



Wikimedia Commons 2010 • EmDee • Creative Commons Attribution-ShareAlike 3.0 Unported

Overview

- History / early approaches
 - Add-on studies
 - Problems with α -inflation
- Uncertain CV ...
- Recent developments
 - Review of guidelines
 - Two-stage sequential designs

Overview

- Open issues
 - Feasibility / futility rules
 - Arbitrary PE and/or power; adaption for stage 1 PE
 - Dropping a candidate formulation from a higher-order X-over
 - Application to replicated designs (for HVDs/HVDPs)

History / early approaches

- Sometimes properly planned studies fail due to
 - Pure chance (producer's risk hit)
 - False assumptions about variability and/or T/R-ratio
 - Poor study conduct (increasing variability)
 - 'True' bioinequivalence
- The patient's risk must be preserved
 - Already noticed at Bio-International Conferences (1989, 1992) and guidelines from the 1990s

History / early approaches

- *'The primary concern in bioequivalence assessment is to limit the risk of erroneously accepting bioequivalence. Only statistical procedures which do not exceed the nominal risk of 5% can be approved, and among them the one with the smallest risk of erroneously rejecting bioequivalence should be selected.'**
- Performing a second study and pooling data with the first's not acceptable
- Performing a (much larger) second study and base BE on this study *alone* was (and is) acceptable

* **CPMP Working Party**
Investigation of Bioavailability and Bioequivalence: Note for Guidance
Section 3.6 Data analysis, Document Ref. III/54/89-EN (1 May 1992)

History / early approaches

- However, naïve pooling (*without* α -adjustment) was performed in the past
 - Statistical model modified in order to include a formulation-by-study interaction factor.
 - Test for homogeneity of error variances between studies
 - Pooling only acceptable if both tests not significant*

* **H Mellander**

Problems and Possibilities with the Add-On Subject Design, in:

Midha KK, Blume HH (eds.)

Bio-International. Bioavailability, Bioequivalence and Pharmacokinetics
medpharm Scientific Publishers, Stuttgart, pp. 85–90 (1993)

Add-on Designs

- Example (acc. to Canada's 1992+ guidances)
 - Second part in at least 12 subjects
Pooling only allowed if both of two consistency tests not significant ($p > 0.05$)
 - Equality of residual mean squares (F -test) of the two parts. Smaller MSE must be used as the denominator.
Example:
0.01321 (1st part: $n=55$, df 53)
0.01718 (2nd part: $n=14$, df 12)

Add-on Designs

Sum of Squares					
Hypothesis	DF	SSE	MSE	F_stat	P_value
Sequence	1	0.0271658	0.0271658	0.173639	0.6786
Sequence*Subject	53	8.29185	0.15645	11.844	<0.0001
Treatment	1	0.211196	0.211196	15.9885	0.0002
Period	1	0.0271536	0.0271536	2.05565	0.1575
Error	53	0.700088	0.0132092		

Sum of Squares					
Hypothesis	DF	SSE	MSE	F_stat	P_value
Sequence	1	0.0489527	0.0489527	0.419204	0.5295
Sequence*Subject	12	1.4013	0.116775	6.79641	0.0012
Treatment	1	0.0349142	0.0349142	2.03203	0.1795
Period	1	0.0839476	0.0839476	4.88581	0.0472
Error	12	0.206183	0.0171819		

$$\hat{F} = \frac{MSE_{large}}{MSE_{small}} = \frac{0.0171819}{0.0132092} = 1.30075$$

$$F_{1-0.05,12,53} = 1.940$$

$$p(\hat{F}) = 0.24595 > 0.05 \quad \checkmark$$

Add-on Designs

- Example (Canada cont'd)
 - Second part in at least 12 subjects. Pooling is only allowed if two consistency tests not significant ($p > 0.05$):
 - Since first test not significant ($p = 0.246$), pool studies
 - Now test for study-by-formulation interaction

Add-on Designs

Tests of Model Effects

Hypothesis	Numer_DF	Denom_DF	F_stat	P_value
int	1	56.5	2144.16	<0.0001
Study	1	56.5	0.0007	0.9784
Treatment	1	64.6	9.9949	0.0024
Treatment*Study	1	64.6	0.1156	0.7349



Bioequivalence statistics

$$F_{1-0.05,1,64.6} = 3.989$$

User-Specified Confidence Level for CI's = 95.0000
 Percent of Reference to Detect for 2-1 Tests = 20.0%
 A.H.Lower = 0.800 A.H.Upper = 1.250

$$p(0.1156) = 0.7349 > 0.05$$

Formulation variable: Treatment

Reference: R	LSMean=	6.088010	SE=	0.132921	GeoLSM=	440.543718
Test: T	LSMean=	6.167145	SE=	0.132921	GeoLSM=	476.822902

Difference = 0.0791, Diff_SE= 0.0250, df= 64.6
 Ratio(%Ref) = 108.2351

CI 90% = (103.8061, 112.8531)
 CI 95% = (102.9556, 113.7853)



Average bioequivalence shown for confidence=95.00 and percent=20.0.

Add-on Designs

- Example (Canada cont'd)
 - Formulation-by-study interaction not significant (p 0.7349), pooled analysis acceptable
 - No α -adjustment mentioned in 1992 guideline, but recommended in 2010 draft (Bonferroni: 95% CI)
 - 2010 draft allows also for group sequential designs
 - Group sequential designs allow better control of patient's risk

Problems with α -inflation

- Patient's risk likely is not preserved
 - The probability to obtain at least one significant result with k independent (!) t -tests (at level α) is

$$P(k) = 1 - (1 - \alpha)^k$$

$$P(2) = 1 - (1 - 0.05)^2 = 0.0975$$

- Bonferroni-correction of two studies would mandate calculation of a 95% confidence interval

$$\alpha_{adj} = \alpha/k$$

$$P_{adj}(2) = 1 - (1 - 0.025)^2 = 0.04938 < 0.05$$

- Applicability doubtful since no *independent* tests!

Problems with α -inflation

- Patient's risk (cont'd)
 - For two repeated tests on accumulating data the overall level is $\sim 8\%$ ¹
 - In naïve pooling the variance will be underestimated²
 - Simulations of BE studies (sample sizes 24 – 48, CV_{intra} 19 – 37%, 1 – 3 interim looks) showed empirical α of up to 5.97%³

¹ **Armitage P, McPherson K, and BC Rowe**

Repeated significance tests on accumulating data

J R Statist Soc A 132, 235–44 (1969)

² **Wittes J, Schabenberger O, Zucker D, Brittain E, and M Proschan**

Internal pilot studies I: type I error rate of the naïve t-test

Statistics in Medicine 18, 3481–91 (1999)

³ **Hauck WW, Preston PE, and FY Bois**

A group sequential approach to crossover trials for average bioequivalence

Journal of Biopharmaceutical Statistics 7(1), 87–96 (1997)

Problems with α -inflation

- Patient's risk (cont'd)
 - Simulations of 1 Mio BE studies (12 subjects in 1st study, CV_{intra} 20%, sample size re-estimation based on PE 0.95 and CV_{intra} of 1st study) showed empirical α of 5.84%¹
 - With two repeated tests at 2.94% overall $\alpha \sim 5\%$ ²
 - Derived for tests assuming normally distributed data with known variances. Approximately valid if sample size not too small.

¹ **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](https://doi.org/10.1002/pst.294)
<http://www3.interscience.wiley.com/cgi-bin/abstract/115805765/ABSTRACT>

² **SJ Pocock**
Group sequential methods in the design and analysis of clinical trials
Biometrika 64, 191–9 (1977)

Uncertain CV ...

- CV_{intra} used in sample size estimation is *not set in stone* but an *estimate*!

- Sample sizes for $1-\beta$ 90%, PE 0.95, CV_{intra} 20% → $n=26$

- **Not done yet!**
What if $CV_{intra} \neq 20\%$?

CV_{intra}	n	power _n	power _{n=26}
15	16	0.92602	0.99153
16	18	0.92685	0.98379
17	20	0.92601	0.97253
18	22	0.92400	0.95763
19	24	0.92114	0.93922
20	26	0.91763	0.91763
21	28	0.91362	0.89329
22	30	0.90919	0.86659
23	32	0.90443	0.83794
24	36	0.91451	0.80767
25	38	0.90889	0.77606



Recent developments

- Review of guidelines
 - New Zealand (Oct 2001)
 - Sequential Designs
 - Declared in the protocol
 - Maximum sample size *a priori* ($\leq 40!$)
 - 'Appropriate statistical tests (e.g., sequential *t*-test)'
 - FDA
 - Sequential Designs: not mentioned in guidances but acceptable (pers. comm. Barbara Davit, Ljubljana, May 2010)
 - EMA (Jan 2010)
 - Sequential Designs: fairly detailed informations given

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.

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
 - stated in Canadian draft guidance (2010) and EMA's BE guideline (2010).

AL Gould

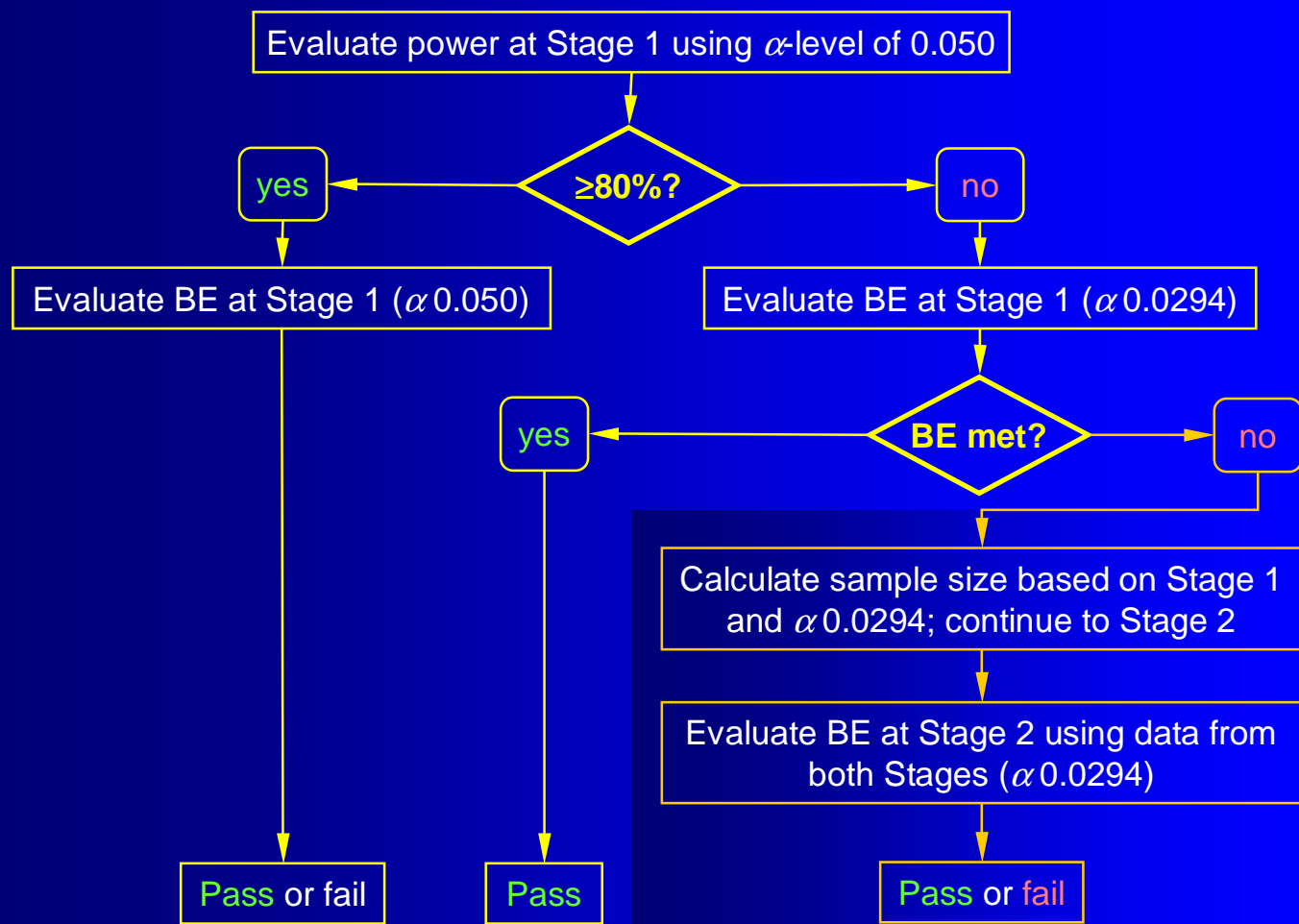
Group Sequential Extension of a Standard Bioequivalence Testing Procedure
J Pharmacokin Biopharm 23/1, 57–86 (1995)

Sequential Designs

- Methods by Potvin *et al.* (2008) promising
 - Supported by 'The Product Quality Research Institute' (members: FDA/CDER, Health Canada, USP, AAPS, PhRMA, ...)
 - Acceptable by US-FDA
 - Canada? Or Gould (1995) mandatory?
 - Acceptable as a Two-Stage Design in the EU
 - Three of BEBAC's protocols already approved by German BfArM

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](https://doi.org/10.1002/pst.294)
<http://www3.interscience.wiley.com/cgi-bin/abstract/115805765/ABSTRACT>

Potvin *et al.* (Method C)



Potvin *et al.* (Method C)

● Technical Aspects

- Only *one* Interim Analysis (after Stage 1)
- If possible, use software (too wide step sizes in Diletti's tables), preferable the exact method (avoid approximations)
- Should be termed 'Power Analysis' *not* 'Bioequivalence Assessment' in the protocol
- No *a-posteriori* Power – only a validated method in the decision tree
- No adjustment for the PE observed in Stage 1

Potvin *et al.* (Method C)

- Technical Aspects (cont'd)
 - No stop criterion (*'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)
 - If power <80% in Stage 1 or in the pooled analysis (data from Stages 1 + 2), Pocock's α 0.0294 is used (*i.e.*, the $1 - 2 \times \alpha = 94.12\%$ CI is calculated)
 - Overall patient's risk preserved at $\sim \leq 0.05$

Potvin *et al.* (Method C)

- 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 for significant differences between Stages

Potvin *et al.* (Method C)

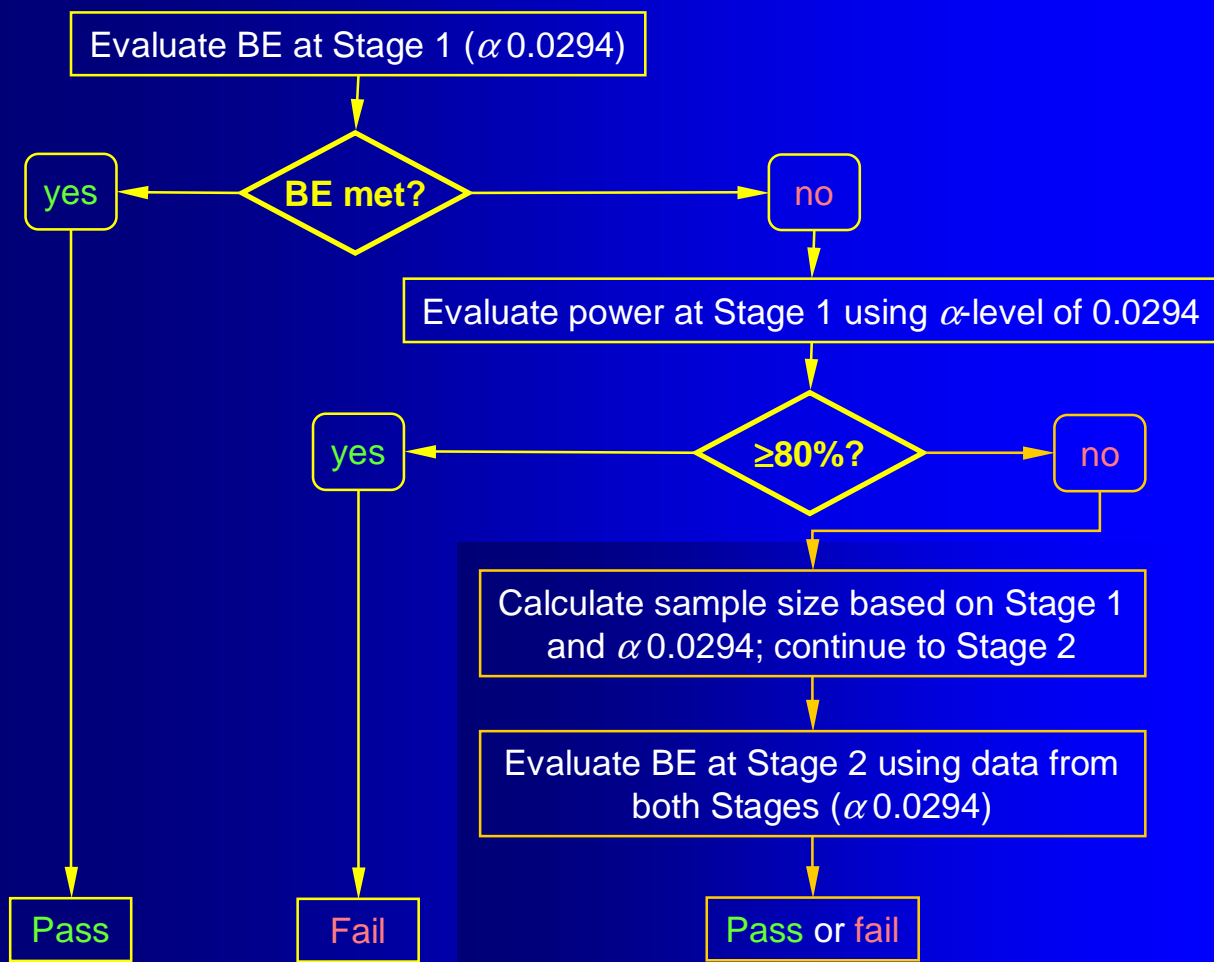
- Technical Aspects (cont'd)

- Potvin *et al.* used a simple approximative power estimation based on the shifted t -distribution (to increase speed in their simulations?)
- If possible use the exact method (Owen; package *PowerTOST* exact = TRUE) or at least the one based on the noncentral t -distribution (*PowerTOST* exact = FALSE)

- Power obtained in Stage 1:

method	power
approx. (shifted t)	64.94%
approx. (noncentral t)	66.45%
exact	66.47%

Potvin *et al.* (Method B)



Potvin et al. (example B/C)

Model Specification and User Settings

Dependent variable : Response
 Transform : LN
 Fixed terms : int+Sequence+Treatment+Period
 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

α 0.0294
(if power <80%)

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 90% = (95.1474, 124.3162)
 CI User = (92.9291, 127.2838)

Failed 90% CI (if power ≥80%)
and 94.12% CI (if power <80%)

Failed to show average bioequivalence for confidence=94.12 and percent=20.0.

Potvin et al. (example B/C)

```
require(PowerTOST)
power.TOST(alpha=0.05, logscale=TRUE,
            theta1=0.8, theta2=1.25, theta0=0.95,
            cv=0.182132, n=12,
            design = "2x2", exact = TRUE)
```

α 0.05 (C), α 0.0294 (B), expected ratio 95% – *not* 108.76% obs. in stage 1! CV_{intra} 18.2%, 12 subjects in Stage 1

[1] 0.6646934

Power 66.5% – 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", exact = TRUE,
              print = TRUE)
```

Calculate total sample size: expected ratio 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 for Stage 2

Potvin et al. (example B/C)

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

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

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.

α 0.0294 in
pooled analysis

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

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

● Pros & cons

- Method C (*if power $\geq 80\%$!*) is a conventional BE study; no penalty in terms of α needs to be applied
- Method C goes 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
- 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 (94.12% CI) or stage 2 as the 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 ~20% higher.

Sequential Designs

- Methods by Potvin *et al.* (2008) limited to point estimate of 0.95 and 80% power
 - Follow-up paper
 - Slight inflation of patient's risk (α 0.0547) observed in Methods B/C if PE 0.90 instead of 0.95 was used
 - Method D (like C, but α 0.0280 instead of α 0.0294)
 - Might be useful if PE 0.95 and power 90% as well; *not validated yet!*

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. (2011), [DOI: 10.1002/pst.483](https://doi.org/10.1002/pst.483)

Sequential Designs

● Caveats

- Methods for 'classical' group-sequential designs derived based on
 - Test for differences (superiority, parallel groups)
 - Large samples (Z test of normal distributed data with known variance)
 - Fixed total sample size (interim analysis at N/k)
 - Balanced case (no drop outs)
- Don't apply any published procedure unquestioned (*i.e.*, if not validated for bioequivalence)
- *Simulations mandatory* to derive an empirical $\alpha (\leq 0.052)$!

Open Issues

- Feasibility / futility rules
 - It would be desirable to stop a study after stage 1 under certain circumstances
 - (1) BE is unlikely to be shown in even very high sample sizes (e.g., CI outside acceptance range)
→ reformulate
 - (2) It turns out that the drug/formulation is highly variable
→ replicate design study in order to perform scaling required
 - (3) The calculated sample size exceeds the budget of the project by far

Open Issues

- Feasibility / futility rules
 - These points are not covered by Potvin *et al.*
 - If you decide to include a rule for early stopping, it's not part of the statistical procedure any more
 - (1) and (2) are ethically justifiable
 - (3) Acceptance?

Open Issues

- Arbitrary PE and/or power
 - Simulations mandatory
 - Set desired PE and power
 - Define maximum α -inflation (≤ 0.052 ?)
 - Simulate sufficiently large number of studies (N)
 - Count number of studies accepted BE at 1.25 (n_1) and number of studies rejected BE at the desired PE (n_2)
 - Empirical $\alpha = n_1/N$
 - Empirical $\beta = n_2/N$; power = $1 - \beta$
 - Start with Pocock's nominal α 0.0294 and decrease stepwise if empirical α too high
 - Compiled language almost necessary (speed!)

Open Issues

- Adaption for stage 1 PE (full adaptive design)
 - If applied naïvely, α -inflation of up to 30%!*
 - Various methods for superiority trials, but nothing in the area of BE published
 - Simulations mandatory

* **Cui L, Hung MJ, and S-J Wang**

Modification of sample size in group sequential clinical trials
Biometrics 55, 853–7 (1999)

Open Issues

- Dropping a candidate formulation from a higher-order cross-over design

Stage 1		
I	II	III
T ₁	T ₂	R
T ₂	R	T ₁
R	T ₁	T ₂
T ₁	R	T ₂
T ₂	T ₁	R
R	T ₂	T ₁
...

Stage 2	
I	II
R	T ₂
T ₂	R
...	...

How to decide which formulation to drop?

- Statistical model of BE assumes IID (common σ^2)
 - Let's assume to continue with T₂
 - If $\sigma^2_{T_1} > \sigma^2_{T_2}$ and/or σ^2_R , the pooled variance in Stage 1 will be inflated. The estimated total sample size will be too high. Expensive, but no influence on α expected.
 - If $\sigma^2_{T_1} < \sigma^2_{T_2}$ and/or σ^2_R , power will be lower – increasing the producer's risk only.

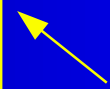
Don't try this at home!

- 6x3 dose proportionality study
 R 20 mg, T₁ 30 mg, T₂ 40 mg; CV_{intra} 8.76%
 - T₂ ÷ 2, all effects fixed (EMA), Method D_B, PE 90%, α 0.028

Stage 1						
	I		II		III	
R	162.28	T ₂	153.44	T ₁	235.62	
T ₁	75.34	R	72.04	T ₂	43.82	
T ₂	64.38	T ₁	78.18	R	71.28	
T ₁	124.06	T ₂	49.06	R	86.42	
T ₂	100.22	R	97.72	T ₁	121.36	
R	32.30	T ₁	63.87	T ₂	70.33	
T ₁	118.74	R	42.25	T ₂	65.97	
T ₂	66.07	T ₁	69.52	R	38.30	

Stage 2				
	I		II	
R	80.23	T ₂	64.26	
T ₂	65.72	R	67.91	
T ₂	19.61	R	20.81	
R	32.55	T ₂	29.35	

Extremely imbalanced due to arbitrary cut of original dataset!
 N=6 (single balanced block) would have zero df for sequences.



Don't try this at home!

Model Specification and User Settings
Dependent variable : Response
Transform : LN
Fixed terms : int+sequence+treatment+period+subject(sequence)

8 subjects in Stage 1,
all effects fixed (EMA)

Final variance parameter estimates:

Var(Residual) 0.0811756

CV_{intra} 8.13%

Bioequivalence Statistics

User-Specified Confidence Level for CI's = 94.4000

α 0.028 (Method B/D)

Percent of Reference to Detect for 2-1 Tests = 20.0%

A.H.Lower = 0.800 A.H.Upper = 1.250

Reference: Reference LSMean= 4.263887 SE= 0.103103 GeoLSM= 71.085745

Test: Test 1 LSMean= 4.686177 SE= 0.103103 GeoLSM= 108.437840

Difference = 0.4223, Diff_SE= 0.1436, df= 12.0

Ratio(%Ref) = 152.5451

CI User = (112.5795, 206.6985)

Failed to show average bioequivalence for confidence=94.40 and percent=20.0.

Test: Test 2 LSMean= 4.318248 SE= 0.103103 GeoLSM= 75.056997

Difference = 0.0544, Diff_SE= 0.1436, df= 12.0

Ratio(%Ref) = 105.5866

CI User = (77.9237, 143.0697)

Failed to show average bioequivalence for confidence=94.40 and percent=20.0.

Don't try this at home!

```
require(PowerTOST)
power.TOST(alpha=0.0280, logscale=TRUE,
            theta1=0.8, theta2=1.25, theta0=0.90,
            CV=se2CV(0.0811756), n=8,
            design="3x6x3", exact=TRUE)
```

α 0.028, expected ratio 90%,
MSE 0.08118 (CV_{intra} 8.13%),
8 subjects in Stage 1, 6x3 design

[1] 0.7776753

Power 77.8% <80% – initiate Stage 2

```
sampleN.TOST(alpha=0.0280, targetpower=0.80, logscale=TRUE,
              theta1=0.8, theta2=1.25, theta0=0.90,
              CV=se2CV(0.0811756), design="3x6x3", exact=TRUE,
              print=TRUE)
```

Calculate total sample size:
expected ratio 90%, CV_{intra} 8.13%,
80% power, keeping 6x3 design

```
+++++ Equivalence test - TOST +++++
      Sample size estimation
```

```
-----
Study design: 3x6x3 crossover
log-transformed data (multiplicative model)
```

```
alpha = 0.0294, target power = 0.8
BE margins      = 0.8 ... 1.25
Null (true) ratio = 0.9, CV = 0.08130951
```

```
Sample size
n      power
12    0.930078
```

Total sample size 12: include another 4 for Stage 2

Don't try this at home!

Model Specification and User Settings

Dependent variable : Response

Transform : LN

Fixed terms : int+Sequence+Stage+Period(Stage)+Treatment

Random/repeated terms : Sequence*Stage*Subject

Final variance parameter estimates:

Var(Residual) 0.0985763

4 subjects in Stage 2 (12 total),
modified model for pooled analysis

Bioequivalence Statistics

User-Specified Confidence Level for CI's = 94.4000

Percent of Reference to Detect for 2-1 Tests = 20.0%

A.H.Lower = 0.800 A.H.Upper = 1.250

Reference: Reference LSMean= 3.888945 SE= 0.216489 GeoLSM= 48.859311

Test: Test 1 LSMean= 4.284496 SE= 0.229396 GeoLSM= 72.565947

Difference = 0.3956, Diff_SE= 0.1256, df= 14.825

Ratio(%Ref) = 148.5202

CI User = (114.4688, 192.7011)

Failed to show average bioequivalence for confidence=94.40 and percent=20.0.

Test: Test 2 LSMean= 3.889827 SE= 0.216489 GeoLSM= 48.902424

Difference = 0.0009, Diff_SE= 0.1069, df= 14.825

Ratio(%Ref) = 100.0882

CI User = (80.1937, 124.9182)

Average bioequivalence shown for confidence=94.40 and percent=20.0.

Don't try this at home!

- Lessons learned, open questions
 - Not validated! Don't think about using it at all!
 - Note that due to the massive imbalance the LSM of Test 1 (although not included in Stage 2) changed from Stage 1 in the pooled analysis!
 - Stage 1: 108.44
 - Pooled: 72.57
 - Drug has low CV_{intra} , but high CV_{inter} – Apples and oranges?

CV%	T ₁	T ₂	R	model
Stage 1	28.61	41.30	70.66	period
Stage 2	–	82.50	85.95	period
Pooled	28.61	56.91	65.87	period

Don't try this at home!

- Lessons learned, open questions
 - Must use software in the power calculation which can handle the degrees of freedom of a Williams' design in Stage 1 correctly (e.g., *PowerTOST*)
 - Obvious which formulation to drop in this example, but what if formulations are similar in PEs?
Keep the one with smaller CV_{inter} ?
 - Design in the sample size estimation of Stage 2?
 - ~~2x2 (block size 2 → 10)~~
 - 3x6 (block size 6 → 12) ✓
 - The former would have failed in the example

Don't try this at home!

- Lessons learned, open questions
 - Tempting idea, but not recommended
 - until a statistical decision tree is developed and
 - suitable simulations have shown that the patient's risk is not inflated

Open Issues

- Replicated designs (HVDs/HVDPs)
 - Nothing published!
 - Statistical model?
 - Although EMA assumes equal variances of formulations (Q&A document Jan 2010) that does not reflect the 'real world' (quite often $\sigma^2_{WR} > \sigma^2_{WT}$)
 - If you set up simulations allow for different variances of test and reference

Congratulations!
**Perfecting (?) the two stage
study design**
Open Questions?



Helmut Schütz

BEBAC

Consultancy Services for
Bioequivalence and Bioavailability Studies
1070 Vienna, Austria

helmut.schuetz@bebac.at

To bear in Remembrance...

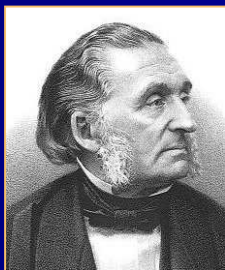
Power. That which statisticians are always calculating but never have.

Power: That which is wielded by the priesthood of clinical trials, the statisticians, and a stick which they use to beta their colleagues.



Power Calculation – A guess masquerading as mathematics.

Stephen Senn



You should treat as many patients as possible with the new drugs while they still have the power to heal.

Armand Trousseau