BE · · BAC

Üdvözöljük!
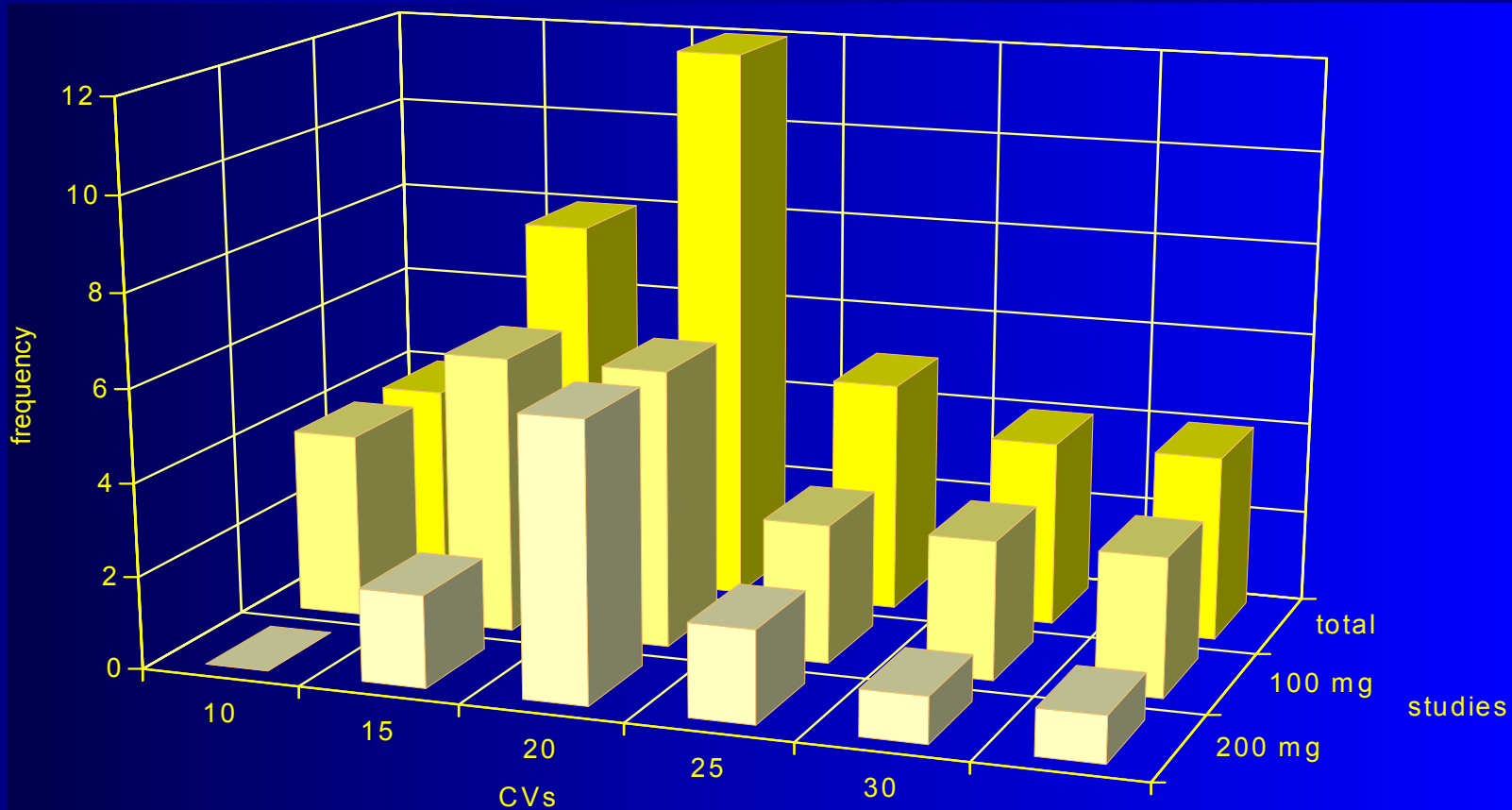
Experiences in Implementing Two-Stage Designs in Europe

Tricks and Traps

Helmut Schütz
BEBAC

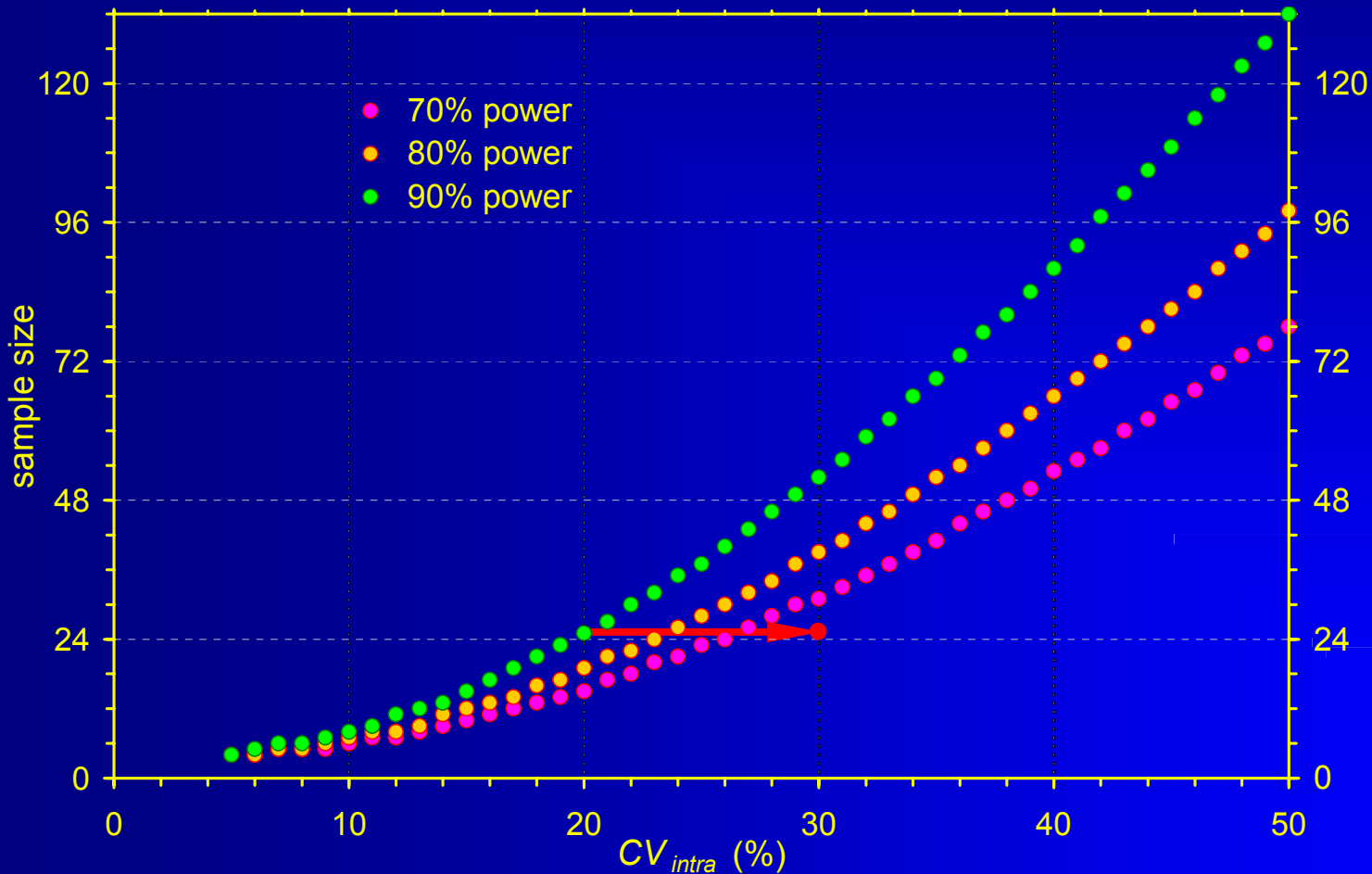Wikimedia Commons • 2005 Dirk Beyer • Creative Commons Attribution-ShareAlike 3.0 Unported

Inconsistent data

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

CV based on assumptions!

2×2 cross-over, T/R 0.95

Study planned on expected 20% CV with 90% power (n 26).

If true CV is 30%, power drops to only 61%…

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) first validated framework in the context of BE
    - 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

Review of Guidelines

- EMA (Jan 2010)

   Acceptable; Potvin *et al.* Method B preferred (?)

- Russia (Draft 2011)

   Acceptable (Methods B and C)

- Canada (May 2012)

   Potvin *et al.* Method C recommended

- FDA (Jun 2012)

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

EMA (TSDs)

- 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*.
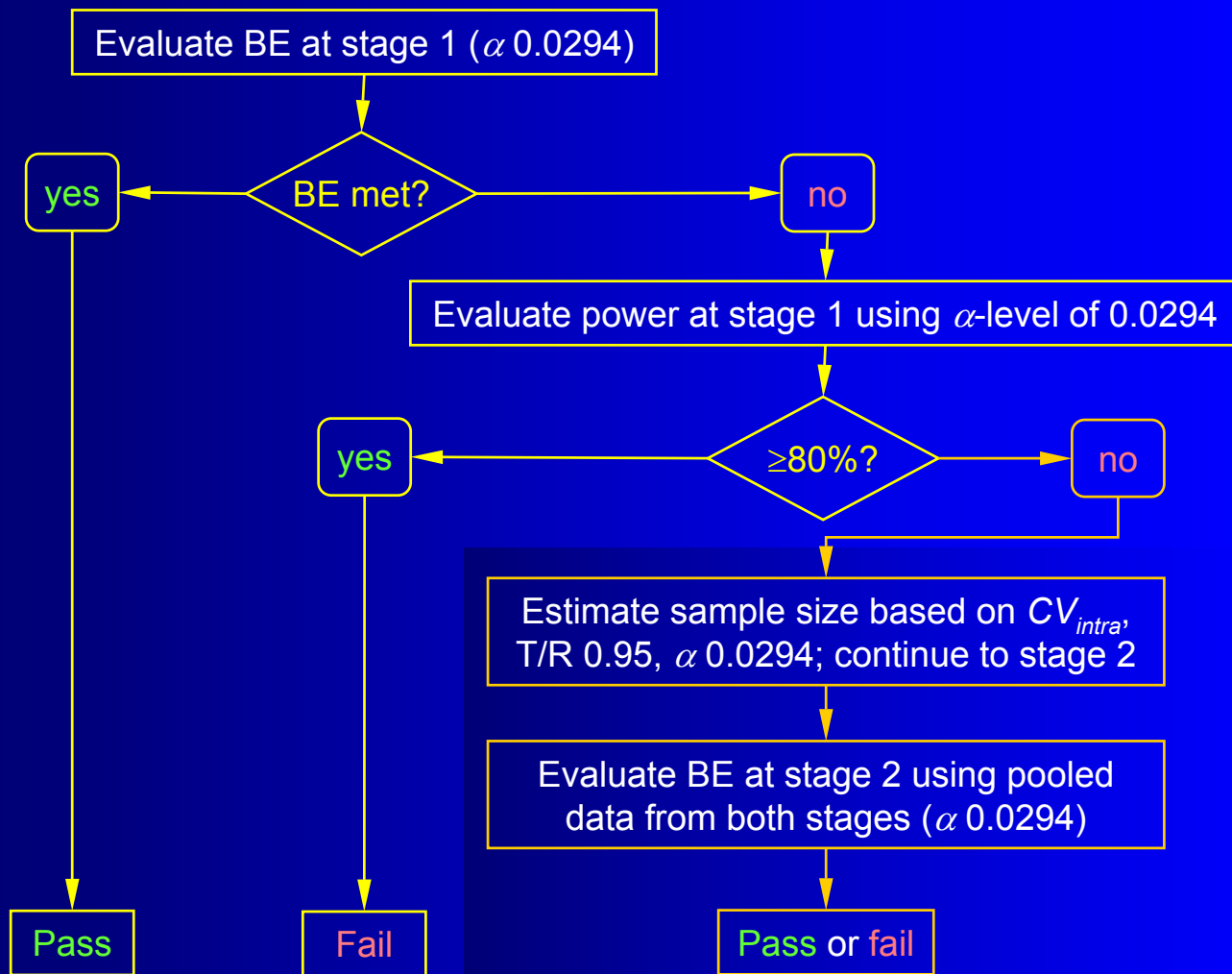 - First stage data should be treated as an interim analysis.

EMA (TSDs)

- 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.
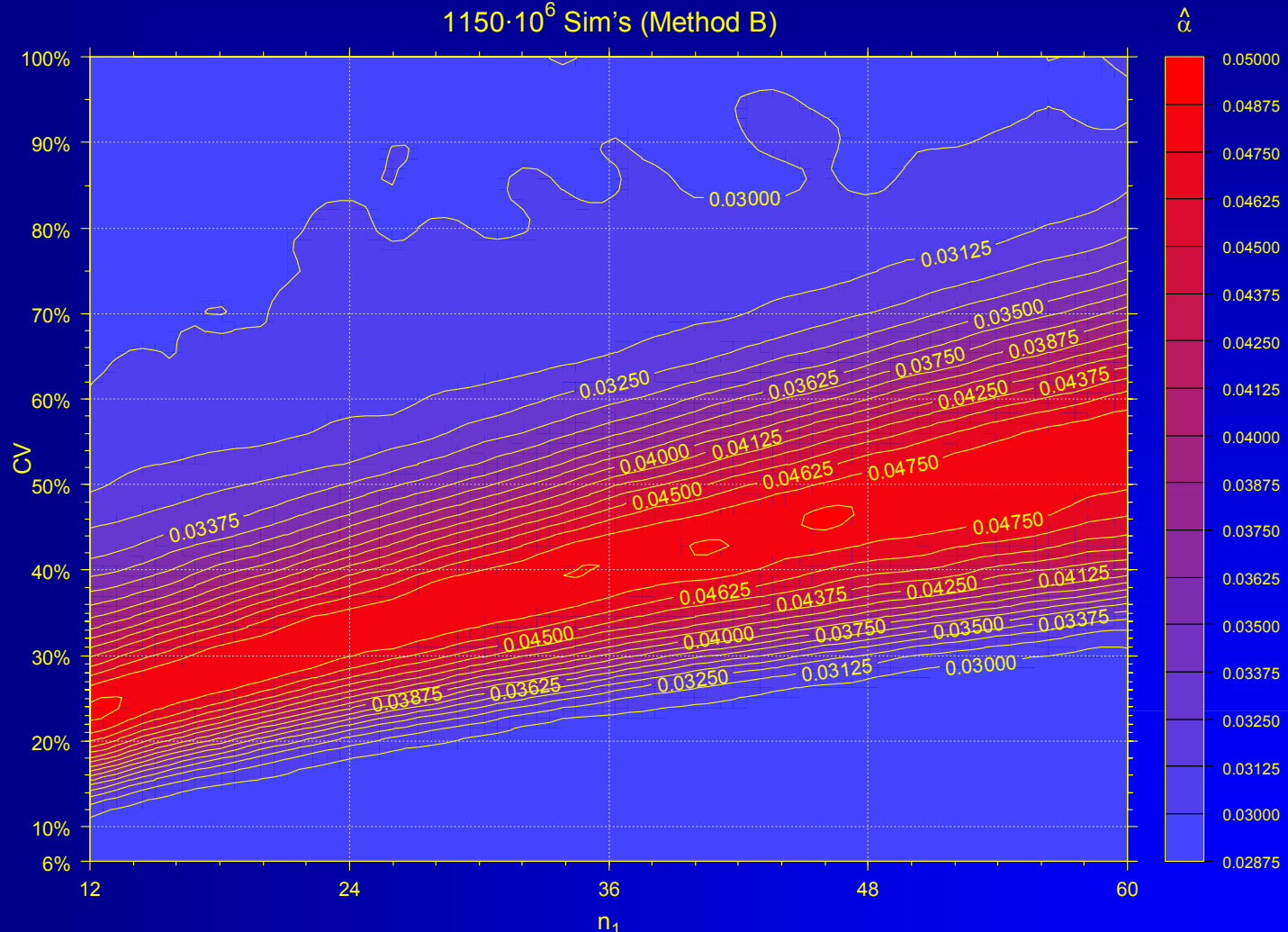
EMA (TSDs)

- 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)
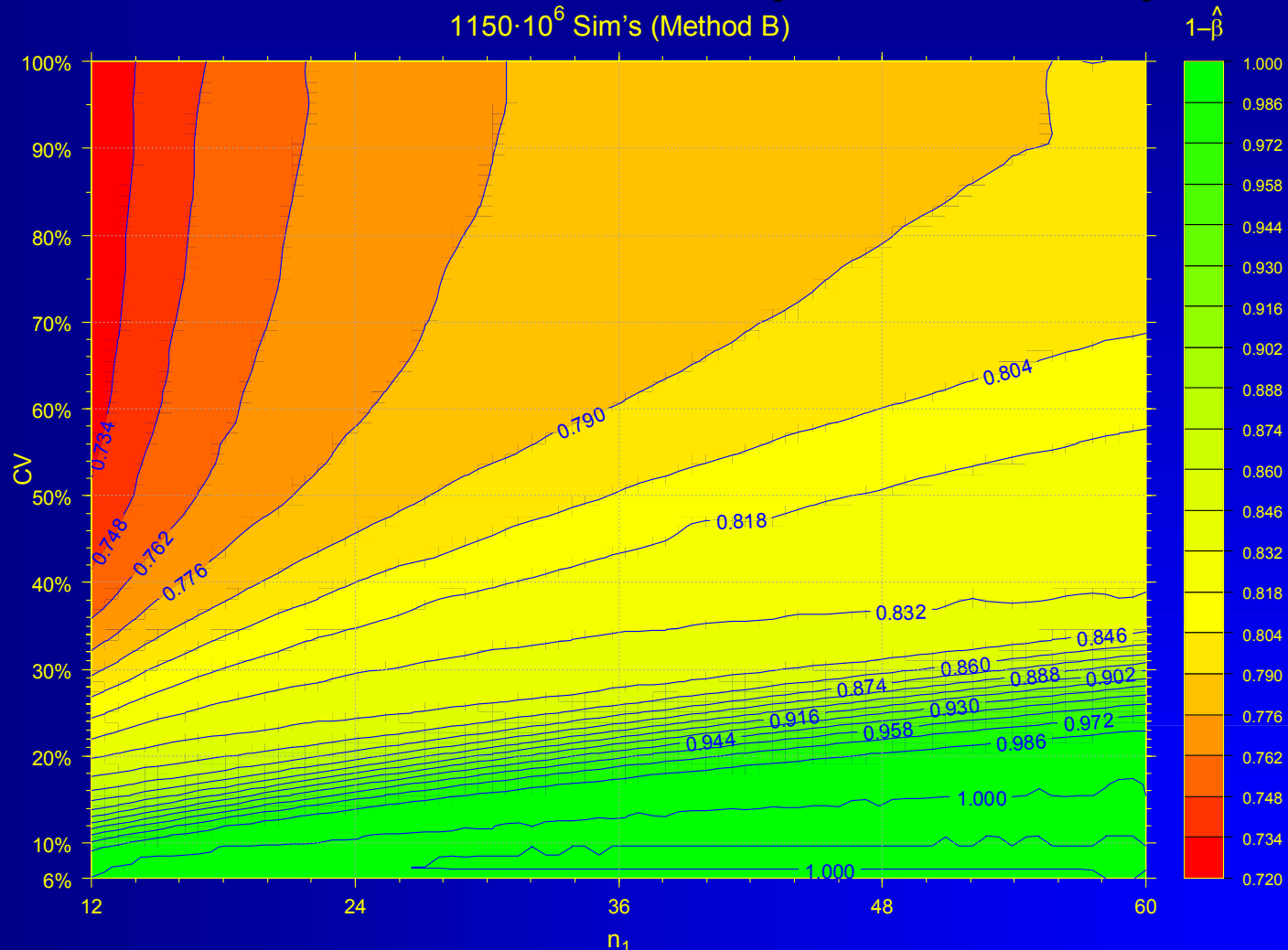
Evaluate BE at stage 1 (α 0.0294)

BE met?

yes → Pass

no → Evaluate power at stage 1 using α-level of 0.0294

≥80%?

yes → Fail

no → Estimate sample size based on CV_{intra}, T/R 0.95, α 0.0294; continue to stage 2

Evaluate BE at stage 2 using pooled data from both stages (α 0.0294)

Pass or fail

Potvin *et al.* (Method B)

$1150 \cdot 10^{6}$ Sim's (Method B)

Potvin *et al.* (Method B)

$1150 \cdot 10^{6}$ Sim's (Method B)

Potvin *et al.* (Method B)

$1150 \cdot 10^6$ Sim's (Method B)

informa
life sciences

Potvin *et al.* (Method B)

Sample size penalty (*CV* 14–40%, 80% power)

- planned for 0.0500
- planned for 0.0294

$n_{total} = 1.084n$

$n_{total} = 1.023n$

n_{total}: average sample size (two-stage)

n: sample size (fixed)
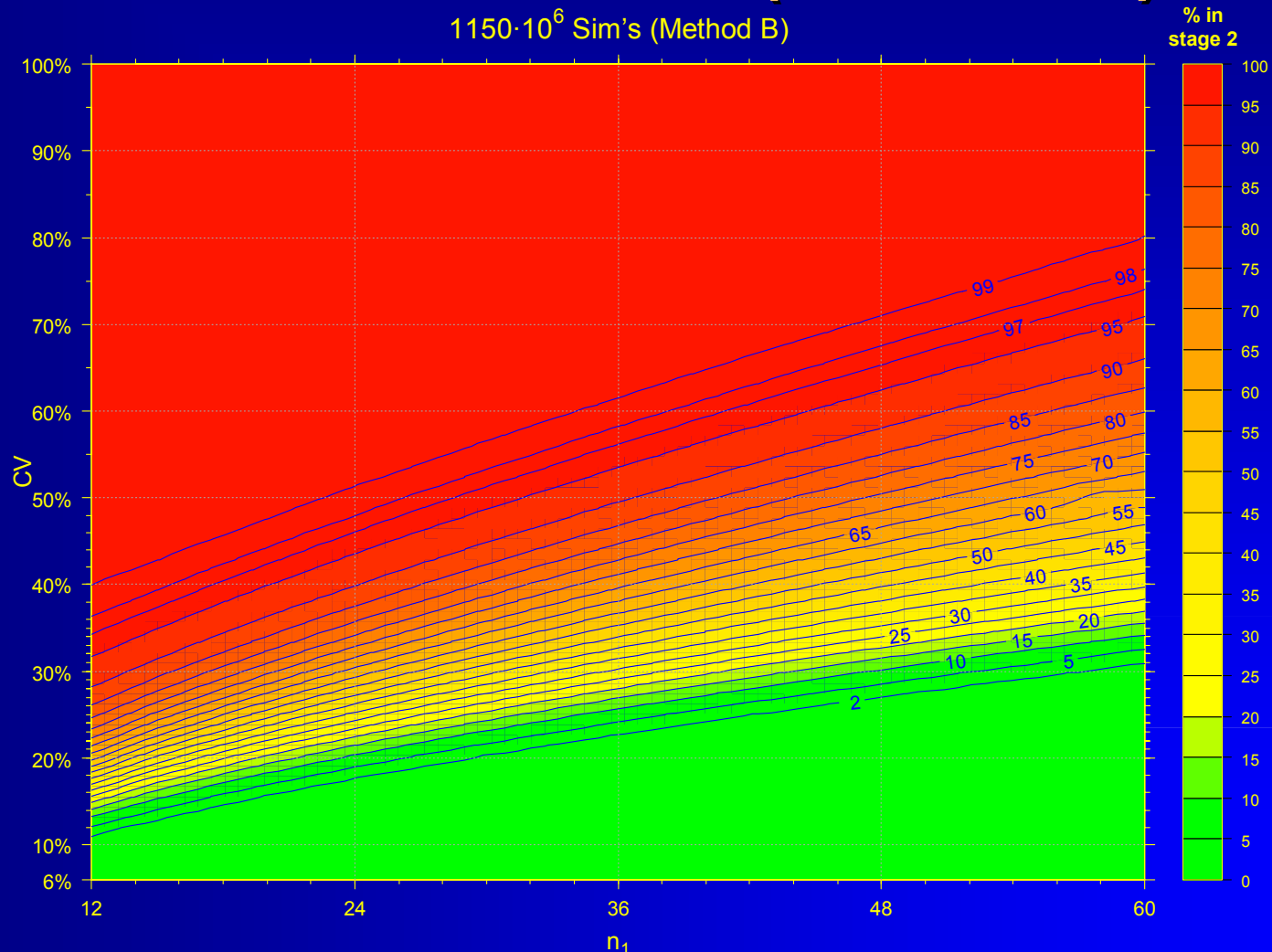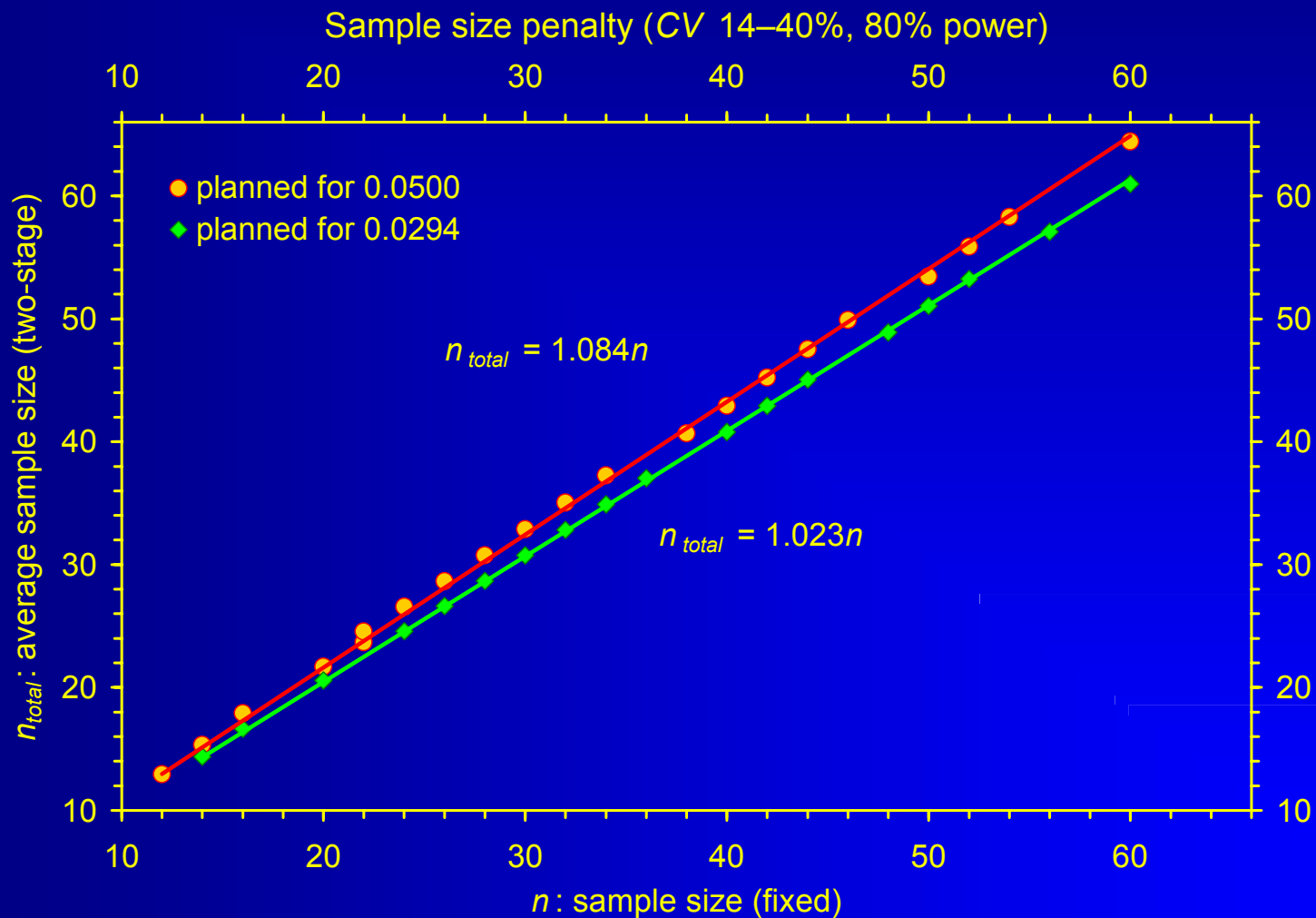
Potvin *et al.* (Method B)

- Technical Aspects
  - Only *one* Interim Analysis (after stage 1).
  - Use software (wide step sizes in Diletti's tables); preferrable the exact method (avoid approximations).
  - 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 $1 - 2 \times \alpha = 94.12\%$ CI is calculated.
  - Overall patient's risk preserved at ≤ 0.05.

Potvin *et al.* (Method B)

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

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

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

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

  - 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%

Example (Potvin Method B)

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

12 subjects in stage 1, conventional BE model

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

CV_{intra} 18.2%

```
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
```

α 0.0294

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

Failed with 94.12% Confidence Interval

Example (Potvin Method B)

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

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

```
[1] 0.5251476
```

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)
```

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

```
+++++++++++ 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
```

Total sample size 20: include another 8 in stage 2

Example (Potvin Method B)

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

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

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

α 0.0294 in pooled analysis

```
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.
```

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

Example (Potvin Method B / EMA)

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

Final variance parameter estimates:
Var(Sequence*Stage*Subject)    0.549653
            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

Q&A Rev. 7 (March 2013)

α 0.0294 in pooled analysis

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

Potvin *et al.* (Method C)

Evaluate power at stage 1 using α-level of 0.050

$\geq 80\%$?

yes

no

Evaluate BE at stage 1 (α 0.050)

Evaluate BE at stage 1 (α 0.0294)

BE met?

yes

no

Estimate sample size based on CV_{intra}, T/R 0.95, α 0.0294; continue to stage 2

Evaluate BE at stage 2 using pooled data from both stages (α 0.0294)

Pass or fail

Pass

Pass or fail

informa
life sciences

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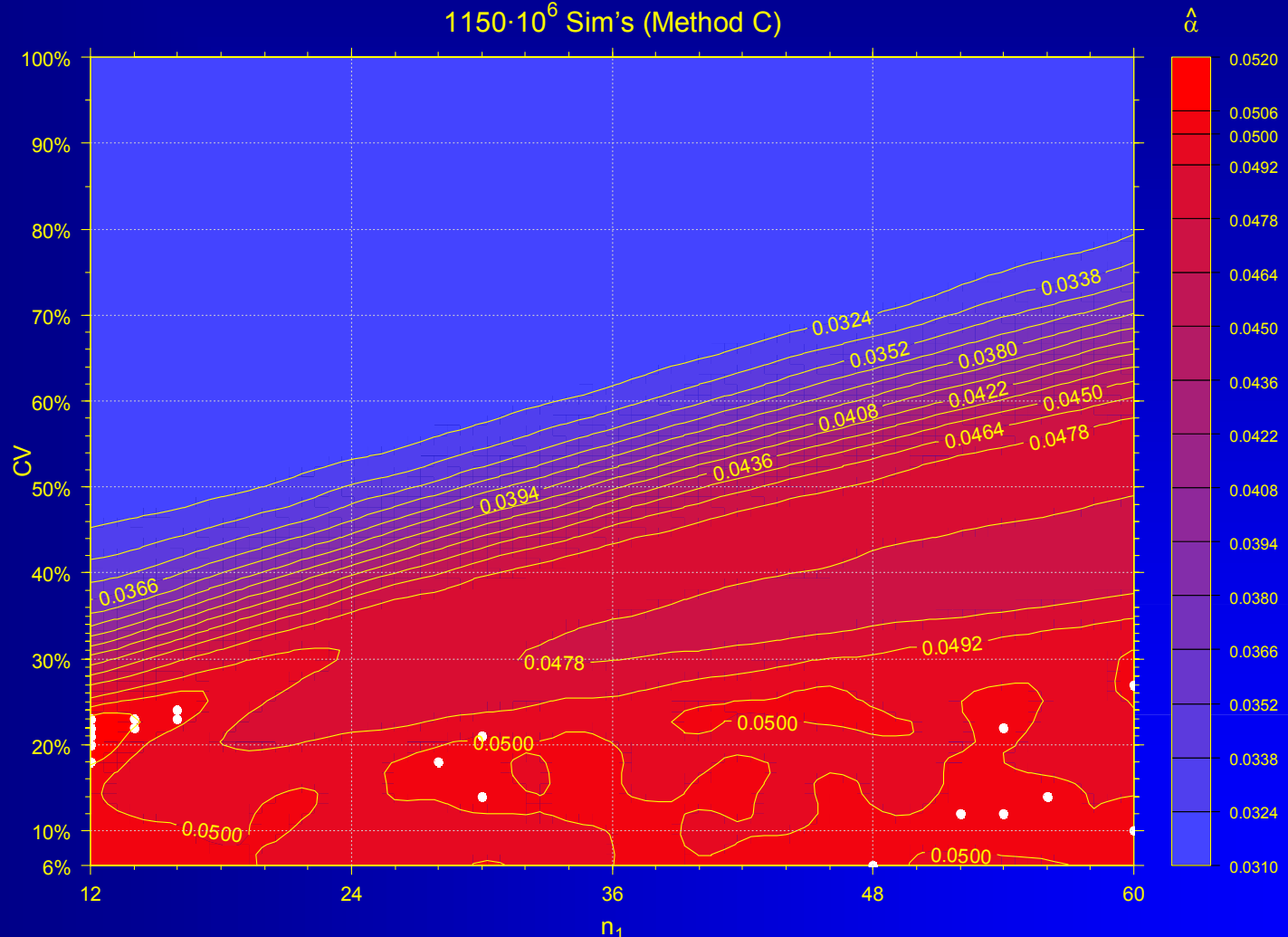
Potvin *et al.* (Method B *vs.* C)

- 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 reason-able 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._ (Method 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.

TSDs: Alternatives

● Methods by Potvin *et al.* (2008) limited to T/R of 0.95 and 80% power

■ Follow-up papers (T/R 0.95…0.90, 80…90% power)

reference	method	T/R	target power	CV	$\alpha_{adj.}$	max. $\alpha_{emp.}$
Potvin *et al.*	B	0.95	80%	10–100%	0.0294	0.0485
	C	0.95				0.0510
Montague *et al.*	D	0.90			0.0280	0.0518
Fuglsang	B	0.95	90%	10–80%	0.0284	0.0501
	D				0.0274	0.0503
	D	0.90			0.0269	0.0501

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
A Fuglsang
Sequential Bioequivalence Trial Designs with Increased Power and Controlled Type I Error Rates
AAPS J 15, pre-print online (2013) DOI: 10.1208/s12248-013-9475-5

Montague *et al.* (Method D)

$1150 \cdot 10^{6}$ Sim's (Method D)
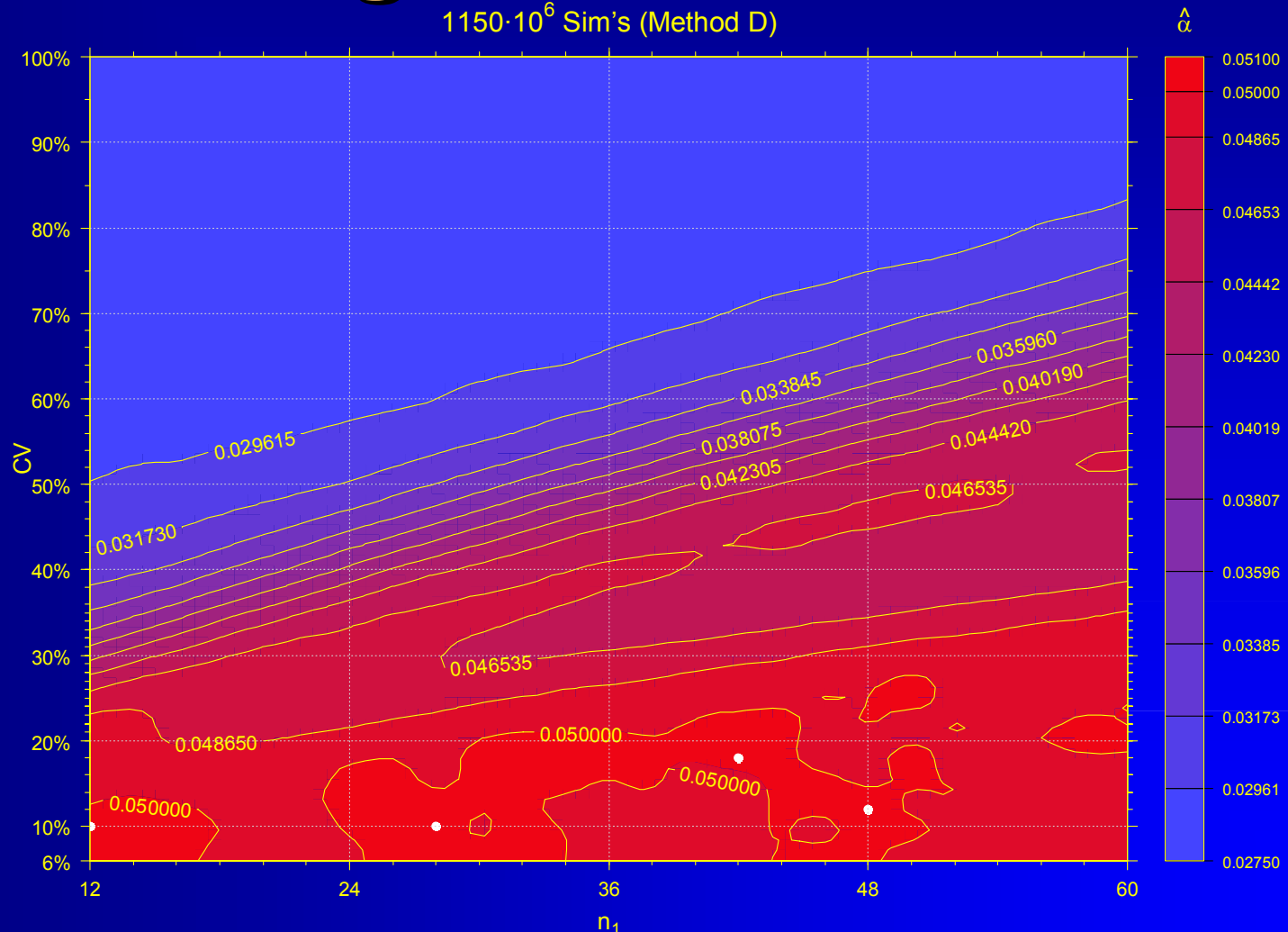
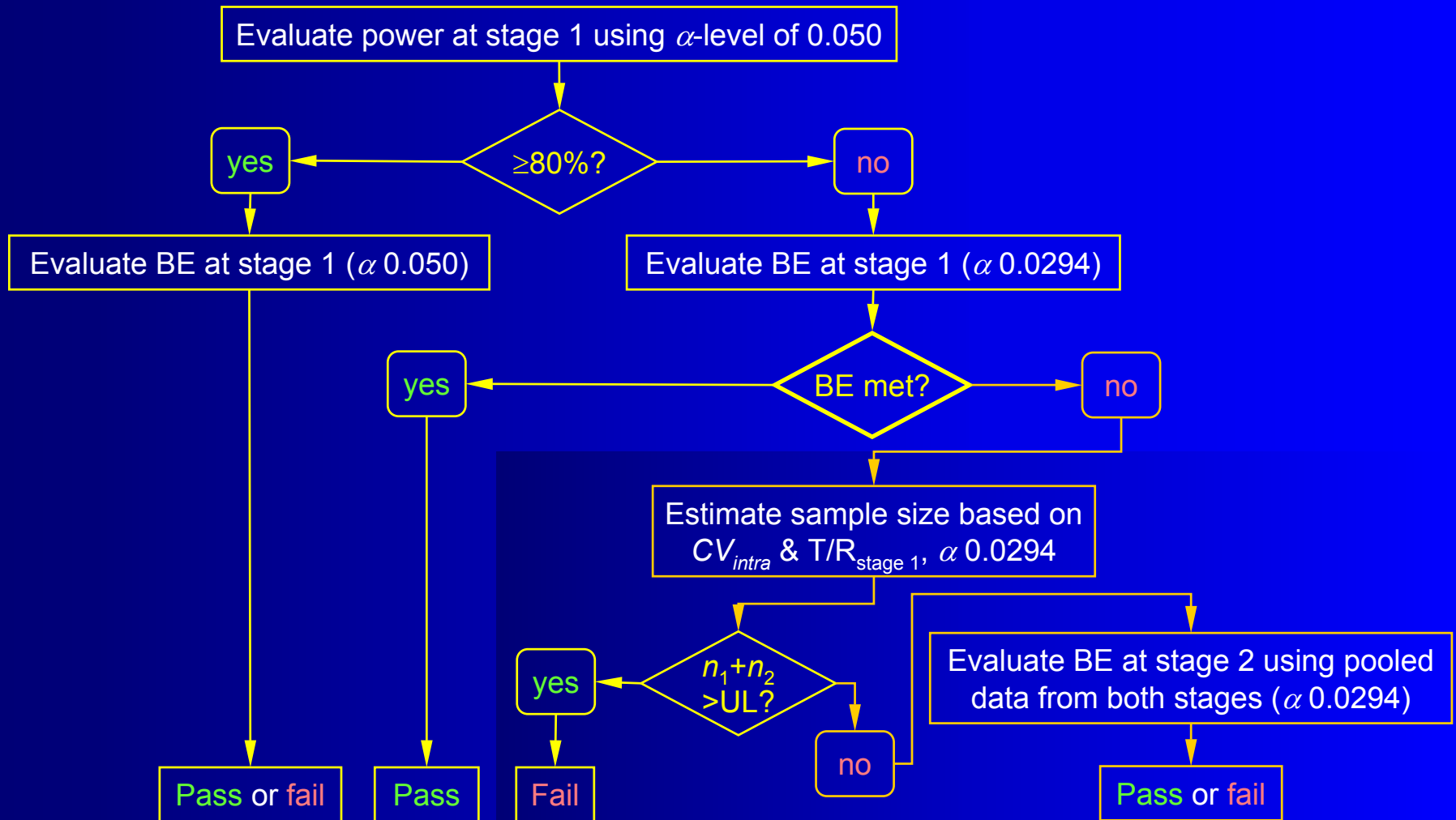# TSDs: Alternatives

- Karalis & Macheras (2013)
    - Based on Method C ($\alpha_{adj.}$ 0.0294)
    - Sample size re-estimation based on observed T/R-ratio in stage 1
    - Upper sample size limit (UL)
    - Frameworks:
        - n_1 12–96, CV 10–60%, $n_1+n_2 \leq$ UL 150
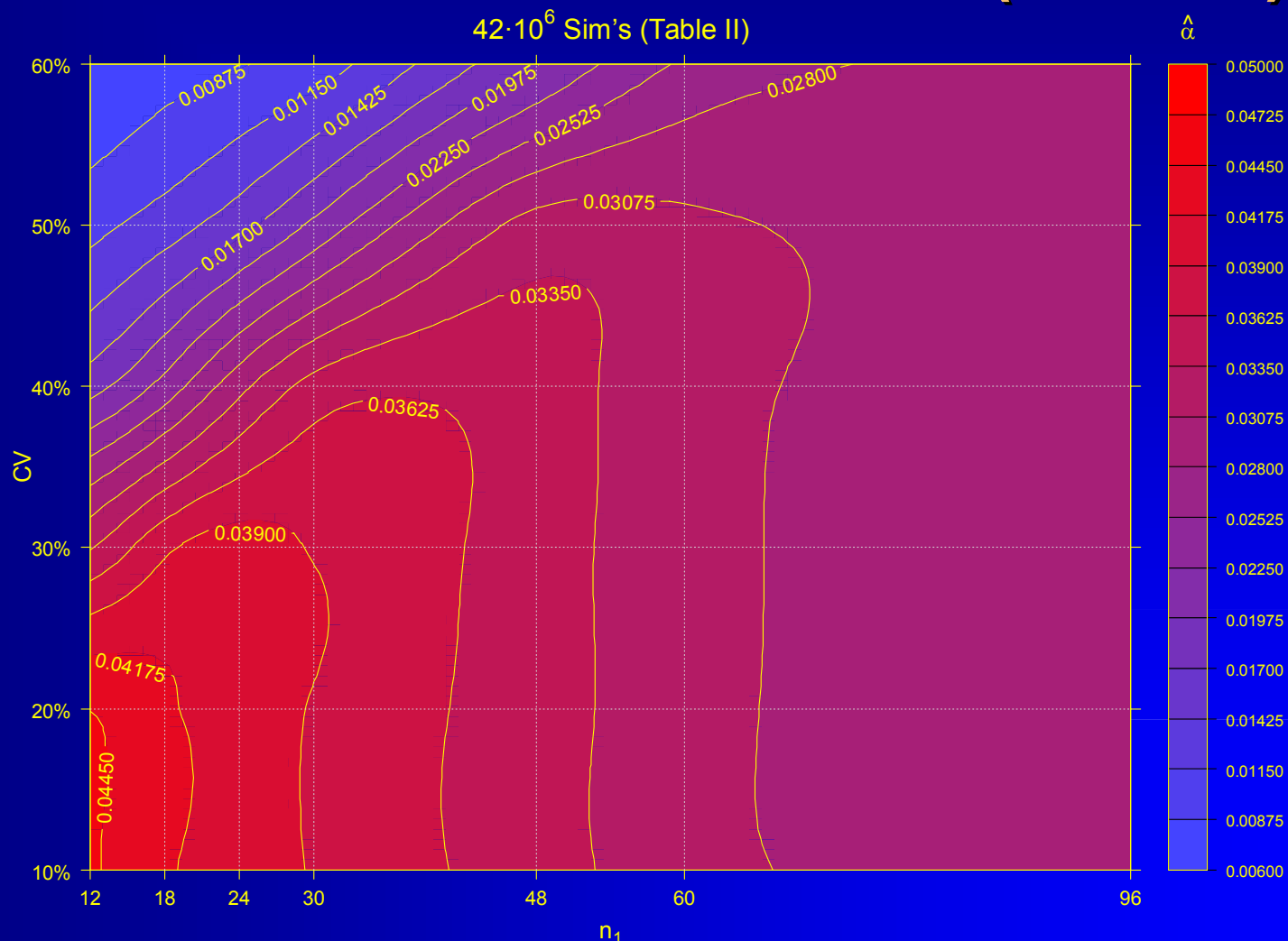   - n_1 18–96, CV 20–40%, $n_1+n_2 \leq$ UL 100

Karalis & Macheras

Evaluate power at stage 1 using α-level of 0.050

$\geq 80\%$?

yes

no

Evaluate BE at stage 1 (α 0.050)

Evaluate BE at stage 1 (α 0.0294)

BE met?

yes

no

Estimate sample size based on CV_{intra} & $T/R_{stage\ 1}$, α 0.0294

$n_1 + n_2 > UL$?

yes

no

Evaluate BE at stage 2 using pooled data from both stages (α 0.0294)

Pass or fail

Pass

Fail

Pass or fail

Karalis & Macheras (UL 150)

$42 \cdot 10^6$ Sim's (Table II)

Karalis & Macheras (UL 150)

α 0.0294, observed T/R 108.76%, CV_{intra} 18.2%, 12 subjects in stage 1

```
require(PowerTOST)
power.TOST(alpha=0.0294, theta0=1.0876,
           CV=0.182132, n=12, design='2x2',
           method='exact')
```

```
[1] 0.4042796
```

Power 40.4% – initiate stage 2

```
sampleN.TOST(alpha=0.0294, targetpower=0.80, logscale=TRUE,
             theta1=0.8, theta2=1.25, theta0=1.0876,
             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 = 1.0876,  CV = 0.182132

Sample size
 n       power
28    0.813921
```

Estimate total sample size: α 0.0294, T/R 108.76%, CV_{intra} 18.2%, 80% power

Total sample size 28 ($\leq$150): include another 14 in stage 2

Case Studies (EMA)

- Method C: Two studies passed in stage 1 (n=15 SD, n=16 MD, C_{max} CV 17.93%, 8.54%, 90% CIs)

- Would have passed with Method B as well; however, 94.12% CIs were not reported.
  - RMS Germany. Accepted by CMSs Austria, Denmark, Sweden, and The Netherlands.
  - Spain: Statistical analysis should be GLM. Please justify.
    - Evaluated with all-fixed effects model. Both studies passed. Response pending (May 2013)

Case Studies (EMA)

- Method C: Study passed in stage 1 (49 subjects, CV 30.65%, 90% CI)
  - UK/Ireland: Unadjusted α in stage 1 not acceptable.
    - Study passed with 94.12% CI (*post hoc* switch to Method B).
  - Austria: 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.
    - One million simulations based on the study's sample size and CV.
    α_{emp} 0.0494 (95% CI: 0.0490–0.0498)

Case Studies (EMA)

- Method C: Study stopped in stage 1
  AUC power >80%, passed with 90% CI
  C_{max} power <80%, passed with 94.12% CI
  - The Netherlands: 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.
    - AUC fails with 94.12% CI…

Outlook

- Feasibility / futility rules.

- Arbitrary expected T/R and/or power.

- Methods without interim power.

- Application to parallel designs.

- Dropping a candidate formulation from a higher-order cross-over; continue with 2×2.

- Exact method (not depending on simulations).

Don't panic!

conventional 2×2 cross-over (fixed sample design)
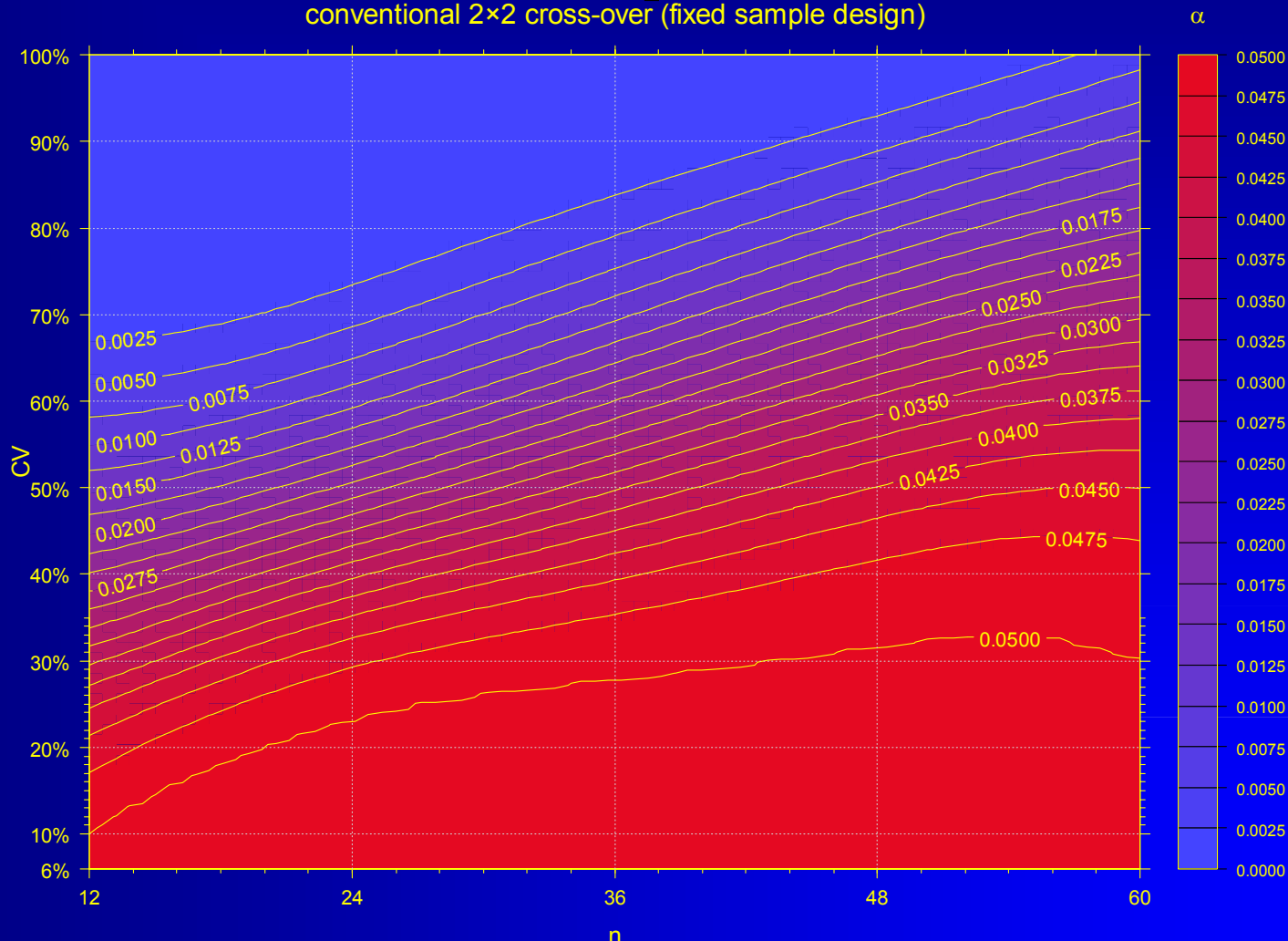
Thank You!

Experiences in Implementing TSDs in Europe: Tricks and Traps
Open Questions?

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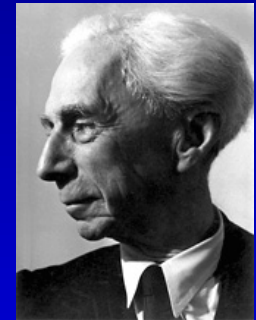1070 Vienna, Austria
helmut.schuetz@bebac.at

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

To bear in Remembrance...

The fundamental cause of trouble in the world today is that the stupid are cocksure
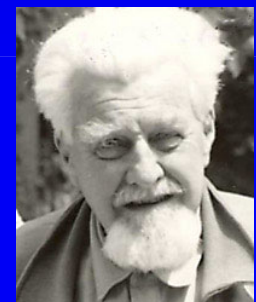while the intelligent are full of doubt.   *Bertrand Russell*

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 (2013)
- 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)
- DB Owen
  A special case of a bivariate non-central t-distribution
  Biometrika 52(3/4), 437–46 (1965)

- 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)
- 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)
- 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)
- AL Gould
  Group Sequential Extension of a Standard Bioequivalence Testing Procedure
  J Pharmacokin Biopharm 23(1), 57–86 (1995)
  DOI: 10.1007/BF02353786

References

- Hauck WW, Preston PE, and FY Bois
  A Group Sequential Approach to Crossover Trials for Average Bioequivalence
  J Biopharm Stat 71, 87–96 (1997)
  DOI: 10.1080/10543409708835171
- Patterson S and B Jones
  Determining Sample Size, in:
  Bioequivalence and Statistics in Clinical Pharmacology
  Chapman & Hall/CRC, Boca Raton (2006)
- SA Julious
  Sample Sizes for Clinical Trials
  Chapman & Hall/CRC, Boca Raton (2010)
- D Labes
  Package 'PowerTOST', Version 1.1-02 (2013-02-28)
  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)
  DOI: 10.1002/pst.294

- Montague TH *et al.*
  Additional results for 'Sequential design approaches for bio-equivalence studies with crossover designs'
  Pharmaceut Statist 11(1), 8–13 (2011) DOI: 10.1002/pst.483
- García-Arieta A and J Gordon
  Bioequivalence Requirements in the European Union: Critical Discussion
  AAPS J 14(4), 738–48 (2012)
  DOI: 10.1208/s12248-012-9382-1
- BM Davit
  Sequential Designs and Interim Analyses in Bioequivalence: FDA's Experience
  Mini-Symposium on Adaptive Study Designs and Assess-ment Approaches for Bioequivalence
  AAPS Annual Meeting, Chicago, IL, October 13–18, 2012
- A Fuglsang
  Sequential Bioequivalence Trial Designs with Increased Power and Controlled Type I Error Rates
  AAPS J 15/3 (2013) DOI: 10.1208/s12248-013-9475-5
- Karalis V and P Macheras
  An Insight into the Properties of a Two-Stage Design in Bioequivalence Studies
  Pharm Res, pre-print online (April 2013)
  DOI: 10.1007/s11095-013-1026-3