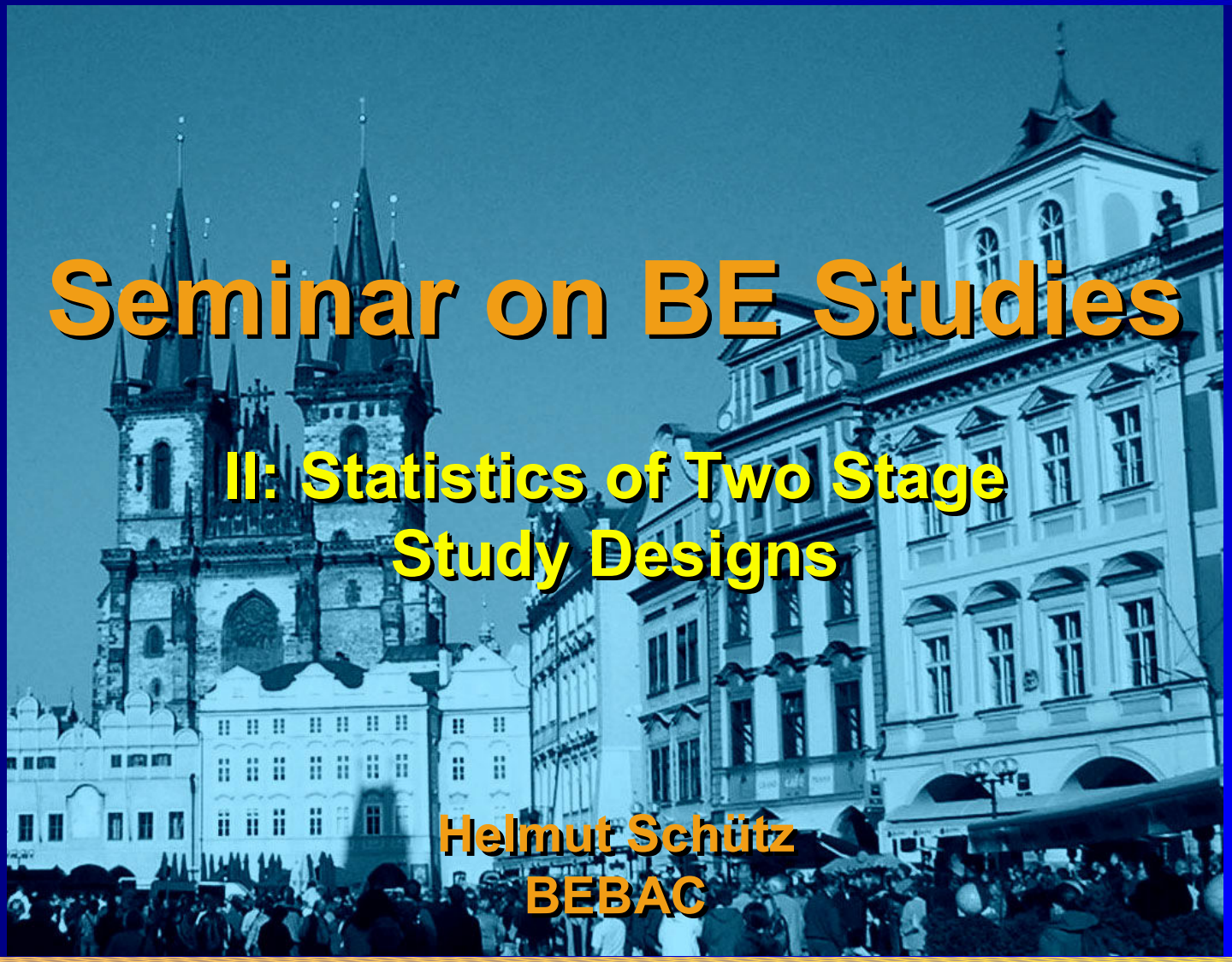


Seminar on BE Studies

II: Statistics of Two Stage Study Designs

Helmut Schütz
BEBAC

α and β

- All formal decisions are subjected to two types of error:
  - α Probability of Error Type I (aka Risk Type I)
  - β Probability of Error Type II (aka Risk Type II)

  Example from the justice system:

Verdict	Defendant innocent	Defendant guilty
Presumption of innocence not accepted (guilty)	**Error type I**	**Correct**
Presumption of innocence accepted (not guilty)	**Correct**	**Error type II**

α and β

- Or in more statistical terms:

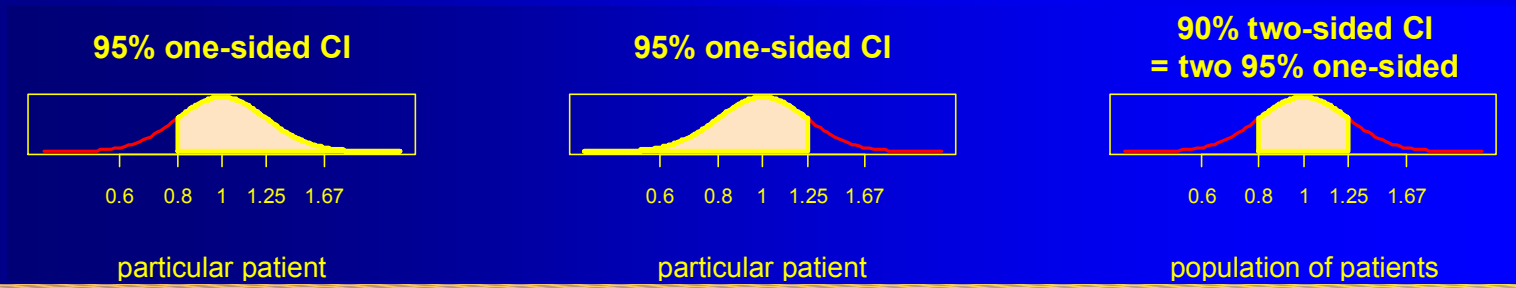
Decision	Null hypothesis true	Null hypothesis false
Null hypothesis rejected	**Error type I**	**Correct (H_a)**
Failed to reject null hypothesis	**Correct (H_0)**	**Error type II**

- In BE-testing the null hypothesis is bioinequivalence ($\mu_1 \neq \mu_2$)!

Decision	Null hypothesis true	Null hypothesis false
Null hypothesis rejected	**Patients' risk**	**Correct (BE)**
Failed to reject null hypothesis	**Correct (not BE)**	**Producer's risk**

α ...

- Patient's Risk to be treated with an inequivalent formulation (H_0 falsely rejected)
  - BA of the test compared to reference in a *particular* patient is risky *either* below 80% *or* above 125%.
  - If we keep the risk of particular patients at α 0.05 (5%), the risk of the entire population of patients (<80% *and* >125%) is 2×α (10%) – expressed as: 90% CI = 1 – 2×α = 0.90



… and β

- Producer's Risk to get no approval for a equivalent formulation (H_0 falsely not rejected)
  - *Set* in study planning to ≤0.2, where power = 1 – β = ≥80%
  - If power is set to 80 %

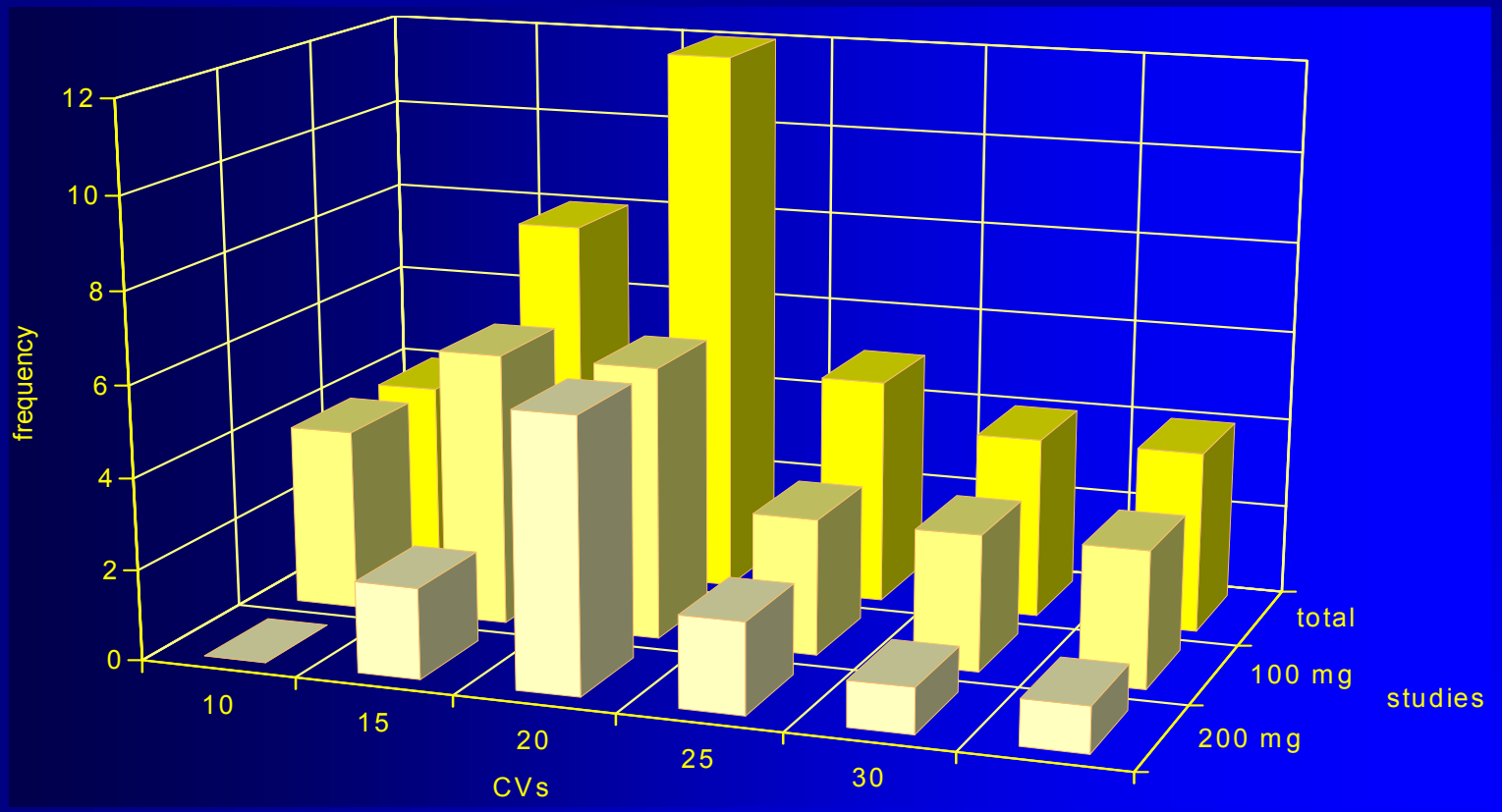
    One out of five studies will fail just by chance!

α 0.05	BE
not BE	β 0.20

  - *A posteriori* (*post hoc*) power does not make sense!

    Either a study has demonstrated BE or not.

Published data



Doxicycline (37 studies from **Blume/Mutschler**, *Bioäquivalenz: Qualitätsbewertung wirkstoffgleicher Fertigarzneimittel*, GOVI-Verlag, Frankfurt am Main/Eschborn, 1989-1996)

Sample Size (Guidelines)

- Recommended minimum
  - 12 WHO, EU, CAN, NZ, AUS, AR, MZ, ASEAN States, RSA, Russia (2011 Draft)
  - 12 USA 'A pilot study that documents BE can be appropriate, provided its design and execution are suitable and a sufficient number of subjects (*e.g.*, 12) have completed the study.'
  - 18 Russia (2008)
  - 20 RSA (MR formulations)
  - 24 Saudia Arabia (12 to 24 if statistically justifiable)
  - 24 Brazil
  - 'Sufficient number' Japan

Sample Size (Limits)

- Maximum
  - NZ: If the calculated number of subjects appears to be higher than is ethically justifiable, it may be necessary to accept a statistical power which is less than desirable. Normally it is not practical to use more than about 40 subjects in a bioavailability study.
  - All others: Not specified (judged by IEC/IRB or local Authorities).
    ICH E9, Section 3.5 applies: *"The number of subjects in a clinical trial should always be large enough to provide a reliable answer to the questions addressed."*

EMEA

- NfG on the Investigation of BA/BE (2001)
  - The number of subjects required is determined by
    - the error variance associated with the primary characteristic to be studied as estimated from
      - a pilot experiment,
      - previous studies, or
      - published data,
    - the significance level desired,
    - the expected deviation (Δ) from the reference product compatible with BE and,
    - the required power.

EMA

- BE Guideline (2010)
  - The number of subjects to be included in the study should be based on an *appropriate* sample size calculation.

Cookbook?

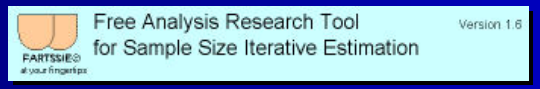
Tools

- Sample Size Tables (Phillips, Diletti, Hauschke, Chow, Julious, …)
- Approximations (Diletti, Chow, Julious, …)
- General purpose (SAS, S+, *R*, StaTable, …)
- Specialized Software (nQuery Advisor, PASS, FARTSSIE, StudySize, …)
- Exact method (Owen – implemented in *R*-package *PowerTOST*)*

* Thanks to Detlew Labes!

Approximations obsolete

- Exact sample size tables still useful in checking plausibility of software's results
- Approximations based on noncentral *t* (FARTSSIE17)



http://individual.utoronto.ca/ddubins/FARTSSIE17.xls

or R / S+ →

- Exact method (Owen) in *R*-package *PowerTOST*

http://cran.r-project.org/web/packages/PowerTOST/

```
require(PowerTOST)
  sampleN.TOST(alpha=0.05,
  targetpower=0.80, theta0=0.95,
  CV=0.30, design='2x2')
```

```
alpha   <- 0.05     # alpha
CV      <- 0.30     # intra-subject CV
theta1  <- 0.80     # lower acceptance limit
theta2  <- 1/theta1 # upper acceptance limit
theta0  <- 0.95     # expected ratio T/R
PwrNeed <- 0.80     # minimum power
Limit   <- 1000     # Upper Limit for Search
n       <- 4        # start value of sample size search
s       <- sqrt(2)*sqrt(log(CV^2+1))
repeat{
  t     <- qt(1-alpha,n-2)
  nc1   <- sqrt(n)*(log(theta0)-log(theta1))/s
  nc2   <- sqrt(n)*(log(theta0)-log(theta2))/s
  prob1 <- pt(+t,n-2,nc1); prob2 <- pt(-t,n-2,nc2)
  power <- prob2-prob1
  n     <- n+2      # increment sample size
  if(power >= PwrNeed | (n-2) >= Limit) break }
Total   <- n-2
if(Total == Limit){
  cat('Search stopped at Limit', Limit,
      ' obtained Power', power*100, '%\n')
  } else
  cat('Sample Size', Total, '(Power', power*100, '%)\n')
```

Which Power?

- Generally Producer's Risk 10–20%
  - Plan for 90% – allowing for contingency *e.g.*,
    - drop-outs,
    - CV_{intra} higher than assumed,
    - deviation of test from reference larger than expected.
  - Power >90% might lead to ethical problems ('forced bioequivalence').
  - FDA (2001): 80–90%
  - EMA (2010): 'appropriate'…
  - Russia (2008, 2011 draft): ≥80%

End of the Story?

- *'Doing the maths'* is just *part* of the job!
  - Does it make sense to rely on studies of different origin and sometimes unknown quality?
    - The reference product may have been subjected to many *(minor only?)* changes from the formulation used in early publications.
    - Different bioanalytical methods are applied. Newer (*e.g.* LC/MS-MS) methods are not *necessarily* better in terms of variability.
    - Generally insufficient information about the clinical setup (*e.g.*, posture control).
    - Review studies critically; don't try to mix oil with water.

Sensitivity Analysis

- ICH E9 (1998)
  - Section 3.5 Sample Size, paragraph 3
    - The method by which the sample size is calculated should be given in the protocol […]. The basis of these estimates should also be given.
    - It is important to investigate the sensitivity of the sample size estimate to a variety of deviations from these assumptions and this may be facilitated by providing a range of sample sizes appropriate for a reasonable range of deviations from assumptions.
    - In confirmatory trials, assumptions should normally be based on published data or on the results of earlier trials.

Sensitivity Analysis

- Example

  nQuery Advisor: $\sigma_w = \sqrt{\ln(CV_{intra}^2 + 1)}; \sqrt{\ln(0.2^2 + 1)} = 0.198042$

nQuery Advisor - [MTE2co-1.nqa]

File Edit View Options Assistants Randomize Plot Window Help

t-tests (TOST) of equivalence in ratio of means for crossover design (natural log scale)

	90% power	25% CV	4 drop outs	25% CV + d.o.	PE 90%	worst case
Test significance levels, α (one-sided)	0.050	0.050	0.050	0.050	0.050	0.050
Lower equivalence limit for μ_T / μ_S, Δ_L	0.800	0.800	0.800	0.800	0.800	0.800
Upper equivalence limit for μ_T / μ_S, Δ_U	1.250	1.250	1.250	1.250	1.250	1.250
Expected ratio, μ_T / μ_S	0.950	0.950	0.950	0.950	0.900	0.900
Crossover ANOVA, sqrt(MSE) (ln scale)	0.198042	0.246221	0.198042	0.246221	0.198042	0.246221
SD differences, σ_d (ln scale)	0.280074	0.348209	0.280074	0.348209	0.280074	0.348209
Power (%)	90.00	77.60	86.88	69.53	66.94	45.09
n per sequence group	13	13	11	11	13	11

20% CV: n=26

25% CV: power 90% → **78%**

20% CV, 4 drop outs: power 90% → **87%**

25% CV, 4 drop outs: power 90% → **70%**

20% CV, PE 90%: power 90% → **67%**

Sensitivity Analysis

- Example

  PowerTOST, function *sampleN.TOST*

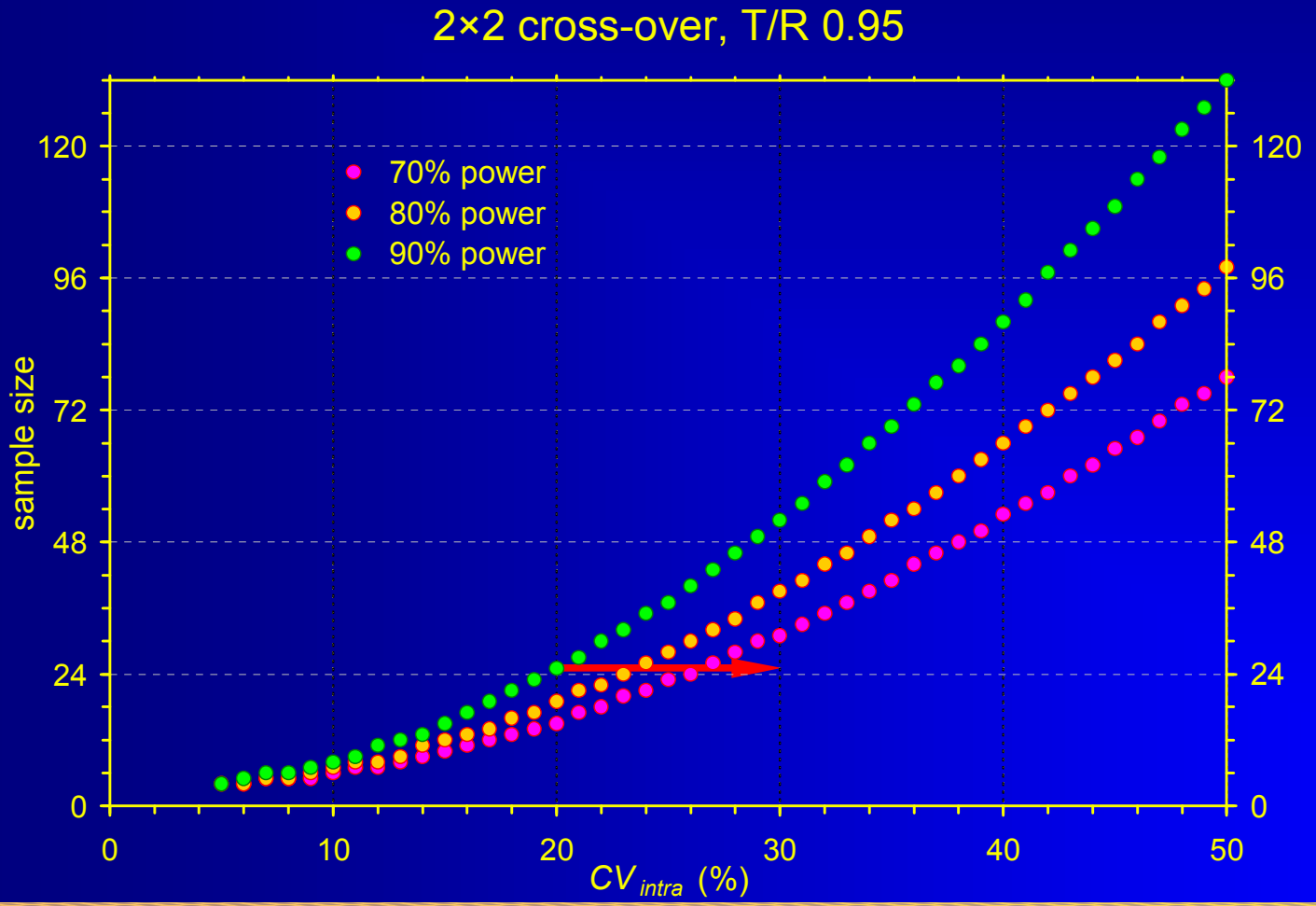
```
require(PowerTOST)
sampleN.TOST(alpha=0.05, targetpower=0.9, theta0=0.95,
             theta1=0.8, theta2=1.25, CV=0.2, design='2x2')

++++++++++++ Equivalence test - TOST ++++++++++++
            Sample size estimation
-----------------------------------------------
Study design:  2x2 crossover
log-transformed data (multiplicative model)

alpha = 0.05, target power = 0.9
BE margins        = 0.8 ... 1.25
Null (true) ratio = 0.95,  CV = 0.2
Sample size
 n     power
26   0.917633
```

Sensitivity Analysis

- To estimate Power for a given sample size, use function *power.TOST*

```
require(PowerTOST)
power.TOST(theta0=0.95, CV=0.25, n=26)
[1] 0.7760553

power.TOST(theta0=0.95, CV=0.20, n=22)
[1] 0.8688866

power.TOST(theta0=0.95, CV=0.25, n=22)
[1] 0.6953401

power.TOST(theta0=0.90, CV=0.20, n=26)
[1] 0.6694514

power.TOST(theta0=0.90, CV=0.25, n=22)
[1] 0.4509864
```

CV based on assumptions!



Example: Study planned on expected 20% CV with 90% power.

If CV is 30%, power drops to only 58%…

Add-on / Two-Stage Designs

- Sometimes properly designed and executed studies fail due to
  - 'true' bioinequivalence,
  - poor study conduct (increasing variability),
  - pure chance (producer's risk hit),
  - false (over-optimistic) assumptions about variability and/or T/R-ratio.
- The patient's risk must be preserved
  - Already noticed at Bio-International Conferences (1989, 1992) and guidelines from the 1990s.

Sequential Designs

- Have a long and accepted tradition in clinical research (mainly phase III)
  - Based on work by Armitage *et al.* (1969), McPherson (1974), Pocock (1977), O'Brien and Fleming (1979), Lan & DeMets (1983), …
    - First proposal by Gould (1995) in the area of BE did not get regulatory acceptance in Europe, but
    - new methods stated in recent guidelines.

AL Gould
Group Sequential Extension of a Standard Bioequivalence Testing Procedure
J Pharmacokin Biopharm 23/1, 57–86 (1995)

Sequential Designs

- Methods by Potvin *et al.* (2008) promising
  - Supported by the 'Product Quality Research Institute' (members: FDA/CDER, Health Canada, USP, AAPS, PhRMA…)
    - Three of BEBAC's protocols accepted by German BfArM, one product approved in 06/2011.

Potvin D, Diliberti CE, Hauck WW, Parr AF, Schuirmann DJ, and RA Smith
Sequential design approaches for bioequivalence studies with crossover designs
Pharmaceut Statist 7/4, 245–62 (2008), DOI: 10.1002/pst.294
http://www3.interscience.wiley.com/cgi-bin/abstract/115805765/ABSTRACT

Review of Guidelines

- Canada (May 2012)

  Potvin *et al.* Method C recommended.

- FDA (Jun 2012)

  Potvin *et al.* Method C recommended.
  API specific guidances: Loteprednol, Dexamethasone / Tobramycin.

- EMA (Jan 2010)

  Acceptable; Potvin *et al.* Method B preferred.

- Russia (Draft 2011)

  Acceptable (Methods B and C).

Two-Stage Design

- EMA GL on BE (2010)
  - Section 4.1.8
    - Initial group of subjects treated and data analysed.
    - If BE not been demonstrated an additional group can be recruited and the results from both groups combined in a final analysis.
    - Appropriate steps to preserve the overall type I error (patient's risk).
    - Stopping criteria should be defined *a priori*.
    - First stage data should be treated as an interim analysis.

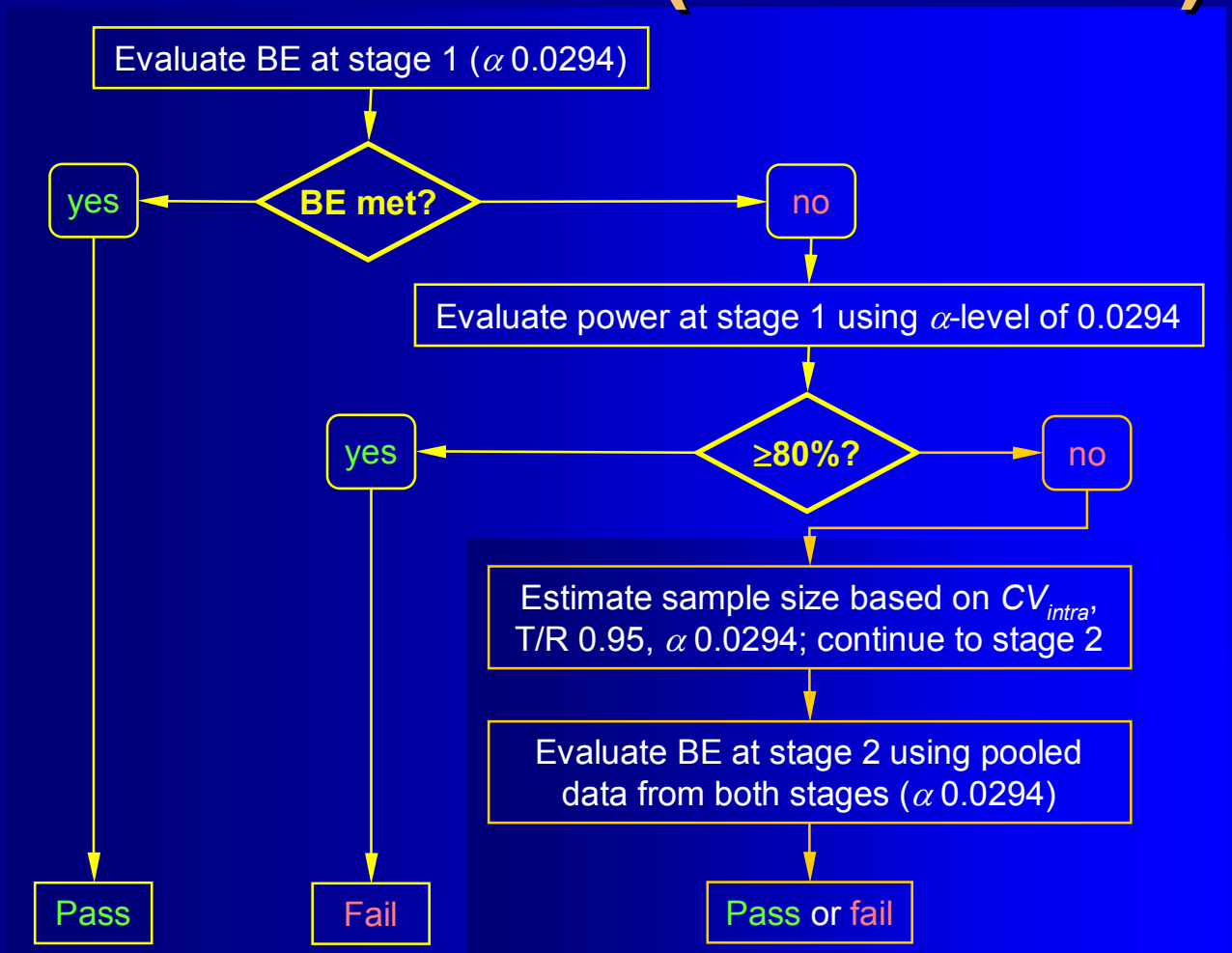
Two-Stage Design

- EMA GL on BE (2010)
  - Section 4.1.8 (cont'd)
    - Both analyses conducted at adjusted significance levels (with the confidence intervals accordingly using an adjusted coverage probability which will be higher than 90%). […] 94.12% confidence intervals for both the analysis of stage 1 and the combined data from stage 1 and stage 2 would be acceptable, but there are many acceptable alternatives and the choice of how much alpha to spend at the interim analysis is at the company's discretion.

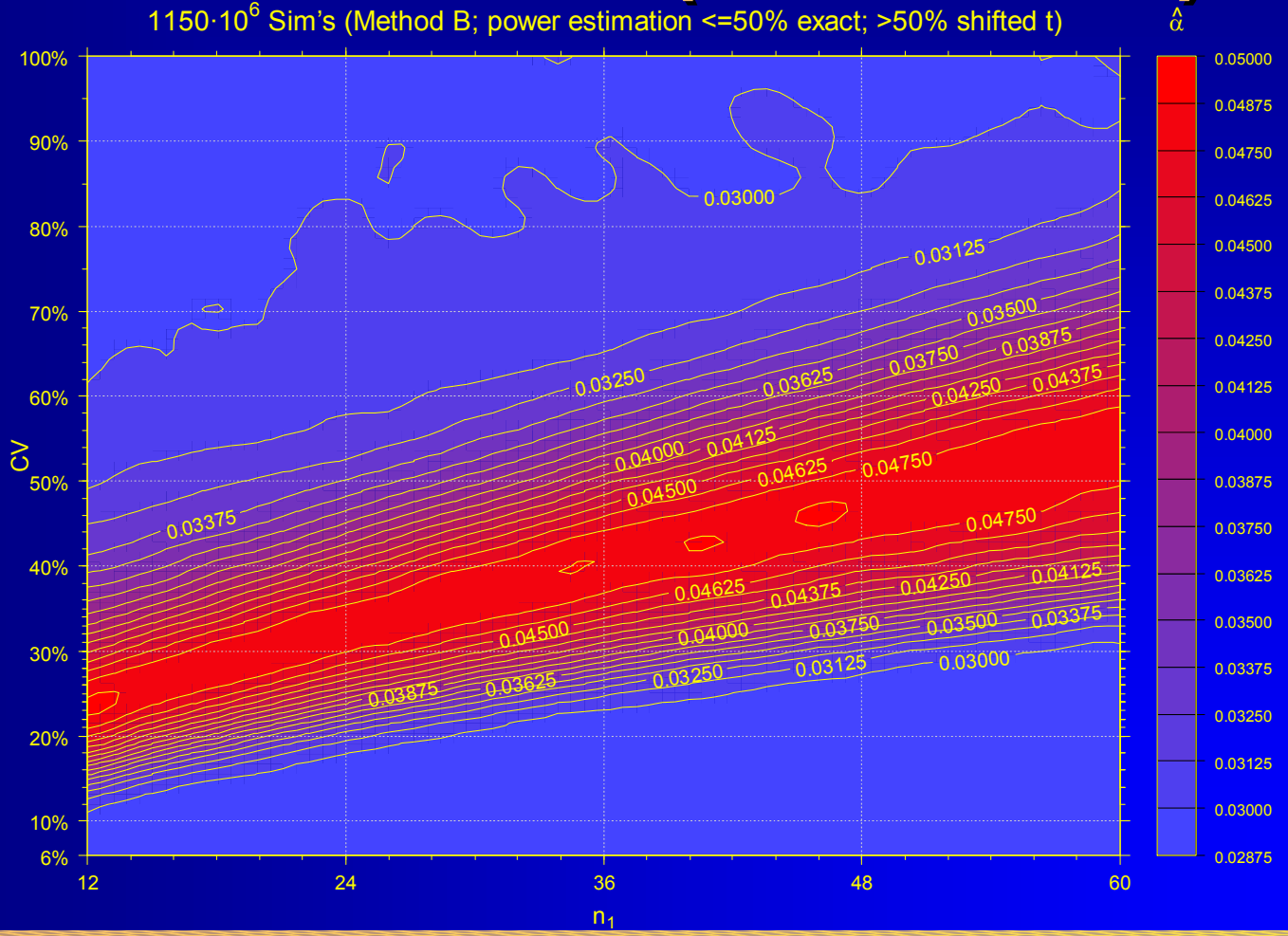
Two-Stage Design

- EMA GL on BE (2010)
  - Section 4.1.8 (cont'd)
    - Plan to use a two-stage approach must be pre-specified in the protocol along with the adjusted significance levels to be used for each of the analyses.
    - When analysing the combined data from the two stages, a term for stage should be included in the ANOVA model.

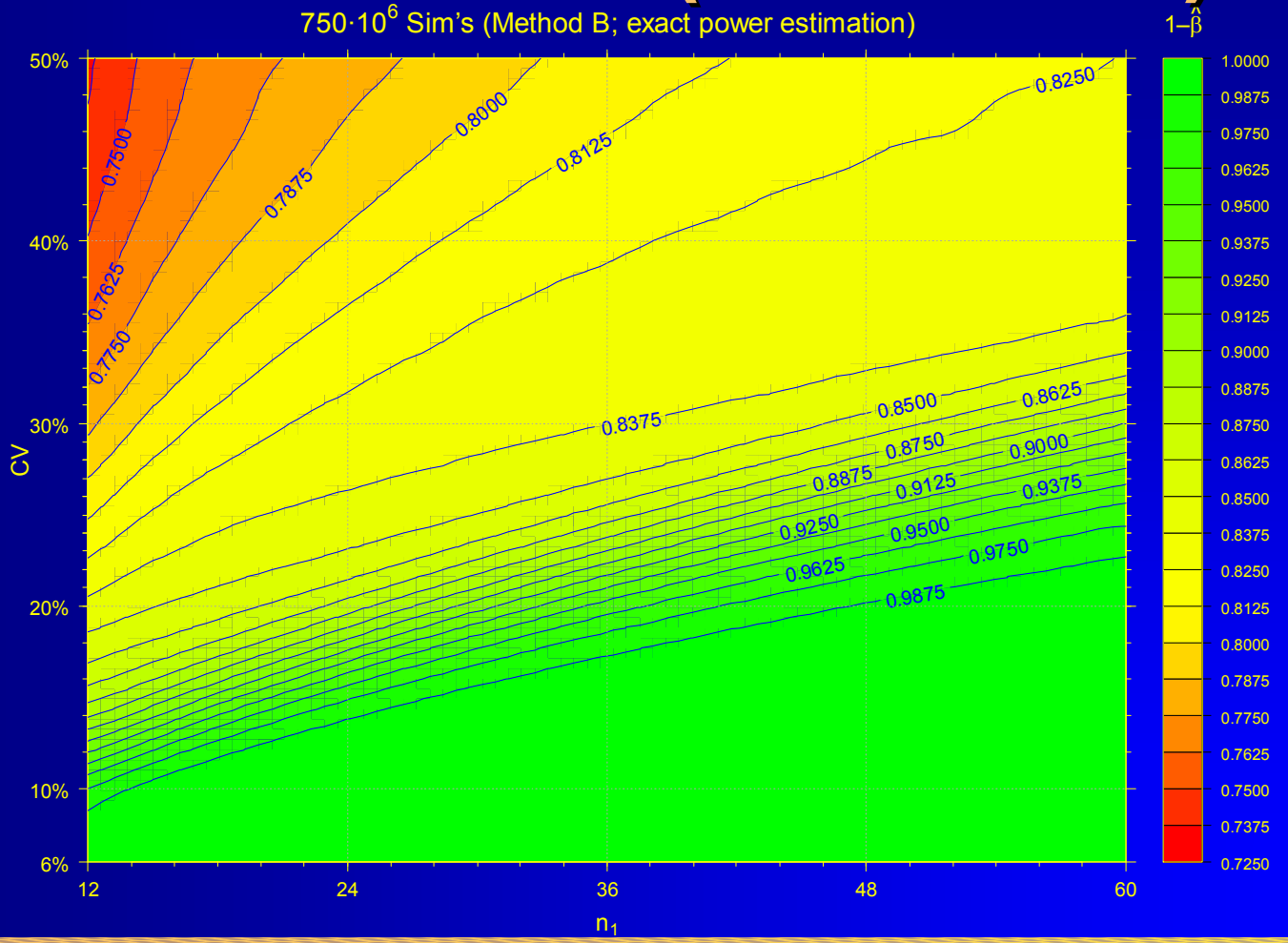
Potvin *et al.* (Method B)



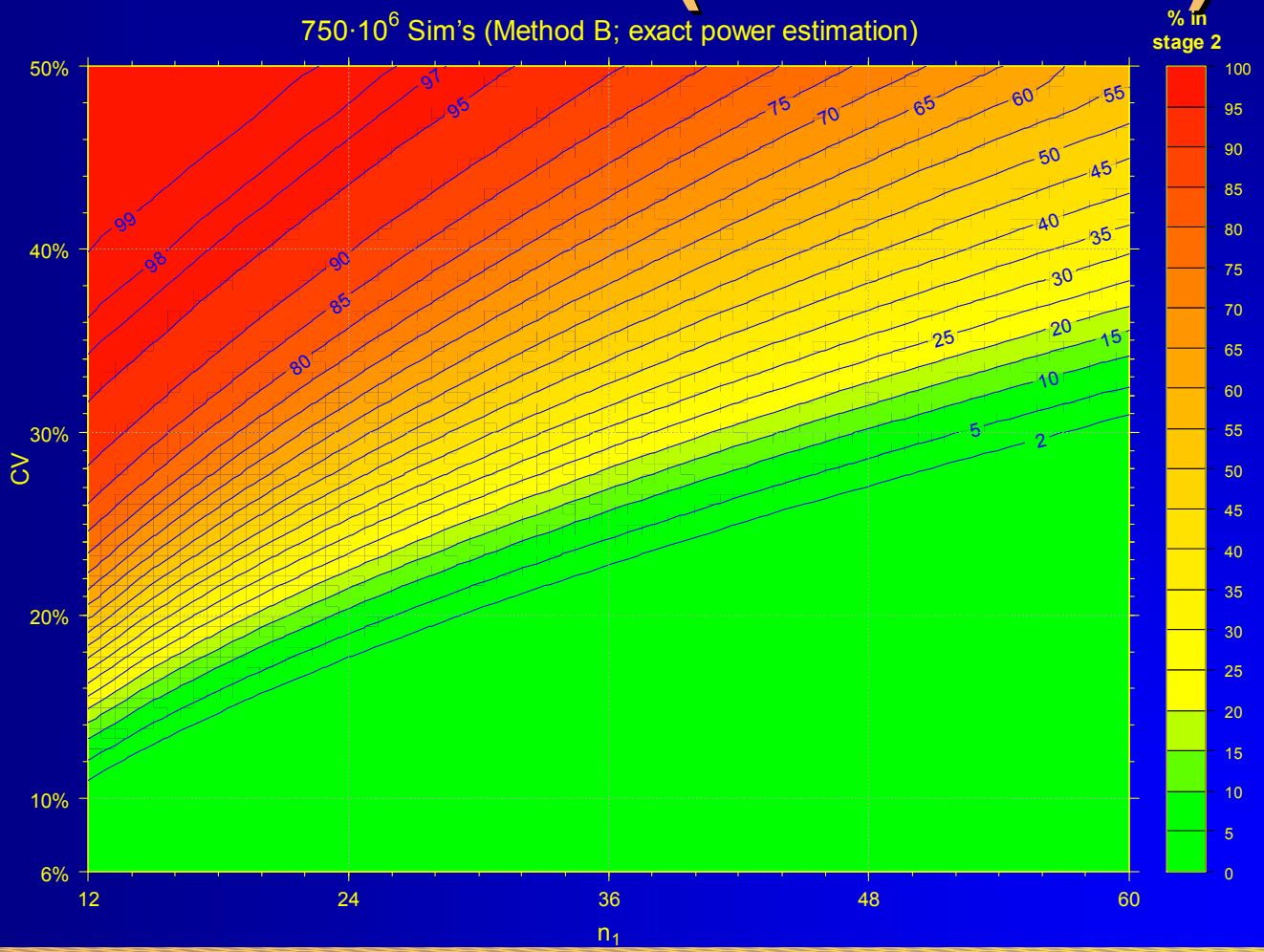
Potvin *et al.* (Method B)



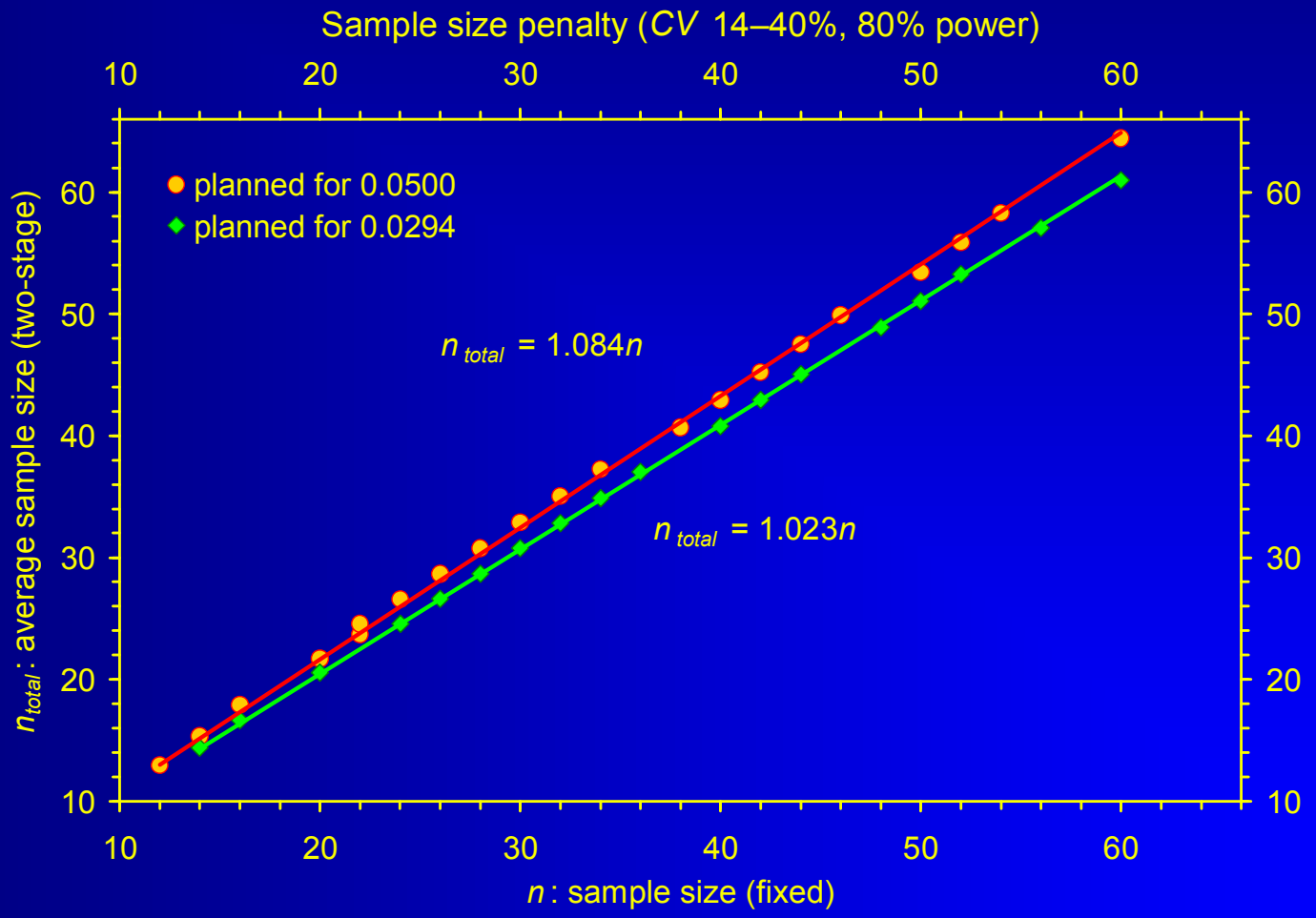
Potvin *et al.* (Method B)



Potvin *et al.* (Method B)



Potvin *et al.* (Method B)



Potvin *et al.* (Method B)

- Technical Aspects
  - Only *one* Interim Analysis (after stage 1).
  - Use software (wide step sizes in Diletti's tables); preferrable the exact method (avoid approxi-mations).
  - Should be termed 'Interim Power Analysis' *not* 'Bioequivalence Assessment' in the protocol.
  - No *a posteriori* Power – only a validated method in the decision tree.
  - No adjustment for T/R observed in stage 1 (not fully adaptive).

Potvin *et al.* (Method B)

- Technical Aspects (cont'd)
  - No futility rule preventing to go into stage 2 with a very high sample size! Must be clearly stated in the protocol (unfamiliar to the IEC because common in Phase III).
  - Pocock's α 0.0294 is used in stage 1 and in the pooled analysis (data from stages 1 + 2), *i.e.*, the 1 – 2×α = 94.12% CI is calculated.
  - Overall patient's risk preserved at ≤0.05.

Potvin *et al.* (Method B)

- Technical Aspects (cont'd)
  - If the study is stopped after stage 1, the (conventional) statistical model is:

    ```
    fixed:  sequence + period + treatment
    random: subject(sequence)
    ```

  - If the study continues to stage 2, the model for the combined analysis is:

    ```
    fixed:  sequence + stage + period(stage) + treatment
    random: subject(sequence × stage)
    ```

  - No poolability criterion!
    Combining is *always allowed* – even if a significant difference between stages is observed. No need to test this effect.

Potvin *et al.* (Method B)

- Technical Aspects (cont'd)
  - Potvin *et al.* used a simple approximative power estimation based on the shifted *t*-distribution.
  - If possible use the exact method (Owen; *R* package *PowerTOST* `method = 'exact'`) or at least one based on the noncentral *t*-distribution (*PowerTOST* `method = 'noncentral'`).
  - Power obtained in stage 1 (example 2 from Potvin):

method	power
approx. (shifted *t*)	50.49%
approx. (noncentral *t*)	52.16%
exact	52.51%

Potvin *et al.* (Method B)

```
Model Specification and User Settings
       Dependent variable : Response
                Transform : LN
            Fixed terms : int+Sequence+Period+Treatment
    Random/repeated terms : Sequence*Subject

Final variance parameter estimates:
    Var(Sequence*Subject)     0.408682
           Var(Residual)     0.0326336
         Intrasubject CV     0.182132

Bioequivalence Statistics
User-Specified Confidence Level for CI's = 94.1200
Percent of Reference to Detect for 2-1 Tests = 20.0%
A.H.Lower =  0.800   A.H.Upper =  1.250
Reference: Reference   LSMean = 0.954668  SE = 0.191772  GeoLSM = 2.597808
---------------------------------------------------------------------------
Test:      Test        LSMean = 1.038626  SE = 0.191772  GeoLSM = 2.825331

    Difference  =   0.0840,  Diff_SE = 0.0737,  df = 10.0
    Ratio(%Ref) = 108.7583

                      Classical
    CI User = (   92.9330,  127.2838)
    Failed to show average bioequivalence for confidence=94.12 and percent=20.0.
```

12 subjects in stage 1, conventional BE model

CV_{intra} 18.2%

α 0.0294

Failed with 94.12% Confidence Interval

Potvin *et al.* (Method B)

```
require(PowerTOST)
power.TOST(alpha=0.0294, theta0=0.95,
           CV=0.182132, n=12, design='2x2',
           method='exact')

[1] 0.5251476
```

α 0.0294, T/R 95% – *not* 108.76% observed in stage 1!
CV_{intra} 18.2%, 12 subjects in stage 1

Power 52.5% – initiate stage 2

```
sampleN.TOST(alpha=0.0294, targetpower=0.80, logscale=TRUE,
             theta1=0.8, theta2=1.25, theta0=0.95,
             CV=0.182132, design='2x2', method='exact',
             print=TRUE)

++++++++++++ Equivalence test - TOST ++++++++++++
             Sample size estimation
-----------------------------------------------
Study design:  2x2 crossover
log-transformed data (multiplicative model)

alpha = 0.0294, target power = 0.8
BE margins        = 0.8 ... 1.25
Null (true) ratio = 0.95,  CV = 0.182132

Sample size
 n     power
20   0.829160
```

Estimate total sample size:
α 0.0294, T/R 95%, CV_{intra} 18.2%, 80% power

Total sample size 20: include another 8 in stage 2

Potvin *et al.* (Method B)

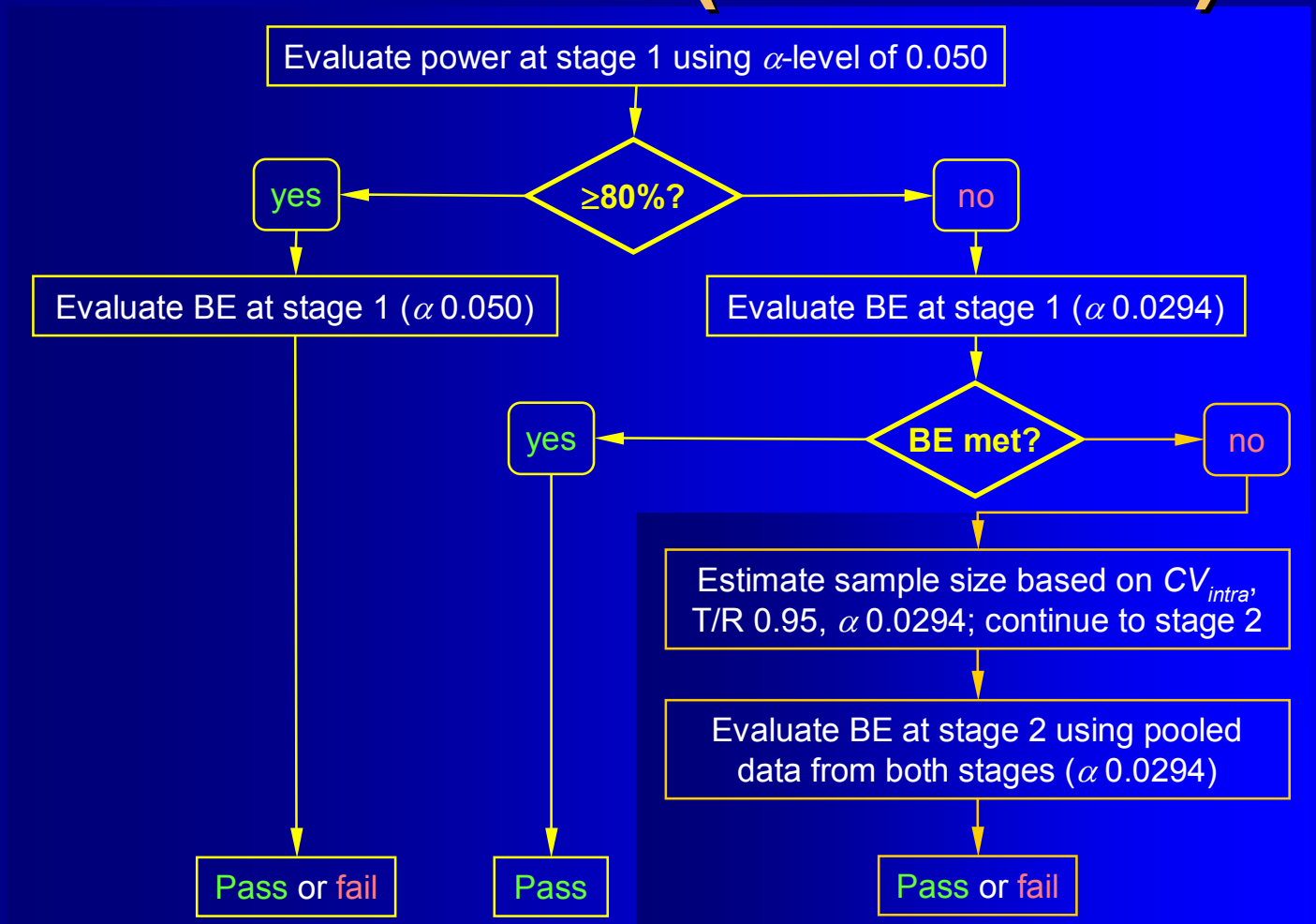
```
Model Specification and User Settings
       Dependent variable : Cmax (ng/mL)
                Transform : LN
            Fixed terms : int+Sequence+Stage+Period(Stage)+Treatment
    Random/repeated terms : Sequence*Stage*Subject

Final variance parameter estimates:
Var(Sequence*Stage*Subject)    0.518978
          Var(Residual)        0.0458956
        Intrasubject CV        0.216714

Bioequivalence Statistics
User-Specified Confidence Level for CI's = 94.1200
Percent of Reference to Detect for 2-1 Tests = 20.0%
A.H.Lower =  0.800   A.H.Upper =  1.250
Formulation variable: Treatment
Reference: Reference   LSMean = 1.133431  SE = 0.171385  GeoLSM = 3.106297
-------------------------------------------------------------------------
Test:      Test        LSMean = 1.147870  SE = 0.171385  GeoLSM = 3.151473

    Difference  =   0.0144,  Diff_SE = 0.0677,  df = 17.0
    Ratio(%Ref) = 101.4544

                      Classical
    CI  90% = (   90.1729,  114.1472)
    CI User = (   88.4422,  116.3810)
    Average bioequivalence shown for confidence=94.12 and percent=20.0.
```

8 subjects in stage 2 (20 total), modified model in pooled analysis

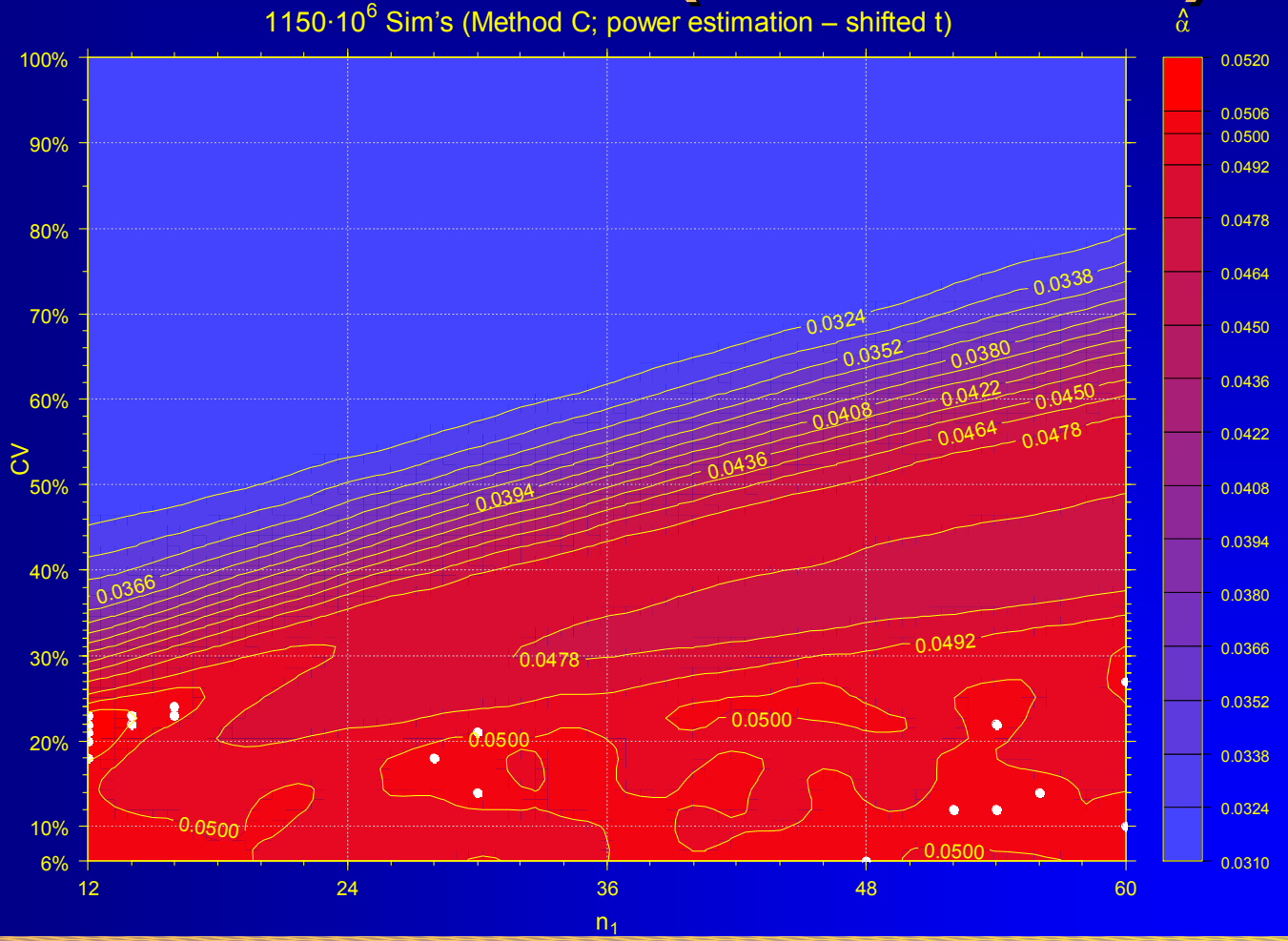
α 0.0294 in pooled analysis

BE shown with 94.12% CI; overall $\alpha \leq 0.05$!

Potvin *et al.* (Method C)



Potvin *et al.* (Method C)



Potvin *et al.* (B *vs.* C)

- Pros & cons
  - Method C (*if power* $\geq$*80%!*) is a conventional BE study; no penality in terms of α needs to be applied.
  - Method C proceeds to stage 2 less often and has smaller average total sample sizes than Method B for cases where the initial sample size is reasonable for the CV.
  - If the size of stage 1 is low for the actual CV both methods go to stage 2 almost all the time; total sizes are similar.
  - Method B slightly more conservative than C.

Potvin *et al.* (B *vs.* C)

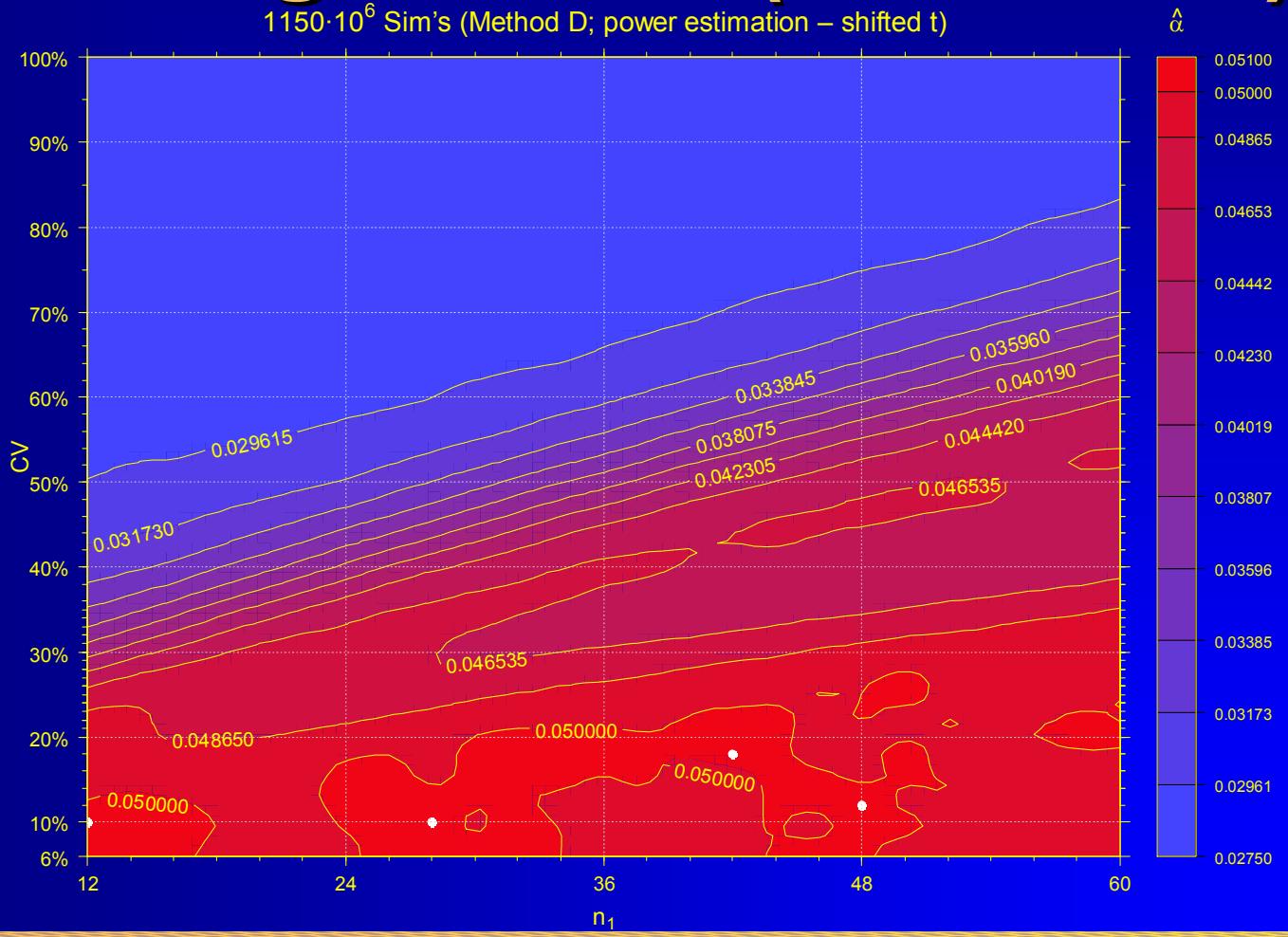
- Recommendations
  - Method C preferred due to slightly higher power than method B (FDA, HPB). Method B for EMA.
  - Plan the study *as if* the CV is known
    - If assumptions turn out to be true = no penalty
    - If lower power (CV_{intra} higher than expected), BE still possible in first stage (penalty; 94.12% CI) or continue to stage 2 as a 'safety net'.
  - Don't jeopardize! Smaller sample sizes in the first stage than in a fixed design don't pay off. Total sample sizes are ~10–20% higher.

Sequential Designs

- Methods by Potvin *et al.* (2008) limited to T/R of 0.95 and 80% power
  - Follow-up paper 2011
    - T/R 0.90 instead of 0.95.
    - Method D (like C, but α 0.0280 instead of α 0.0294).
    - Might be useful if T/R 0.95 and power 90% as well; *not validated yet!* Simulations required.

Montague TH, Potvin D, DiLiberti CE, Hauck WW, Parr AF, and DJ Schuirmann
Additional results for 'Sequential design approaches for bioequivalence studies with crossover designs'
Pharmaceut Statist 11/1, 8–13 (2011), DOI: 10.1002/pst.483

Montague *et al.* (Method D)



Regulatory Acceptance

- Method C: Study passed in first stage (49 subjects, *CV* 30.65%, 90% CI)
  - Deficiency 1: Unadjusted α in stage 1 not acceptable
  - Response 1: Study passed with 94.12% CI (*post hoc* switch to Method B).
  - Deficiency 2: The Applicant should demonstrate that the type I error inflation which can be expected from the chosen approach, did not impact on the decision of bioequivalence.
  - Response 2: One million simulations based on study's sample size and *CV*.
    α_{emp} 0.0494 (95% CI: 0.0490 – 0.0498)

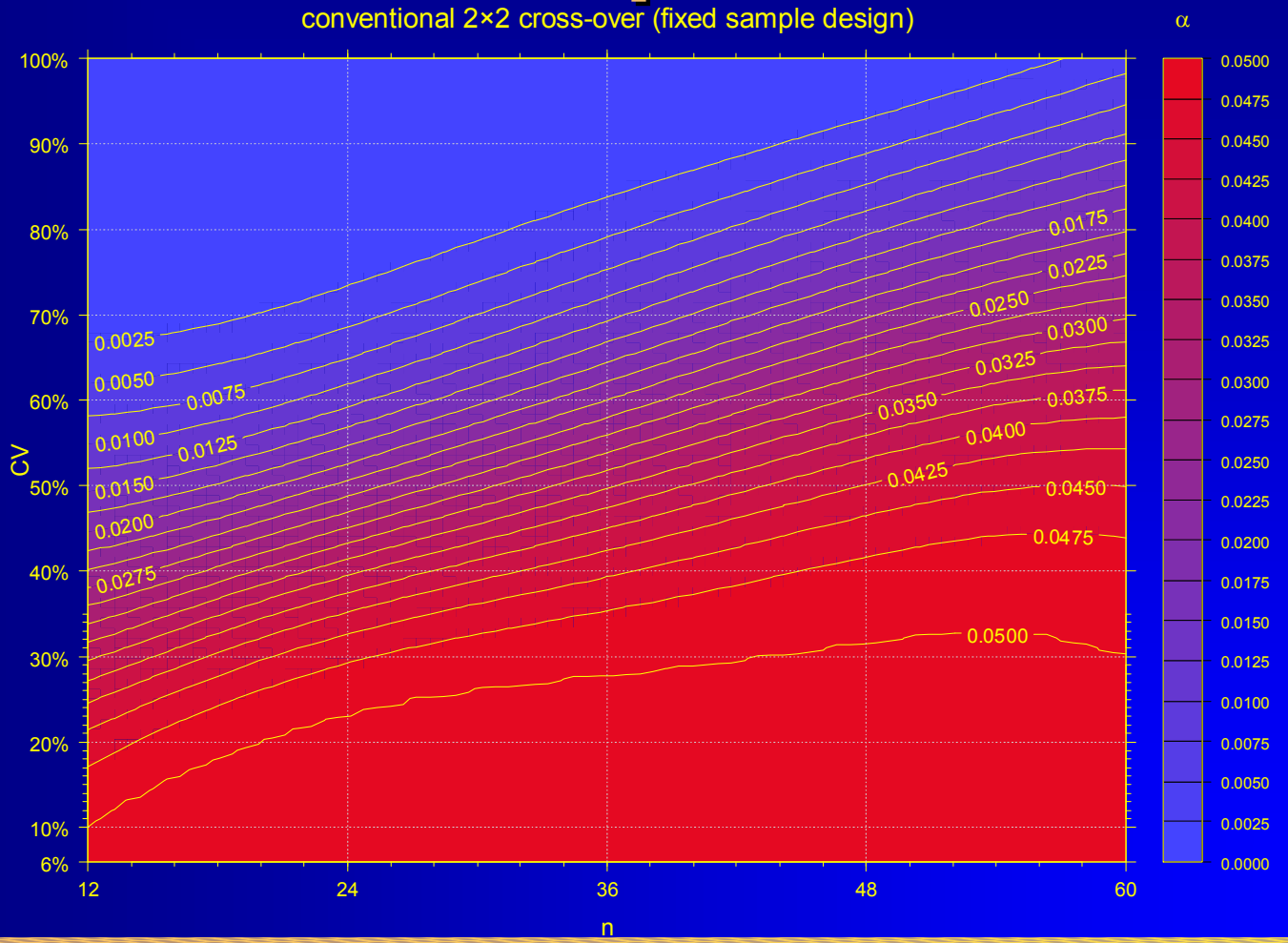
Regulatory Acceptance

- Method C: Study stopped in first stage
  AUC power >80%, passed with 90% CI
  C_{max} power <80%, passed with 94.12% CI
  - Deficiency: Adapting the confidence intervals based upon power is not acceptable and also not in accordance with the EMA guideline. Confidence intervals should be selected *a priori*, without evaluation of the power. Therefore, the applicant should submit the 94.12% confidence intervals for AUC.
  - Pending: AUC would fail with 94.12% CI.

Outlook

- Feasibility / futility rules.
- Arbitrary expected T/R and/or power.
- Adaption for T/R observed in stage 1 (full adaptive design).
- Methods without interim power.
- Application to parallel designs (patients, long half-life drugs).
- Dropping a candidate formulation from a higher-order cross-over; continue with 2×2.

Don’t panic!


conventional 2×2 cross-over (fixed sample design)
α
CV
n
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
6%
12
24
36
48
60
0.0025
0.0050
0.0075
0.0100
0.0125
0.0150
0.0175
0.0200
0.0225
0.0250
0.0275
0.0300
0.0325
0.0350
0.0375
0.0400
0.0425
0.0450
0.0475
0.0500
0.0500
0.0475
0.0450
0.0425
0.0400
0.0375
0.0350
0.0325
0.0300
0.0275
0.0250
0.0225
0.0200
0.0175
0.0150
0.0125
0.0100
0.0075
0.0050
0.0025
0.0000

Thank You!

Statistics of Two Stage Study Designs

Open Questions?

Helmut Schütz

BEBAC

Consultancy Services for
Bioequivalence and Bioavailability Studies
1070 Vienna, Austria
helmut.schuetz@bebac.at

Dedicated to the memory of Dirk Maarten Barends (1945 – 2012).

To bear in Remembrance...

Power. That which statisticians are always calculating but never have.

Power Calculation – A guess masquerading as mathematics.

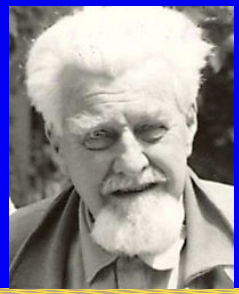
Stephen Senn

In bioequivalence we must not forget the only important – *the patient*! He/she is living person, not just α 0.05.

Dirk Marteen Barends

It is a good morning exercise for a research scientist to discard a pet hypothesis every day before breakfast.
It keeps him young.

Konrad Lorenz



References

- ICH
  - E9: Statistical Principles for Clinical Trials (1998)
- EMA-CPMP/CHMP/EWP
  - Points to Consider on Multiplicity Issues in Clinical Trials (2002)
  - Guideline on the Investigation of BE (2010)
  - Questions & Answers: Positions on specific questions addressed to the EWP therapeutic subgroup on Pharmacokinetics (2011, 2012)
- US-FDA
  - Center for Drug Evaluation and Research (CDER)
    - Statistical Approaches Establishing Bioequivalence (2001)
    - Bioequivalence Recommendations for Specific Products (2007–2012):
      Draft Guidance on Lotepredenol (Jun 2012)
      Draft Guidance on Dexamethasone/Tobramycin (Jun 2012)
- Midha KK *et al.*
  Logarithmic Transformation in Bioequivalence: Application with Two Formulations of Perphenazine
  J Pharm Sci 82/2, 138–44 (1993)
- Hauschke D, Steinijans VW, and E Diletti
  Presentation of the intrasubject coefficient of variation for sample size planning in bioequivalence studies
  Int J Clin Pharmacol Ther 32/7, 376–8 (1994)
- Diletti E, Hauschke D, and VW Steinijans
  Sample size determination for bioequivalence assessment by means of confidence intervals
  Int J Clin Pharm Ther Toxicol 29/1, 1–8 (1991)
- Hauschke D *et al.*
  Sample Size Determination for Bioequivalence Assessment Using a Multiplicative Model
  J Pharmacokin Biopharm 20/5, 557–61 (1992)
- Chow S-C and H Wang
  On Sample Size Calculation in Bioequivalence Trials
  J Pharmacokin Pharmacodyn 28/2, 155–69 (2001)
  Errata: J Pharmacokin Pharmacodyn 29/2, 101–2 (2002)
- DB Owen
  A special case of a bivariate non-central t-distribution
  Biometrika 52, 3/4, 437–46 (1965)
- AL Gould
  Group Sequential Extension of a Standard Bioequivalence Testing Procedure
  J Pharmacokin Biopharm 23/1, 57–86 (1995)
  DOI: 10.1007/BF02353786

References

- Jones B and MG Kenward
  Design and Analysis of Cross-Over Trials
  Chapman & Hall/CRC, Boca Raton (2nd Edition 2000)
- Patterson S and B Jones
  Determining Sample Size, in:
  Bioequivalence and Statistics in Clinical Pharmacology
  Chapman & Hall/CRC, Boca Raton (2006)
- SA Julious
  Tutorial in Biostatistics. Sample sizes for clinical trials with Normal data
  Statistics in Medicine 23/12, 1921–86 (2004)
- SA Julious
  Sample Sizes for Clinical Trials
  Chapman & Hall/CRC, Boca Raton (2010)
- D Labes
  Package 'PowerTOST', Version 1.0-00 (2012-10-26)
  http://cran.r-project.org/web/packages/PowerTOST/PowerTOST.pdf
- Potvin D *et al.*
  Sequential design approaches for bioequivalence studies with crossover designs
  Pharmaceut Statist 7/4, 245–62 (2008)
- Montague TH *et al.*
  Additional results for 'Sequential design approaches for bioequivalence studies with crossover designs'
  Pharmaceut Statist 11/1, 8–13 (2011)
- García-Arieta A and J Gordon
  Bioequivalence Requirements in the European Union: Critical Discussion
  The AAPS Journal 14/4, 738–48 (2012)