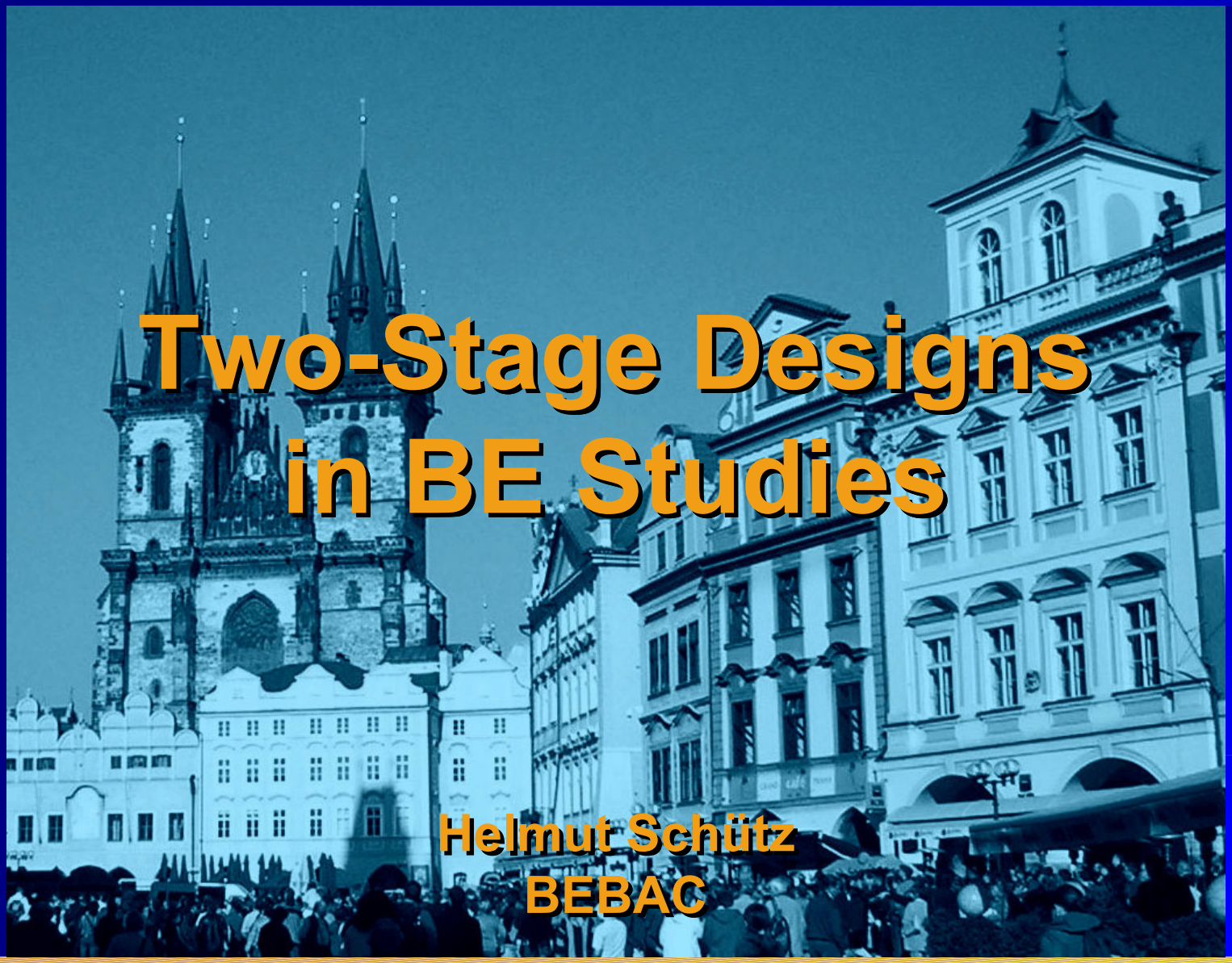
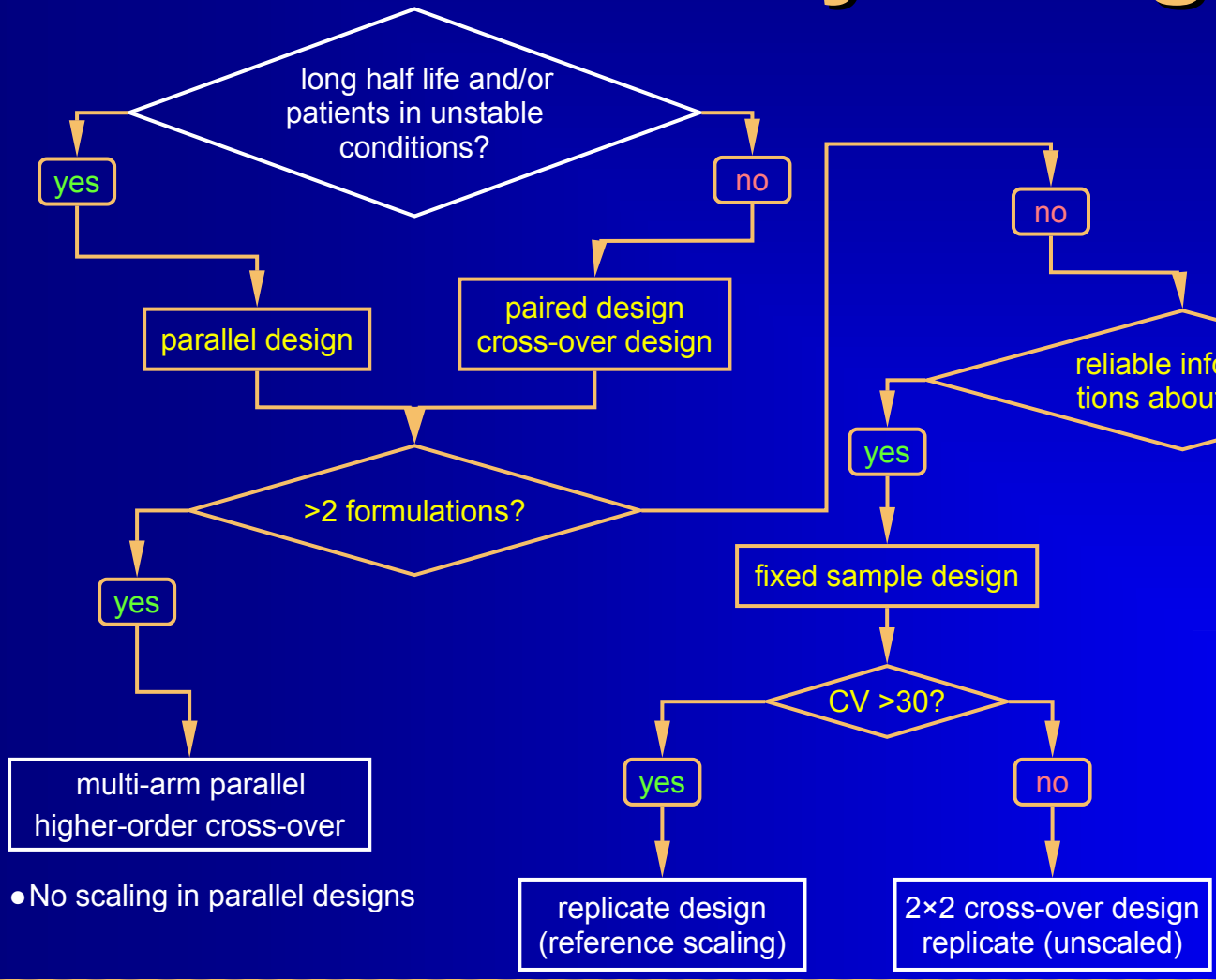


Two-Stage Designs in BE Studies

Helmut Schütz
BEBAC

BE Study Designs



- No scaling in parallel designs

- *Currently* no two-stage design if
  - Parallel design
  - >2 formulations
  - Replicate design
- Futility rules (*e.g.*, maximum sample size) in TSDs problematic.

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 (mainly over-optimistic) assumptions about *CV* 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)
DOI: 10.1007/BF02353786

Sequential Designs

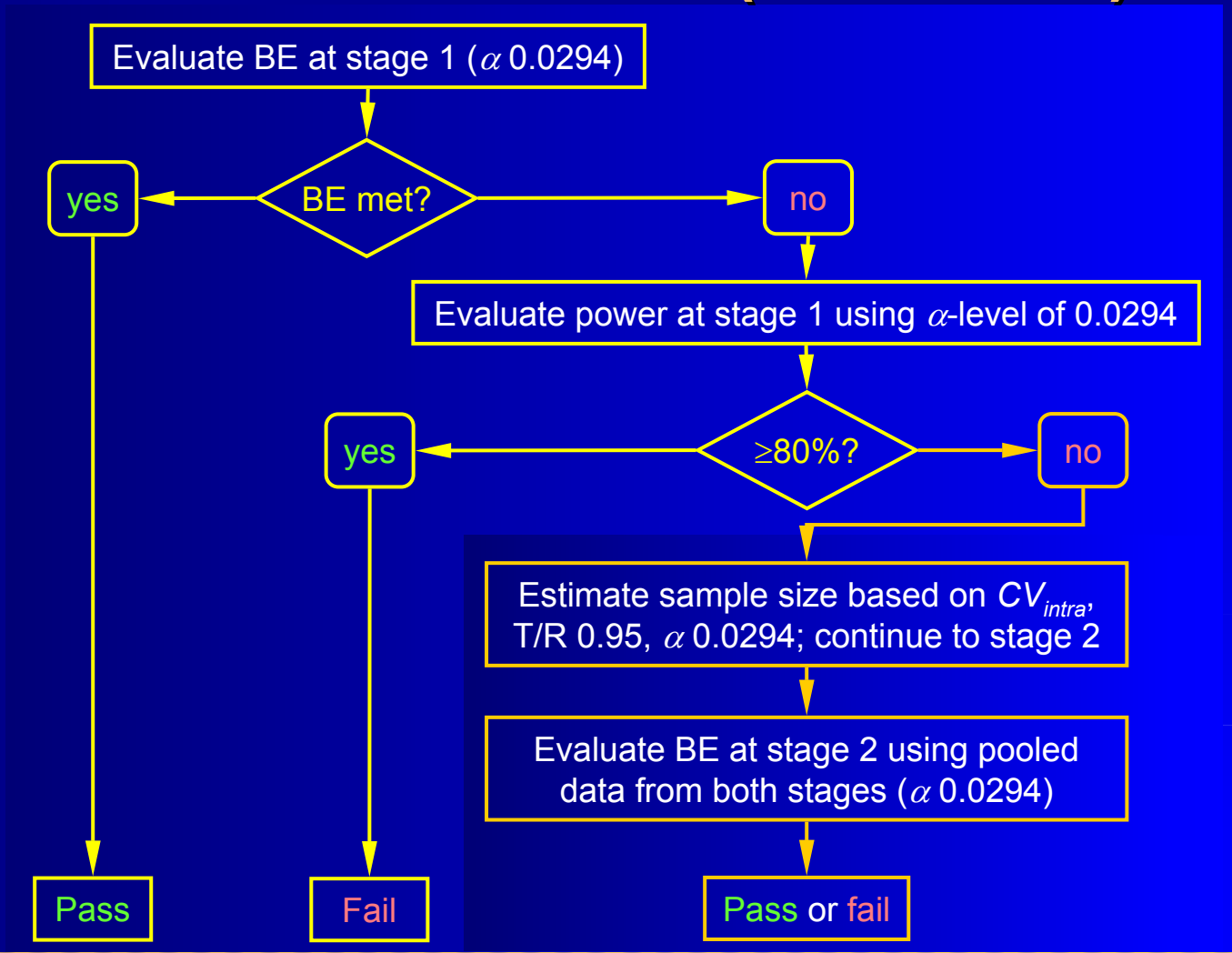
- Methods by Potvin *et al.* (2008) first validated framework in the context of BE
  - Supported by the 'Product Quality Research Institute' (members: FDA/CDER, Health Canada, USP, AAPS, PhRMA…)
    - Three of BEBAC's protocols accepted by German BfArM, first product approved in 06/2011.

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

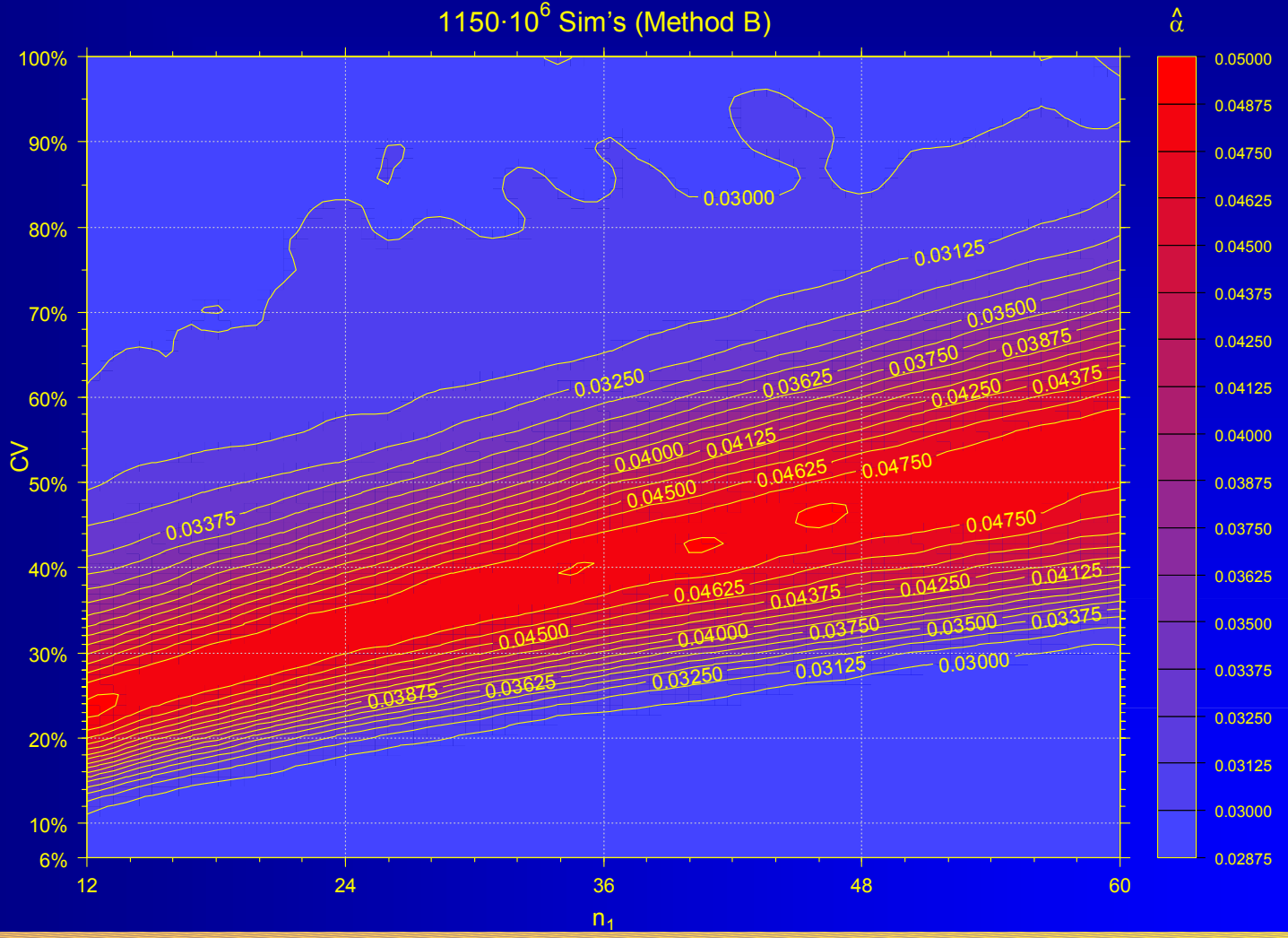
Review of Guidelines

- 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)

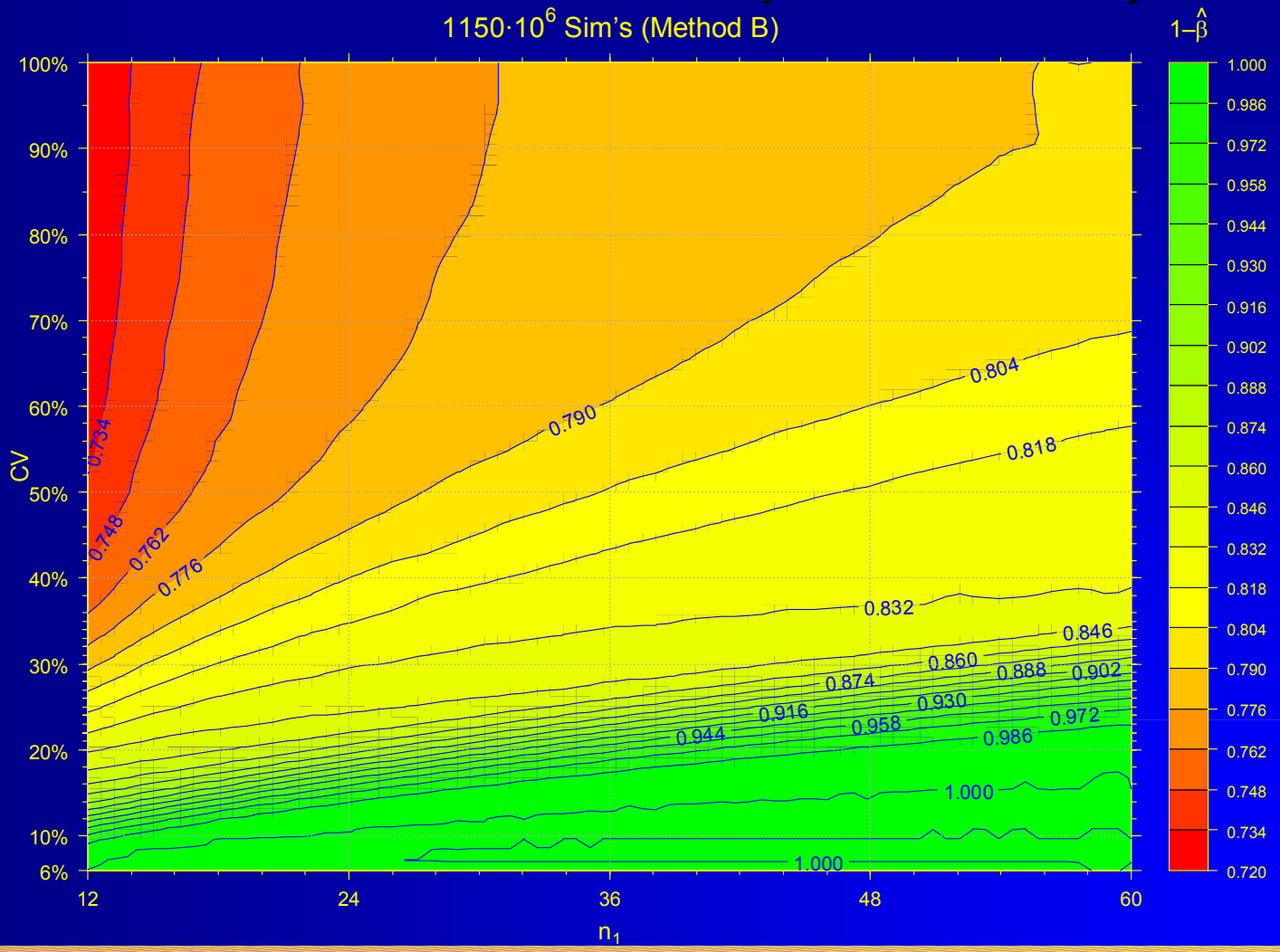
Potvin *et al.* (Method B)



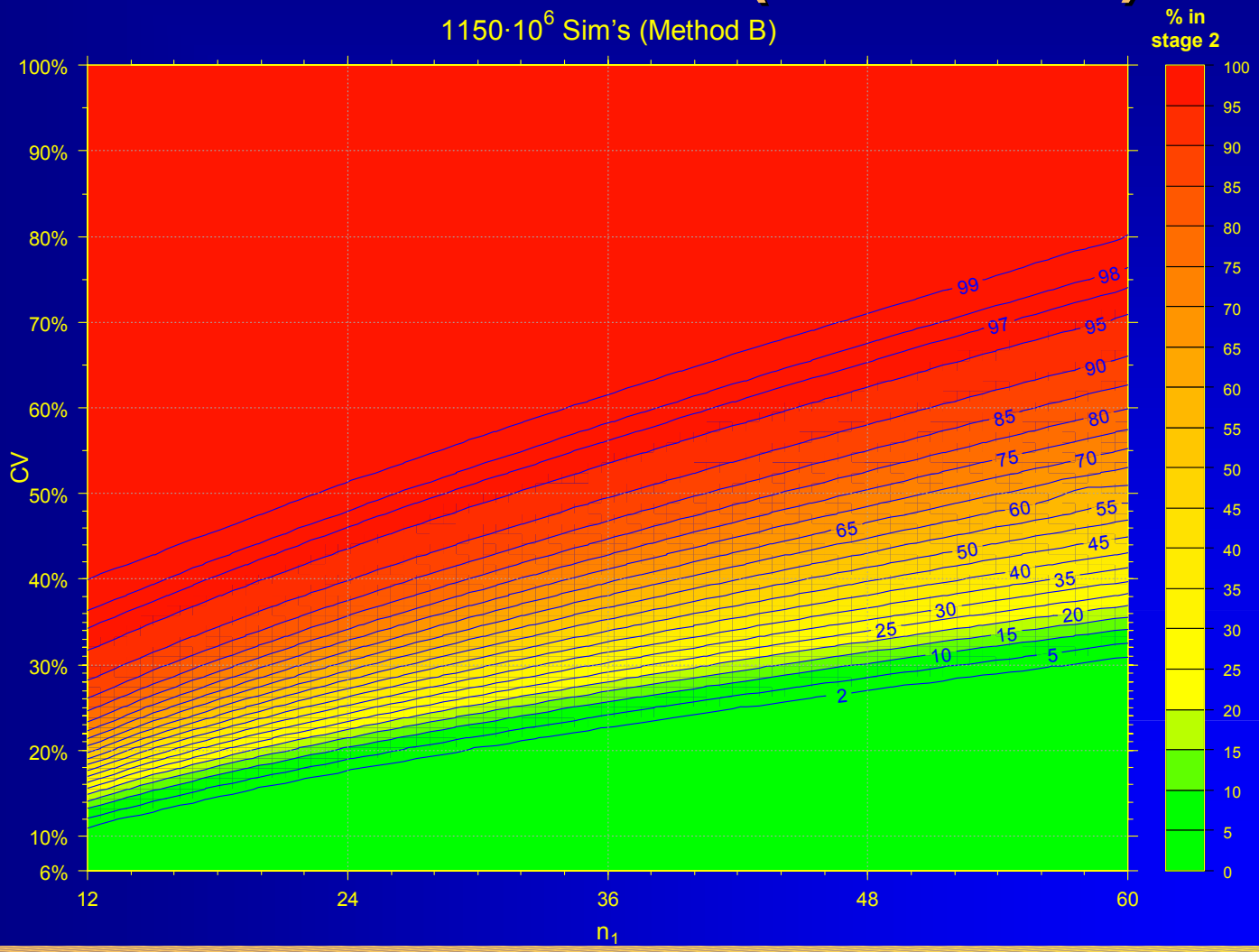
Potvin *et al.* (Method B)



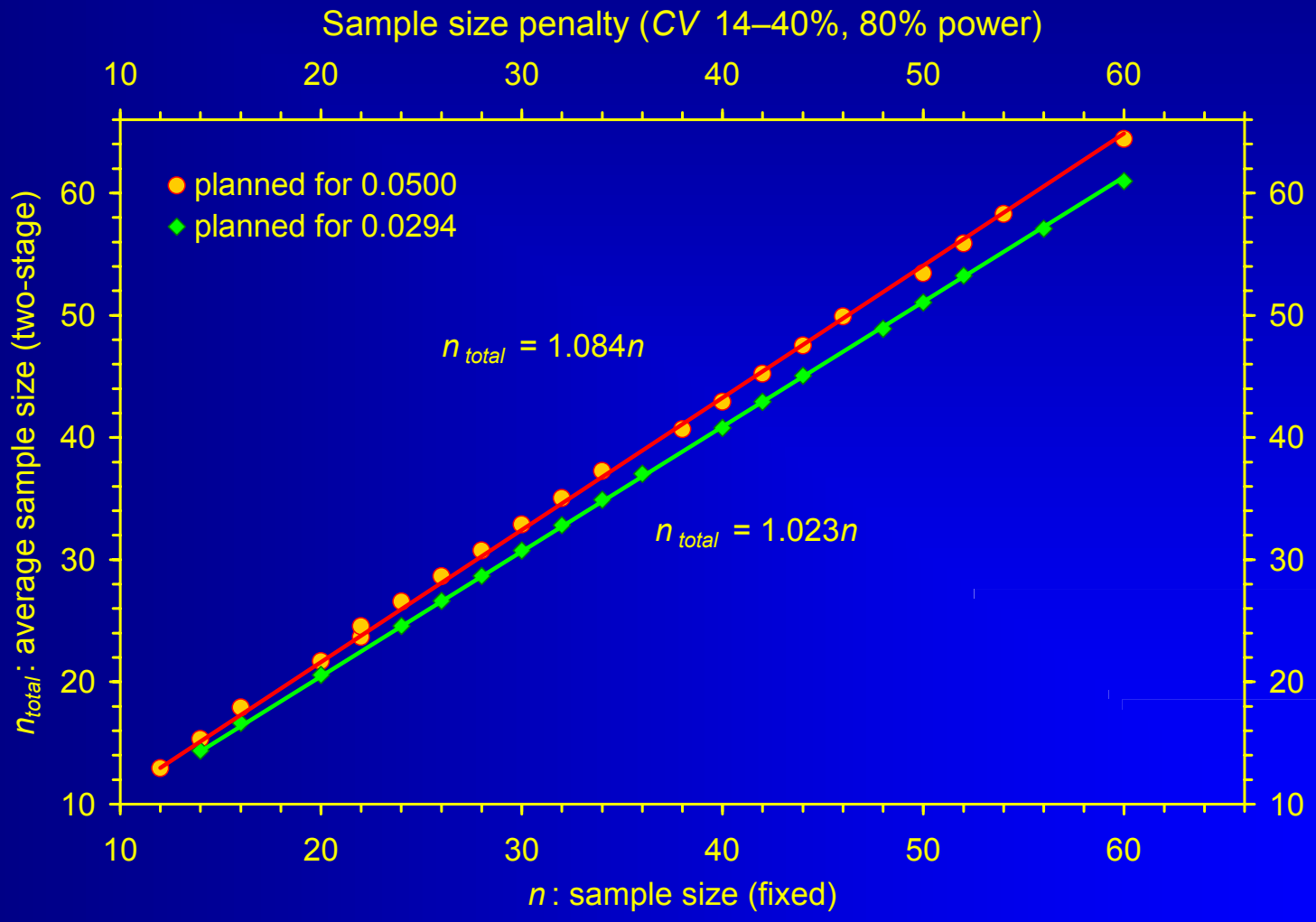
Potvin *et al.* (Method B)



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) + 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 central *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 centr. *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

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

Example (Potvin Method B)

```
library(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,
             theta0=0.95, CV=0.182132, design='2x2',
             method='exact')

++++++++++++ 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

Simulations (n_1 12, *CV* 18.2%)
- α_{emp} 0.042635
- power 85.3%

Total sample size 20: include another 8 in stage 2

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.
```

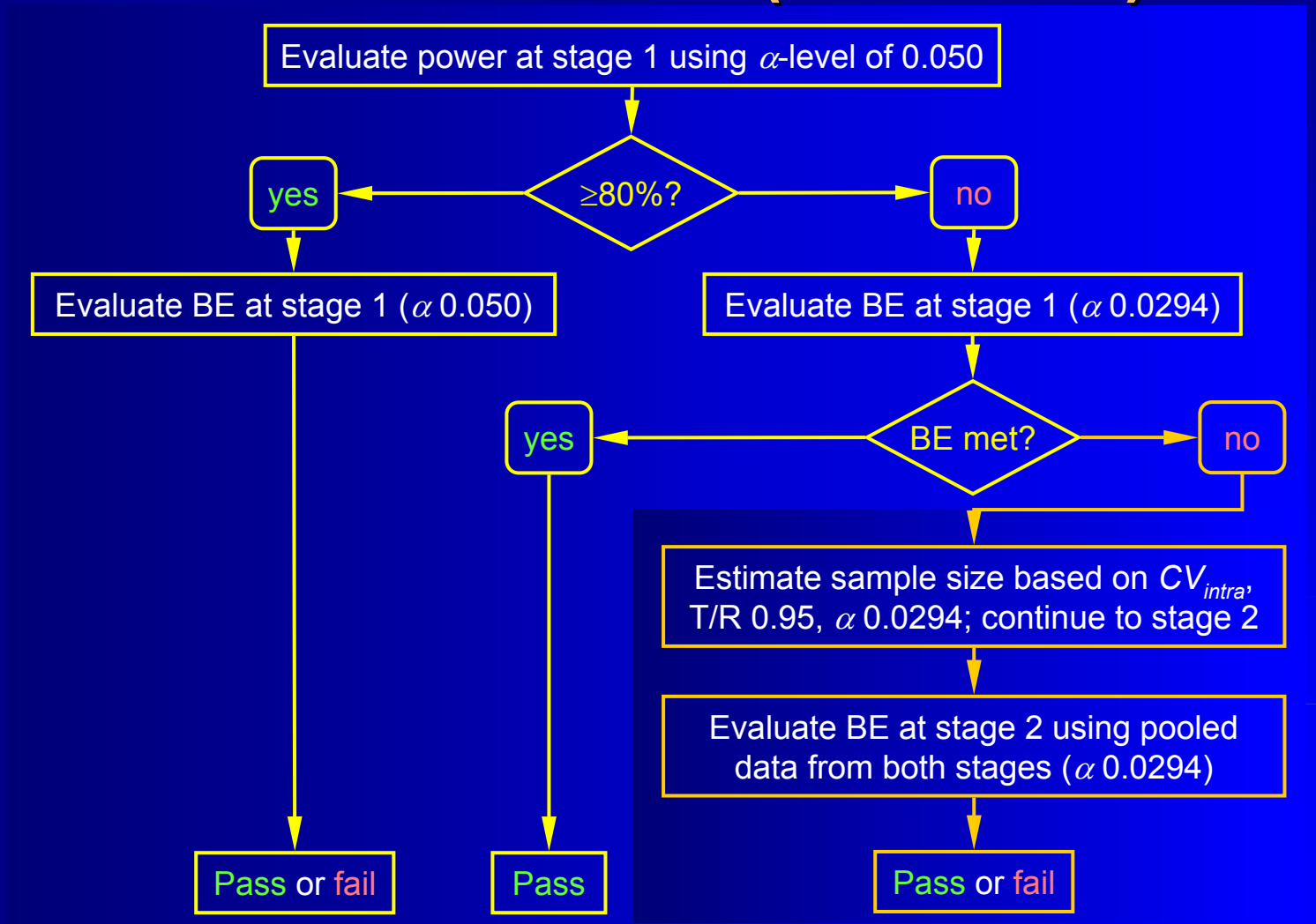
8 subjects in stage 2 (20 total), modified model in pooled analysis

Q&A Rev. 7 (March 2013)

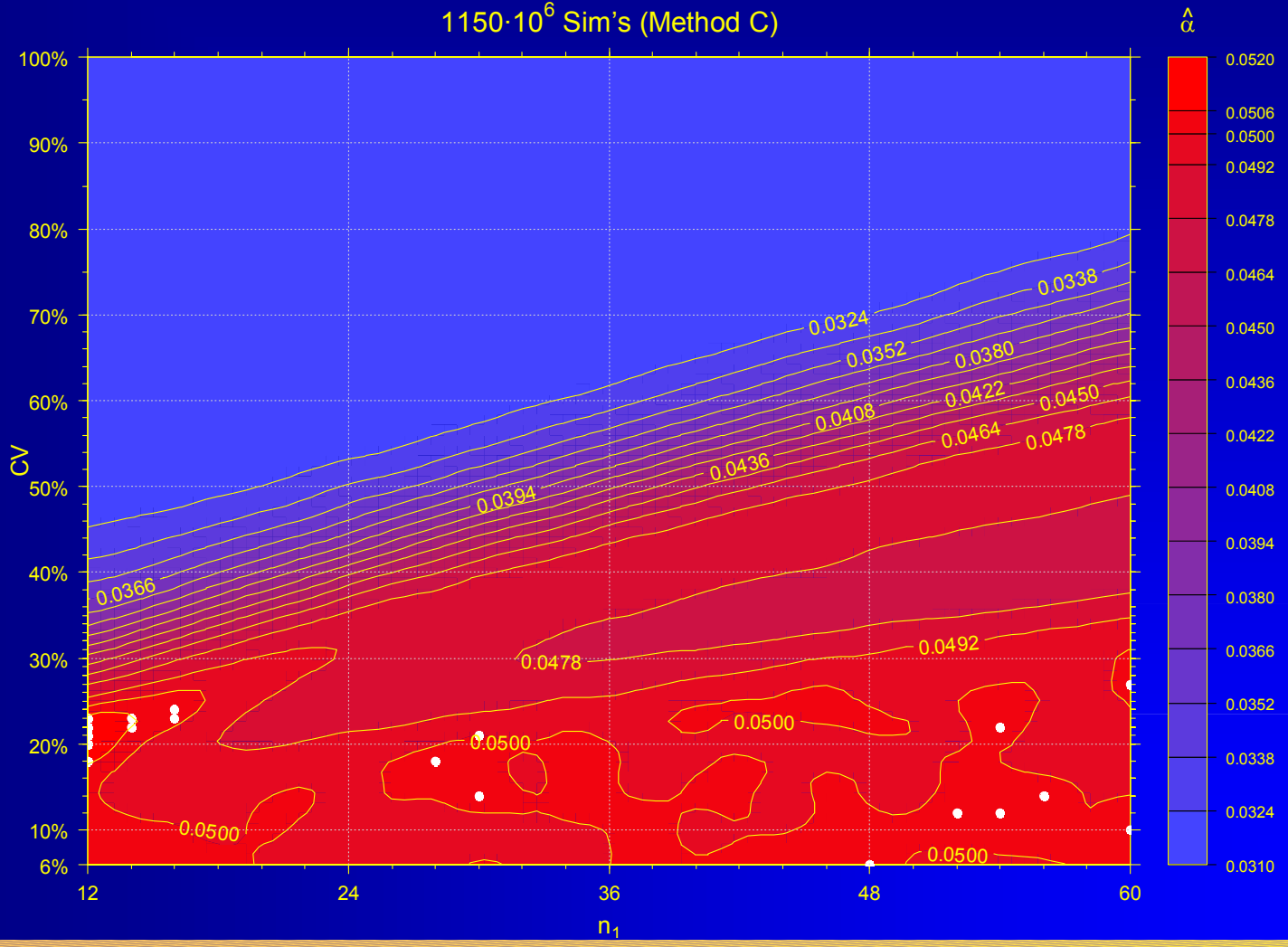
α 0.0294 in pooled analysis

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

Potvin *et al.* (Method C)



Potvin *et al.* (Method C)



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

- Pros & cons
  - Method C (*if power* ≥*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 sample 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, HPFB). 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* 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.

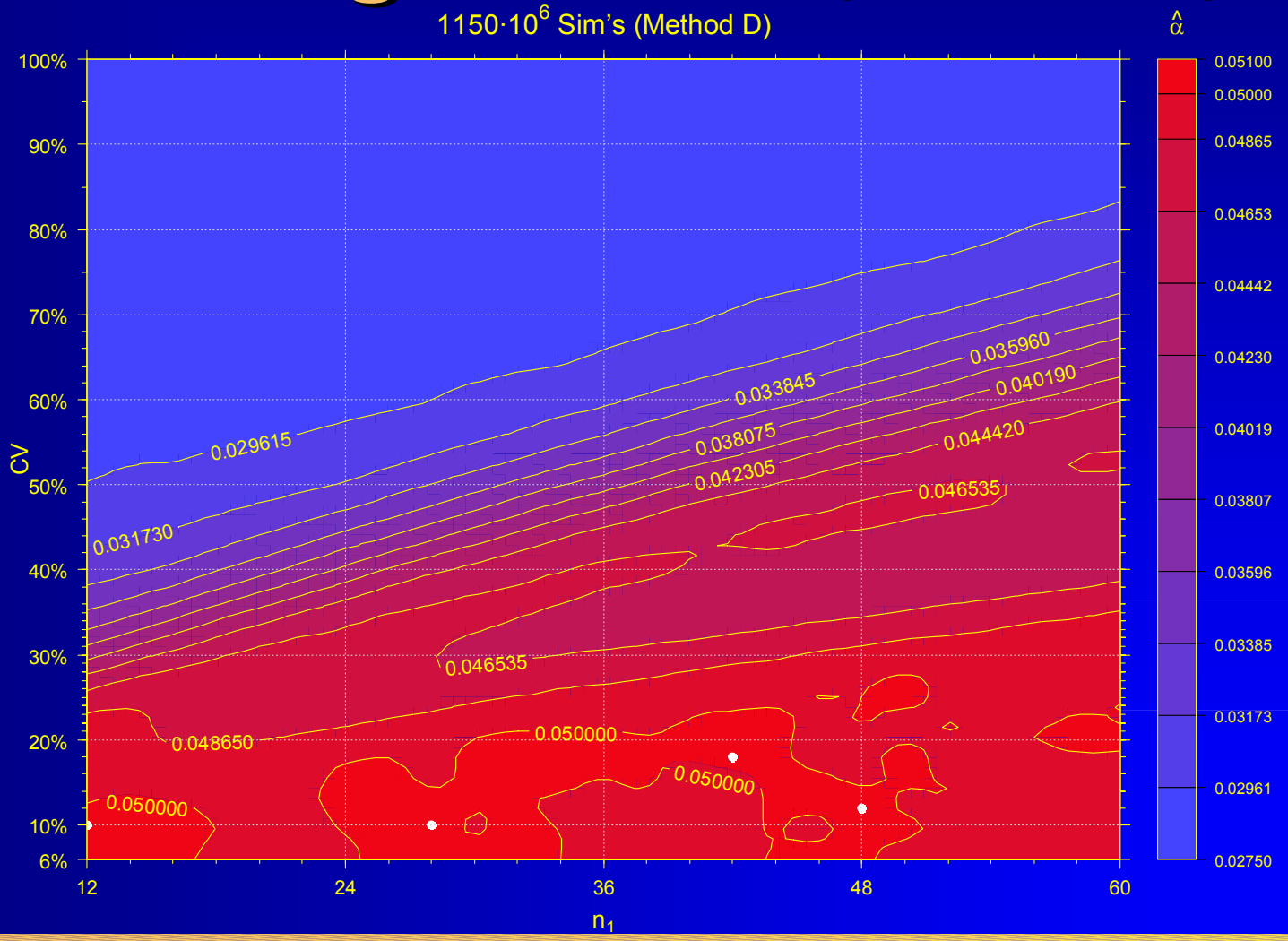
TSDs: Alternatives

- 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(3), 659–61 (2013) DOI: 10.1208/s12248-013-9475-5

Montague *et al.* (Method D)

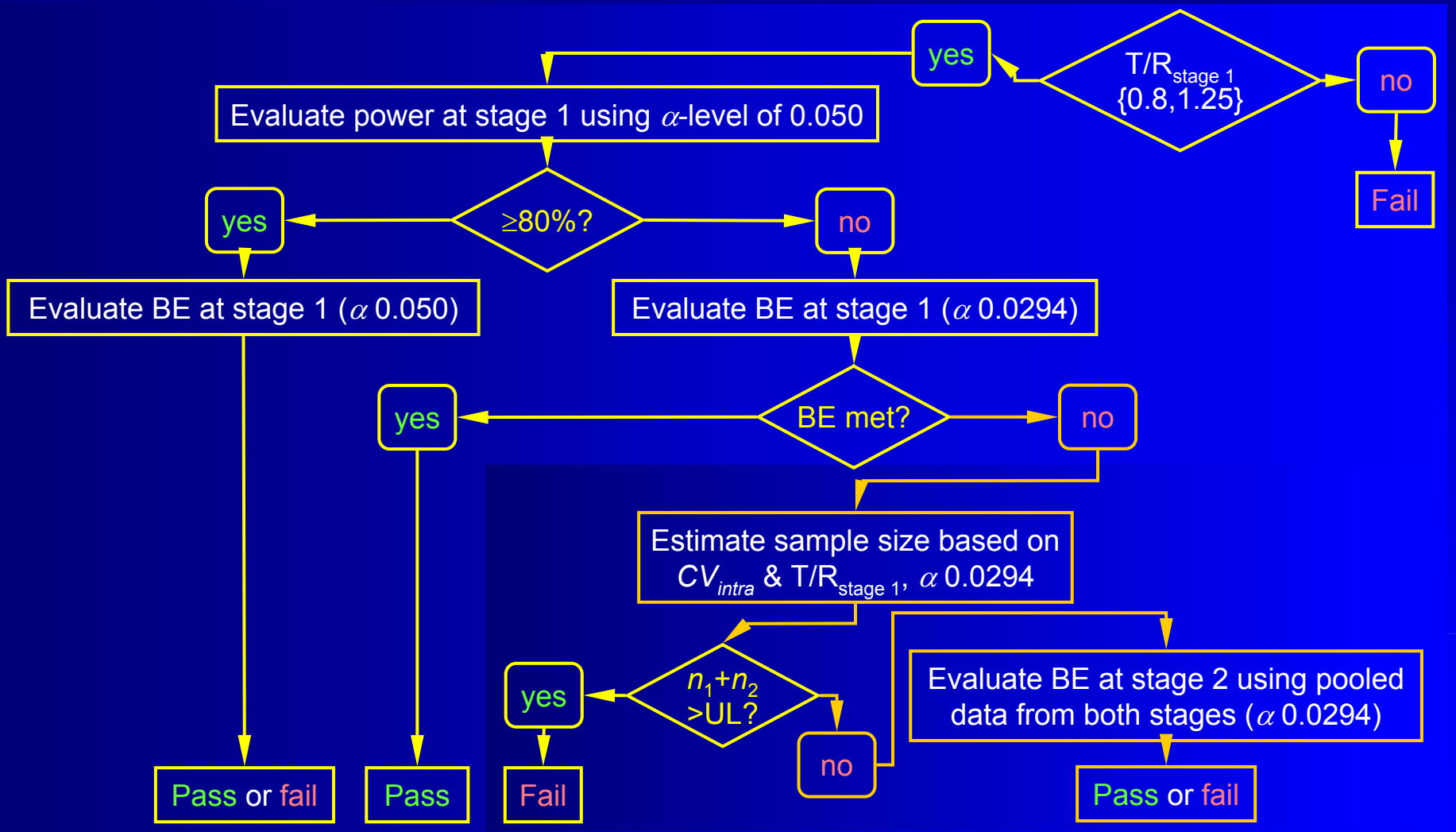


TSDs: Alternatives

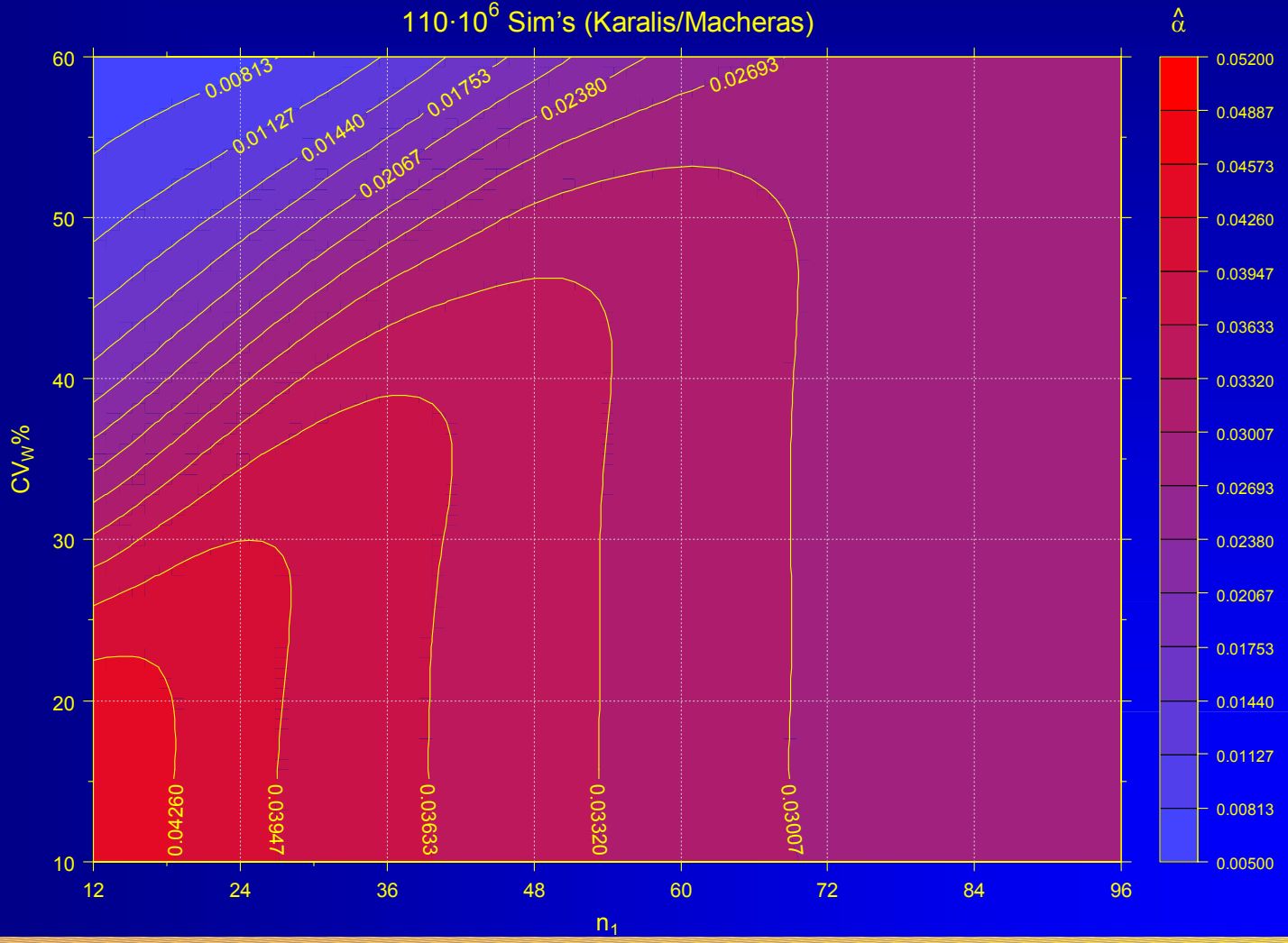
- Karalis & Macheras (2013)
  - Based on Method C ($\alpha_{adj.}$ 0.0294)
  - Sample size re-estimation based on *observed* T/R-ratio in stage 1
  - Upper sample size limit (UL)
  - Frameworks:
    - n_1 12–96, *CV* 10–60%, $n_1+n_2 \leq$ UL 150
    - n_1 18–96, *CV* 20–40%, $n_1+n_2 \leq$ UL 100

Karalis V and P Macheras
An Insight into the Properties of a Two-Stage Design in Bioequivalence Studies
Pharm Res 30(7), 1824–35 (2013), DOI: 10.1007/s11095-013-1026-3

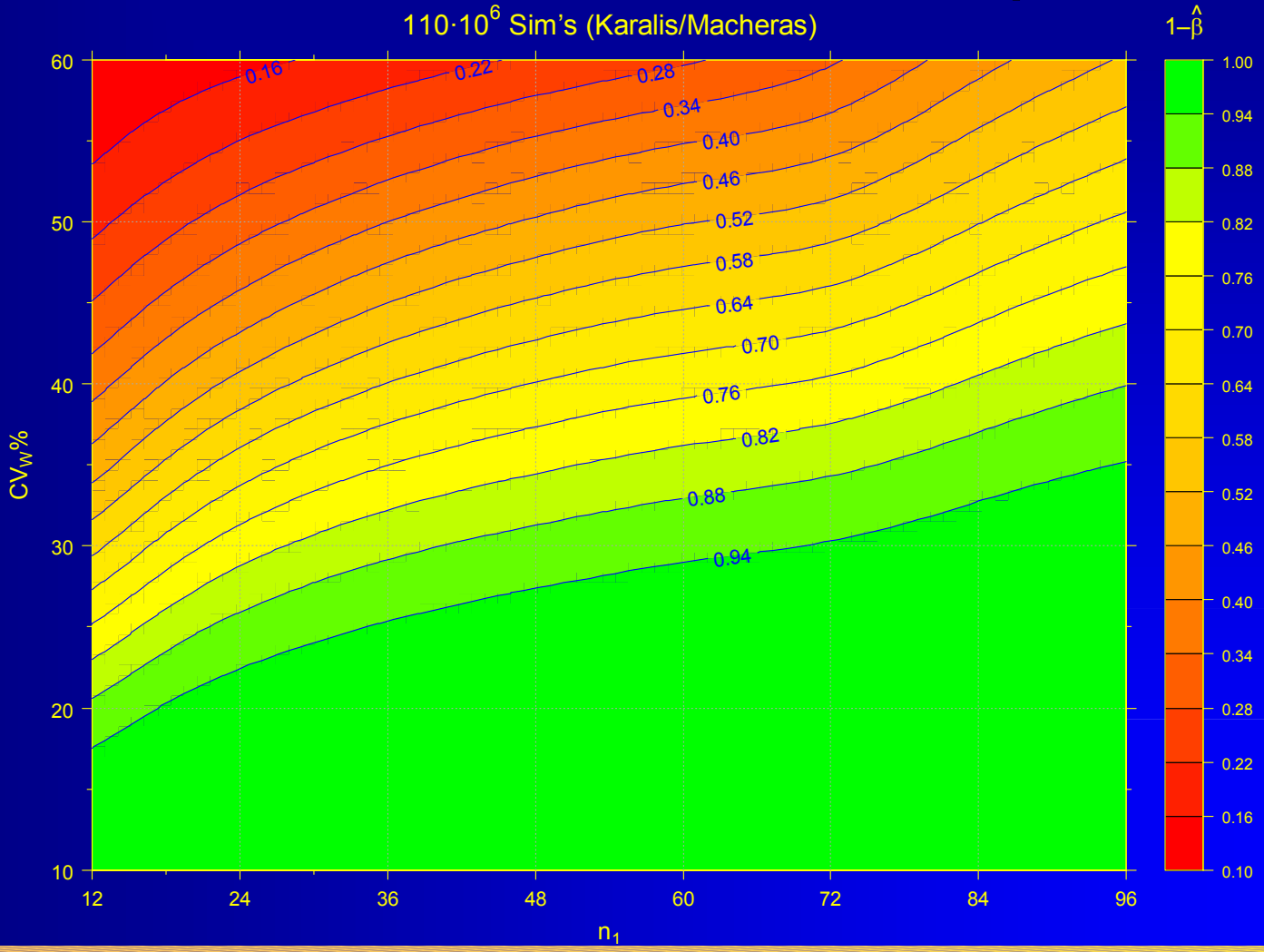
Karalis & Macheras



Karalis & Macheras (n ≤150)



Karalis & Macheras (n ≤150)



Karalis & Macheras (n ≤150)

```
library(PowerTOST)
power.TOST(alpha=0.05, theta0=1.0876,
           CV=0.182132, n=12, design='2x2',
           method='exact')

[1] 0.531698
```

α 0.05, observed T/R 108.76%, CV_{intra} 18.2%, 12 subjects in stage 1

Power 53.2% – initiate stage 2

```
sampleN.TOST(alpha=0.0294, targetpower=0.80,
             theta0=1.0876, CV=0.182132, design='2x2',
             method='exact')

++++++++++ 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

Simulations (n_1 12, *CV* 18.2%, UL 150)
- α_{emp} 0.049681
- power 89.1%

Total sample size 28 (≤150): include another 16 in stage 2

Karalis & Macheras (Expl. a)

- *CV* assumed as 20%, T/R 95%
  - In a fixed sample design for 80% power sample sizes would be 20 (α 0.05) or 24 (α 0.0294).
  - The sponsor chooses n_1 24 and UL 100.
  - 10^6 simulations (Potvin C), 10^5 (K/M)

method	(overall) power	power (stage 1)	% studies to stage 2	$n_{95\%}$
Potvin *et al.*	90.1	88.1	4.2	24
Karalis/Macheras	94.8	83.5	11.4	66

  - ~Three times as many studies forced to stage 2 with a high probability of large sample sizes.

Karalis & Macheras (Expl. b)

- *CV* assumed as 40%, T/R 95%
  - Fixed sample design n 66 (α 0.05) or 80 (α 0.0294).
  - The sponsor chooses n_1 60 and UL 150.
  - 10^6 simulations (Potvin C), 10^5 (K/M)

method	(overall) power	power (stage 1)	% studies to stage 2	$n_{95\%}$
Potvin *et al.*	83.6	69.7	23.8	98
Karalis/Macheras	74.2	67.2	7.2	130

  - Power <80%; only ~⅓ of studies proceed to stage 2, although with considerably larger sample sizes.

Labes D and H Schütz
An Insight into the Properties of a Two-Stage Design in Bioequivalence Studies: A Rejoinder
Pharm Res (submitted September 2013)

Futility Rules revised

- EMA GL Section 4.1.8 ‘Two-stage design’
  “[…] the stopping criteria should be clearly defined prior to the study.”
  - What does that mean?
    - Failing in stage 1 or the pooled analysis according to the chosen method.
      → Part of the validated frameworks.
    - Early stopping for futility (*e.g.*, ‘bad’ ratio, extreme stage 2 sample size caused by high *CV* – better to opt for reference-scaling…).
      → Not validated. A misunderstanding by regulators (stopping criterion ≠ futility rule).

Futility Rules revised

- Introduction of a futility rule does not inflate the patient's risk, but power may drop substantially!
  - State unambiguously in the protocol what the stopping criteria are.
  - If you want to introduce a futility rule, simulations are required in order to maintain sufficient power.

> *"Introduction of […] futility rules may severely impact power in trials with sequential designs and under some circumstances such trials might be unethical."*

A Fuglsang
Futility Rules in Bioequivalence Trials with Sequential Designs
APPS J (accepted October 2013)

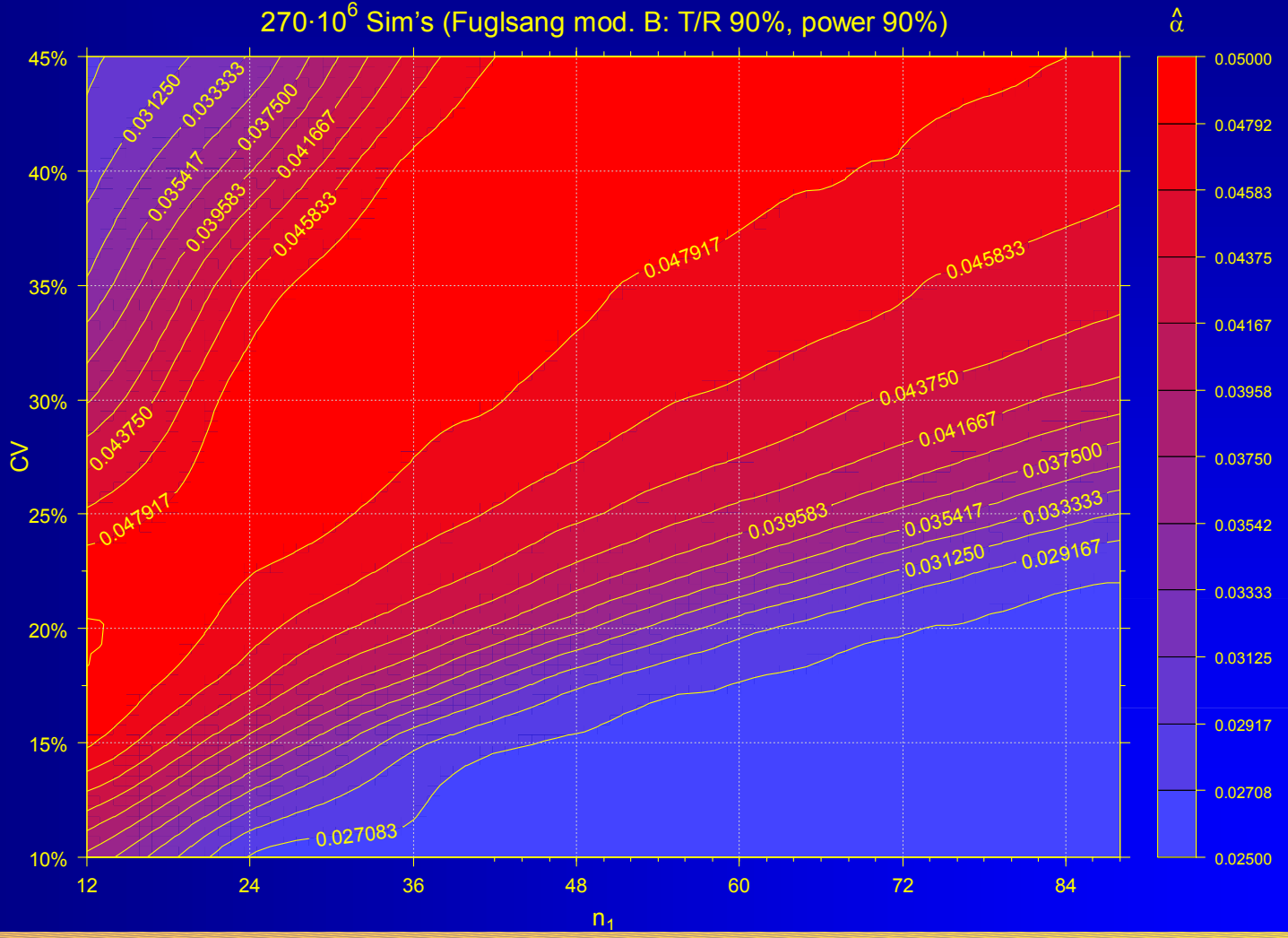
Adventurous Stuff 1

- ‘Must pass’ BE in stage 1 (first to file)
  - Fixed T/R 90% (pessimistic; very likely better)
  - Expected *CV* 20% (pilot study with two references)
  - ~30% drop-out rate; start with 88 to have $n_1 \geq 60$
  - Targets
    - >90% power for n_1 60 – even for extreme *CV* of 45%
    - 90% power for $n_1 \geq 60$ (*CV* 20%) in stage 1
    - Not <80% power for *CV* ≥25% in stage 1
    - Low probability to proceed to stage 2

Adventurous Stuff 1

- 'Must pass' BE in stage 1 (first to file)
  - EMA submission; sponsor wants Method B
  - Fuglsang published $\alpha_{adj.}$ 0.0269 for T/R 0.90 and 90% power – but only for Method C…
  - Same $\alpha_{adj.}$ applicable?
  - Likely…
    - Potvin *et al.* showed less inflation with Method B.
    - Fuglsang needed less adjustment in Method B.
    - But we have to justify that!
  - 10^6 sim's for α and 10^5 for power

Adventurous Stuff 1



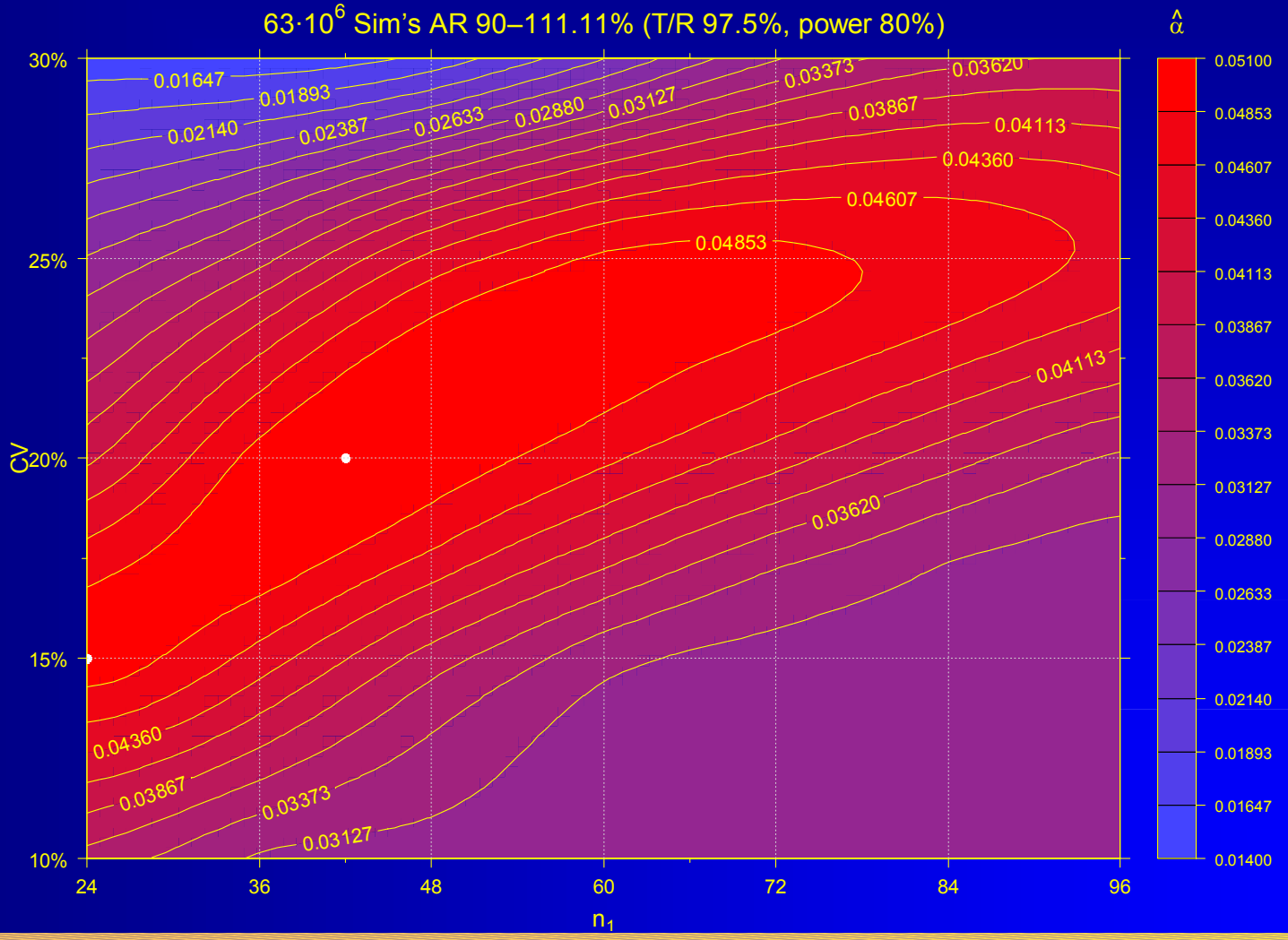
Adventurous Stuff 1

- ‘Must pass’ BE in stage 1 (first to file)
  - Targets met
    - 93% power for n_1 60 (*CV* 20%) and 90% for extreme *CV* of 45%
    - 90% power for n_1 ≥60 (*CV* 20%) in stage 1
    - Low chances to proceed to stage 2 with *CV* 20%: n_1 60: 6%, n_1 72: 1%
    - ≥80% power for *CV* ≥20%, even for a more extreme drop-out rate
    - $\alpha_{adj.}$ 0.0271 would work as well (with 0.0278 <0.052)
  - Study started in September 2013

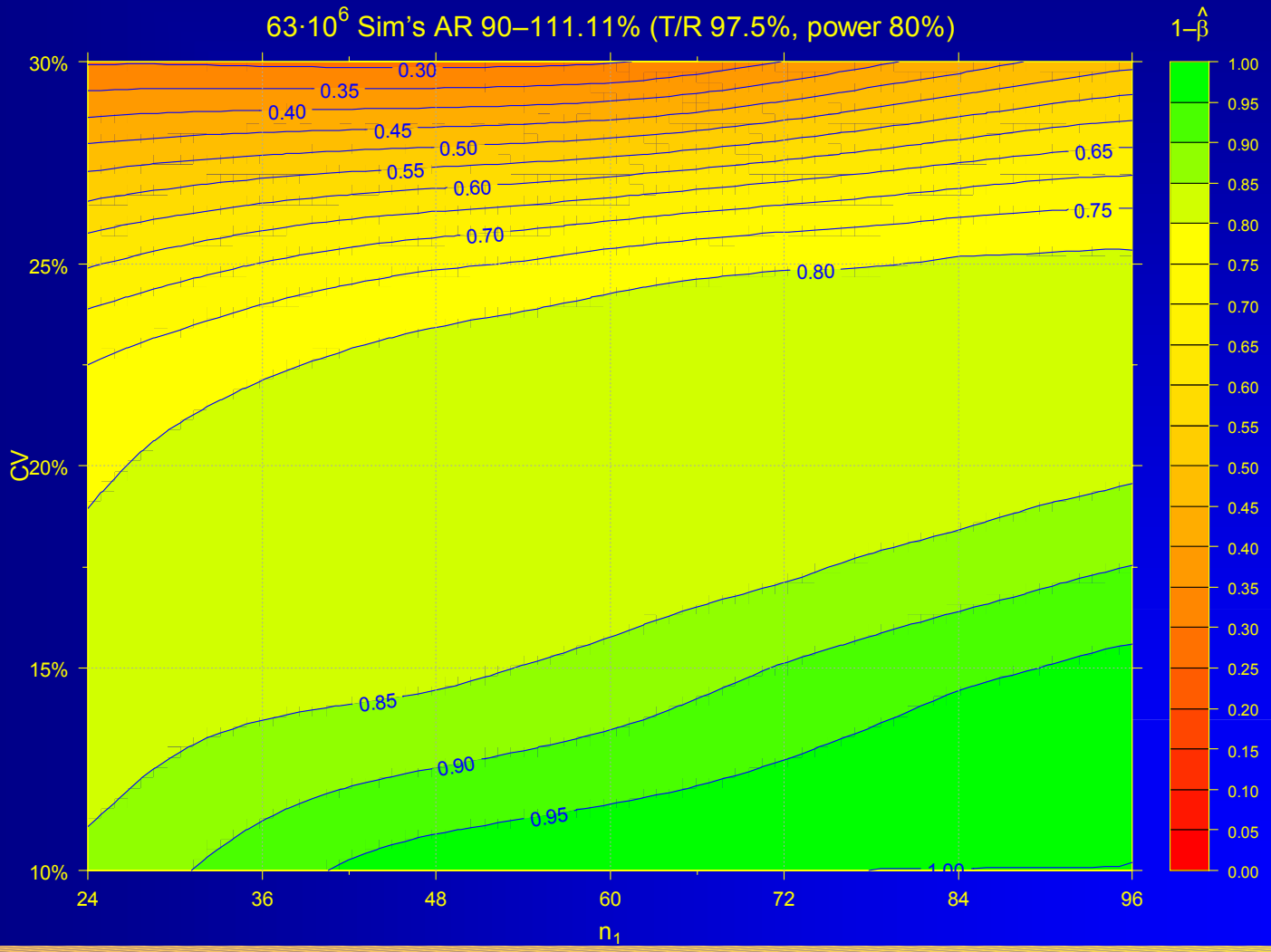
Adventurous Stuff 2

- NTID (EMA AR 90.00 – 111.11%)
  - Fixed T/R 97.5% (tighter; similar to FDA)
  - Expected *CV* <18%
  - Upper sample size limit (n_{max}) based on $2\times n_{fixed}$ for 'pessimistic' *CV* of 20% and 80% power (188)
  - Targets
    - >80% power for n_1 72 (*CV* 18%) and ~80% (*CV* 20%)
    - High chance to show BE already in stage 1
    - Not less than 75% power for *CV* 25%
  - Based on Method B ($\alpha_{adj.}$ 0.0306)
    10^6 sim's for α and power

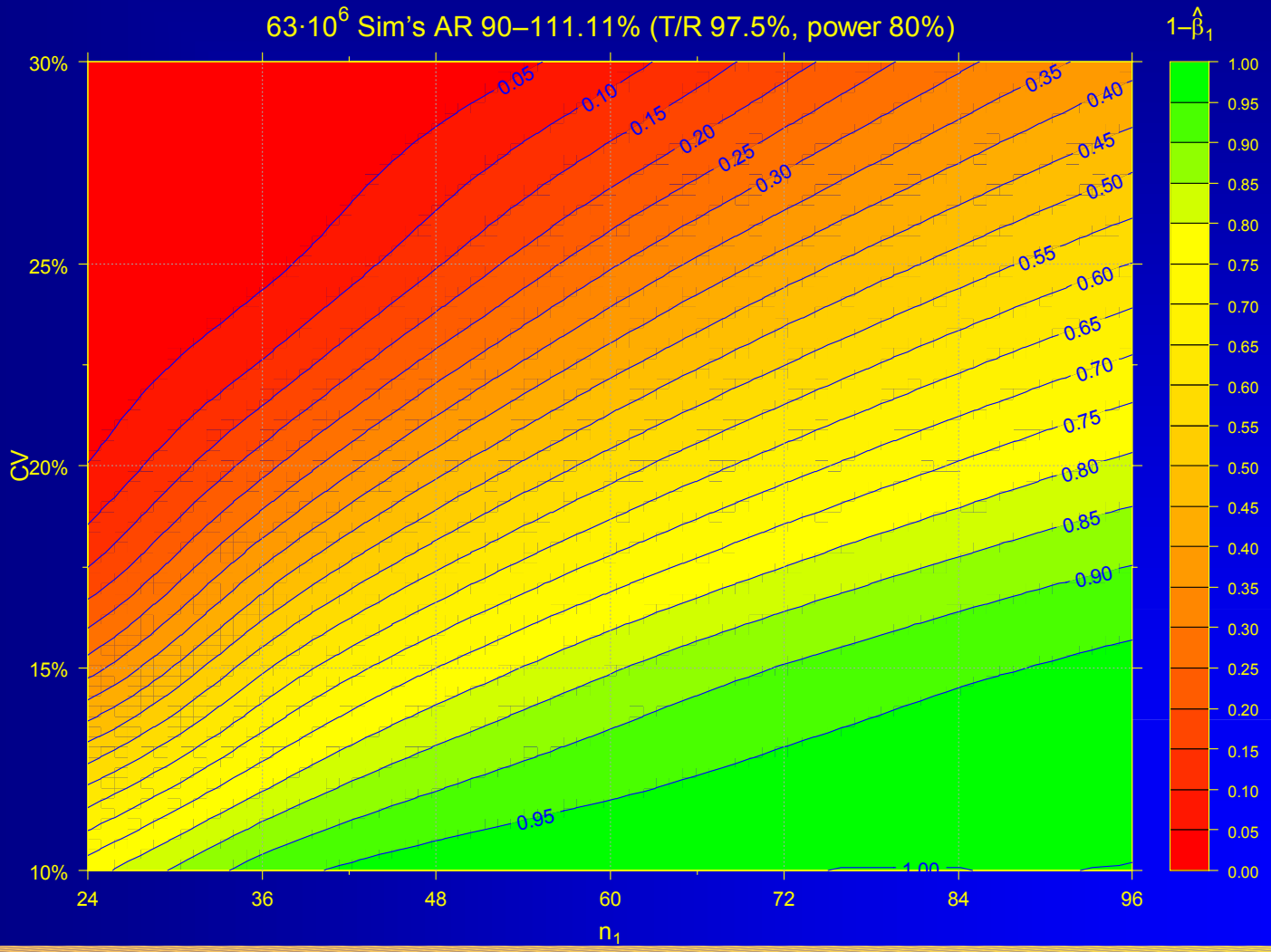
Adventurous Stuff 2



Adventurous Stuff 2



Adventurous Stuff 2



Adventurous Stuff 2

- NTID (EMA AR 90.00–111.11%)
  - Mission *theoretically* accomplished
  - No relevant α inflation ($<$0.051) within n_1 24–96 and *CV* 10–30%
  - Targets met
    - n_1 72 (*CV* 17.5%) 84% power (at n_1 48 still 83%) 83% power (*CV* 20%)
    - 80% power to show BE already in stage 1
    - 80% power for *CV* 25%
  - Sponsor wasn't sure about the ratio (really ±2.5%?) – decided to run a large (!) pilot study.

Case Study 1 (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 as well (*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 Study 2 (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
    - Sponsor repeated the study with a very (!) large sample size and failed on C_{max}. Project cancelled.

Case Study 3 (EMA)

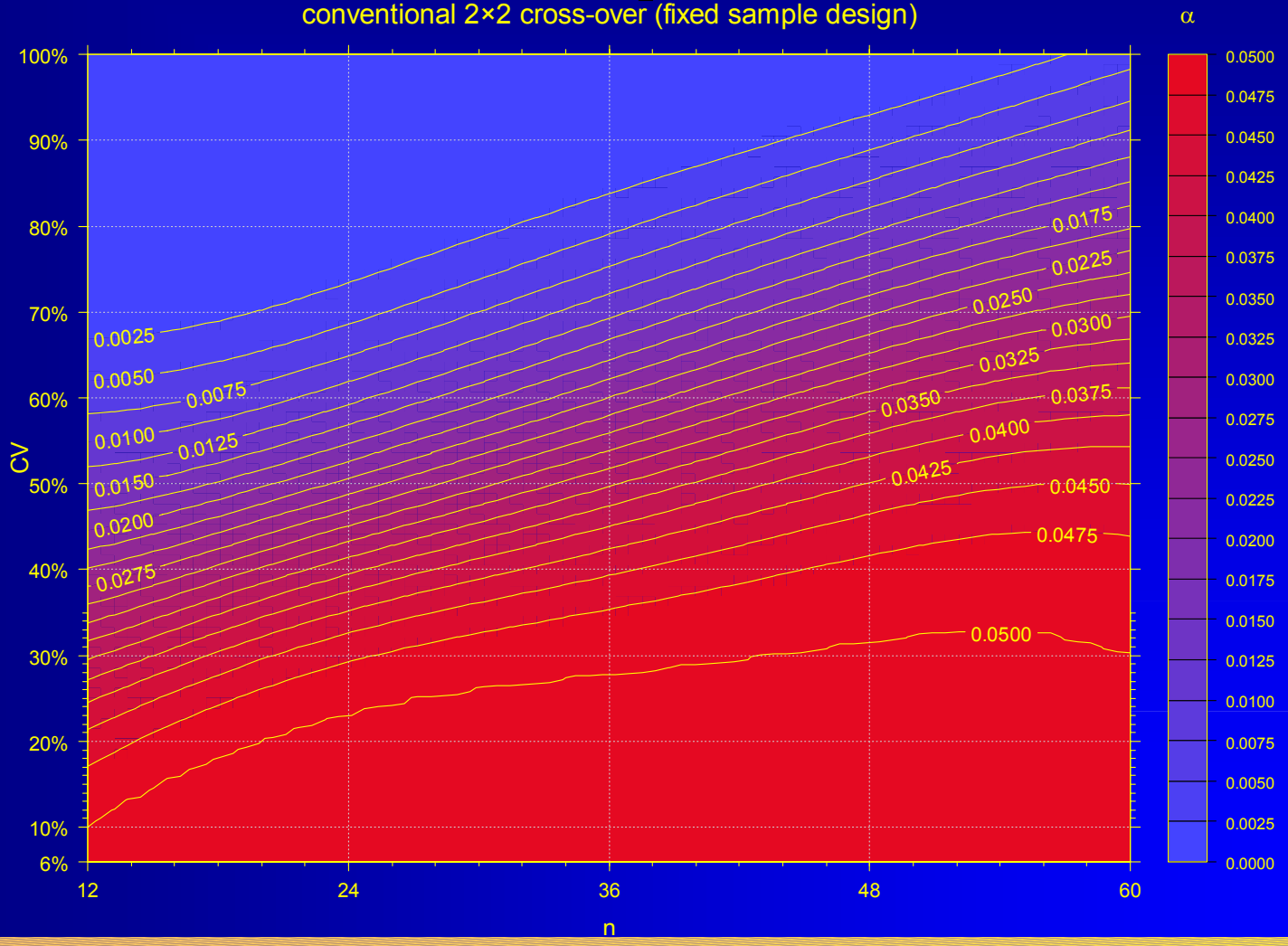
- Method C: Two studies passed in stage 1 (n=15 SD, n=16 MD, C_{max} *CV* 17.93%, 8.54%, 90% CIs)
- Would have passed with Method B as well; however, 94.12% CIs were not reported.
  - RMS Germany. Accepted by CMSs Austria, Denmark, Sweden, and The Netherlands.
  - Spain: Statistical analysis should be GLM. Please justify.
    - Evaluated with all-fixed effects model.
      Both studies passed.
      Issue resolved (September 2013)

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.
- Full adaptive methods.
- Exact method (not depending on simulations).
- Application to replicate designs / scaling.

Don’t panic!

conventional 2×2 cross-over (fixed sample design)



Thank You!

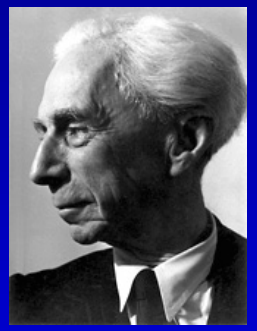
Two-Stage Designs in BE Studies

Open Questions?

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

To bear in Remembrance...

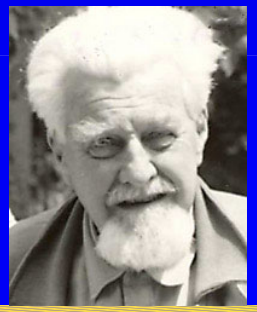
The fundamental cause of trouble in the world today is that the stupid are cocksure
while the intelligent are full of doubt. *Bertrand Russell*



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 (2007–2013):
      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(1), 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 1.1-07 (2013-09-02)
  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
  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) 659–61 (2013)
  DOI: 10.1208/s12248-013-9475-5
- Karalis V and P Macheras
  An Insight into the Properties of a Two-Stage Design in Bio-equivalence Studies
  Pharm Res 30(7), 1824–35 (2013)
  DOI: 10.1007/s11095-013-1026-3