anticipated T/R-ratio θ
  - (intraindividual) Coefficient of variation CV
  - Total sample size N

- **Which value is the best assumption for θ ?**

Sample size determination



Sample size determination in the literature

- Which value to choose for θ ?

A sample size of 50 subjects completing the study was determined to have at least 90% power to show bioequivalence between the 2 somatropin treatments, assuming a 20% intrasubject %CV for the C_{max} of somatropin. A sample size of 60 was planned to allow

The sample size was calculated on the basis of a cross-over design with log-transformed data according to Hauschke et al. [12], considering an intra-individual variation coefficient of 22.7 %, a power of the test of 80 %, a confidence interval of 80–125 %, an expected ratio μT/μR (mean of test versus reference) of 1.10, based

The sample size for this trial was derived from data characterising the variability of metformin products [17-19]. The intraindividual geometric coefficient of variation (gCV) of the C_{max} of metformin is approximately 23.5%. The variability of AUC is lower than that of C_{max}, and did not need to be considered in the sample size assessment. The trial was, therefore, planned to accommodate a true geometric mean ratio of the pharmacokinetic endpoints within the range of 97% to 103% with at least 88% power. These assumptions led to a sample size of 26 evaluable subjects per study part, and an enrolment of 28 subjects per study part was planned to allow for potential drop-outs.

To achieve 85% power (1-sided α = 0.05; true mean ratio = 1.05), 24 participants (4 per sequence) were required to com-

collection times were used for analysis. For each type of organ transplant, a sample size of 24 individuals was required to achieve 90% statistical power for concluding bioequivalence in crossover trials at an alpha of 0.05 using standard bioequivalence limits of 80%–125% and assuming a true difference of 0 [16,33,34].

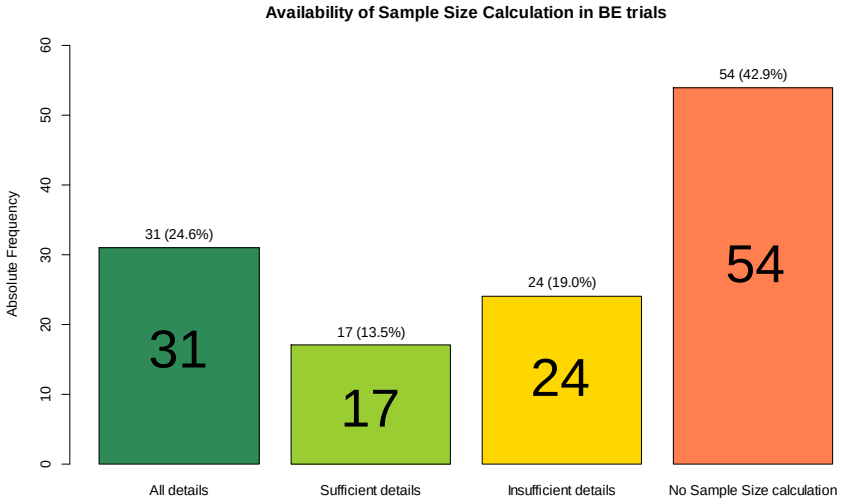
Findings from different studies

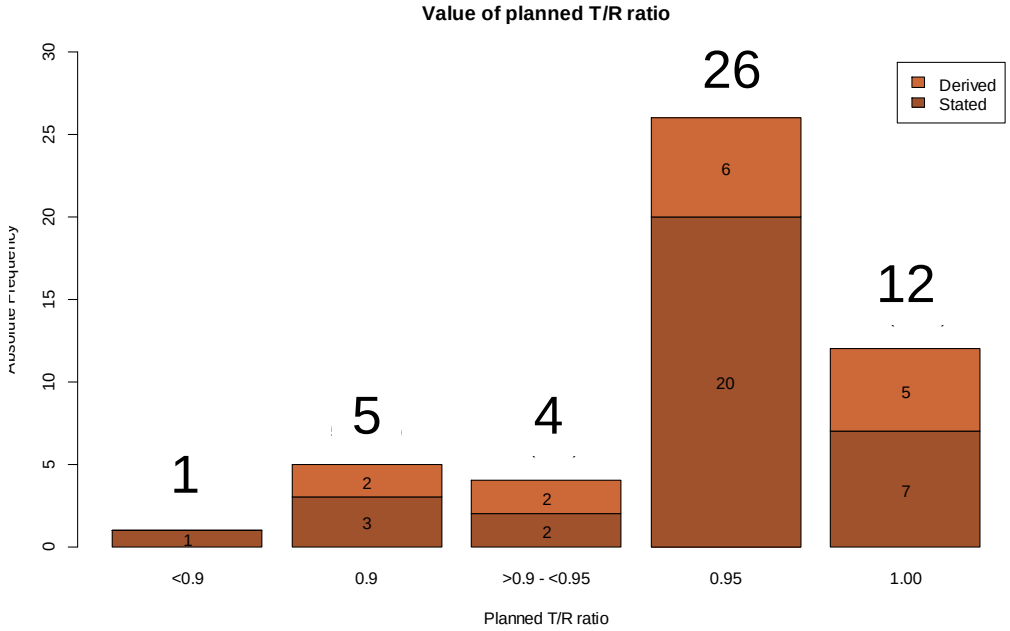
Reference	θ_0	CV	π	n
Radicioni et al. (2017)	1	0.383	0.8	50
Bosilkovska et al. (2016)	1	0.3	0.8	30
Navarro et al. (2016)	1.1	0.227	0.8	40
Ermer et al. (2016)	1.05	0.215	0.85	24
Luo et al. (2016)	0.95 – 1.05	0.25	0.8	24
Boudriau et al. (2016)	0.925 – 1.075	0.2	0.8	33

with paired means option. Fixing the significance level α at 5% and the hypothesized test/reference mean ratio to 1, 50 subjects were considered sufficient to attain a power of 80% to correctly conclude the bioequivalence between the

Systematic review on planning BE trials

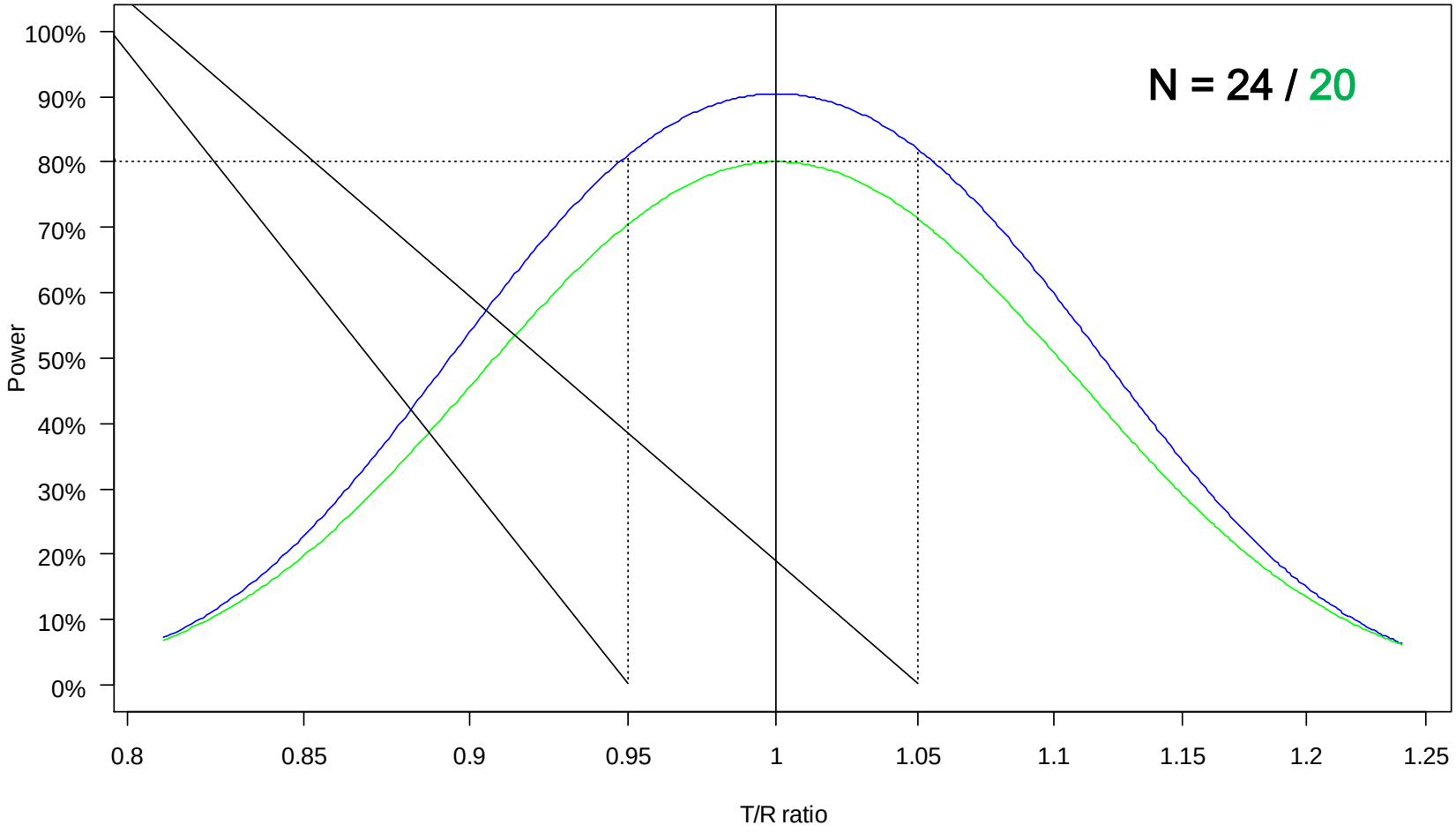
- 126 identified articles of BE trials
  - 48 reported sufficient details for sample size considerations
  - Of those, 12 (25%) assumed a T/R ratio of 1.00





Awounvo S, Ring A, BPS 2018.

Sample size determination



Power approach

- Sample size determination using the power approach
  - Requires one fixed expected value for the T/R ratio θ
  - Most likely value (if no difference in dissolution) is 1.00
  - However any deviation from 1.00 would lead to loss of power
  - Anticipating a different value would be more conservative
  - However **no direction** (0.95 vs. 1.05) could be justified

- Is this the best statistical approach?

Assurance (expected power)

- Previously, many authors have used a single value $\theta \neq 1.00$
  - Most common were $\theta = 1.05$ or $\theta = 0.95$

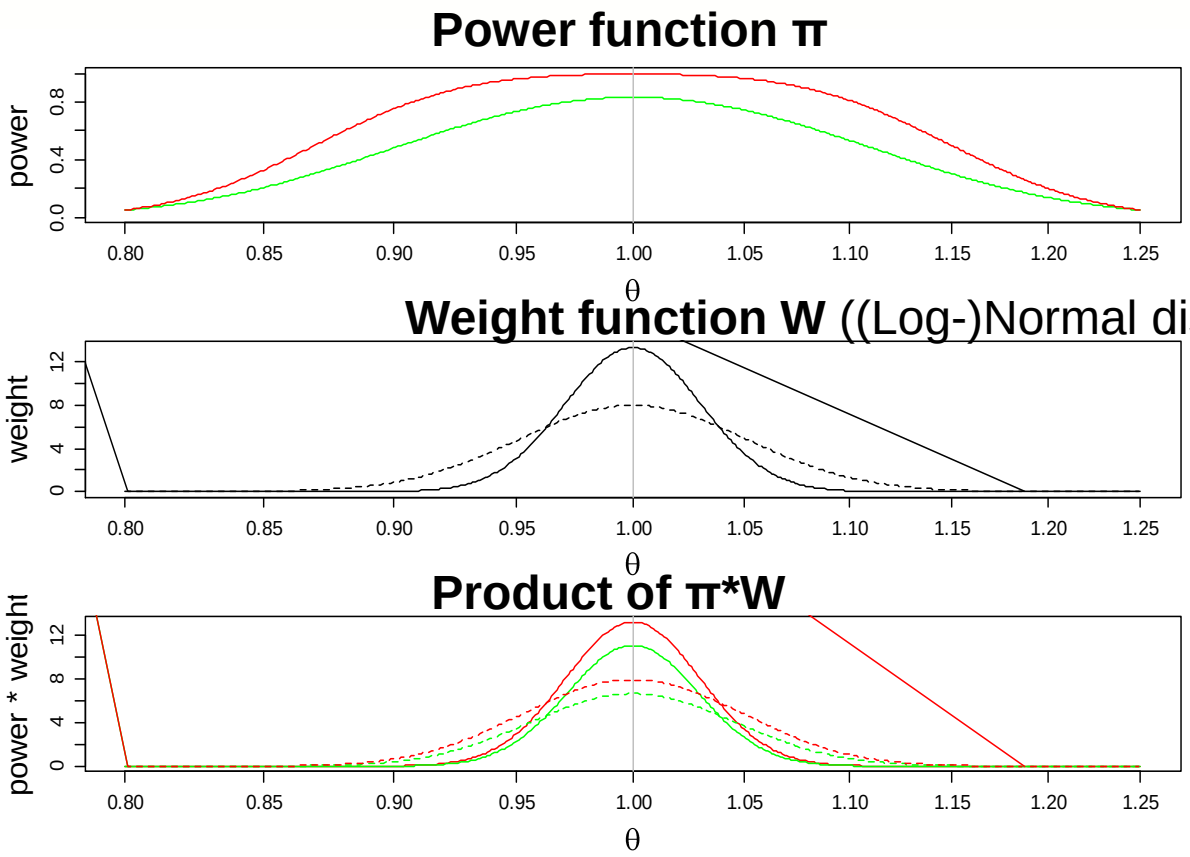
- Instead of taking a single value of θ,
  consider a distribution of possible values of θ

- Assume a (Log-)Normal distribution $(0, \sigma)$
  - Symmetric (on log-scale) around T/R ratio of 1.00
  - Uses a new parameter σ, which characterises the uncertainty
  - Which value of σ would be appropriate?

Assurance (expected power) for the T/R ratio

- Instead of taking a single value of θ, consider a distribution

$$\gamma = P(\text{Successful trial}) = \int \pi(\theta) W(\eta) d\eta \qquad \text{with} \quad \eta = \log(\theta)$$



32
16
} subjects

$$\sigma = \begin{cases} 0.03 \\ 0.05 \end{cases}$$

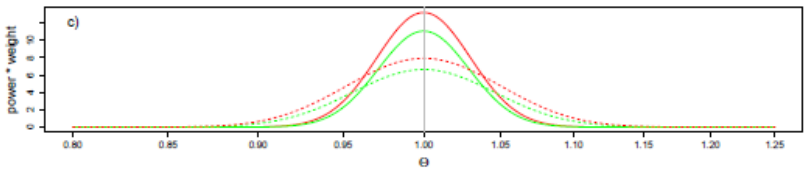
Assurance (expected power)

- Instead of taking a single value of θ, consider a distribution of possible values of θ

- Power function π for various values of θ
  - Maximum at most likely value 1.00
- Weight function W ((Log-)Normal distribution $(0, \sigma)$)
  - Symmetric (on log-scale) around T/R ratio of 1.00

- Assurance is the **area under the product curve**

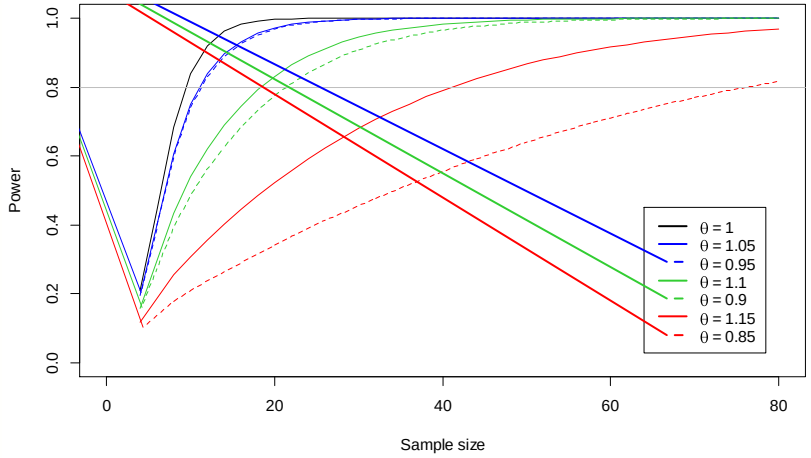
	N=16	N=32
Power	*83%*	*99%*
$\sigma = 0.03$	80%	98%
$\sigma = 0.05$	74%	95%

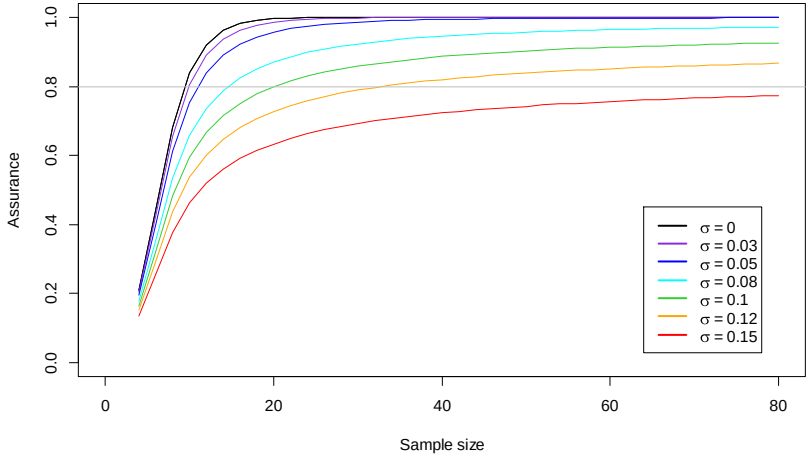


Comparison power vs. Assurance (T/R ratio)

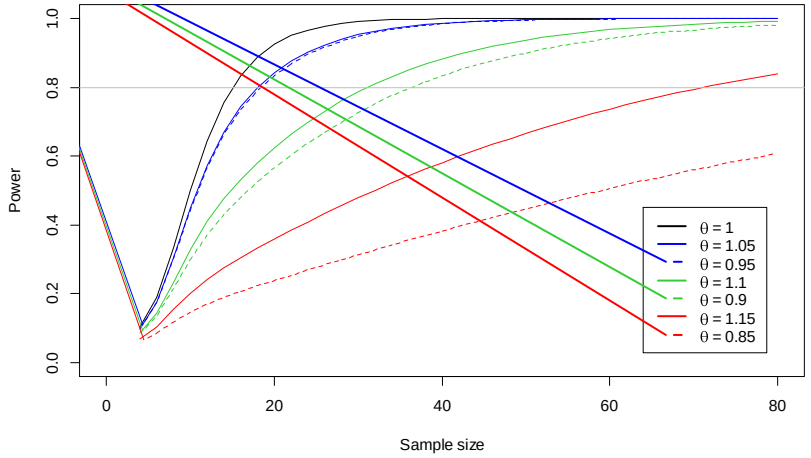
Power **Assurance**

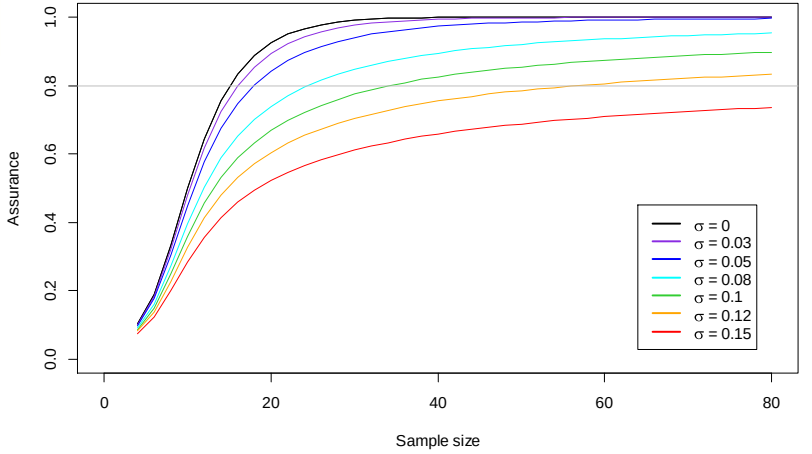
CV=15%





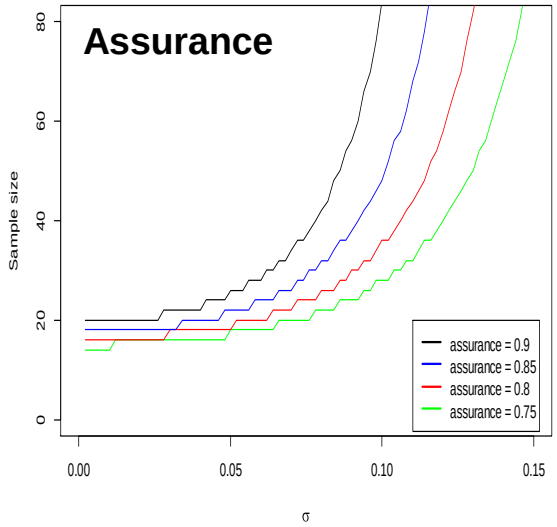
CV=25%

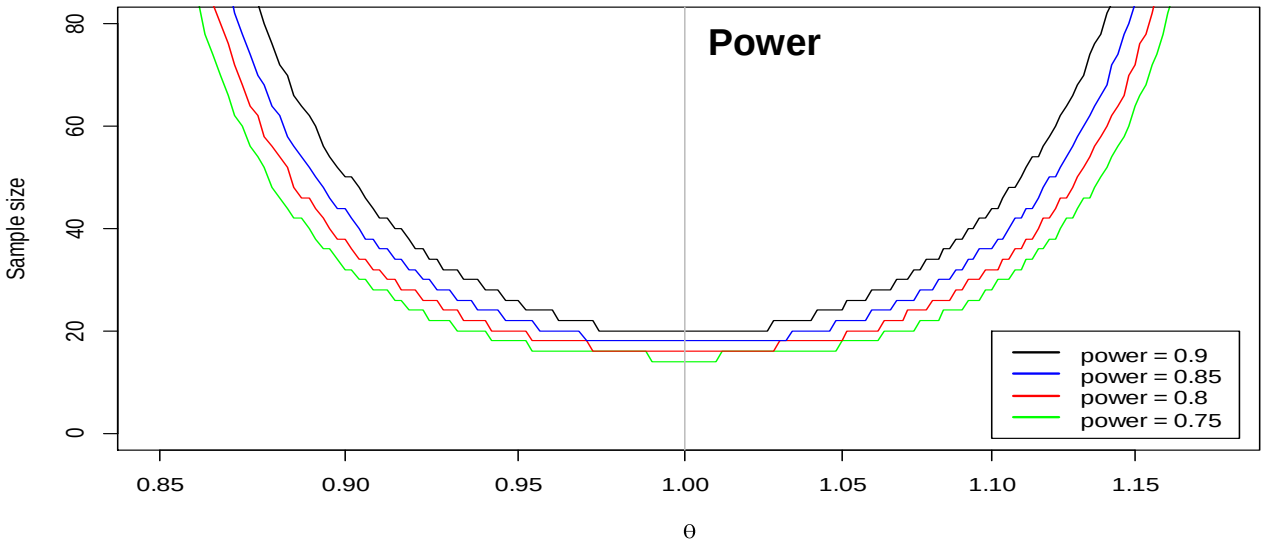


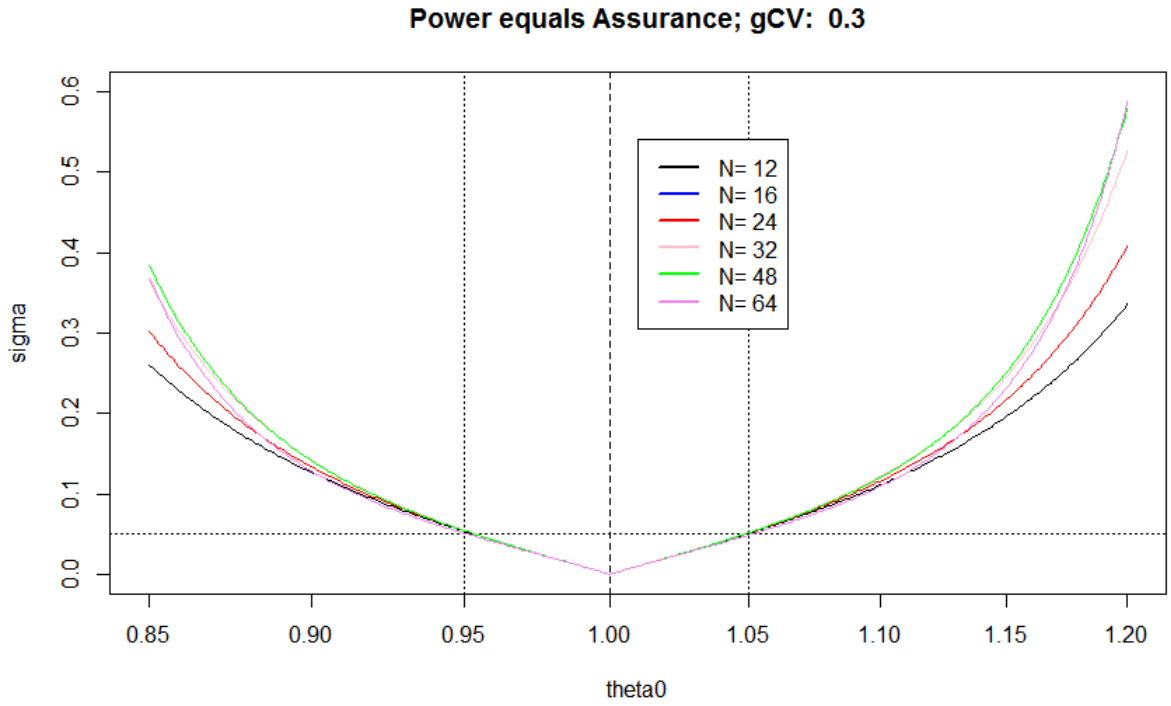


Comparison power vs. assurance

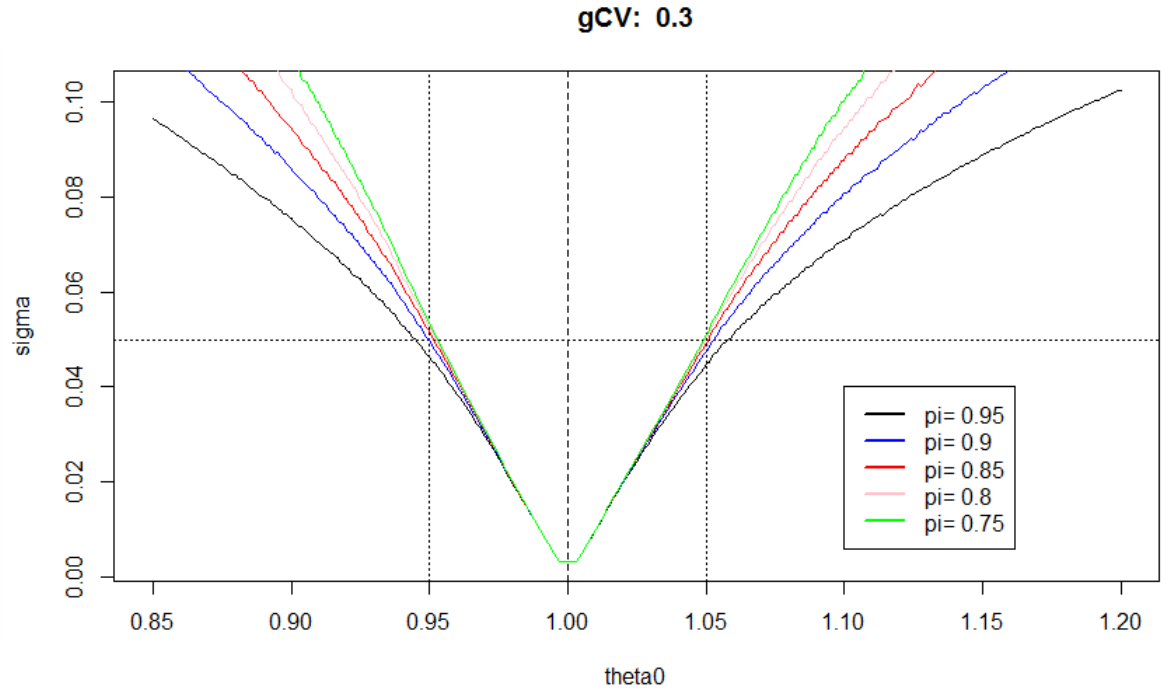
- Sample size for same value of power/assurance depending θ and σ






Power equals Assurance; gCV: 0.3
sigma
theta0
N= 12
N= 16
N= 24
N= 32
N= 48
N= 64
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.85
0.90
0.95
1.00
1.05
1.10
1.15
1.20


medac


gCV: 0.3
sigma
theta0
pi= 0.95
pi= 0.9
pi= 0.85
pi= 0.8
pi= 0.75
0.00
0.02
0.04
0.06
0.08
0.10
0.85
0.90
0.95
1.00
1.05
1.10
1.15
1.20

Additional considerations

- Which value should we select for σ ?
  - Maybe 0.05 is too small?
- Specifications of pharmaceutical drug formulations leave room for slight deviations
  - Typical limits for batch-to-batch variability are in the range of 5%.
  - This variability is typically not covered by the drug specific CV
  - Most clinical pharmacology studies use a single product batch.

- See also
  - Burmeister Getz E et al. Clin Pharmacol Ther. 2017 **Between-Batch Pharmacokinetic Variability** Inflates Type I Error Rate in Conventional Bioequivalence Trials …
  - Burmeister Getz E et al. Clin Pharmacol Ther. 2016 **Batch-to-Batch Pharmacokinetic Variability** Confounds Current Bioequivalence Regulations …

Conclusion

- In general, the focus of Statisticians is on the variability, not on the mean
  - Assurance appears to be a better concept to handle uncertainty than power calculations with $\theta \neq 1.00$
- Traditional power estimations using deviations of θ from unity of up to 5% are similar to assurance with σ<4-6%
  - In this range, relationship between θ and σ almost linear and independent of *N* and *CV*
- In contrast to most superiority trials, the assurance for BE trials in the range of 80%-90% can be achieved with practical sample sizes