Statistical Design and Analysis II

Helmut Schütz
BEBAC

Minimum Sample Size

- 12 – WHO, EU, CAN, NZ, AUS, Malaysia, Argentina, ASEAN States, South Africa (20 for MR)
- FDA – 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.
- 24 – Saudia Arabia (12 to 24 if statistically justifiable)
- 24 – Brazil

Maximum Sample Size

- New Zealand
  - ‘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 in Guidelines (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.’

EU

- 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.

EU

- NfG on the Investigation of BA/BE (2001)
  - Problems/solutions
    - … the error variance associated with the *primary characteristic* to be studied …
      - Since BE must be shown both for AUC and C_{max}, and,
      - if you plan your sample size only for the 'primary characteristic' (*e.g.*, AUC), in many cases you will fail for the secondary parameter (*e.g.*, C_{max}), which most likely shows higher variability – your study will be 'underpowered'.
      - Based on the assumption, that CV is identical for test and reference (what if only the reference formulation has high variability, *e.g.*, some formulations of PPIs?).

EU

- NfG on the Investigation of BA/BE (2001)
  - Problems/solutions
    - … as estimated from
      - a *pilot experiment*,
      - *previous studies*, or
      - *published data*,
    - The correct order should read:

      1. previous studies → 2. pilot study → 3. published data
         - Only in the first case you ‘know’ all constraints resulting in variability
         - Pilot studies are often too small to get *reliable* estimates of variability
         - Advisable only if you have data from a couple of studies

EU

- NfG on the Investigation of BA/BE (2001)
  - Problems/solutions
    - … the *significance level desired* …
      - Throughout the NfG the significance level (α, error type I: patient's risk to be treated with a bio***in***equivalent drug) is fixed to 5% (corresponding to a 90% confidence interval)
      - You may *desire* a higher significance level, but such a procedure is not considered acceptable
      - In special cases (*e.g.*, dose proportionality testing), a correction for multiplicity may be necessary
      - In some legislations (*e.g.*, Brazil's ANVISA), α must be tightened to 2.5% for NTIDs (95% confidence interval)

EU

- NfG on the Investigation of BA/BE (2001)
  - Problems/solutions
    - … the *required power*.
      - Generally the power is set to at least 80 % (β, error type II: producers's risk to get no approval for a bioequivalent drug; power = 1 – β).
        Remember: *1 out of 5 studies will fail just by chance!*
      - If you plan for power of less than 70 %, problems with the ethics committee are likely (ICH E9).
      - If you plan for power of more than 90 % (especially with low variability drugs), problems with the regulator are possible ('forced bioequivalence').
      - Add subjects ('alternates') according to the expected drop-out rate!

EU

- NfG on the Investigation of BA/BE (2001)
  - Problems/solutions
    - … the *expected deviation (Δ) from the reference* …
      - Reliable estimate only from a previous full-sized study
      - If you are using data from a pilot study, allow for a safety margin
      - If no data are available, commonly a GMR (geometric test/reference-ratio) of 0.95 (Δ = 5%) is used
      - If more than Δ = 10% is expected, questions from the ethics committee are likely
      - BE GL (2010) batches must not differ more than 5%.

EU

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

Cookbook?

Hints

- Literature search for CV%
  - Preferably other BE studies (the bigger, the better!)
  - PK interaction studies (Cave: mainly in steady state! Generally lower CV than after SD)
  - Food studies (CV higher/lower than fasted!)
  - If CV_{intra} is not given (quite often), a little algebra helps. All you need is the 90% geometric confidence interval and the sample size.
    - Point estimate (PE) from the CI

$$PE = \sqrt{CL_{lo} \cdot CL_{hi}}$$

Algebra…

- Calculation of CV_{intra} from CI
  - Estimate the number of subjects / sequence (example 2×2 cross-over)
    - If total sample size (N) is an even number, *assume* (!) $n_1 = n_2 = ½N$
    - If N is an odd number, *assume* (!) $n_1 = ½N + ½$, $n_2 = ½N - ½$ (*not* $n_1 = n_2 = ½N$!)
  - Difference between one CL and the PE in log-scale; use the CL which is given with more significant digits

$$\Delta_{CL} = \ln PE - \ln CL_{lo} \quad or \quad \Delta_{CL} = \ln CL_{hi} - \ln PE$$

Algebra...

- Calculation of CV_{intra} from CI
  - Calculate the Mean Square Error (MSE)

$$MSE = 2\left(\frac{\Delta_{CL}}{\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}\cdot t_{1-2\cdot\alpha,n_1+n_2-2}}\right)^2$$

  - CV_{intra} from MSE as usual

$$CV_{\text{intra}}\% = 100\cdot\sqrt{e^{MSE}-1}$$

Algebra…

- Calculation of CV_{intra} from CI
  - Example: 90% CI [0.91 – 1.15], N 21 (n_1 = 11, n_2 = 10)

$$PE = \sqrt{0.91 \cdot 1.15} = 1.023$$

$$\Delta_{CL} = \ln 1.15 - \ln 1.023 = 0.11702$$

$$MSE = 2\left(\frac{0.11702}{\sqrt{\left(\frac{1}{11}+\frac{1}{10}\right)\times 1.729}}\right)^2 = 0.04798$$

$$CV_{\text{intra}}\% = 100 \times \sqrt{e^{0.04798} - 1} = 22.2\%$$

Algebra…

- Proof: CI from calculated values
  - Example: 90% CI [0.91 – 1.15], N 21 (n_1 = 11, n_2 = 10)

$$\ln PE = \ln\sqrt{CL_{lo} \cdot CL_{hi}} = \ln\sqrt{0.91 \times 1.15} = 0.02274$$

$$SE_{\Delta} = \sqrt{\frac{2 \cdot MSE}{N}} = \sqrt{\frac{2 \times 0.04798}{21}} = 0.067598$$

$$CI = e^{\ln PE \pm t \cdot SE_{\Delta}} = e^{0.02274 \pm 1.729 \times 0.067598}$$

$$CI_{lo} = e^{0.02274 - 1.729 \times 0.067598} = 0.91$$

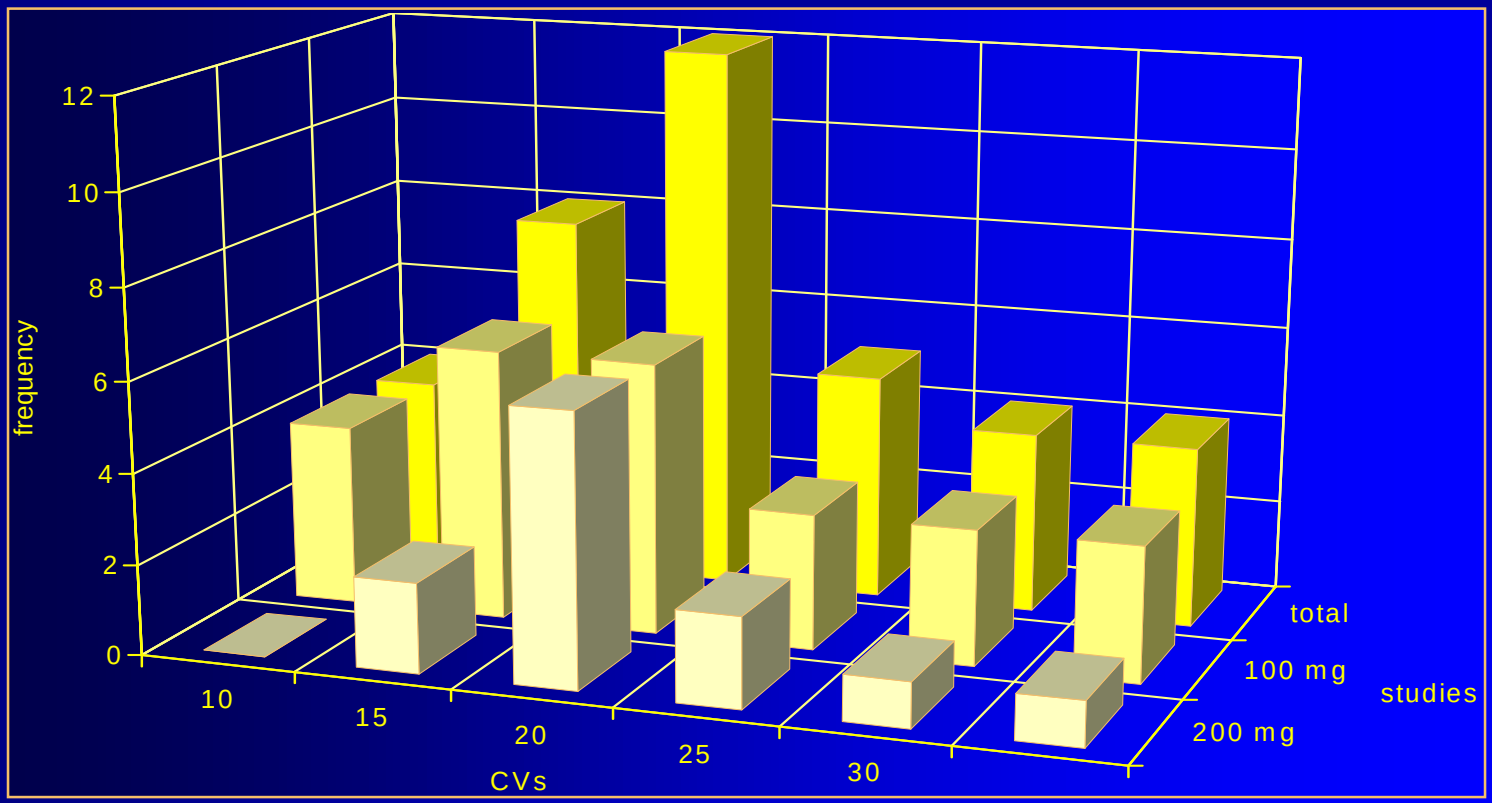
$$CI_{hi} = e^{0.02274 + 1.729 \times 0.067598} = 1.15$$

Sensitivity to Imbalance

- If the study was more imbalanced that you assumed, the estimated CV is conservative
  - Example: 90% CI [0.89 – 1.15], N 24 (n_1 = 16, n_2 = 8, but not reported as such); CV 24.74% in the study

n_1	n_2	CV%
12	12	26.29
13	11	26.20
14	10	25.91
15	9	25.43
16	8	24.74

Literature data



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

Pooling of CV%

- Intra-subject CV from different studies can be pooled
  - Do not use the arithmetic mean (or the geometric mean either) of CVs
  - In the parametric model of log-transformed data, additivity of variances (not of CVs!) apply
  - Before pooling variances must be weighted acccording to the sample size
    - Calculate the variance from CV

$$\sigma_W^2 = \ln(CV_{\text{intra}}^2 + 1)$$

Pooling of CV%

- Intra-subject CV from different studies
  - Calculate the total variance weighted by degrees of freedom

$$\sum \sigma_w^2 df$$

  - Calculate the pooled CV from total variance

$$CV = \sqrt{e^{\sum \sigma_w^2 df / \sum df} - 1}$$

  - Optionally calculate an upper (1-α) % confidence limit on the pooled CV (recommended α=0.20)

$$CL_{CV} = \sqrt{e^{\sum \sigma_w^2 df / \chi^2_{1-\alpha, \sum df}} - 1}$$

Pooling of CV%

- Example 1: $n_1=n_2$;
  $CV_{Study1} < CV_{Study2}$

studies	N
2	24

df (total)	α	1-α	total	CV_{pooled}	CV_{mean}
20	0.2	0.8	1.2540	**0.254**	~~0.245~~
		$\chi^2_{(1-\alpha,df)}$	14.578	0.300	+17.8%

CV_{intra}	n	seq.	df (mj)	σw	$\sigma^2 w$	$\sigma^2 w \times df$	CV_{intra} / pooled	$>CL_{upper}$
0.200	**12**	**2**	10	0.198	0.0392	0.3922	78.6%	no
0.300	**12**	**2**	10	0.294	0.0862	0.8618	117.9%	yes

Pooling of CV%

- Example 2: $n_1 < n_2$;
  $CV_{Study1} < CV_{Study2}$

studies	N
2	36

df (total)	α	1-α	total	CV_{pooled}	CV_{mean}
32	0.2	0.8	2.2881	**0.272**	~~0.245~~
		$\chi^2_{(1-\alpha,df)}$	25.148	0.309	+13.4%

CV_{intra}	n	seq.	df (mj)	σw	$\sigma^2 w$	$\sigma^2 w \times df$	$CV_{intra\ /\ pooled}$	$>CL_{upper}$
0.200	**12**	**2**	10	0.198	0.0392	0.3922	73.5%	no
0.300	**24**	**2**	22	0.294	0.0862	1.8959	110.2%	no

Pooling of CV%

- Example 3: $n_1 > n_2$;
  $CV_{Study1} < CV_{Study2}$

studies	N
2	36

df (total)	α	1-α	total	CV_{pooled}	CV_{mean}
32	0.2	0.8	1.7246	**0.235**	~~0.245~~
		$\chi^2_{(1-\alpha,df)}$	25.148	0.266	+13.2%

CV_{intra}	n	seq.	df (mj)	σ_w	σ^2_w	$\sigma^2_w \times df$	$CV_{intra\ /\ pooled}$	$>CL_{upper}$
0.200	**24**	**2**	22	0.198	0.0392	0.8629	85.0%	no
0.300	**12**	**2**	10	0.294	0.0862	0.8618	127.5%	yes

α- *vs.* β-Error

- α-Error (aka error type I): patient's risk to be treated with a bioinequivalent formulation
  - Reminder: BA in a particular patient can be *either* below 80% *or* above 125%.
  - If we keep the risk of particular patients at 0.05 (5%), the risk of entire the population of patients (<80% *and* >125%) is 2×α (10%)
    That's where the 90% confidence interval comes from (CI = 1 – 2×α = 0.90)
  - Although α is generally set to 0.05, sometimes <0.05 (*e.g.*, NTDIs in Brazil, multiplicity, interim analyses).

α- *vs.* β-Error

- β-Error (aka error type II): producer's risk to get no approval for a bioequivalent formulation
  - Generally *set* in study planning to ≤0.2, where power = 1 - β = ≥80%
  - No guidelines about power ('appropriate'), but
    - 70% only in exceptional cases
    - >90% *may* raise questions from the Ethics Committee (suspection of 'forced bioequivalence')
  - There is no *a posteriori* (aka *post hoc*) power!
    Either a study has shown BE or not.
    Phoenix'/WinNonlin's output is statistical nonsense!

Power *vs.* Sample Size

- It is not possible to *directly* calculate the needed sample size.
- Power is calculated instead, and the lowest sample size which fulfills the minimum target power is used.
  - Example: α 0.05, target power 80% (β=0.2), T/R 0.95, CV_{intra} 20% $\rightarrow$ minimum sample size 19 (power 81%), rounded *up* to the next even number in a 2×2 study (power 83%).

n	power
16	73.54%
17	76.51%
18	79.12%
19	81.43%
20	83.47%

Tools

- Specialized Software (nQuery Advisor, PASS, FARTSSIE, StudySize, …)
- General purpose (SAS, R, S+, StaTable, …)
- Sample Size Tables (Phillips, Diletti, Hauschke, Chow, Julious, …)
- Approximations (Diletti, Chow, Julious, …)

Background

- Reminder: Sample Size is not directly obtained; only power
- Solution given by DB Owen (1965) as a difference of two bivariate noncentral *t*-distributions
  - Definite integrals cannot be solved in closed form
    - 'Exact' methods rely on numerical methods (currently the most advanced is AS 243 of RV Lenth; implemented in R, FARTSSIE, EFG). nQuery uses an earlier version (AS 184).

Background

- Power calculations…
  - ‘Brute force’ methods (also called ‘resampling’ or ‘Monte Carlo’) converge asymptotically to the true power; need a good random number generator (*e.g.*, Mersenne Twister) and may be time-consuming
  - ‘Asymptotic’ methods use large sample approximations
  - Approximations provide algorithms which should converge to the desired power based on the *t*-distribution

Comparison

original values	Method	Algorithm	CV% 5.	7.5	10.	12.	12.5	14.	15.	16.	17.5	18.	20.	22.
Patterson & Jones (2006)	exact	AS 243	4	5	7	8	9	11	12	13	15	16	19	22
Diletti *et al.* (1991)	exact	?	4	5	7		9		12		15		19	
nQuery Advisor 7 (2007)	exact	AS 184	4	6	8	8	10	12	12	14	16	16	20	22
FARTSSIE 1.6 (2008)	exact	AS 243	4	5	7	8	9	11	12	13	15	16	19	22
EFG 1.01 (2009)	exact	AS 243	4	5	7	8	9	11	12	13	15	16	19	22
	brute force	ElMaestro	4	5	7	8	9	11	12	13	15	16	19	22
StudySize 2.0.1 (2006)	asympt.	?		5	7	8	9	11	12	13	15	16	19	22
Hauschke *et al.* (1992)	approx.				8	8	10	12	12	14	16	16	20	22
Chow & Wang (2001)	approx.			6	6	8	8	10	12	12	14	16	18	22
Kieser & Hauschke (1999)	approx.		2		6	8		10	12	14		16	20	24

original values	Method	Algorithm	CV% 22.5	24.	25.	26.	27.5	28.	30.	32.	34.	36.	38.	40.
Patterson & Jones (2006)	exact	AS 243	23	26	28	30	33	34	39	44	49	54	60	66
Diletti *et al.* (1991)	exact	?	23		28		33		39					
nQuery Advisor 7 (2007)	exact	AS 184	24	26	28	30	34	34	40	44	50	54	60	66
FARTSSIE 1.6 (2008)	exact	AS 243	23	26	28	30	33	34	39	44	49	54	60	66
EFG 1.01 (2009)	exact	AS 243	23	26	28	30	33	34	39	44	49	54	60	66
	brute force	ElMaestro	23	26	28	30	33	34	39	44	49	54	60	66
StudySize 2.0.1 (2006)	asympt.	?	23	26	28	30	33	34	39	44	49	54	60	66
Hauschke *et al.* (1992)	approx.		24	26	28	30	34	36	40	46	50	56	64	70
Chow & Wang (2001)	approx.		24	26	28	30	34	34	38	44	50	56	62	68
Kieser & Hauschke (1999)	approx.			28	30	32		38	42	48	54	60	66	74

Approximations

Hauschke D, Steinijans VW, Diletti E, and M Burke
Sample Size Determination for Bioequivalence Assessment Using a Multiplicative Model
J Pharmacokin Biopharm 20/5, 557-561 (1992)

Patient's risk α 0.05, Power 80% (Producer's risk β 0.2), AR [0.80 – 1.25], CV 0.2 (20%), T/R 0.95

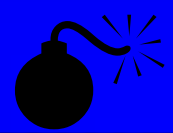
1. $\Delta = \ln(0.8)-\ln(T/R) = -0.1719$
2. Start with e.g. n=8/sequence
   1. $df = n \cdot 2 - 1 = 8 \times 2 - 1 = 14$
   2. $t_{\alpha,df} = 1.7613$
   3. $t_{\beta,df} = 0.8681$
   4. new $n = [(t_{\alpha,df} + t_{\beta,df})^2 \cdot (CV/\Delta)]^2 =$ $(1.7613+0.8681)^2 \times (-0.2/0.1719)^2 = 9.3580$
3. Continue with n=9.3580/sequence (N=18.716 → 19)
   1. df = 16.716; roundup to the next integer 17
   2. $t_{\alpha,df} = 1.7396$
   3. $t_{\beta,df} = 0.8633$
   4. new $n = [(t_{\alpha,df} + t_{\beta,df})^2 \cdot (CV/\Delta)]^2 =$ $(1.7396+0.8633)^2 \times (-0.2/0.1719)^2 = 9.1711$
4. Continue with n=9.1711/sequence (N=18.3422 → 19)
   1. df = 17.342; roundup to the next integer 18
   2. $t_{\alpha,df} = 1.7341$
   3. $t_{\beta,df} = 0.8620$
   4. new $n = [(t_{\alpha,df} + t_{\beta,df})^2 \cdot (CV/\Delta)]^2 =$ $(1.7341+0.8620)^2 \times (-0.2/0.1719)^2 = 9.1233$
5. Convergence reached (N=18.2466 → 19): Use 10 subjects/sequence (20 total)

S-C Chow and H Wang
On Sample Size Calculation in Bioequivalence Trials
J Pharmacokin Pharmacodyn 28/2, 155-169 (2001)

Patient's risk α 0.05, Power 80% (Producer's risk β 0.2), AR [0.80 – 1.25], CV 0.2 (20%), T/R 0.95

1. $\Delta = \ln(T/R) - \ln(1.25) = 0.1719$
2. Start with e.g. n=8/sequence
   1. $df_\alpha = \text{roundup}(2 \cdot n-2) \cdot 2-2 = (2 \times 8-2) \times 2-2 = 26$
   2. $df_\beta = \text{roundup}(4 \cdot n-2) = 4 \times 8-2 = 30$
   3. $t_{\alpha,df} = 1.7056$
   4. $t_{\beta/2,df} = 0.8538$
   5. new $n = \beta^2 \cdot [(t_{\alpha,df} + t_{\beta/2,df})^2/\Delta^2 =$ $0.2^2 \times (1.7056+0.8538)^2 / 0.1719^2 = 8.8723$
3. Continue with n=8.8723/sequence (N=17.7446 → 18)
   1. $df_\alpha = \text{roundup}(2 \cdot n-2) \cdot 2-2 = (2 \times 8.8723-2) \times 2-2 = 30$
   2. $df_\beta = \text{roundup}(4 \cdot n-2) = 4 \times 8.8723-2 = 34$
   3. $t_{\alpha,df} = 1.6973$
   4. $t_{\beta/2,df} = 0.8523$
   5. new $n = \beta^2 \cdot [(t_{\alpha,df} + t_{\beta/2,df})^2/\Delta^2 =$ $0.2^2 \times (1.6973+0.8538)^2 / 0.1719^2 = 8.8045$
4. Convergence reached (N=17.6090 → 18): Use 9 subjects/sequence (18 total)

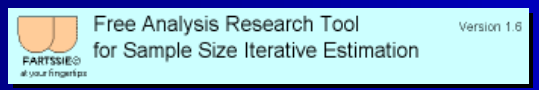
sample size	18	19	20
power %	79.124	81.428	83.468



Approximations obsolete

- Exact sample size tables still useful in checking the plausibility of software's results
- Approximations obsolete, since exact method available (FARTSSIE17) or …

```
alpha   <- 0.05     # alpha
CV      <- 0.30     # intra-subject CV
theta1  <- 0.80     # lower acceptance limit
theta2  <- 1/theta1 # upper acceptance limit
ratio   <- 0.95     # expected ratio T/R
PwrNeed <- 0.80     # minimum power
Limit   <- 500      # Upper Limit for Search
n       <- 3        # start value of sample size search
s       <- sqrt(2)*sqrt(log(1+CV^2))
repeat{
  t     <- qt(1-alpha,n-2)
  nc1   <- sqrt(n)*(log(ratio)-log(theta1))/s
  nc2   <- sqrt(n)*(log(ratio)-log(theta2))/s
  prob1 <- pt(+t,n-2,nc1); prob2 <- pt(-t,n-2,nc2)
  power <- prob2-prob1
  n     <- n+1        # increment sample
  if(power >= PwrNeed | (n-1) >= Limit) break }
Total   <- n-1
if(Total == Limit){
  cat("Search stopped at Limit",Limit,
      " obtained Power",power*100,"%\n")
  } else
  cat("Sample Size",Total,"(Power",power*100,"%)\n")
```



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

Power Curves

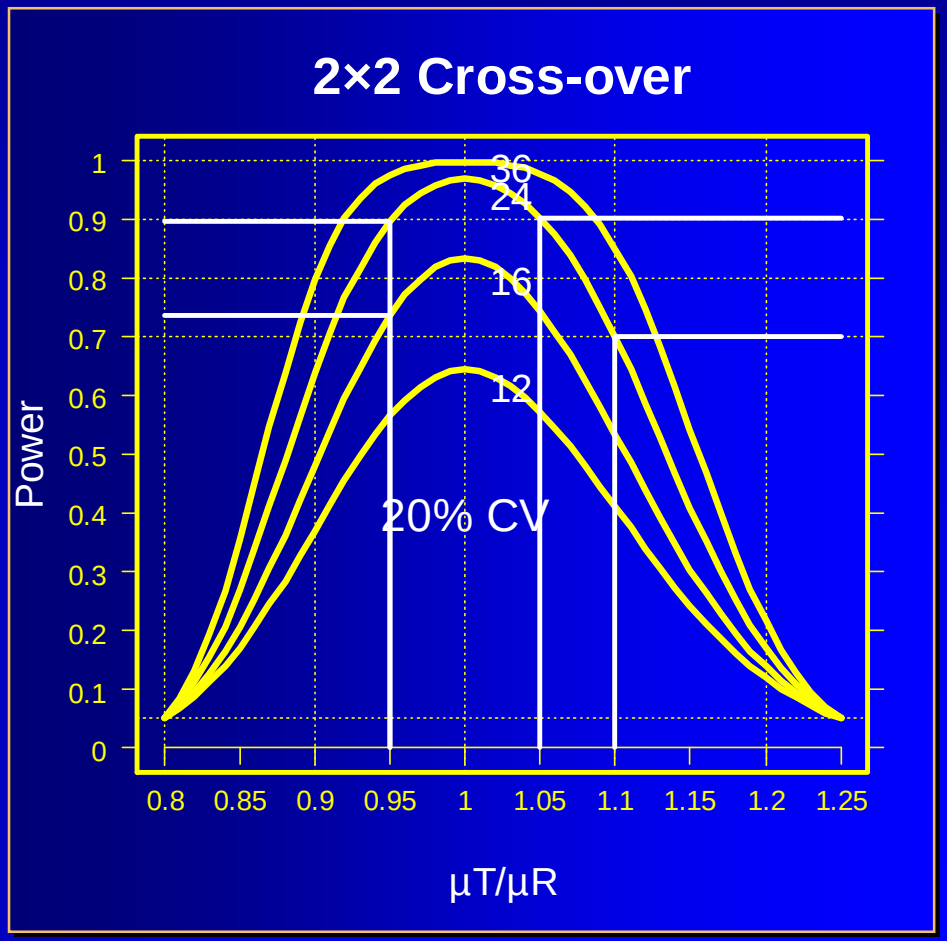
Power to show

BE with 12 – 36 subjects for

CV_{intra} = 20%

n 24 → 16:
power 0.896 → 0.735

μ_T/μ_R 1.05 → 1.10:
power 0.903 → 0.700



HVDs/HVDPs

- EU GL on BE (2010)
  - The regulatory switching condition θ_s is derived from the regulatory standardized variation σ_0. For CV_{WR} = 30% we get

$$\sigma_0 = \sqrt{\ln(0.3^2 + 1)} = 0.2936$$

and

$$\theta_s = \frac{\ln(1.25)}{\sigma_0} = 0.7601$$

Tothfalusi L, Endrenyi L and A Garcia Arieta
Evaluation of Bioequivalence for Highly Variable Drugs with Scaled Average Bioequivalence
Clin Pharmacokinet 48/11, 725-743 (2009)

HVDs/HVDPs

- EU GL on BE (2010)
  - Average Bioequivalence (ABE) with Expanding Limits (ABEL)
    - If you have σ_{WR} (the intra-subject standard deviation of the reference formulation) go to the next step; if not, calculate it from CV_{WR}:

$$\sigma_{WR} = \sqrt{\ln(CV_{WR}^2 + 1)}$$

    - Calculate the scaled acceptance range based on the regulatory constant *k* (0.7601):

$$[U, L] = e^{\pm k \cdot \sigma_{WR}}$$

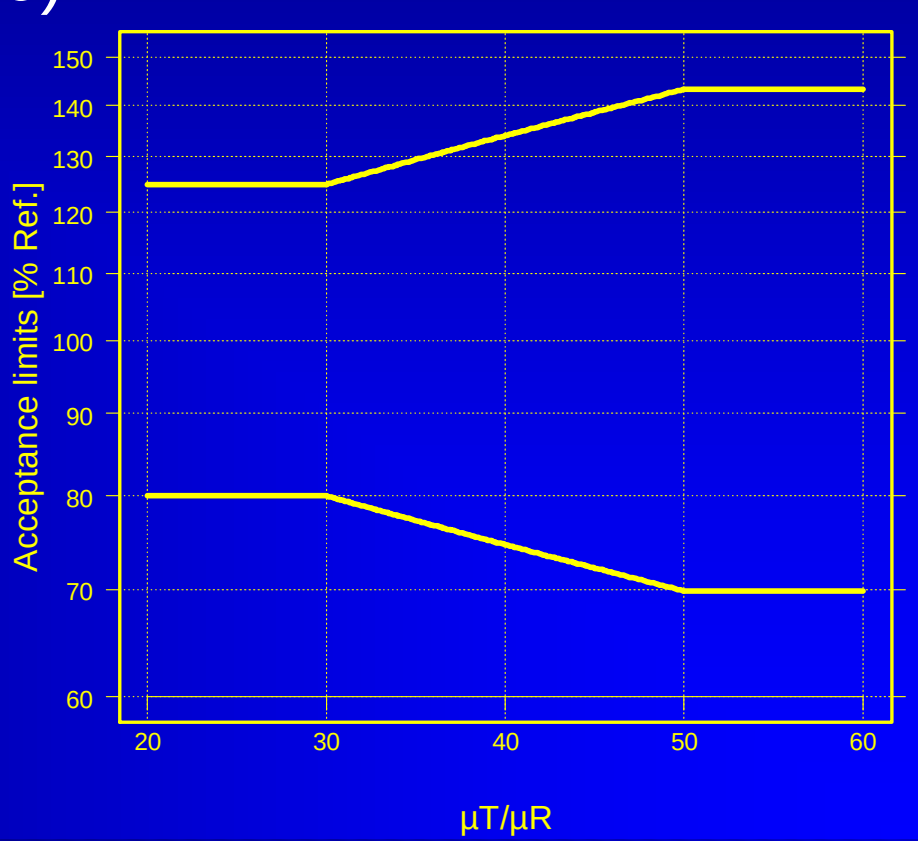
HVDs/HVDPs

- EU GL on BE (2010)
  - Scaling allowed for C_{max} only (not AUC!) based on CV_{WR} >30% in the study.
  - Limited to a maximum of CV_{WR} 50% (*i.e.*, higher CVs are treated *as if* CV = 50%).
  - PE restricted with 80% – 125% in any case.
  - No commercial software for sample size estimation can handle the PE restriction.
  - Expect a solution from the  community soon…

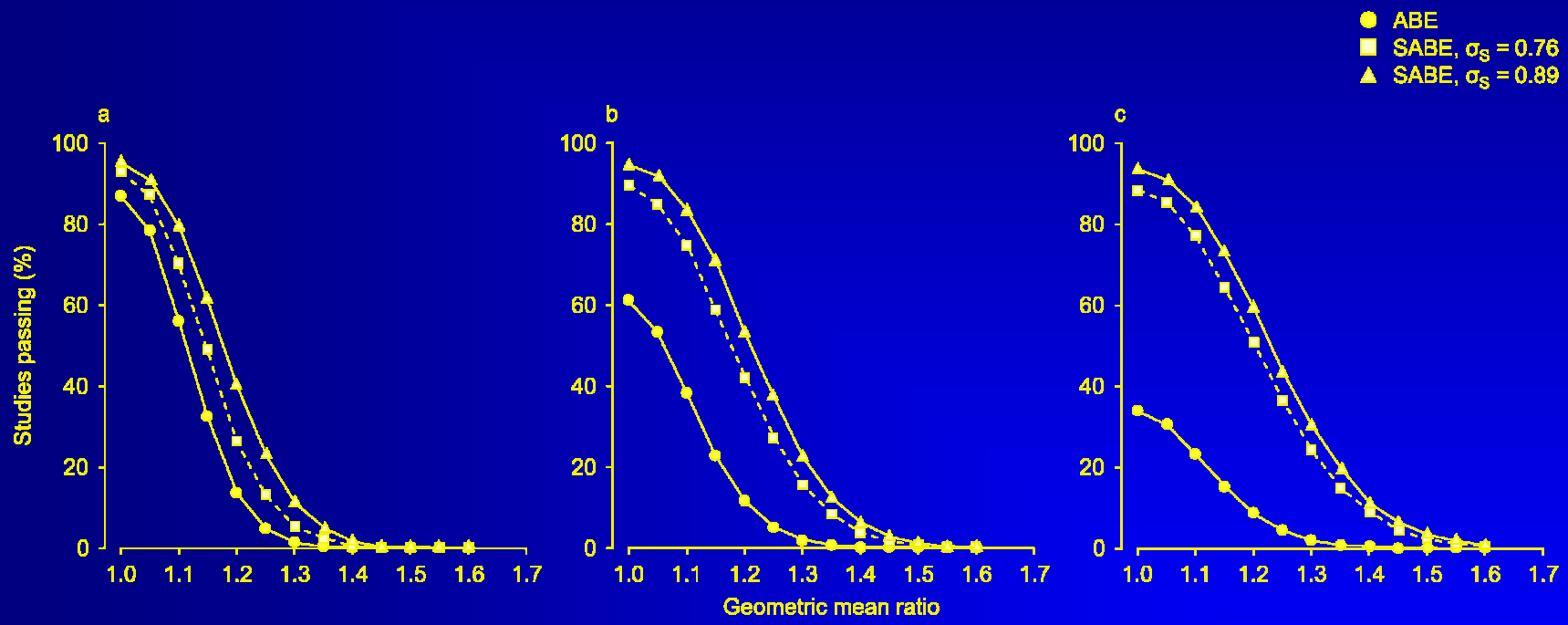
HVDs/HVDPs

- EU GL on BE (2010)

CV%	L%	U%
30	80.00	125.00
32	78.87	126.79
34	77.77	128.58
36	76.69	130.39
38	75.64	132.20
40	74.61	134.02
42	73.61	135.85
44	72.63	137.68
46	71.68	139.52
48	70.74	141.36
50	69.83	143.20



HVDs/HVDPs



Totfalushi *et al.* (2009), Fig. 3
Simulated (n=10000) three-period replicate design studies (TRT-RTR) in 36 subjects;
GMR restriction 0.80–1.25. (a) CV=35%, (b) CV=45%, (c) CV=55%.
ABE: Conventional Average Bioequivalence, SABE: Scaled Average Bioequivalence,
0.76: EU criterion, 0.89: FDA criterion.

HVDs/HVDPs

- Replicate designs
  - 4-period replicate designs:
    sample size = ½ of 2×2 study's sample size
  - 3-period replicate designs:
    sample size = ¾ of 2×2 study's sample size
  - Reminder: number of treatments (and biosamples) is identical to the concentional 2×2 cross-over.
  - Allow for a safety margin – expect a higher number of drop-outs due to the additional period(s).
  - Consider increased blood loss (ethics!)
    Eventually bioanalytics has to be improved.

Example ABEL

- RTR–TRT Replicate Design, n=18

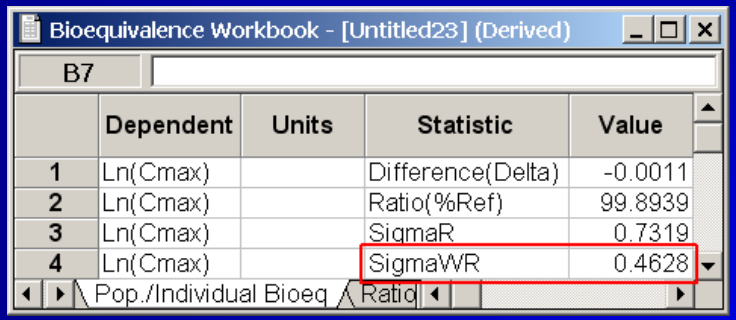
Subj	Seq	Per	Trt	Cmax
1	1	1	R	209.91
1	1	2	T	111.05
1	1	3	R	116.36
2	1	1	R	101.16
2	1	2	T	100.31
2	1	3	R	31.71
3	1	1	R	14.83
3	1	2	T	57.10
3	1	3	R	21.47
4	1	1	R	118.71
4	1	2	T	37.34
4	1	3	R	52.29
5	1	1	R	36.11
5	1	2	T	83.95
5	1	3	R	17.76
6	1	1	R	146.44
6	1	2	T	40.45
6	1	3	R	38.34

Subj	Seq	Per	Trt	Cmax
7	1	1	R	58.49
7	1	2	T	62.80
7	1	3	R	123.23
8	1	1	R	105.34
8	1	2	T	103.32
8	1	3	R	43.67
9	1	1	R	59.73
9	1	2	T	169.03
9	1	3	R	48.26
10	1	1	R	38.34
10	1	2	T	31.19
10	1	3	R	19.43
11	2	1	T	51.95
11	2	2	R	195.71
11	2	3	T	65.87
12	2	1	T	18.72
12	2	2	R	20.63
12	2	3	T	7.45

Subj	Seq	Per	Trt	Cmax
13	2	1	T	92.76
13	2	2	R	59.54
13	2	3	T	56.84
14	2	1	T	159.20
14	2	2	R	155.50
14	2	3	T	165.31
15	2	1	T	162.41
15	2	2	R	47.31
15	2	3	T	88.23
16	2	1	T	19.44
16	2	2	R	42.80
16	2	3	T	18.93
17	2	1	T	90.58
17	2	2	R	42.39
17	2	3	T	54.57
18	2	1	T	42.96
18	2	2	R	171.86
18	2	3	T	59.15

Example ABEL

- σ_{WR} (WinNonlin)



Calculate the scaled acceptance range based on the regulatory constant *k* (0.7601) and the limiting CV:

$$[U, L] = e^{\pm k \cdot \sigma_{WR}} \quad CV_{WR} = \sqrt{\exp(\sigma_{WR}^2 - 1)}$$

σ_{WR}	0.4628
CV_{WR}	0.4887
L	0.7154
U	1.3977

30%<CV_{WR}<50%: use calculated limits.

Example ABEL

- ABE

  PE: 99.89

  90% CI:

  72.04, 138.52

  failed ABE

  failed 75 – 133

  $30<CV_{WR}<50$

  [L,U]

  71.54, 139.77

  passed ABEL

  (90% CI within [L,U], PE within 80.00 – 125.00)

Bioequivalence Text - [Untitled4] (Read-only) (Derived)

```
Bioequivalence Statistics

User-Specified Confidence Level for CI's and Power = 90.0000
Percent of Reference to Detect for 2-1 Tests and Power = 20.0%
A.H.Lower =  0.800    A.H.Upper =  1.250

Formulation variable: Trt
Reference: R    LSMean=     4.069159   SE=     0.173739   GeoLSM=    58.507730
------------------------------------------------------------------------------

Test:       T    LSMean=     4.068098   SE=     0.174718   GeoLSM=    58.445673

    Difference =     -0.0011,   Diff_SE=     0.1876,   df= 16.5
    Ratio(%Ref) =     99.8939

                        Classical                  Westlake
    CI  80% = (    77.7639,   128.3217)   (    75.1692,   124.8308)
    CI  90% = (    72.0378,   138.5217)   (    67.3124,   132.6876)
    CI  95% = (    67.1817,   148.5344)   (    59.4138,   140.5862)
    Failed to show average bioequivalence for confidence=90.00 and percent=20.0.

                      Two One-Sided T-tests

    Prob(< 80%)=0.1266  Prob(> 125%)=0.1244  Max=0.1266  Total=0.2510
```

Read Only | Line 187 / 189

Sensitivity Analysis

- ICH E9
  - 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: $\sigma_w = \sqrt{\ln(CV_{\text{intra}}^2 + 1)}; \sqrt{\ln(0.2^2 + 1)} = 0.198042$

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% → 69%

20% CV, PE 90%: power 90% → 69%

Sensitivity Analysis

- Must be done *before* the study *(a priori)*
- *A posteriori* power:
  - High values do not support the claim of already demonstrated bioequivalence
  - Low values do not invalidate a bioequivalent formulation
  - Further reader:

JM Hoenig and DM Heisey
The Abuse of Power: The Pervasive Fallacy of Power Calculations for Data Analysis
The American Statistician 55/1, 19–24 (2001)
http://www.math.uiowa.edu/~rlenth/Power/2badHabits.pdf
RV Lenth
Two Sample-Size Practices that I don't recommend
Joint Statistical Meetings, Indianapolis (2000)
http://www.math.uiowa.edu/~rlenth/Power/2badHabits.pdf

Sample Size: Pilot Studies

- Pilot Studies
  - Small pilot studies (sample size <12)
    - are useful in checking the sampling schedule and
    - the appropriateness of the analytical method, but
    - are not suitable for the purpose of sample size planning.

Justification

- Good Scientific Practice!
  - Every influental factor can be *tested* in a pilot study.
    - Sampling schedule: matching C_{max}, lag-time (first point C_{max} problem), reliable estimate of λ_z
    - Bioanalytical method: LLOQ, ULOQ, linear range, metabolite interferences, ICSR
    - Food, posture,…
    - Variabilty of PK metrics
    - Location of PE

Justification

- Best description by FDA (2003)
  - The study can be used to validate analytical methodology, assess variability, optimize sample collection time intervals, and provide other information. For example, for conventional immediate-release products, careful timing of initial samples may avoid a subsequent finding in a full-scale study that the first sample collection occurs after the plasma concentration peak. For modified-release products, a pilot study can help determine the sampling schedule to assess lag time and dose dumping.

Application

- Most common to assess CV and PE needed in sample size estimation for a pivotal BE study
  - To select between candidate test formulations compared to one reference
  - To find a suitable reference
  - If design issues (clinical performance, bioanalytics) are already known, a two-stage sequential design would be a better alternative!

Drawbacks

- CV-estimates have a high degree of uncertainty (in the pivotal study it is more likely that you will be able to reproduce the PE, than the CV)
  - The smaller the size of the pilot, the more uncertain the outcome.
  - The more formulations you have tested, lesser degrees of freedom will result in worse estimates.

Solutions

- *Do not* use the pilot study's CV, but calculate an upper confidence interval!
  - Gould recommends a 75% CI (*i.e.*, a producer's risk of 25%).
  - Unless you are under time pressure, a two-stage design will help in dealing with the uncertain estimate from the pilot.

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

Sample Size: Pilot Studies

- Pilot Studies (cont'd)
  - Moderate sized pilot studies (sample size ~12–24) lead to more consistent results (both CV_{intra} and PE).
    - If you stated a procedure in your protocol, even BE may be claimed in the pilot study, and no further study will be necessary.
    - If you have some previous hints of high intra-subject variability (>30%), a pilot study size of *at least* 24 subjects is reasonable.
      A Sequential Design may also avoid an unnecessary large pivotal study.

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.

'Internal Pilot Study Design'

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 later phases of clinical research (mainly Phase III)
  - Based on work by Armitage *et al.* (1969), McPherson (1974), Pocock (1977), O'Brien and Fleming (1979) and others
    - First proposal by LA Gould (1995) in the area of BE did not get regulatory acceptance in Europe, but
    - Stated in the current Canadian Draft Guidance (November 2009).

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

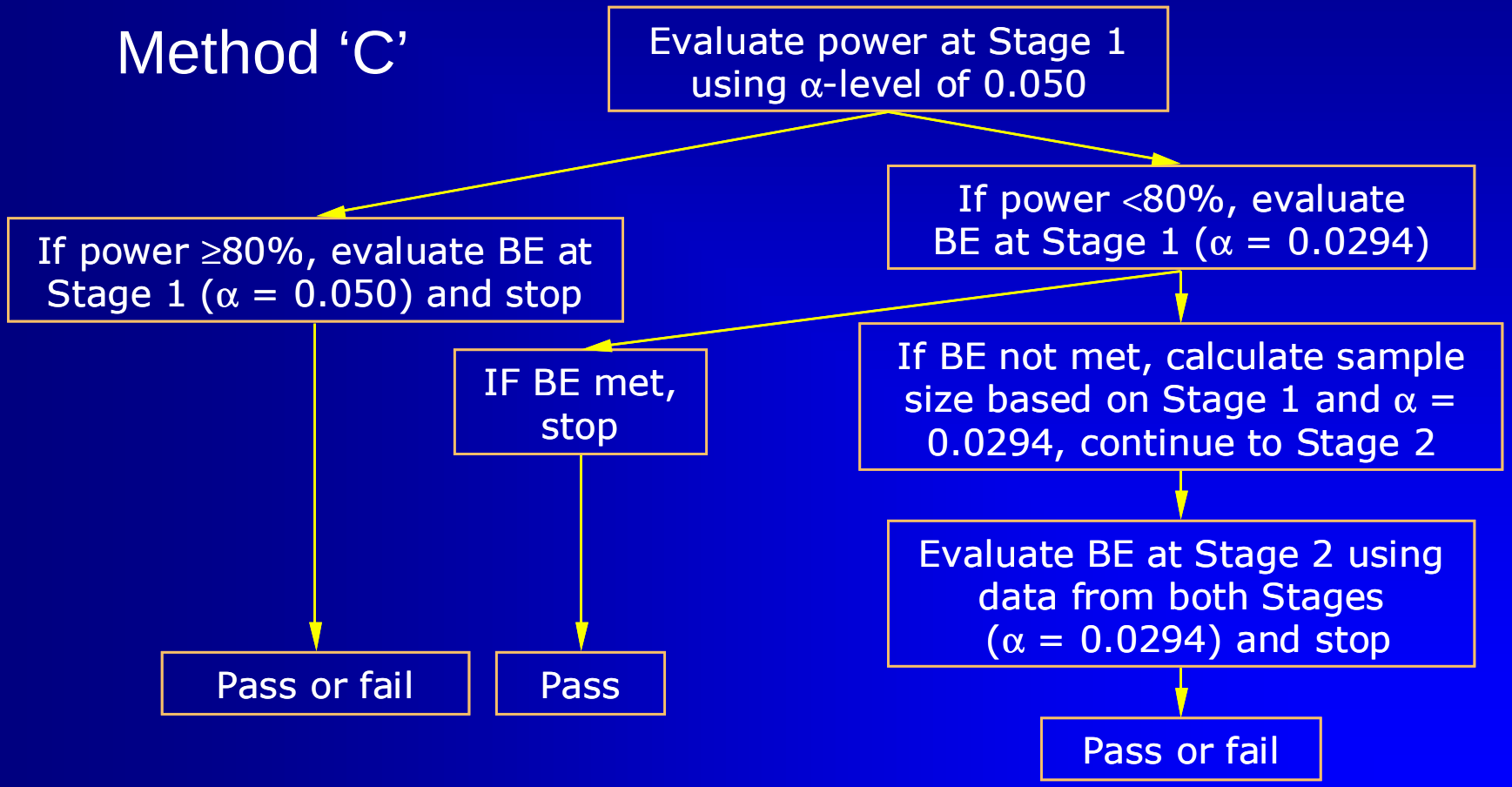
Sequential Designs

- Current work by D Potvin *et al.* (2007) seems promising
  - Work backed by support form FDA, USP, Health Canada,…
    - Likely to be implemented by US-FDA
    - Should be acceptable as a Two-Stage Design in the EU
    - Two of BEBAC's protocols approved by BfArM and competent EC in May and December 2009

Potvin D, Diliberti CE, Hauck WW, Parr AF, Schuirmann DJ, and RA Smith
Sequential design approaches for bioequivalence studies with crossover designs
Pharmaceut Statist (2007), DOI: 10.1002/pst.294
http://www3.interscience.wiley.com/cgi-bin/abstract/115805765/ABSTRACT

Potvin *et al.* (2007)

Method 'C'

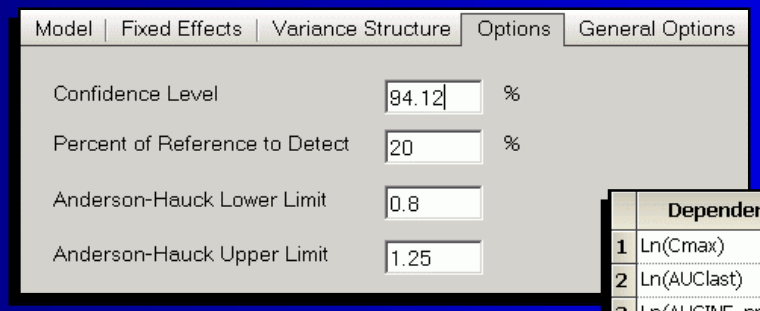


Potvin *et al.* (2007)

- Technical Aspects
  - Only *one* Interim Analysis (after Stage 1)
  - If possible, use software (too wide step sizes in Diletti's tables)
  - Should be called '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
  - No stop criterion for Stage 2! Must be clearly stated in the protocol (may be unfamiliar to the IEC, because standard in Phase III)

Potvin *et al.* (2007)

- Technical Aspects (cont'd)
  - Adjusted α of 0.0294 based on Pocock 1977
  - In the pooled analysis (data from Stages 1 + 2), α 0.0294 is used (*i.e.*, the 94.12% Confidence Interval is calculated)



	Dependent	Ratio__Ref_	CI_User_Lower	CI_User_Upper
1	Ln(Cmax)	105.26625	98.075483	112.98422
2	Ln(AUClast)	99.163003	96.124777	102.29726
3	Ln(AUCINF_pred)	98.82479	95.726146	102.02374

  - Overall patient's risk is ≤0.05

Potvin *et al.* (2007)

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

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

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

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

  - No poolability criterion; combining is *always allowed* – even for significant differences between Stages.

Potvin *et al.* (2007)

- Advantage
  - Currently the only *validated* procedure for BE!
- Drawbacks
  - *Not validated* for a correction of effect size (PE) observed in Stage 1 (must continue with the one used in sample size planing).
  - No stop criterion (EMA GL on BE?)
  - Not validated for any other design than the conventional 2×2 crossover (no higher order crossovers, no replicate designs).

Nuisance: group effect

- More than one group of subjects
  - ‘If a crossover study is carried out in two or more groups of subjects (*e.g.*, if for logistical reasons only a limited number of subjects can be studied at one time), the statistical model should be modified to reflect the multigroup nature of the study. In particular, the model should reflect the fact that the periods for the first group are different from the periods for the second group.’

FDA, Center for Drug Evaluation and Research (CDER)
Guidance for Industry: Statistical Approaches to Establishing Bioequivalence (2001)

Nuisance: group effect

- More than one group of subjects
  - Cases where '… the study is carried out in two or more groups and those groups are studied at different clinical sites, or at the same site but greatly separated in time (months apart, for example) […] should be discussed with the appropriate CDER review division.'
  - EMEA BA/BE (2001), BE GL (2010)
    - The study should be designed in such a way that the formulation effect can be distinguished from other effects.

Nuisance: group effect

- Increasing number of referrals (deficiency letters) from
  - Canada
  - Gulf States (Saudia Arabia, Emirates, Oman)
- Extended Statistical model (fixed effects in ANOVA)
  - Group
  - Group × Treatment Interaction
  - If both terms are not significant ($p>0.05$), pooling of groups is justified.

Nuisance: group effect

- Recommendations
  - If possible, multiple groups should be avoided.
  - Keep the time interval between groups as short as possible.
  - Do not split the study into equally sized groups.
    - Perform at least one group in the maximum capacity of the clinical site (*e.g.*, 24+8 instead of 16+16 for a total of 32).
    - If a significant group and/or group × treatment interaction is found (preventing a pooled analysis), it may still be possible to demonstrate BE in the largest group only.

Are we making progress?

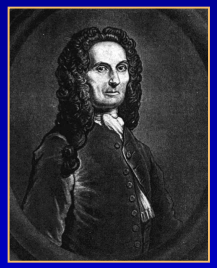
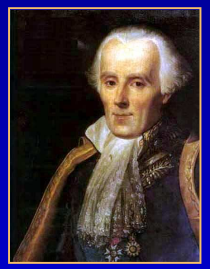
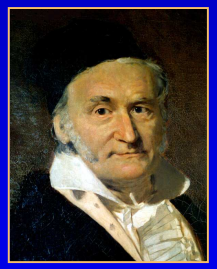
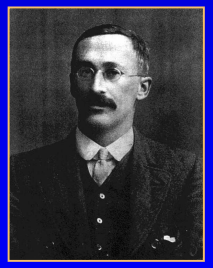
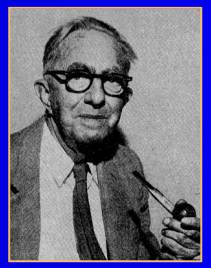
- About 3 000 – 10 000 BE studies / year are conducted worldwide; only ~ 1 – 5% of them are published.
- Although a standard for publishing data of BE studies was already suggested in 1992,[1)]
  - a review in 2002 found only 17 complete data sets on AUC and 12 on C_{max}.[2)]
  - Since no 'real world' data are available, proposed methods (*e.g.*, reference-scaled ABE) rely *entirely* on simulations!
  - Studies seen by regulators are 'selection biased'.

1) **Sauter R, Steinijans VW, Diletti E, Böhm E and H-U Schulz**
Int J Clin Pharm Ther Toxicol 30/Suppl.1, S7-S30 (1992)

2) **Nakai K, Fujita M and M Tomita**
Int J Clin Pharmacol Ther 40, 431-438 (2002)

Bell curve (and beyond?)

- Abraham de Moivre (1667-1754), Pierre-Simon Laplace (1749-1827)
  Central limit theorem 1733, 1812
- Carl F. Gauß (1777-1855)
  Normal distribution 1795
- William S. Gosset, aka Student (1876-1937)
  t-distribution 1908
- Frank Wilcoxon (1892-1965)
  Nonparametric tests 1945

Congratulations!
Statistical Design and Analysis II
Open Questions?

(References in your Handouts)

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

Quote from RV Lenth

There is simple intuition behind results like these: If my car made it to the top of the hill, then it is powerful enough to climb that hill; if it didn't, then it obviously isn't powerful enough. Retrospective power is an obvious answer to a rather uninteresting question. A more meaningful question is to ask whether the car is powerful enough to climb a particular hill never climbed before; or whether a different car can climb that new hill. Such questions are prospective, not retrospective.

The fact that retrospective power adds no new information is harmless in its own right. However, in typical practice, it is used to exaggerate the validity of a significant result ("not only is it significant, but the test is really powerful!"), or to make excuses for a nonsignificant one ("well, P is .38, but that's only because the test isn't very powerful"). The latter case is like blaming the messenger.

RV Lenth
Two Sample-Size Practices that I don't recommend
http://www.math.uiowa.edu/~rlenth/Power/2badHabits.pdf

References

- Collection of links to global documents
  http://bebac.at/Guidelines.htm
- ICH
  - E3: Structure and Content of Clinical Study Reports (1995)
  - E6: Good Clinical Practice (1996)
  - E8: General Considerations for Clinical Trials (1997)
  - E9: Statistical Principles for Clinical Trials (1998)
- WHO
  - Guidelines for GCP for trials on pharmaceutical products (WHO Technical Report Series No. 850, Annex 3, 1995)
  - Handbook for GCP (2005)
  - WHO Expert Committee on Specifications for Pharmaceutical Preparations, Fortieth Report (WHO Technical Report Series No. 937, Annex 9: Additional guidance for organizations performing *in vivo* bioequivalence studies. 2006)
- US FDA
  - 21CFR320: BA and BE Requirements (Revision 2008)
  - Center for Drug Evaluation and Research (CDER) CDER's Manual of Policies and Procedures
    - Review of BE Study Protocols (2006)
    - Review of BE Studies with Clinical Endpoints in ANDAs (2006)
  - Center for Drug Evaluation and Research (CDER)
    - Statistical Approaches Establishing Bioequivalence (2001)
    - Bioavailability and Bioequivalence Studies for Orally Administered Drug Products – General Considerations (Rev.1 2003)
    - ANDA Checklist for Completeness and Acceptability (2006)
    - Bioequivalence Recommendations for Specific Products (2007)
    - ANDA Checklist for Completeness and Accept-ability (2006)
    - Submission of Summary BE Data for ANDAs (2009)

References

- EudraLex – The Rules Governing Medicinal Products in the European Union
  http://ec.europa.eu/enterprise/pharmaceuticals/eudralex/
- Directive 2001/20/EC: Implementation of GCP in the Conduct of Clinical Trials on Medicinal Products for Human Use (2001)
- EMEA GCP Inspector's Group
  Procedure for Conducting GCP Inspections requested by the EMEA
  - Annex I: Investigator Site (2007)
  - Annex IV: Sponsor Site and/or Contract Research Organisations (CRO) (2007)
  - Annex V: Bioanalytical part, Pharmacokinetic and Statistical analyses of Bioequivalence Trials (2008)
- EMEA/CPMP/CHMP
  - NfG on the Investigation of BA/BE (2001)
  - Points to Consider on Multiplicity Issues in Clinical Trials (2002)
  - BA/BE for HVDs/HVDPs: Concept Paper (2006); removed form EMEA's website in Oct 2007. Available at http://bebac.at/downloads/14723106en.pdf
  - Questions & Answers on the BA and BE Guideline (2006)
  - Draft Guideline on the Investigation of BE (2008)
  - Questions & Answers : Positions on specific questions addressed to the EWP therapeutic subgroup on Pharmacokinetics (2009)
  - Guideline on the Investigation of BE (2010)