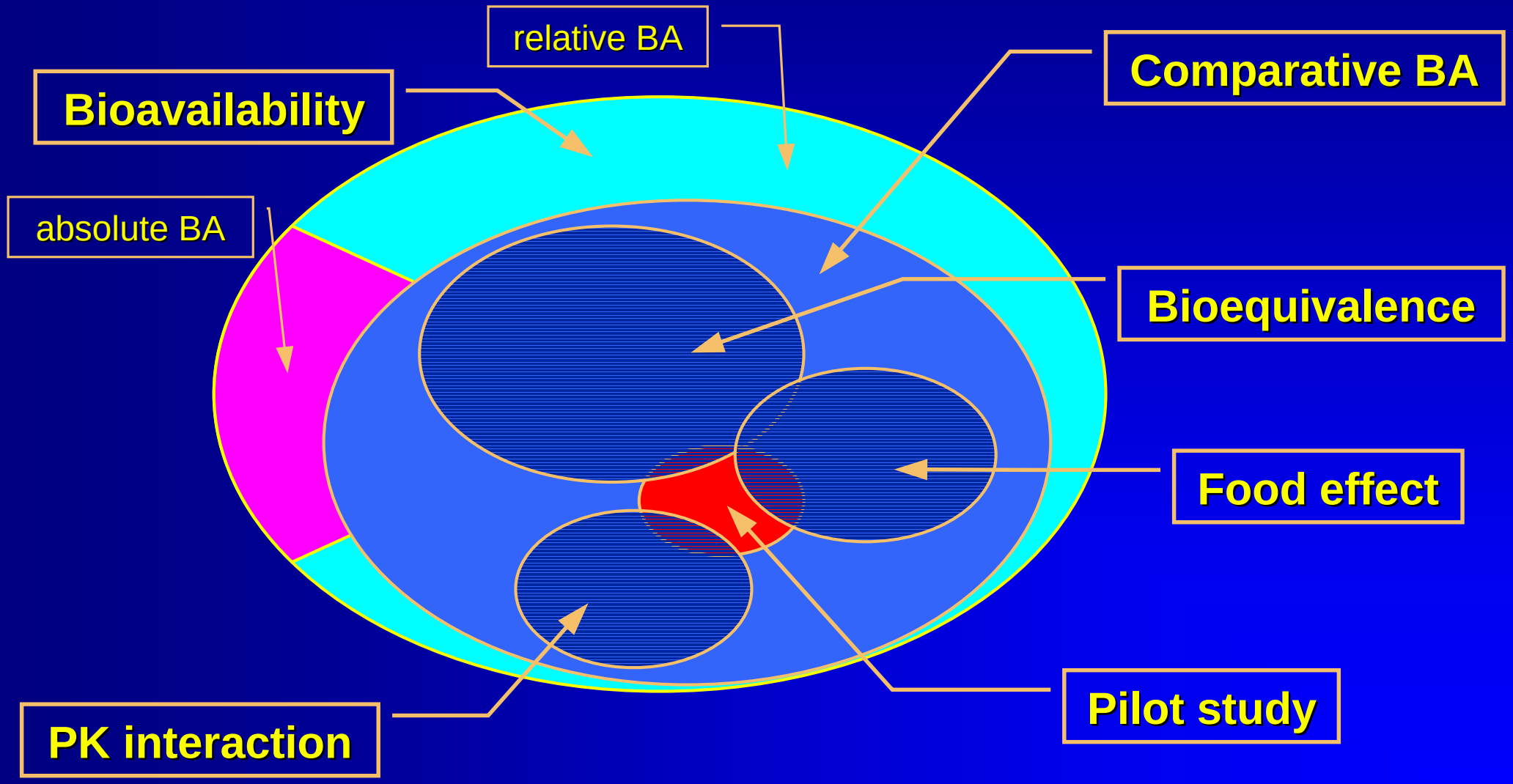
Study Design and Evaluation Issues 3/3

Statistical Design and Analysis

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

Terminology



Defining Study Objectives

- According to the EU NfG (3. Design and Conduct of Studies, paragraph 2):

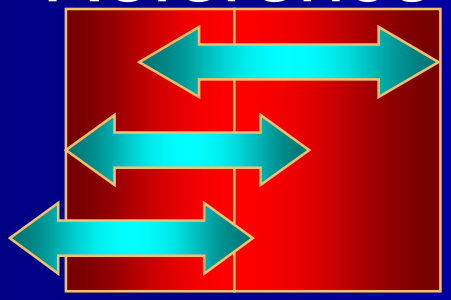
  'A bioequivalence study is basically a comparative bioavailability study designed to establish equivalence between test and reference products.'

  - Comparative BA,
  - designed to demonstrate BE,
  - reference = innovator's product.

EMEA Human Medicines Evaluation Unit / CPMP
Note for Guidance on the Investigation of Bioavailability and Bioequivalence (2001)
http://www.emea.eu.int/pdfs/human/ewp/140198en.pdf#page=6

Defining Study Objectives

- Comparative BA
  - true experiment; no bibliographic comp.
- Designed to demonstrate BE
  - variability,
  - deviation of test from reference,
  - drop-out rate,…
    - ➔to be able (statistical power!) to demonstrate BE
- Reference = Innovator's product



#1: BE [90%–125%]

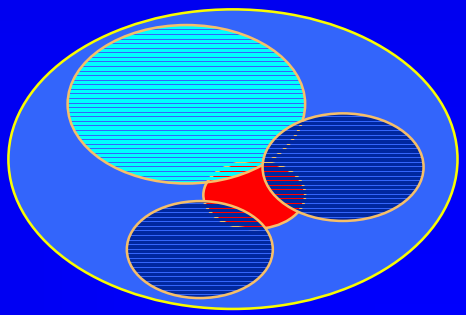
#2: BE [80%–110%]

#3: not BE [76%–103%]; (but 'BE' to #2)

Defining Study Objectives

- Definition of BE (EU NfG, Section 2.4)

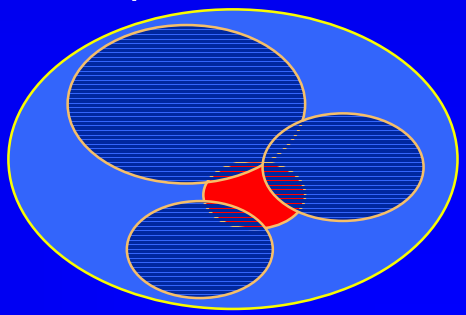
  'Two medicinal products are bioequivalent if they are pharmaceutically equivalent or pharmaceutical alternatives and if their bio-availabilities after administration in the same molar dose are similar to such degree that their effects, with respect to both efficacy and safety, will be essentially the same.'



Defining Study Objectives

- *In vivo* BE mandatory, if
  - Waiving (NfG Section 5.1.1) not possible
    - in MA of Generics
    - Manufacturing changes (EU Major variation type II(d)-(f) ~ FDA SUPAC Level 3)
  - Pharmacokinetic interaction studies,
  - Studies of fixed-combination products.

'[…] are similar to such degree that their effects, with respect to both efficacy and safety, will be essentially the same.'

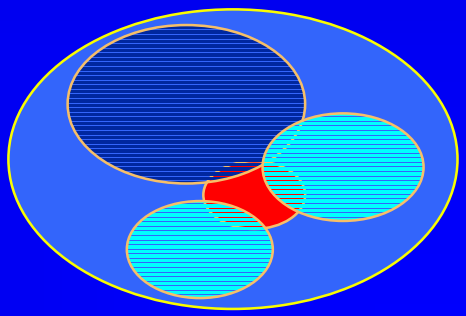


Defining Study Objectives

- Statistical concept of BE also applicable to
  - Food effect studies,
  - Pharmacokinetic interaction studies,
  - Studies of fixed-combination products.

'[…] are similar to such degree that their effects, with respect to both efficacy and safety, will be essentially the same.'

EMEA Human Medicines Evaluation Unit / CPMP
Modified Release Oral and Transdermal Dosage Forms: Section II (Quality)
CPMP/EWP/280/96 (1999)
EMEA Human Medicines Evaluation Unit / CPMP
The Investigation of Drug Interactions
CPMP/EWP/560/95 (1997)
EMEA
Fixed Combination Medicinal Products
CPMP/EWP/240/95 Rev. 1 (2008)



Defining Study Objectives

- Since *in vivo* BE relies on 'rich' PK data:
  - Sufficient number of blood samples (C_{max}!) / urine collection periods
  - Sampling long enough to cover ≥80 % of AUC_{∞}
  - Wash-out ≥3× $t_{½}$ (recommended ≥5× $t_{½}$)
  - Saturation phase long enough to reach steady-state: ≥5× $t_{½}$ (recomm. ≥7× $t_{½}$)
  - Pre-dose samples (carry-over, compliance)

EU Draft NfG (2008): for IR formulations no more sampling beyond 72 hours!

Defining Study Objectives

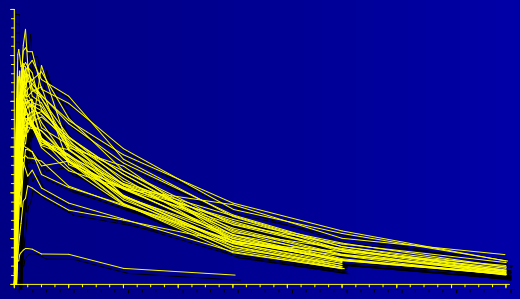
- PK metrics
  - Extent of bioavailability / Total exposure
    - single dose
      - AUC_t, AUC_∞ (plasma)
      - Ae_t, Ae_∞ (urine)
    - steady state
      - AUC_τ, AUC_{24h}, (plasma)
      - Ae_τ, Ae_{24h}, (urine)

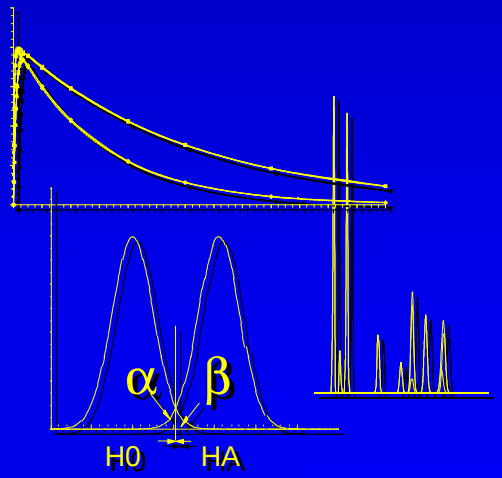
Defining Study Objectives

- PK metrics
  - Rate of bioavailability / Peak exposure / Early exposure
    - single dose
      - C_{max}, (t_{max}, partial AUC) (plasma)
      - ΔAe_{max} (urine)
    - steady state
      - as above
      - Fluctuation [$PTF = (C_{max} - C_{min})/C_{av}$]
    - MR formulations
      - MRT, HVD, $t_{75\%}$

Assumptions: General

World *'Reality'*





Model *'Data'*

Theory *'Truth'*

Assumptions: Pharmacokinetics

$$\frac{F_1 \cdot AUC_1}{\cancel{D_1 \cdot CL_1}} , \frac{F_2 \cdot AUC_2}{\cancel{D_2 \cdot CL_2}}$$

$$F_{rel}(BA) = \frac{AUC_1}{AUC_2}$$

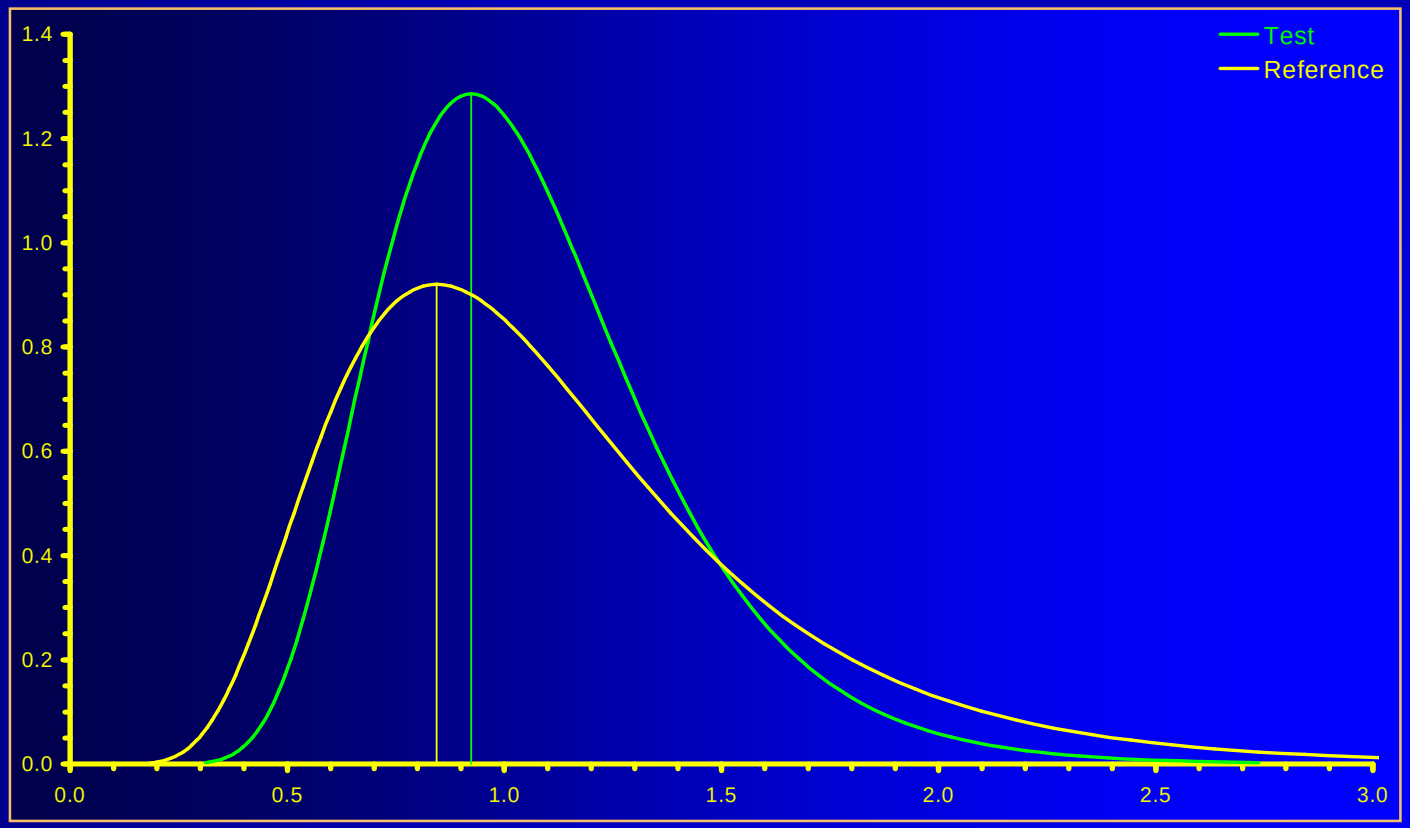
Assumption 1: $D_1 = D_2$ ($D_1/D_2 = 1^*$)

Assumption 2: $CL_1 = CL_2$

Assumptions: Statistics

Distribution

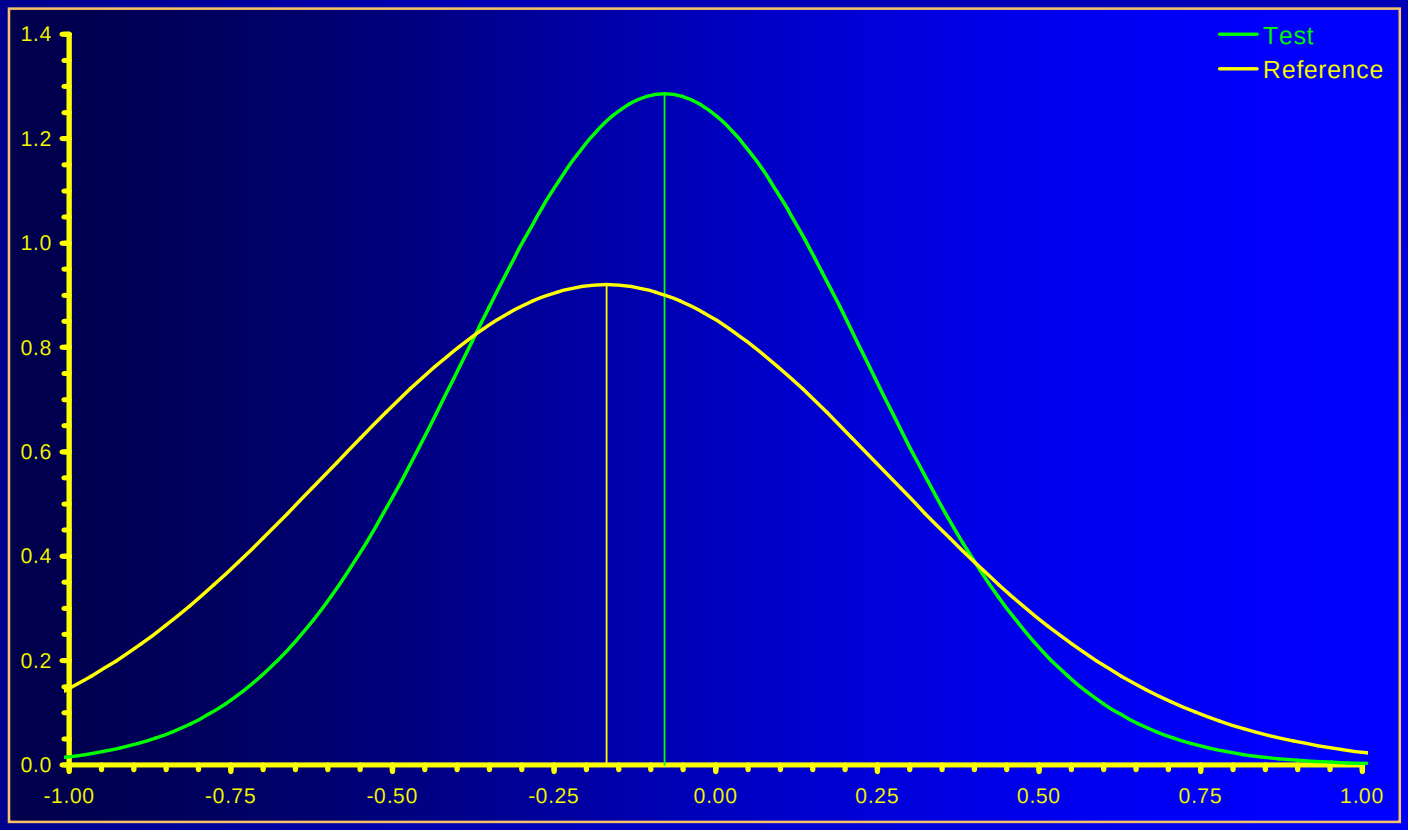
- IDD (Independent Identically Distribution)



Assumptions: Statistics

Multiplicative Model

- Log-Transformation (PK, Analytics)



Assumptions: Statistics

Multiplicative Model (X-over without carryover)

$$X_{ijk} = \mu \cdot \pi_k \cdot \Phi_l \cdot s_{ik} \cdot e_{ijk}$$

X_{ijk}: *ln*-transformed response of j-th subject ($j=1,\ldots,n_i$) in i-th sequence ($i=1,2$) and k-th period ($k=1,2$), μ: global mean, μ_l: expected formulation means ($l=1,2$: $\mu_1=\mu_{test}$, $\mu_2=\mu_{ref.}$), π_k: fixed period effects, Φ_l: fixed formulation effects ($l=1,2$: $\Phi_1=\Phi_{test}$, $\Phi_2=\Phi_{ref.}$)

Assumptions: Statistics

Multiplicative Model (X-over without carryover)

$$X_{ijk} = \mu \cdot \pi_k \cdot \Phi_l \cdot s_{ik} \cdot e_{ijk}$$

s_{ik}: random subject effect, e_{ijk}: random error

Main Assumptions:

- All $ln\{s_{ik}\}$ and $ln\{e_{ijk}\}$ are independently and normally distributed about unity with variances σ^2_s and σ^2_e.
- All observations made on different subjects are independent.

Global Harmonization?

Transformations (e.g. […], logarithm) should be specified in the protocol and a rationale provided […]. The general principles guiding the use of transformations to ensure that the assumptions underlying the statistical methods are met are to be found in standard texts […].

In the choice of statistical methods due attention should be paid to the statistical distribution […]. When making this choice (for example between parametric and non-parametric methods) it is important to bear in mind the need to provide statistical estimates of the size of treatment effects together with confidence intervals […].

ICH Topic E 9
Statistical Principles for Clinical Trials (1998)

Global Harmonization?

No analysis is complete until the assumptions that have been made in the modeling have been checked. Among the assumptions are that the repeated measurements on each subject are independent, normally distributed random variables with equal variances. Perhaps the most important advantage of formally fitting a linear model is that diagnostic information on the validity of the assumed model can be obtained. These assumptions can be most easily checked by analyzing the residuals.

Jones B and MG Kenward
Design and Analysis of Cross-Over Trials
Chapman & Hall, Boca Raton (2nd ed 2003)

Nonparametrics

The limited sample size in a typical BE study precludes a reliable determination of the distribution of the data set. Sponsors and/or applicants are not encouraged to test for normality of error distribution after log-transformation […].

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

But: acceptable in
Turkey (MOH, November 2005)
Saudia Arabia (SFDA, May 2005)

Nonparametrics

5. In which cases may a non-parametric statistical model be used?

The NfG states under 3.6.1–Statistical analysis: "*AUC and C_{max} should be analysed using ANOVA after log transformation.*"

The reasons for this request are the following:

a) the AUC and C_{max} values as biological parameters are usually not normally distributed;
b) a multiplicative model may be plausible;
c) after log transformation the distribution may allow a parametric analysis.

Comments:

a) – true b) – true c) – maybe, but may also terribly fail

EMEA/CHMP/EWP/40326/2006
Questions & Answers on the BA and BE Guideline (2006)

Nonparametrics

5. In which cases may a non-parametric statistical model be used?

However, the true distribution in a pharmacokinetic data set usually cannot be characterised due to the small sample size, so it is not recommended to have the analysis strategy depend on a pre-test for normality. Parametric testing using ANOVA on log-transformed data should be the rule. Results from non-parametric statistical methods or other statistical approaches are nevertheless welcome as sensitivity analyses. Such analyses can provide reassurance that conclusions from the experiment are robust against violations of the assumptions underlying the analysis strategy.

Comment: It is well known that the efficiency of *e.g.*, the Wilcoxon-Mann-Whitney test for normal distributed data is $3/\pi \approx 95.5$ %; for *not normal distributed data* the efficiency is >100 %!

Nonparametrics

4.1.8 Evaluation / Statistical analysis

The pharmacokinetic parameters under consideration should be analysed using ANOVA (or equivalent parametric method). The data should be transformed prior to analysis using a logarithmic transformation. A confidence interval for the difference between formulations on the log-transformed scale is obtained from the ANOVA model. This confidence interval is then back-transformed to obtain the desired confidence interval for the ratio on the original scale. A non-parametric analysis is not acceptable.

EMEA/CPMP/EWP/QWP/1401/98 Rev. 1
Draft Guideline on the Investigation of Bioequivalence (2008)

Walter Hauck: ‘Also interesting that they now say they will not accept nonparametric analyses. That seems a step backwards.’ (personal communication Oct 2008)

Global Harmonization?

FDA, EMEA (Q&A, BE Draft)

ln-Transformation (based on PK, analytics)

→ Parametric Evaluation (*e.g.*, ANOVA)

↓

Data and Residuals normally distributed ?

yes → Parametric Evaluation (*e.g.*, ANOVA)

no ↓

Nonparametric Evaluation (*e.g.*, WMW)

ICH
Good Statistical Practice

Global Harmonization?

- In almost all regulations two metrics are necessary to demonstrate BE, namely
  - extent (*e.g.*, AUC_t, AUC_∞, Ae), and
  - rate (*e.g.*, C_{max}, PTF) of exposure.
- One exception: US-FDA (where AUC_∞ *and* AUC_t must demonstrate extent of BE)
  - Although stated in the Guideline, such a requirement is statistically flawed.
    - Multiplicity issues (what is the patient's risk?)
    - Impossible α-adjustment (interdependence)

There can be only one!

Basic Designs

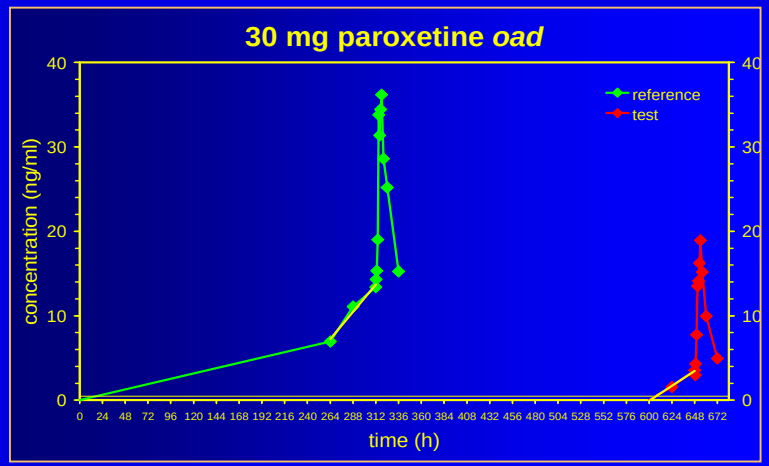
- Single Dose / Multiple Dose
  - Cross-over
    - Standard 2×2
    - Higher Order Designs (for more than 2 treatments)
      - Latin Squares
      - Variance Balanced Designs (Williams' Designs)
      - Incomplete Block Designs
    - Replicate designs
  - Parallel Groups

Single Dose / Multiple Dose

- Single Dose recommended in most Guide-lines, but steady-state studies
  - may be required:
    - in the case of dose- or time-dependent pharmacokinetics
    - for some modified release products (additionally to single dose BE)
  - may be considered:
    - if problems of sensitivity preclude sufficiently precise plasma concentration measurements after SD administration.
      With current developments in bioanalytical methodology, you should have strong evidence of infeasibility if you claim the necessity of a Multiple Dose study based on lacking methods.
      Regulators are concerned with efficacy/safety issues – not with the budget of pharmaceutical companies!

Single Dose / Multiple Dose

- Steady-state studies
  - No Wash-out between Periods (Switch-Over)!
  - In order to fulfil the superposition principle of linear pharmacokinetics ($AUC_\tau = AUC_\infty$), you must demonstrate achievement of steady-state
    - Linear-regression of pre-dose values in saturation phase
      - slope (from at least the last three values) should not significantly ($p>0.05$, two-sided) differ from zero,
      - subjects not in steady-state at begin of sampling of the profile should be excluded from the evaluation – if stated in protocol!



Single Dose / Multiple Dose

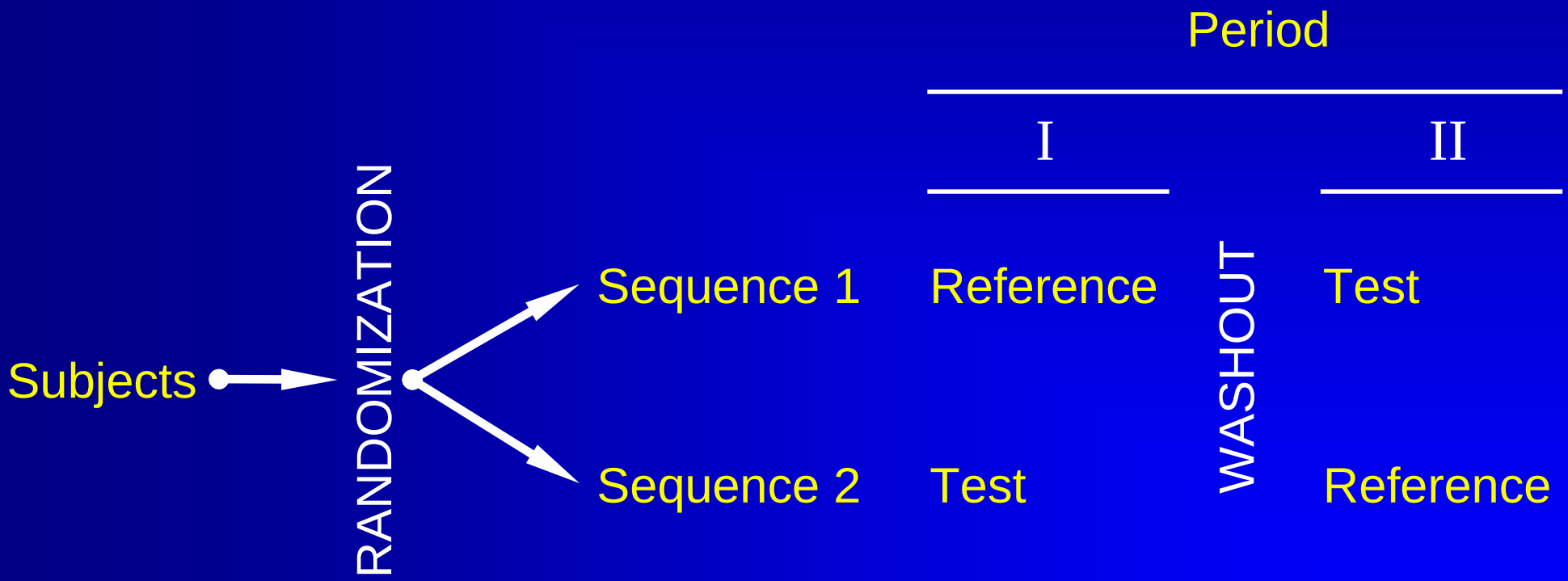
- Steady-state studies
  - Demonstration of steady-state (cont'd)
    - Multivariate method (simultaneous testing of all pre-dose values in all subjects)
      - *E.g.*, Hotellings T^2
      - Benefit additional statement possible *when* steady-state was obtained
      - Drawback: if significant result, no possibility to exclude particular subjects (rendering the entire study worthless).
    - *t*-test of last two pre-dose values
      - Pro: most easy to perform, relatively insensitive to outliers
      - Con: as above

Cross-over designs

- Standard 2×2×2 (two-treatment two-sequence two-period) design
  - Each subject is randomly assigned to either sequence RT or sequence TR at two treatment periods
    - Dosing periods are separated by a washout period of sufficient length for the drug received in the first period to be completely metabolized or excreted from the circulation.
    - Smaller subject numbers compared to a parallel design, since the *within-subject* variability determines sample size (rather than *between-subject* variability).

Cross-over designs

- Standard 2×2×2 design



Assumptions: Cross-over

Multiplicative Model (X-over without carryover)

$$X_{ijk} = \mu \cdot \pi_k \cdot \Phi_l \cdot s_{ik} \cdot e_{ijk}$$

- All $ln\{s_{ik}\}$ and $ln\{e_{ijk}\}$ are independently and normally distributed about unity with variances σ^2_s and σ^2_e.
  - ➔ This assumption may not hold true for all formulations; if the reference formulation shows higher variability than the test formulation, a *'good'* test will be penalized for the *'bad'* reference.
- All observations made on different subjects are independent.
  - ➔ This assumption should not be a problem, unless you plan to include twins or triplets in your study…

Cross-over designs

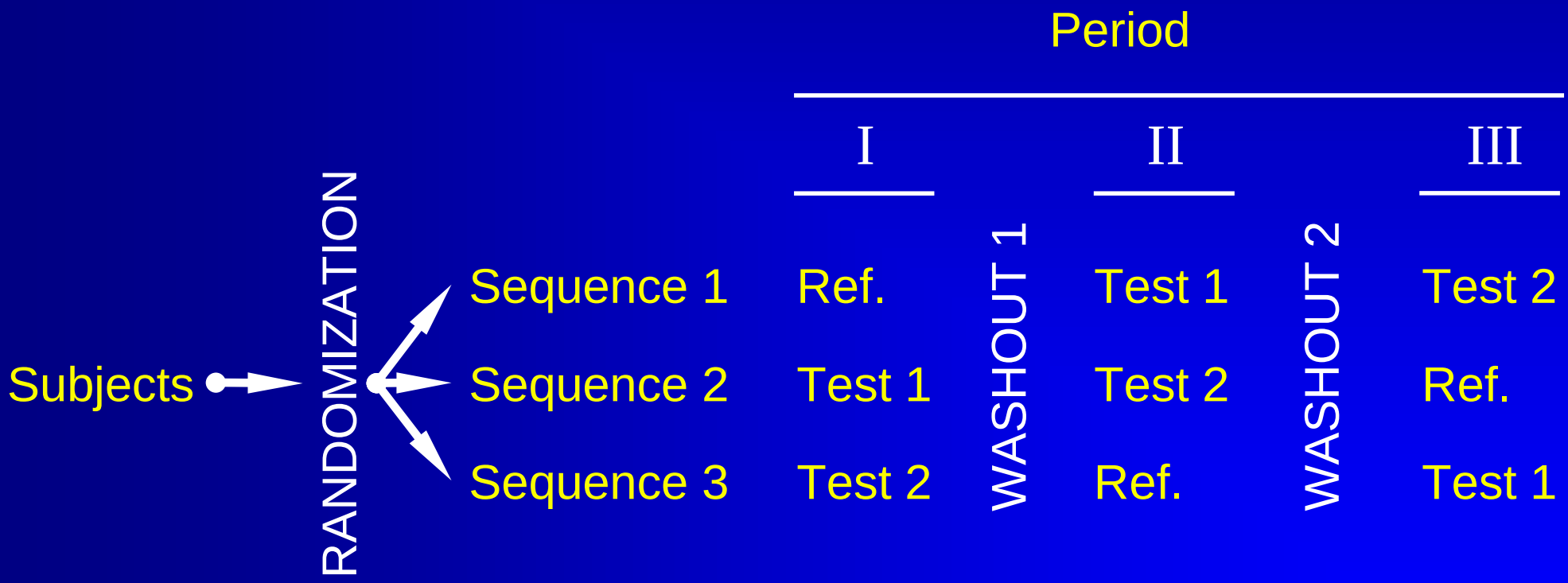
- Standard 2×2×2 design
  - Advantages
    - Globally applied standard protocol for BE
    - Straigthforward statistical analysis
  - Disadvantages
    - Not suitable for drugs with long half life (→ parallel groups)
    - Not optimal for studies in patients with instable diseases (→ parallel groups)
    - Not optimal for HVDs (→ Replicate Designs)

Cross-over designs

- Higher Order Designs (for more than two treatments)
  - Latin Squares
    Each subject is randomly assigned to sequences, where number of treatments = number of sequences = number of periods.

Cross-over designs

- 3×3×3 Latin Square design



Cross-over designs

- 3×3×3 Latin Square design
  - Advantages
    - Allows to choose between two candidate test formulations or comparison of a test formulation with two references
    - Easy to adapt
    - Number of subjects in the study is a multiplicative of three
    - Design for establishment of Dose Proportionality
  - Disadvantages
    - Statistical analysis more complicated (especially in the case of drop-outs and a small sample size) – not available in some softwares
    - Extracted pairwise comparisons are imbalanced
    - May need measures against multiplicity (increasing the sample size)
    - Not mentioned in any guideline

Cross-over designs

- Higher Order Designs (for more than two treatments)
  - Variance Balanced Designs (Williams' Designs)
    - For *e.g.*, three formulations there are three possible pairwise differences among formulation means (*i.e.*, form. 1 *vs.* form. 2., form 2 *vs.* form. 3, and form. 1 *vs.* form. 3)
    - It is desirable to estimate these pairwise effects with the same degree of precision (there is a common variance for each pair)
      - Each formulation occurs only once with each subject
      - Each formulation occurs the same number of times in each period
      - The number of subjects who receive formulation *i* in some period followed by formulation *j* in the next period is the same for all *i # j*
    - Such a design for three formulations is the three-treatment six-sequence three-period Williams' Design

Cross-over designs

- Williams' Design for three treatments

Sequence	Period		
	I	II	III
1	R	T_2	T_1
2	T_1	R	T_2
3	T_2	T_1	R
4	T_1	T_2	R
5	T_2	R	T_1
6	R	T_1	T_2

Cross-over designs

- Williams' Design for four treatments

Sequence	Period			
	I	II	III	IV
1	R	T_3	T_1	T_2
2	T_1	R	T_2	T_3
3	T_2	T_1	T_3	R
4	T_3	T_2	R	T_1

Cross-over designs

- Williams' Designs
  - Advantages
    - Allows to choose between two candidate test formulations or comparison of a test formulation with two references
    - Design for establishment of Dose Proportionality
    - Paired comparisons (*e.g.*, for a nonparametric method) can be extracted, which are also balanced
    - Mentioned in Brazil's (ANVISA) guideline
  - Disadvantages
    - Mores sequences for an *odd* number of treatment needed than in a Latin Squares design (but equal for even number)
    - Statistical analysis more complicated (especially in the case of drop-outs) – not available in some softwares
    - May need measures against multiplicity (increasing the sample size)

Cross-over designs

- Extraction of 2×2 comparisons (T_1/R, T_2/R)
  - Latin Squares

Seq.	P_1	P_2	P_3
1	T_1	T_2	R
2	T_2	R	T_1
3	R	T_1	T_2

Seq.	P_1'	P_2'
1	T_1	R
2	R	T_1
3	R	T_1

Seq.	P_1"	P_2"
1	T_2	R
2	T_2	R
3	R	T_2

imbalanced

  - Williams' design

Seq.	P_1	P_2	P_3
1	T_1	T_2	R
2	T_2	R	T_1
3	R	T_1	T_2
4	T_1	R	T_2
5	T_2	T_1	R
6	R	T_2	T_1

Seq.	P_1'	P_2'
1	T_1	R
2	R	T_1
3	R	T_1
4	T_1	R
5	T_1	R
6	R	T_1

Seq.	P_1"	P_2"
1	T_2	R
2	T_2	R
3	R	T_2
4	R	T_2
5	T_2	R
6	R	T_2

balanced

Cross-over designs

- Higher Order Designs (cont'd)
  - Bonferroni-correction needed (sample size!)
    - *If more than one formulation will be marketed* (for three simultaneous comparisons without correction patients' risk increases from 5 % to 14 %).
    - *Sometimes* requested by regulators in dose proportionality.

k	$P_{\alpha=0.05}$	$P_{\alpha=0.10}$	$\alpha_{adj.}$	$P_{\alpha adj.}$	$\alpha_{adj.}$	$P_{\alpha adj.}$
1	5.00%	10.00%	0.0500	5.00%	0.100	10.00%
2	9.75%	19.00%	0.0250	4.94%	0.050	9.75%
3	14.26%	27.10%	0.0167	4.92%	0.033	6.67%
4	18.55%	34.39%	0.0125	4.91%	0.025	9.63%
5	22.62%	40.95%	0.0100	4.90%	0.020	9.61%
6	26.49%	46.86%	0.0083	4.90%	0.017	9.59%

Cross-over designs

- Replicate designs
  - Each subject is randomly assigned to sequences, where at least one of the treatments is administered at least twice.
    - Not only the global within-subject variability, but also the within-subject variability per treatment may be estimated.
    - Smaller subject numbers compared to a standard 2×2×2 design – but outweighed by the increased number of periods.
    - Same overall number of individual treatments!
    - Mandatory in the EU if an extended acceptance range for C_{max} (0.75–1.33) is aimed at (HVDP must be demonstrated in advance)

Cross-over designs

- Replicate designs
  - Advantages
    - Some experience from FDA's initiative on population BE (PBE) and individual BE (IBE)
    - Reference scaling average bioequivalence (RSABE)
    - Handling of outliers (subject-by-formulation interaction may be ruled out)
  - Disadvantages
    - Statistical analysis complicated (especially in the case of drop-outs and if RSABE is the target) – not available in standard software
    - Many publications, but still no agreement on methodology
    - Mentioned only in South African GL; will be adopted by FDA

Cross-over designs

- Replicate designs
  - Examples
    - Two-sequence three-period
      T R T
      R T R
      Sample size to obtain the same power as a 2×2×2 study: 75 %
    - Two-sequence four-period
      T R T R
      R T R T
      Sample size to obtain the same power as a 2×2×2 study: 50 %
    - and many others… (FDA for RSABE: TRR–RTR–RRT)
    - The statistical model is a little bit complicated – and dependent on the actual design

$$X_{ijkl} = \mu \cdot \pi_k \cdot \Phi_l \cdot s_{ij} \cdot e_{ijkl}$$

HVDs/HVDPs

- Highly Variable Drugs / Drug Products (intra-subject variability >30 %)
  - ✓ USA Replicate Design recommended.
    Reference Scaled Average Bioequivalence under discussion: minimum number of subjects (24 or 36), restriction on GMR (0.8–1.25)
  - ± EU […] under certain circumstances […] alternative well-established designs could be considered such as […] replicate designs for substances with highly variable disposition.
    Widening of acceptance range in a pivotal BEstudy (for C_{max} only) after demonstration of reference HVDP (pilot replicate design).
    RSABE according to the Draft GL not acceptable.

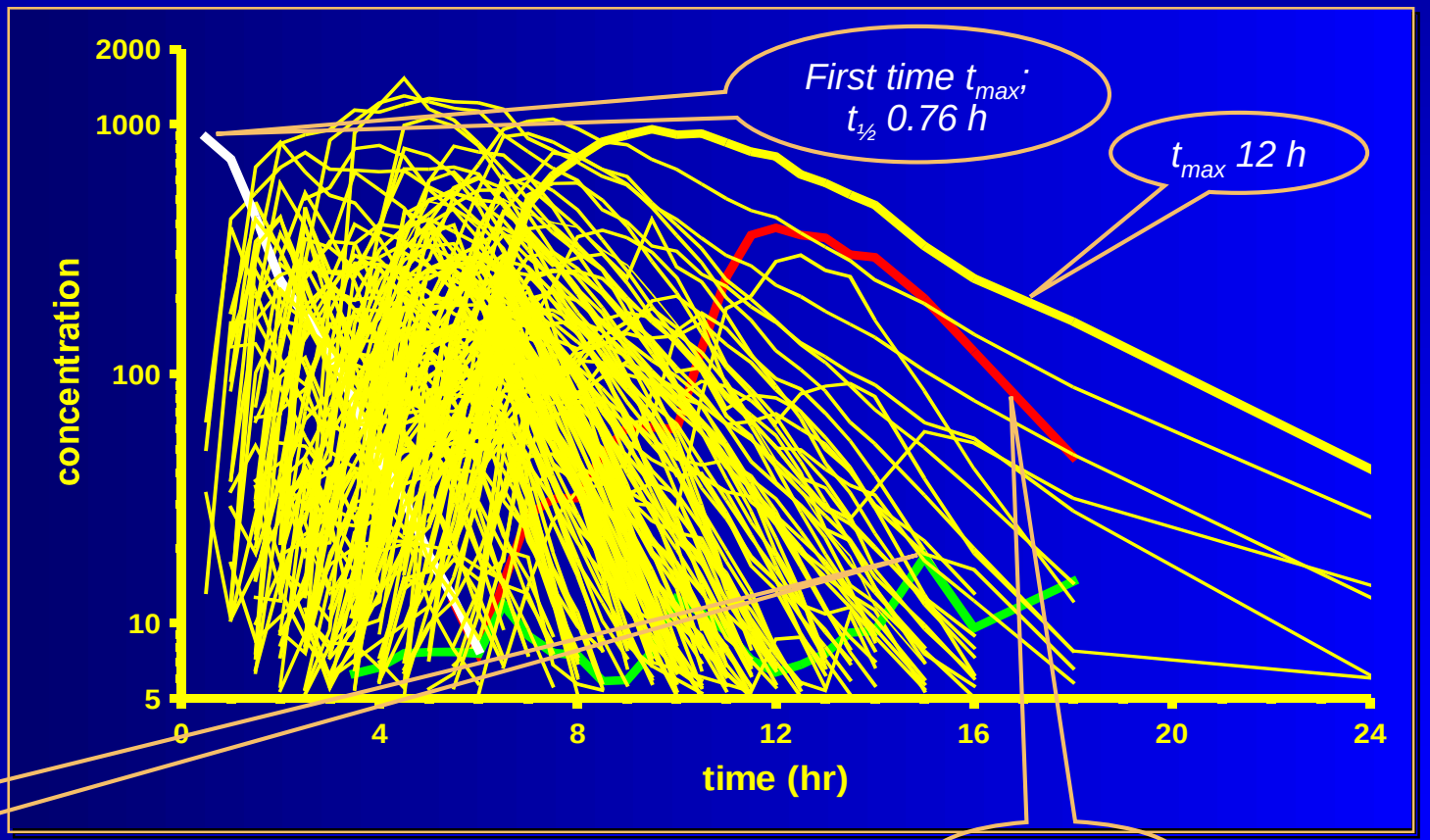
HVDs/HVDPs

- Does knowledge of the PK profile always help in demonstrating bioequivalence when a conventional BE study is unsuitable?
  - Omeprazole: Highly Variable Drug Product (HVDP), higher variability in fed state as compared to fasted state commonly observed, sensitive to low pH, breakdown of gastric resistant coating (especially of the reference product) not unusual, high variability in C_{max}/t_{max} due to gastric emptying, …

HVDs/HVDPs

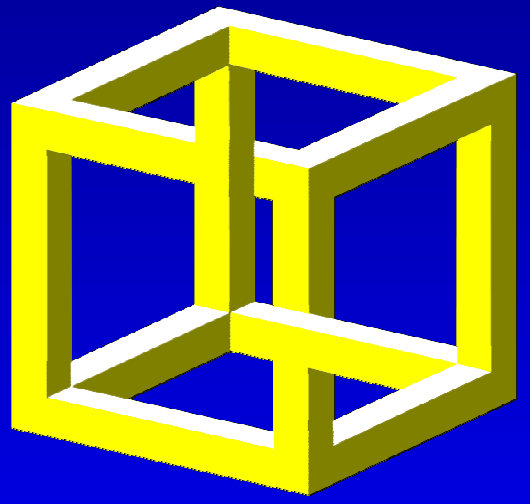
- Attempt to deal with high variability in C_{max}

Powered to 90% according to CV from previous studies; 140 (!) subjects and to 80% for expect-ed dropout rate. Sampling every 30 min up to 14 hours (7785 total).



HVDs/HVDPs

- Ways out?
  - Replicate designs could be considered e.g. for substances with highly variable pharmacokinetic characteristics. (EU BE Draft, Section 4.1.2)
  - Nonparametric methods
    A non-parametric analysis is **not acceptable**. (BE Draft, Section 4.1.8)
  - Compartmental (Population PK) methods
    The use of compartmental methods for the estimation of parameters is **not acceptable**. (BE Draft, Section 4.1.5)



HVDPs

- All (!) ANDAs submitted to FDA/OGD 2003–2005 (1010 studies, 180 drugs)
  - 31% (57/180) highly variable (CV ≥30%)
  - of these HVDs/HVDPs,
    - 60% due to PK (*e.g.*, first pass metabol.)
    - 20% formulation performance
    - 20% unclear

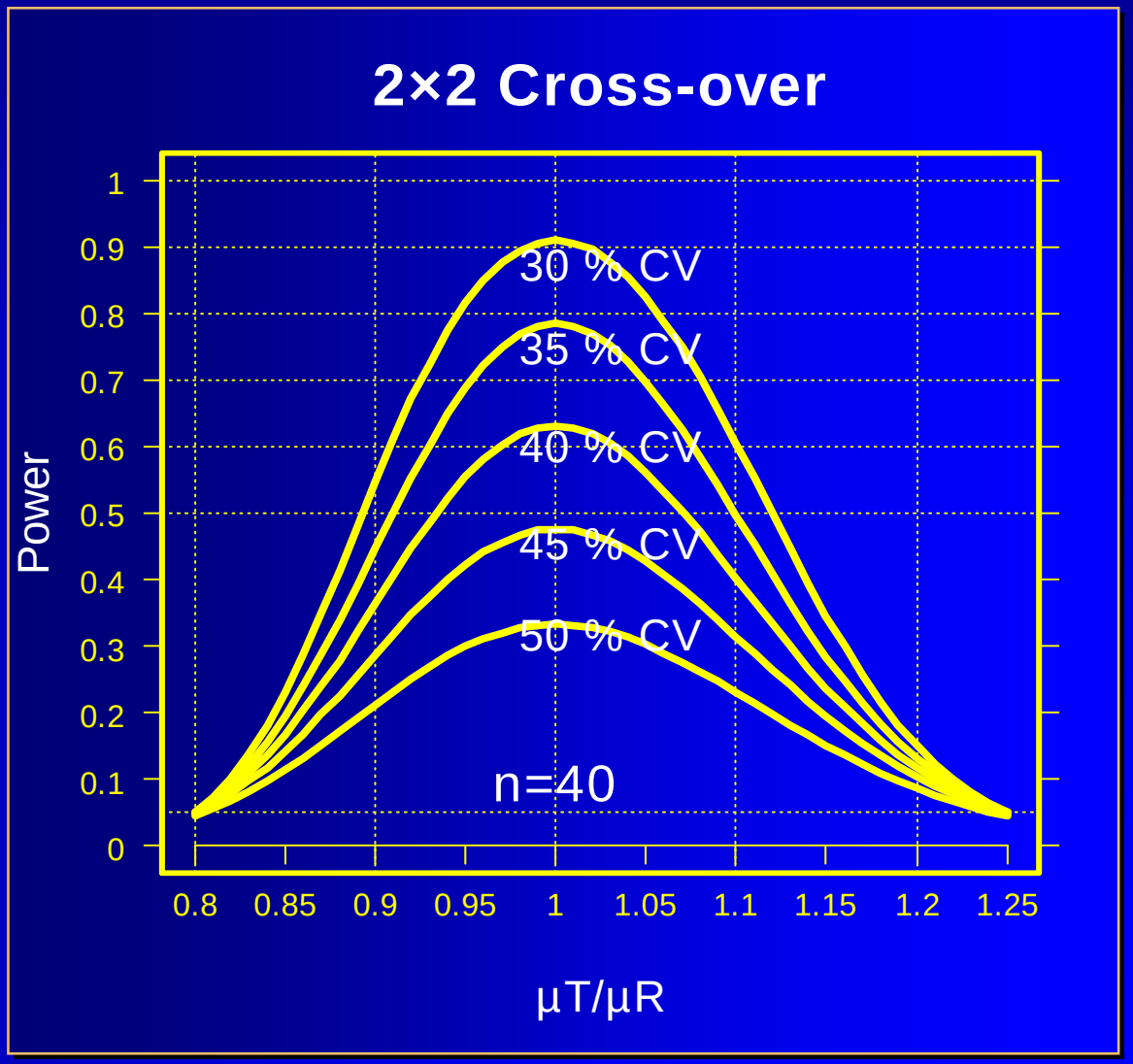
Davit BM, Conner DP, Fabian-Fritsch B, Haidar SH, Jiang X, Patel DT, Seo PR, Suh K, Thompson CL, and LX Yu
Highly variable drugs: observations from bioequivalence data submitted to the FDA for new generic drug applications
AAPS J 10(1): 148-56 (2008)

HVDPs

Power to show BE with 40 subjects for CV_{intra} = 30–50%

μT/μR 0.95, CV_{intra} 30%
→ power 0.816
μT/μR 1.00, CV_{intra} 45%
→ power 0.476 <
Roulette 0.486 (!)

μT/μR 0.95, CV_{intra} 45%
→ n=82 (power 0.807)

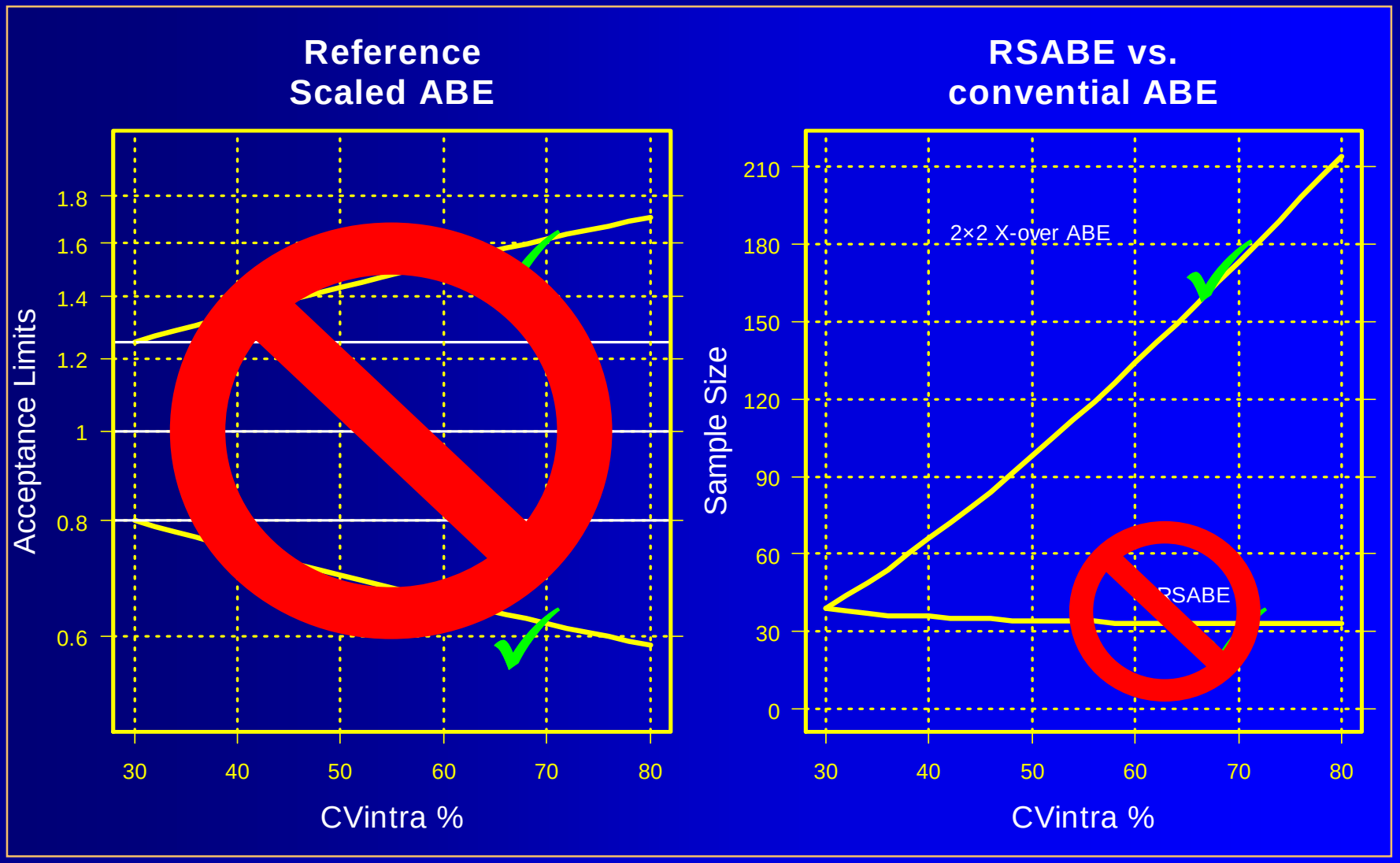


HVDPs (US/EU)

- Advisory Committee for Pharmaceutical Sciences (ACPS) to FDA (10/2006) on HVDs
- Follow-up paper in 2008 (likely to be implemented in next Guideline)
  - Replicate study design [TRR–RTR–RRT]
  - Reference Scaled Average Bioequivalence (RSABE)
  - Minimum sample size 24 subjects
  - Point estimate restricted to [0.80,1.25]

Haidar SH, Davit B, Chen M-L, Conner D, Lee LM, Li QH, Lionberger R, Makhlouf F, Patel D, Schuirmann DJ, and LX Yu
Bioequivalence Approaches for Highly Variable Drugs and Drug Products
Pharmaceutical Research 25/1, 237-241 (2008)
http://www.springerlink.com/content/u503p62056413677/fulltext.pdf

HVDPs (US/EU)


Reference
Scaled ABE
Acceptance Limits
1.8
1.6
1.4
1.2
1
0.8
0.6
30
40
50
60
70
80
CVintra %
RSABE vs.
convential ABE
Sample Size
210
180
150
120
90
60
30
0
2×2 X-over ABE
RSABE
30
40
50
60
70
80
CVintra %

HVDs/HVDPs

- Is suggested EU-method of any good?
  - Replicate designs … (BE Draft, Section 4.1.2) **without scaling**
    - reduce the number of subjects (to 75% for a 3-period design and to 50% for a 4-period design as compared to a conventional 2×2),
    - but keep the *theoretical* number of treatments constant:
      - The potentional drop-out rate increases.
      - Practically *more* treatments must be administered in order to maintain the desired power!

HVDs/HVDPs

- Example
  - AR [0.80,1.25], CV_{intra} 49.5%, T/R 0.95%, power 80%, $n_{2\times2}$ 96
  - expected dropout rate of 10% per washout
    - 2×2 study: 96+10=106 subjects, 212 treatments
    - 4×2 study: 48+16=64 subjects, 256 treatments

Ethical?

  - Proposed FDA Scaling-Method: AR [0.7006,1.4273], PE [0.80,1.25], n 34 (!)

HVDPs: $C_{ss,min}$

- EMEA Draft BE Guideline (2008)
  - Acceptance limits
    - [...] at steady state AUCτ, $C_{max,ss}$, and $C_{min,ss}$ should be analysed using the same acceptance interval as stated above.
      - $C_{min,ss}$ was added probably after concerns for oxyco-done, but this metric will be rather tough to meet for some drugs.
      - Since scaling is not allowed, sample sizes are expected to be *very* high (for HVDPs even in steady state the variability of $C_{ss,min}$ » $C_{ss,max}$).

Early Exposure

- Partial AUCs for Rapid Onset Drugs
  - US-FDA 2003 (III.A.8.a.)
    - […] that the partial area be truncated at the *population median of T_{max} values for the reference formulation*. We also recommend that at least two quantifiable samples be collected before the expected peak time to allow adequate estimation of the partial area.
  - Canada-TGD 2005
    - […] $AUC_{Reftmax}$ for a test product is defined as the area under the curve to the time of the maximum concentration of the reference product, *calculated for each study subject*.

Early Exposure

- Partial AUCs for Rapid Onset Drugs (cont'd)
  - EU-EMEA BE Draft 2008
    - When partial AUC is to be determined, frequent early sampling is recommended with preferably at least two quantifiable samples before expected t_{max}. […] partial AUCs can be used as a measure of early exposure. The partial area can in most cases be truncated at the population median of t_{max} values for the reference formulation. However, an alternative time point for truncating the partial AUC can be used when clinically relevant. *The time point for truncating the partial AUC should be pre-specified and justified in the study protocol*.

Early Exposure (HVDP?)

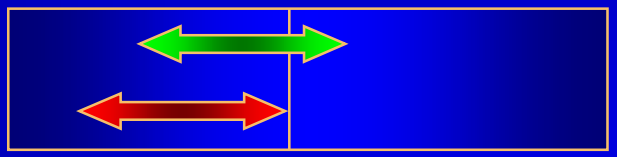
- Partial AUCs for Rapid Onset Drugs (cont'd)

Example	median t_{maxref}	PE	nonparametric CI		BE	FDA	parametric CI		BE	TGD	parametric CI		BE
1	1.5 h	±0.00 h	-0.25 h (85%)	+0.25 h (115%)	yes	90.1%	75.0%	110.1%	no (CV 26.4%)	85.7%	72.3%	102.3%	no (CV 23.8%)
2	1.5 h	+0.26 h	±0.00 h (100%)	+0.50 h (130%)	no	66.1%	53.1%	82.0%	no (CV 29.7%)	62.4%	46.2%	84.3%	no (CV 42.4%)

- Even for formulations with low intra-subject variability
  - Example 1: AUC_t 13.3% C_{max} 17.0%
  - Example 2: AUC_t 6.33% C_{max} 9.43%
- it was not possible to demonstrate BE due to high variability of this metric. It's unclear how median t_{maxref} can be stated in the protocol (EMEA) – the innovator's SmPC (=label) often states the *arithmetic mean* only.

Low Variability

- Drugs / Drug Products with CV_{intra} <10%
  - No specific statements in any guideline.
  - Problems may arise according to significant treatment effects in ANOVA (*i.e.*, although the 90% CI is within the acceptance range – 100% is not included) – even for the minimum sample size of 12.



  - Denmark
    - DKMA considers that the 90% CI for the ratio test versus reference should include 100% […].
    - Deviations may be accepted if they can be adequately justified not to have impact on either the overall thera-peutic effect or safety profile of the product.

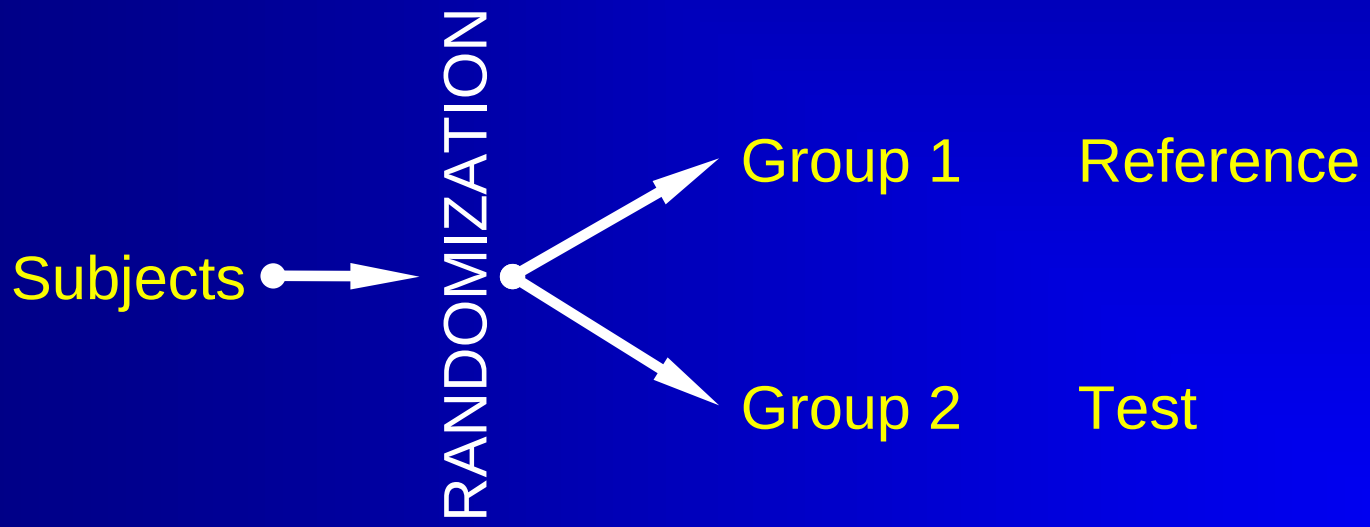
Danish Medicines Agency (DKMA)
Bioequivalence and labelling of medicinal products with regard to generic substitution (Jan 2006)
http://www.dkma.dk/1024/visUKLSArtikel.asp?artikelID=6437

Parallel Groups

- Two-group parallel design
  - Each subject receives one – and only one –treatment in a random fashion
    - Usually each group contains the same number of subjects.
    - Higher subject numbers compared to a cross-over design, since the *between-subject* variability determines sample size (rather than *within-subject* variability)

Parallel Groups

- Two-group parallel design



Parallel Groups

- Two-group parallel design
  - Advantages
    - Clinical part – *sometimes* – faster than X-over
    - Straigthforward statistical analysis
    - Drugs with long half life
    - Potentially toxic drugs or effect and/or AEs unacceptable in healthy subjects
    - Studies in patients, where the condition of the disease irreversibly changes
  - Disadvantages
    - Lower statistical power than X-over (*rule of thumb:* sample size should at least be doubled)
    - Phenotyping mandatory for drugs showing polymorphism

Parallel Groups

- Design Issues
  - EMEA NfG on BA/BE (2001)
    - 3.2.4 Genetic phenotyping
      ‘Phenotyping and/or genotyping of subjects should be considered for […] all studies using parallel group design.
      If a drug is known to be subject to major genetic polymorphism, studies could be performed in panels of subjects of known phenotype or genotype for the polymorphism in question.’
  - Since the comparison is based on *inter-subject* effects
    - One study of the major phenotype/genotype
    - Two studies of the respective phenotype/genotype – only if requested!

Parallel Groups

- Evaluation
  - FDA/CDER, Statistical Approaches to Establishing Bioequivalence (2001)
    - Section VI. B.1.d. Parallel Designs
      'For parallel designs, the confidence interval for the difference of means in the log scale can be computed using the total between-subject variance. As in the analysis for replicated designs (section VI. B.1.b), equal variances should not be assumed.'
  - The conventional *t*-test depends on the assumption that samples come from populations that have identical variances
    - 'Naive pooling' of variances is relatively robust against unequal variances, but rather sensitive to inbalanced data
    - If assumptions are violated, the conventional *t*-test becomes liberal (*i.e.*, the CI is too tight; patient's risk > 5 %).

Sample Data Set

- Will be used throughout the lecture
- 2×2×2 Cross-over Study
  - 24 subjects (balanced: TR=RT=12)
  - Single dose
  - Target parameter: AUC_{0-t}
  - CV_{intra} 20.0 %
  - CV_{inter} 32.6 %
  - http://bebac.at/downloads/24sub.txt (CSV-format)

Trt	Rand	Sub	P_1	P_2
1	RT	1	44.1	39.1
1	RT	2	33.6	23.8
1	RT	3	45.5	40.8
2	TR	4	19.5	21.1
2	TR	5	67.2	51.5
2	TR	6	25.7	30.1
1	RT	7	35.3	26.7
1	RT	8	26.0	36.5
1	RT	9	38.2	57.8
2	TR	10	33.6	32.5
2	TR	11	25.1	36.8
2	TR	12	44.1	42.9
1	RT	13	25.6	20.1
1	RT	14	58.0	45.3
1	RT	15	47.2	51.8
2	TR	16	16.5	21.4
2	TR	17	47.3	39.4
2	TR	18	22.6	17.3
1	RT	19	17.5	30.1
1	RT	20	51.7	36.0
1	RT	21	24.5	18.2
2	TR	22	36.3	27.2
2	TR	23	29.4	39.6
2	TR	24	18.3	20.7

Parallel Groups: Example

- Evaluation (sample data set, period 1 only)
  - Original data set
    - Balanced (T 12, R 12)
    - Equal variances (s^2_R 0.1292, s^2_T 0.1796)
      F-ratio test *p* 0.5947
      Levene test *p* 0.5867
  - Modified data set
    - Values of subjects 4 – 6 multiplied by three
    - Subjects 22 – 24 removed
    - Inbalanced (T 9, R 12)
    - Unequal variances (s^2_R 0.1292, s^2_T 0.5639)
      F-ratio test *p* 0.0272
      Levene test *p* 0.1070

Parallel Groups: Example

- Evaluation (original data set)
  - Is your software able to give the correct answer?

Software / Method	equal variances	unequal variances
'manual' (Excel 2000)	63.51% – 110.19%	63.48% – 110.25%
R 2.7.0 (2008)	63.51% – 110.19%	63.49% – 110.22%
NCSS 2001 (2001)	63.51% – 110.19%	63.49% – 110.22%
STATISTICA 5.1H (1997)	63.51% – 110.19%	63.49% – 110.22%
WinNonlin 5.2.1 (2008)	63.51% – 110.20%	*not implemented!*
Kinetica 4.4.1 (2007)	63.51% – 110.19%	*not implemented!*
EquivTest/PK (2006)	63.51% – 110.18%	*not implemented!*

Parallel Groups: Example

- Evaluation (modified data set)

Software	equal variances	unequal variances
R 2.7.0 (2008)	81.21% – 190.41%	76.36% – 202.51%
NCSS 2001 (2001)	81.21% – 190.41%	76.36% – 202.51%

  - Inflated α-risk in 'conventional' *t*-test (naive pooling) is reflected in a tighter confidence interval.
  - Preliminary testing for equality in variances is flawed*) and should be avoided (FDA).
  - Approximations (*e.g.*, Satterthwaite, Aspin-Welch, Howe, Milliken-Johnson) are currently *not implemented* in packages 'specialized' in bioequivalence testing (WinNonlin, Kinetica, EquivTest/PK)!

*) **Moser BK and GR Stevens**
Homogeneity of variance in the two-sample means test
Amer Statist 46:19-21 (1992)

Sample Size Estimation

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

Sample Size Estimation

- Rationale for Pilot Studies (FDA/CDER, BA/BE Studies – General Considerations, 2003)
  - Validation of analytical methodology
  - Assessment of variability
  - Optimization of sample collection time intervals
  - 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.

Sample Size Estimation

- 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.'

Sample Size Estimation

- 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.

Sample Size Estimation

- NfG on the Investigation of BA/BE
  - 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.*, *prazoles?).

Sample Size Estimation

- NfG on the Investigation of BA/BE
  - 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

Sample Size Estimation

- NfG on the Investigation of BA/BE
  - 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 restrictive legislations (*e.g.*, Brazil's ANVISA), α must be tightened to 2.5 % for NTIDs (95 % confidence interval)

Sample Size Estimation

- NfG on the Investigation of BA/BE
  - 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!

Sample Size Estimation

- NfG on the Investigation of BA/BE
  - 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

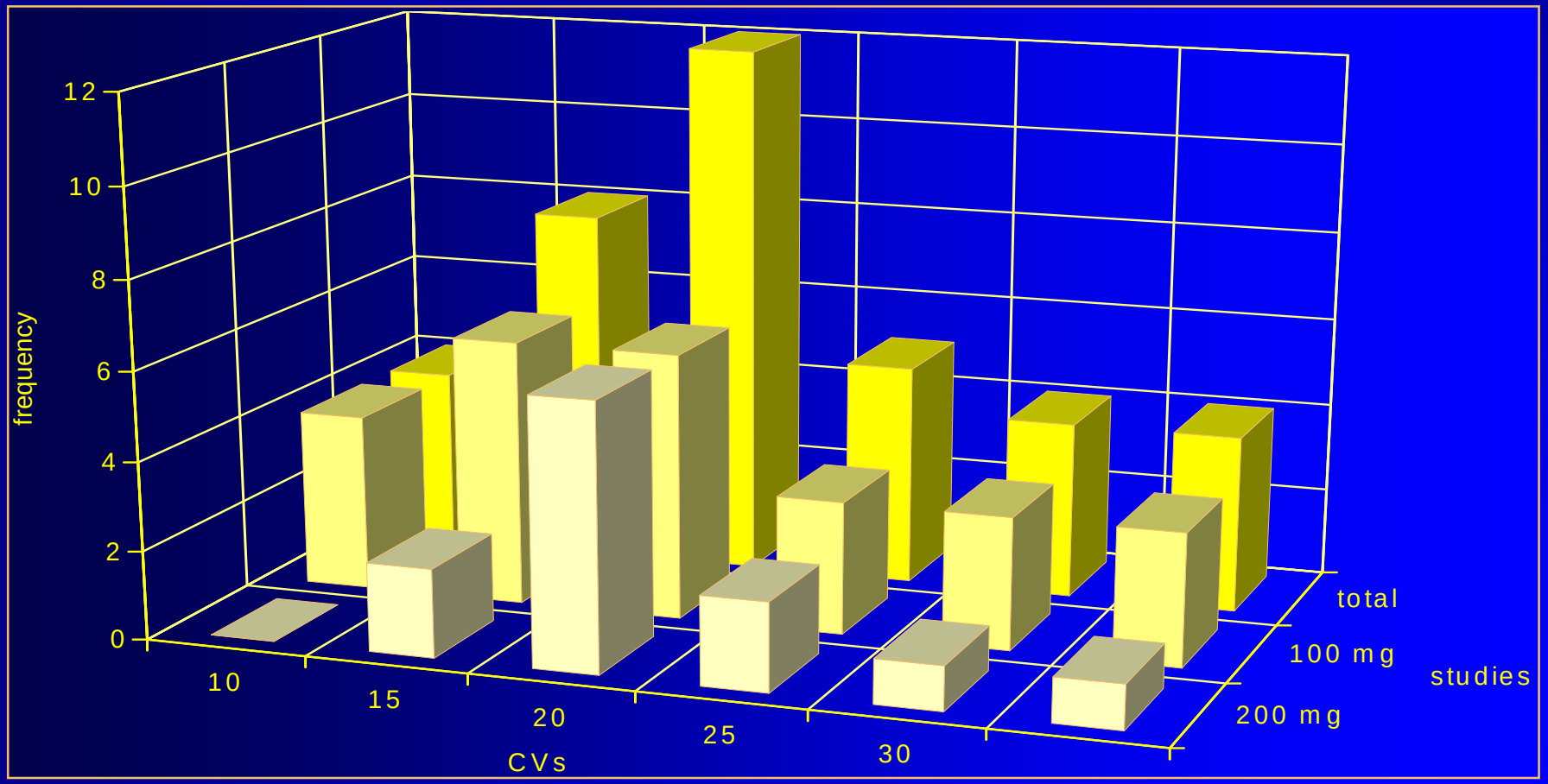
Sample Size Estimation

- Sample size planning (EMEA Draft BE Guideline, 2008)
  - The number of subjects to be included in the study should be based on an ***appropriate*** sample size calculation.

Cookbook?

Sample Size Estimation

Literature data…



Doxicycline *(37 studies ref. by Blume/Mutschler, 1996)*

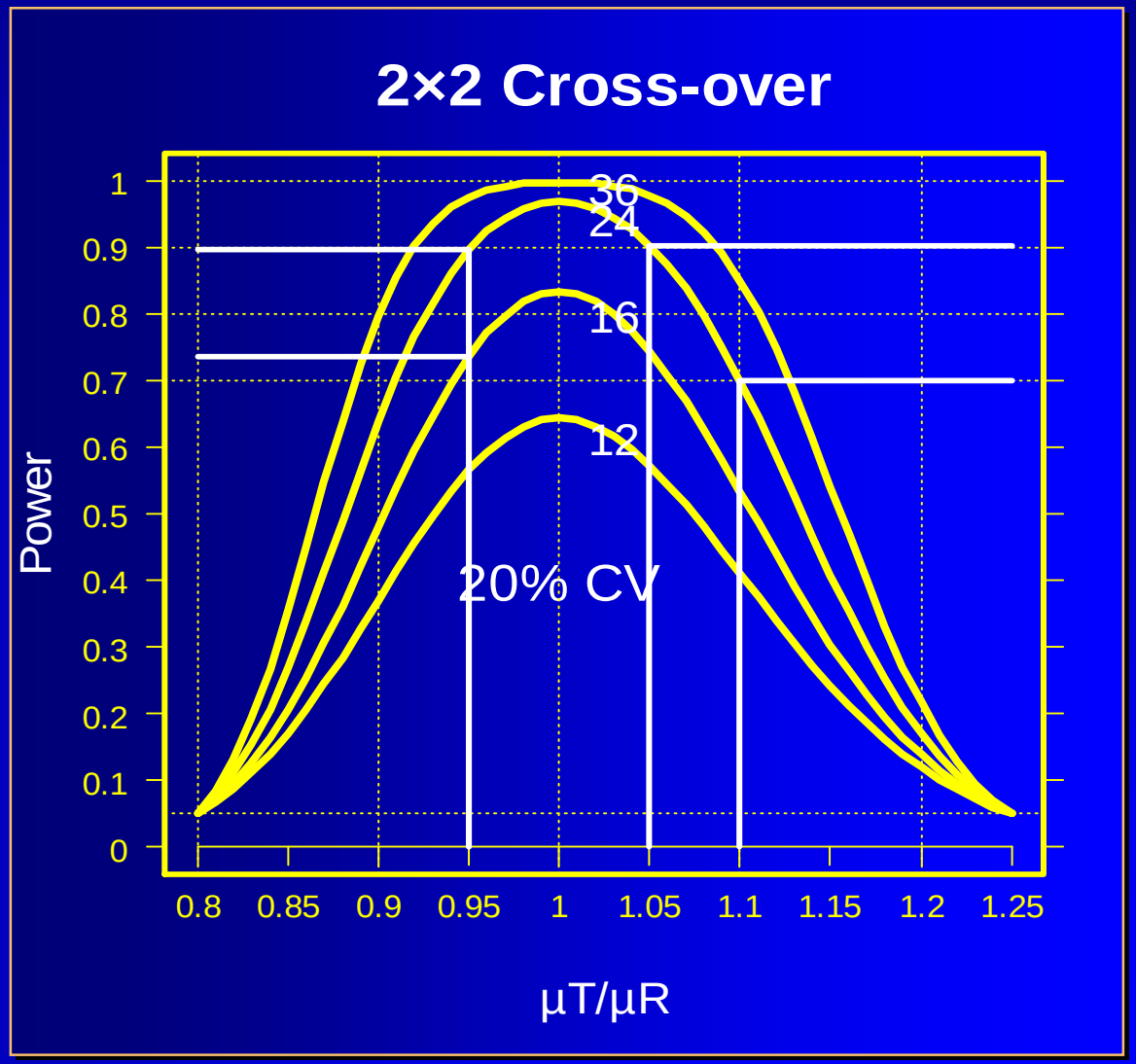
Sample Size Estimation

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



Sample Size: 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.

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.
  - 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.
    - You may also use an upper confidence limit of CV_{intra} in sample size estimation.
    - 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

- EMEA Draft BE Guideline(2008)
  - 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).
    - First stage data should be treated as an interim analysis.

'Internal Pilot Study Design'

Two-Stage Design

- EMEA Draft BE Guideline (2008)
  - Section 4.1.8 (cont'd)
    - Both analyses conducted at adjusted signifi-cance levels (with the confidence intervals accordingly using an adjusted coverage proba-bility which will be higher than 90%).
    - 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.

Two-Stage Design

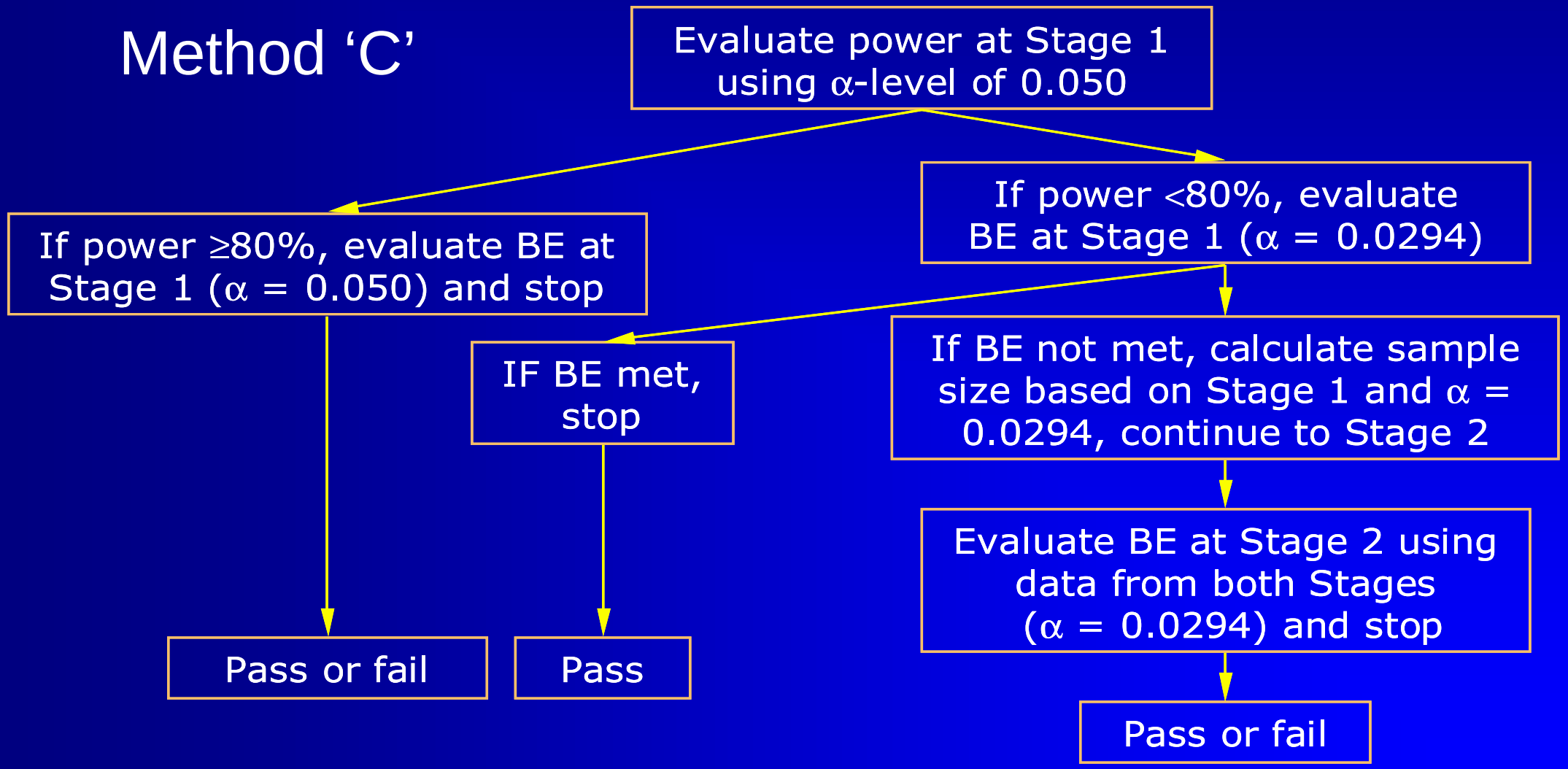
- Critical Remarks
  - ‘BE not been demonstrated’ in initial group:
    If test at $\alpha \leq 0.05$, patient’s risk already ‘spent’!
  - ‘Adjusted significance levels’:
    Bonferroni not validated in BE setting; patient’s risk may be inflated (>0.05)!

likely to be implemented by the FDA

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

Sequential Design

Method 'C'



Outliers

- Problems
  - Parametric methods (ANOVA, GLM) are very sensitive to outliers
    - A single outlier may underpower a properly sized study
    - Exclusion of outliers only possible if procedure stated in the protocol, and reason is justified, *e.g.*,
      - Lacking compliance (subject did not take the medication),
      - Vomiting (up to 2 × t_{max} for IR, at all times for MR),
      - Analytical problems (*e.g.*, interferences in chromatography);
      - Not acceptable if only based on statistical grounds.

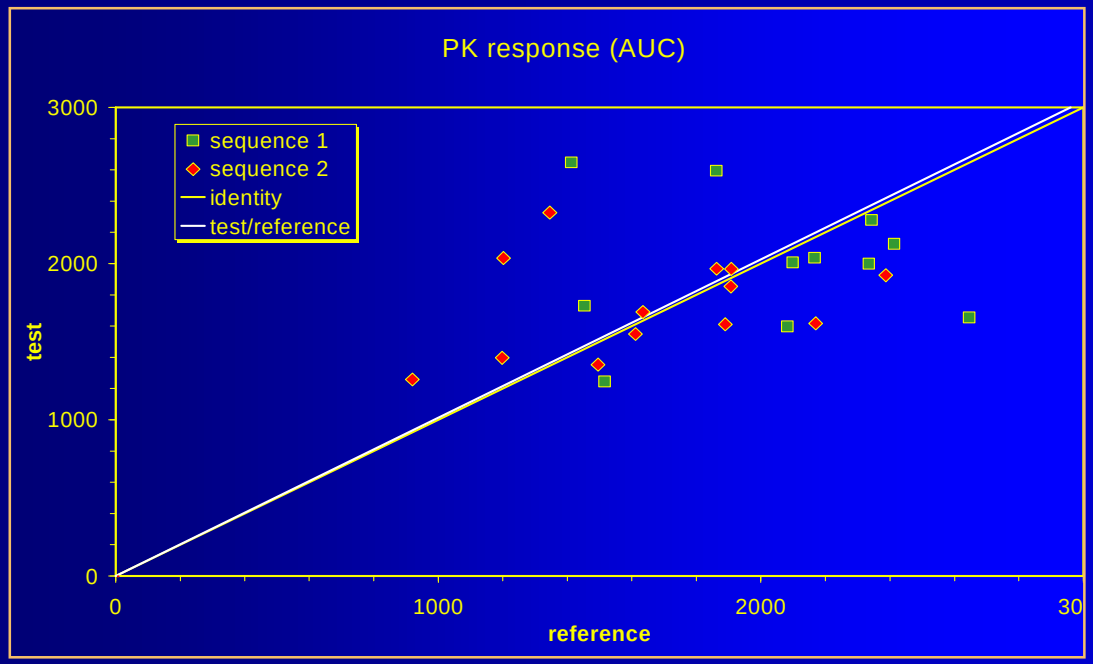
Outliers

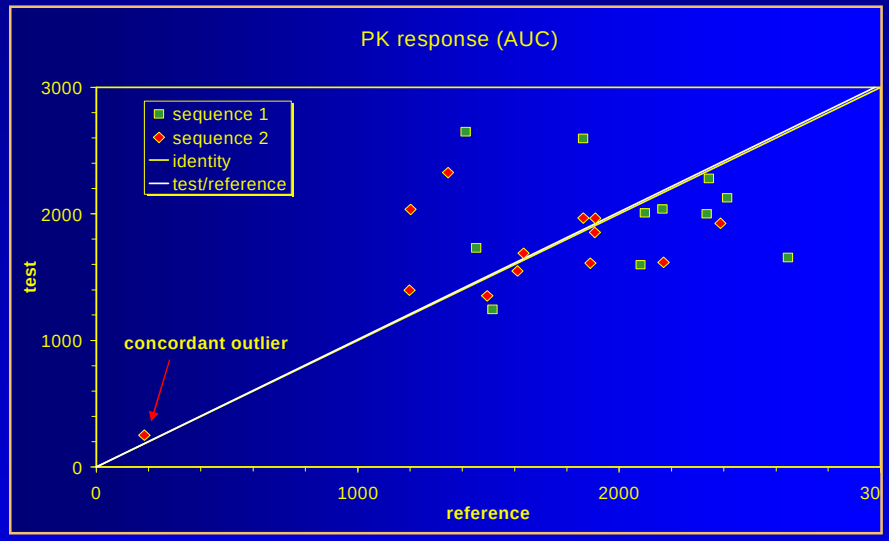
- Types
  - I: Concordant outlier

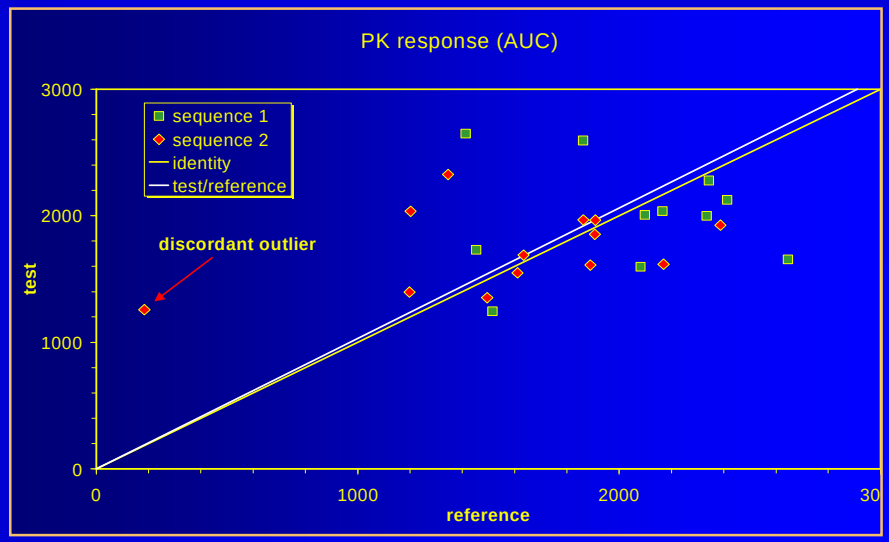
    The PK response for both test and reference deviates from the majority of the study sample.
    - Poor metabolizers may lead to high concentrations in 5-10% of subjects.
    - Does not effect the BE-assessment, but should be discussed (polymorphism known?)
  - II: Discordant outlier

    The PK response of either test or reference deviates form the majority of the study sample.

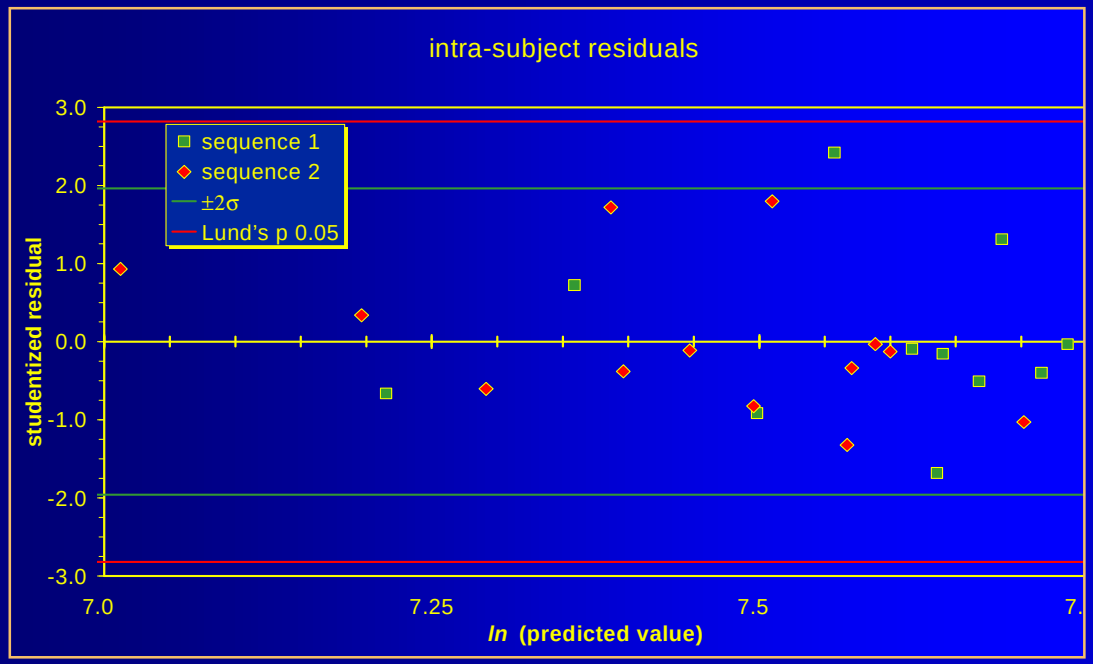
Outliers

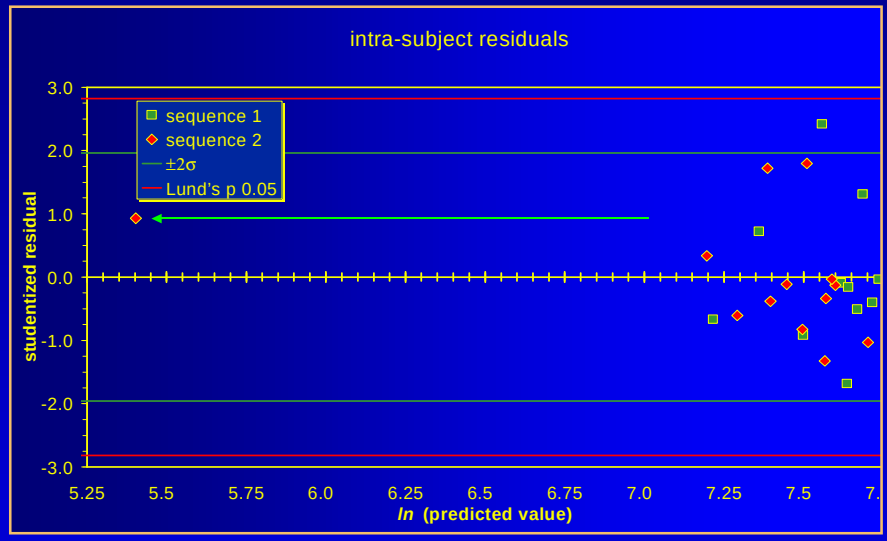

PK response (AUC)
sequence 1
sequence 2
identity
test/reference
test
reference
0
1000
2000
3000

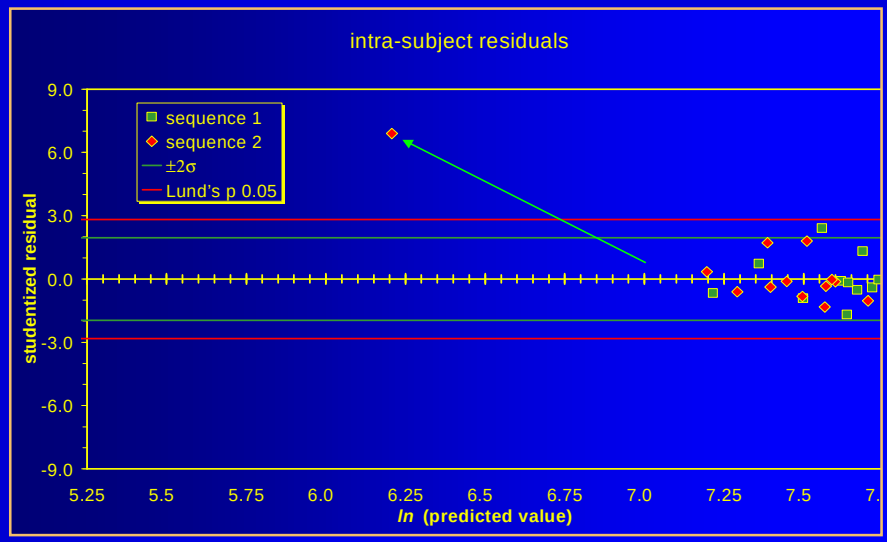

PK response (AUC)
sequence 1
sequence 2
identity
test/reference
concordant outlier
test
reference
0
1000
2000
3000


PK response (AUC)
sequence 1
sequence 2
identity
test/reference
discordant outlier
test
reference
0
1000
2000
3000

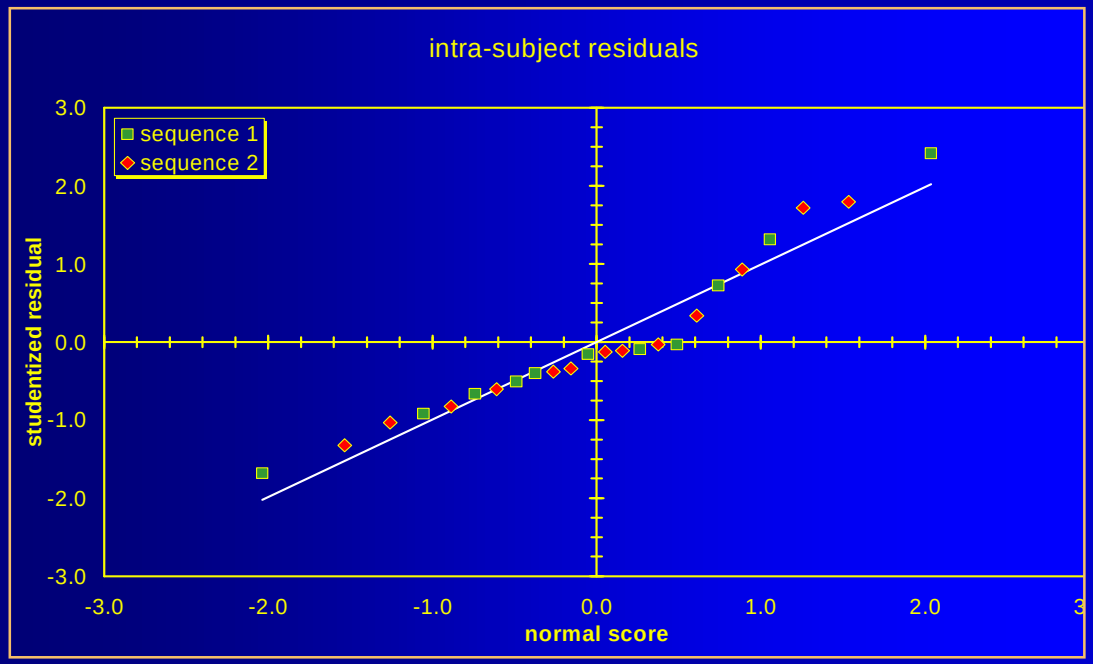
Outliers

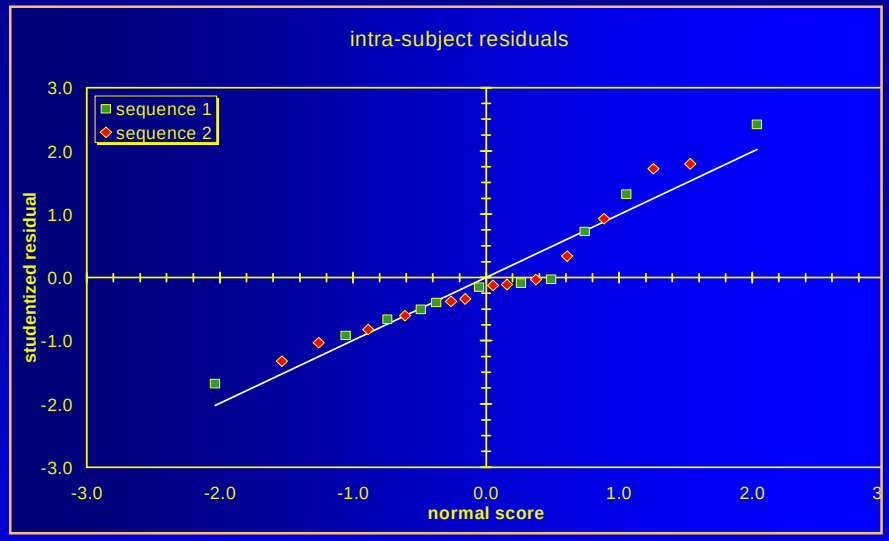

intra-subject residuals
sequence 1
sequence 2
±2σ
Lund's p 0.05
studentized residual
3.0
2.0
1.0
0.0
-1.0
-2.0
-3.0
7.0
7.25
7.5
ln (predicted value)

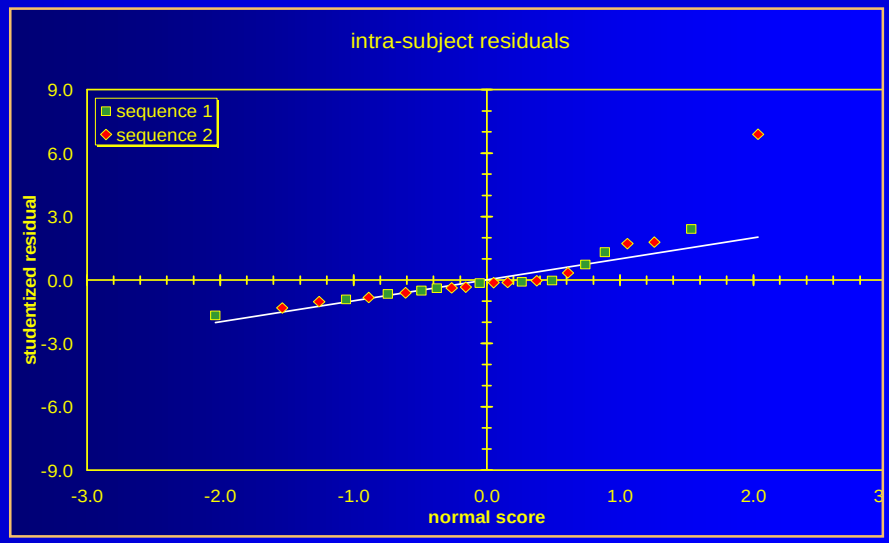

intra-subject residuals
sequence 1
sequence 2
±2σ
Lund's p 0.05
studentized residual
3.0
2.0
1.0
0.0
-1.0
-2.0
-3.0
5.25
5.5
5.75
6.0
6.25
6.5
6.75
7.0
7.25
7.5
ln (predicted value)


intra-subject residuals
sequence 1
sequence 2
±2σ
Lund's p 0.05
studentized residual
9.0
6.0
3.0
0.0
-3.0
-6.0
-9.0
5.25
5.5
5.75
6.0
6.25
6.5
6.75
7.0
7.25
7.5
ln (predicted value)

Outliers


intra-subject residuals
sequence 1
sequence 2
studentized residual
normal score
3.0
2.0
1.0
0.0
-1.0
-2.0
-3.0
-3.0
-2.0
-1.0
0.0
1.0
2.0


intra-subject residuals
sequence 1
sequence 2
studentized residual
normal score
3.0
2.0
1.0
0.0
-1.0
-2.0
-3.0
-3.0
-2.0
-1.0
0.0
1.0
2.0


intra-subject residuals
sequence 1
sequence 2
studentized residual
normal score
9.0
6.0
3.0
0.0
-3.0
-6.0
-9.0
-3.0
-2.0
-1.0
0.0
1.0
2.0

Outliers

- Strategies / Solutions
  - Be prepared to face the unexpected!
  - Examples of drugs/formulations with documented product failures:
    - Drugs sensitive to low pH (gastric resistance!),
    - Monolithic MR products,
    - …
  - Include available information (PK, literature, former studies) in the protocol.
  - Develop a statistical contingency plan.

Outliers

- Solution I
  - Since assumptions of the parametric statistical model are violated, you may apply a statistical method which does not rely on those!
  - Drawback: Lacking regulatory acceptance of nonparametric methods in many countries…
    - ☺ WHO (Technical Report Series No. 937, Annex 9, Section 6.8, May 2006)
    - ☺ Japan NIHS (Bioequivalence Studies for Generic Products, Q&A Document, November 2006)
    - ☹ All other regulatory agencies

Outliers

- Solution II
  - Stay with the parametric method, but
    - evaluate both the full data set and the reduced data set (outliers excluded) and discuss influence on the outcome of the study.
  - In accordance with EMEA's Q&A #3:
    - Exceptional reasons may justify post-hoc data exclusion […]. In such a case, the *applicant must demonstrate that the condition stated to cause the deviation is present in the outlier(s) only* and absence of this condition has been investigated using the same criteria for all other subjects.
    - Results of statistical analyses with and without the group of excluded subjects should be provided.

Practically impossible!

Re-testing of subjects

- If you suspect a product failure of the reference formulation, you may consider re-testing;
  - the outlying subject should be re-tested
    - with both the test and reference.
  - Include ≥5 subjects, who showed a ‘normal’ response in the main study (*i.e.*, size of re-tested group ≥6 or 20 % of subjects, whichever is larger).
  - Expect questions anyway (although *sometimes* suggested by the FDA, not covered in any guideline; statistical evaluation not trivial…)

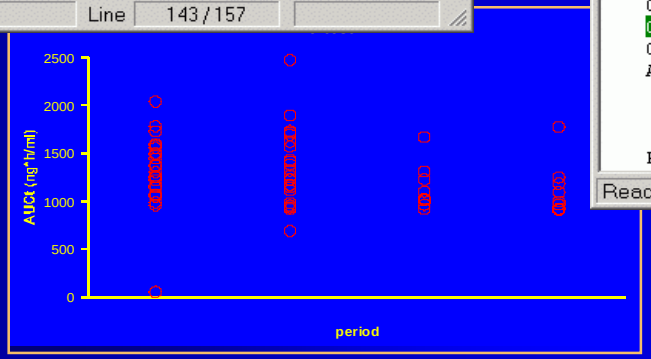
Re-testing of subjects

n=24: 83.3%–131.1% → +n=6: 86.7%–122.5%

Bioequivalence Text - [Untitled11]

```
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: formulation
Reference: R   LSMean=    7.085698  SE=    0.104041  GeoLSM= 1194.756651
--------------------------------------------------------------------------

Test:      T   LSMean=    7.129902  SE=    0.104041  GeoLSM= 1248.754167

    Difference =     0.0442,  Diff_SE=    0.1321,  df= 22.0
    Ratio(%Ref) =   104.5195

                        Classical                Westlake
    CI  80% = (   87.7732,  124.4610)  (   80.9169,  119.0831)
    CI  90% = (   83.2969,  131.1493)  (   74.5939,  125.4061)
    CI  95% = (   79.4604,  137.4815)  (   68.4657,  131.5343)
    Failed to show average bioequivalence for confidence=90.00 and percent=20.0.

                      Two One-Sided T-tests

    Prob(< 80%)=0.0277  Prob(> 125%)=0.0947  Max=0.0947  Total=0.1224
```

Read Only | Line 143 / 157

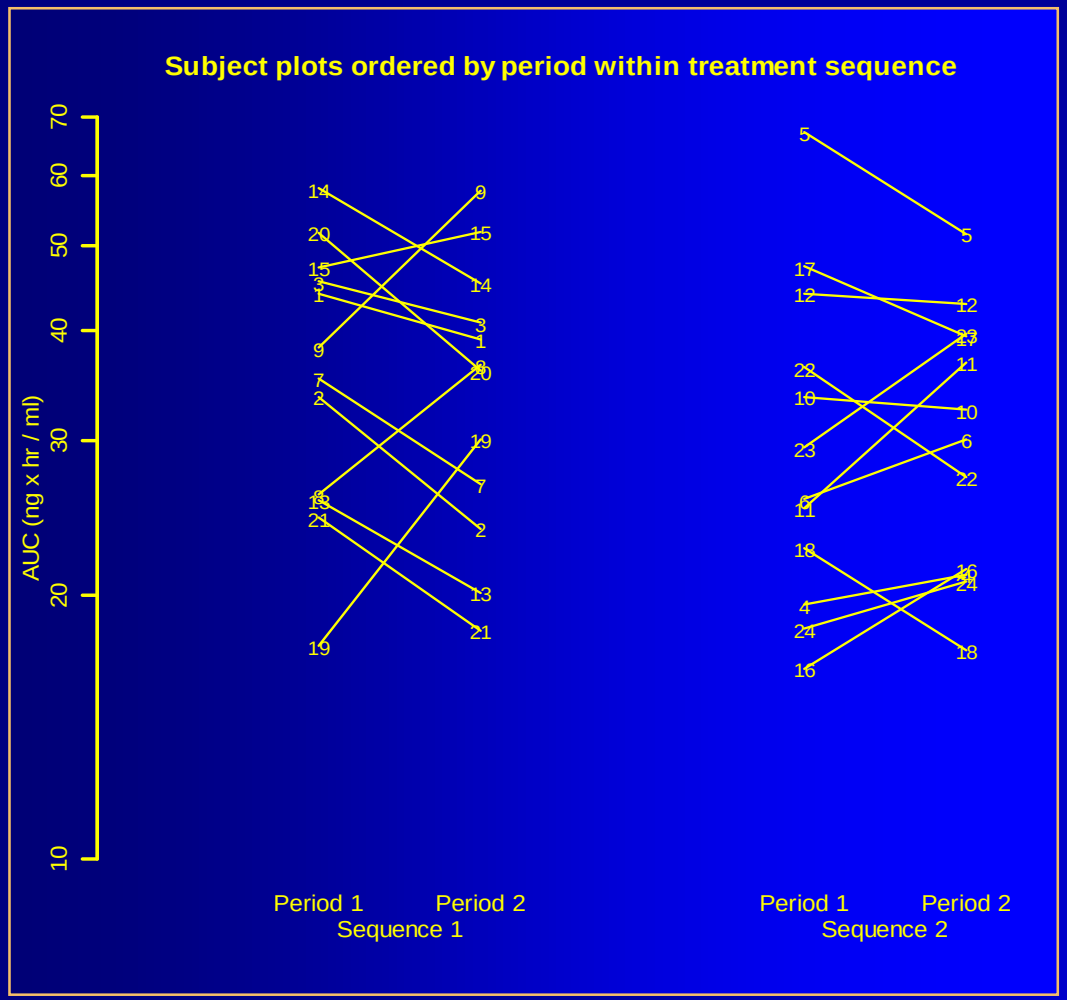


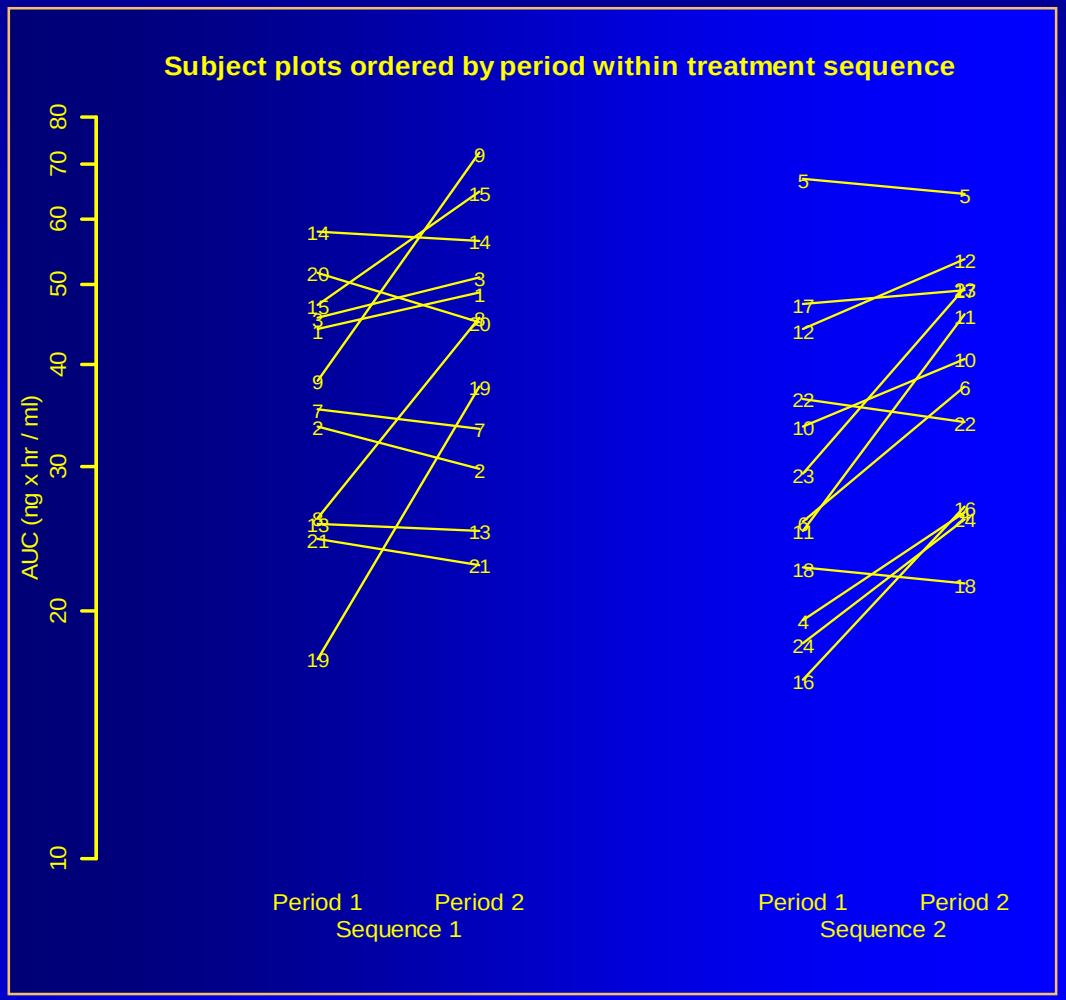
Bioequivalence Text - [Untitled10]

```
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: formulation
Reference: R   LSMean=    7.106774  SE=    0.109442  GeoLSM= 1220.204475
--------------------------------------------------------------------------

Test:      T   LSMean=    7.137202  SE=    0.052715  GeoLSM= 1257.903871

    Difference =     0.0304,  Diff_SE=    0.1017,  df= 28.7
    Ratio(%Ref) =   103.0896

                        Classical                Westlake
    CI  80% = (   90.2172,  117.7987)  (   85.7333,  114.2667)
    CI  90% = (   86.7263,  122.5403)  (   81.2107,  118.7893)
    CI  95% = (   83.7240,  126.9345)  (   76.9552,  123.0448)
    Average bioequivalence shown for confidence=90.00 and percent=20.0.

                      Two One-Sided T-tests

    Prob(< 80%)=0.0093  Prob(> 125%)=0.0341  Max=0.0341  Total=0.0434
```

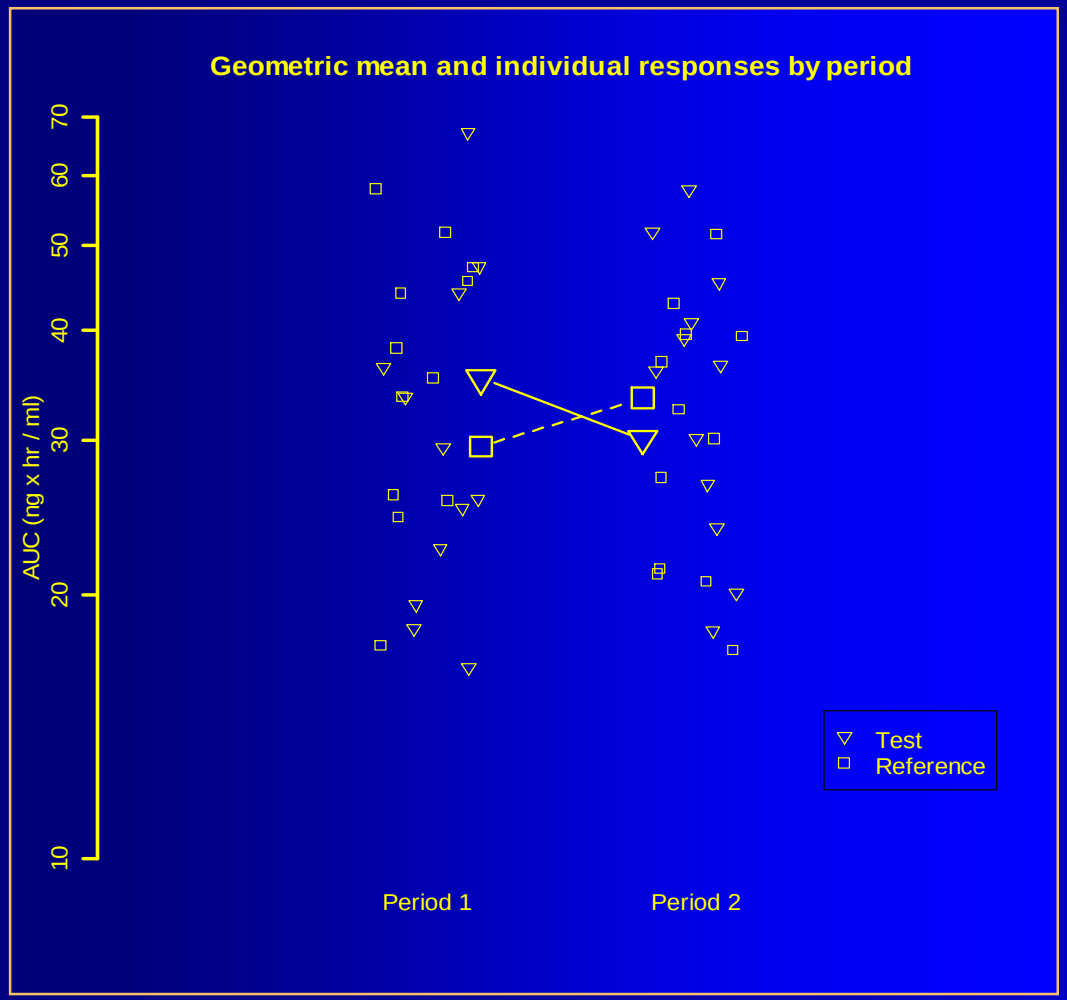
Read Only | Line 211 / 225

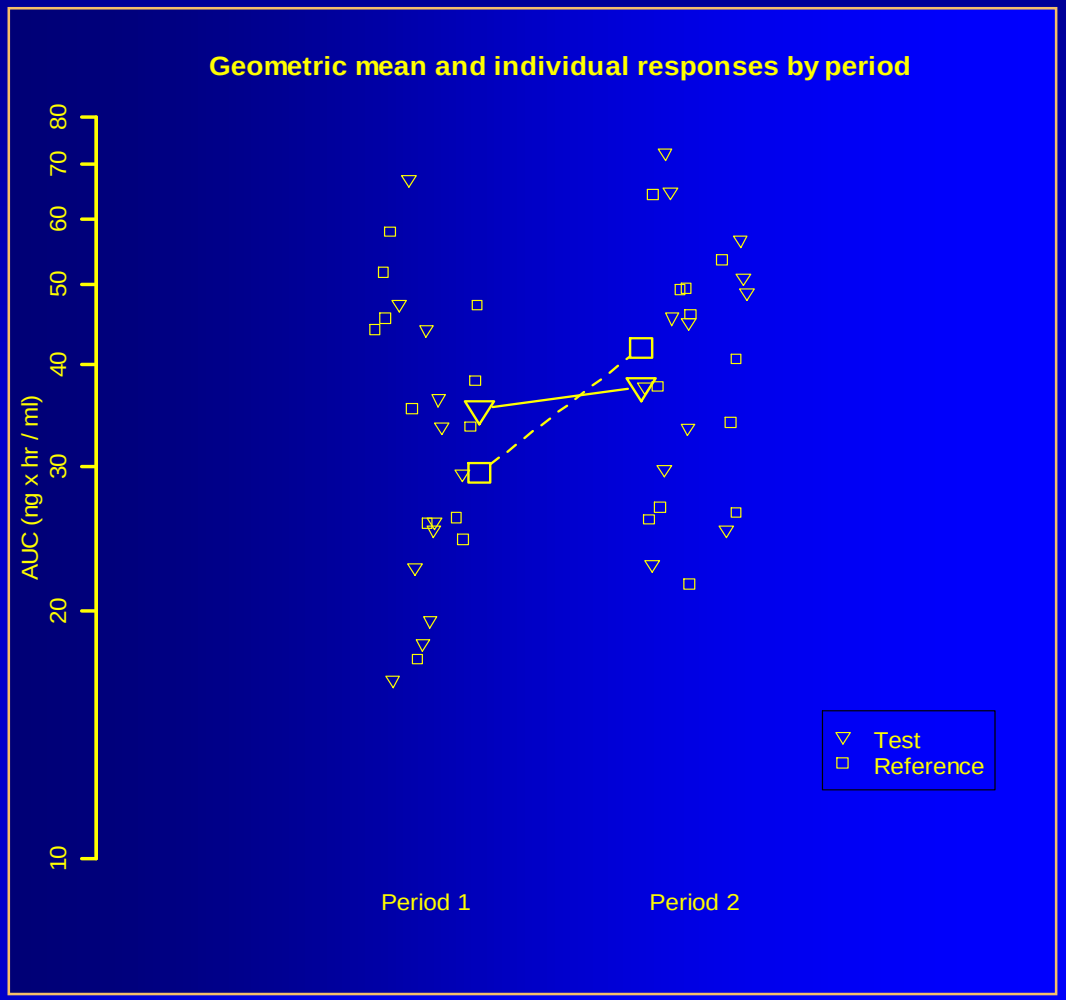
Nuisance: period effect





Nuisance: period effect

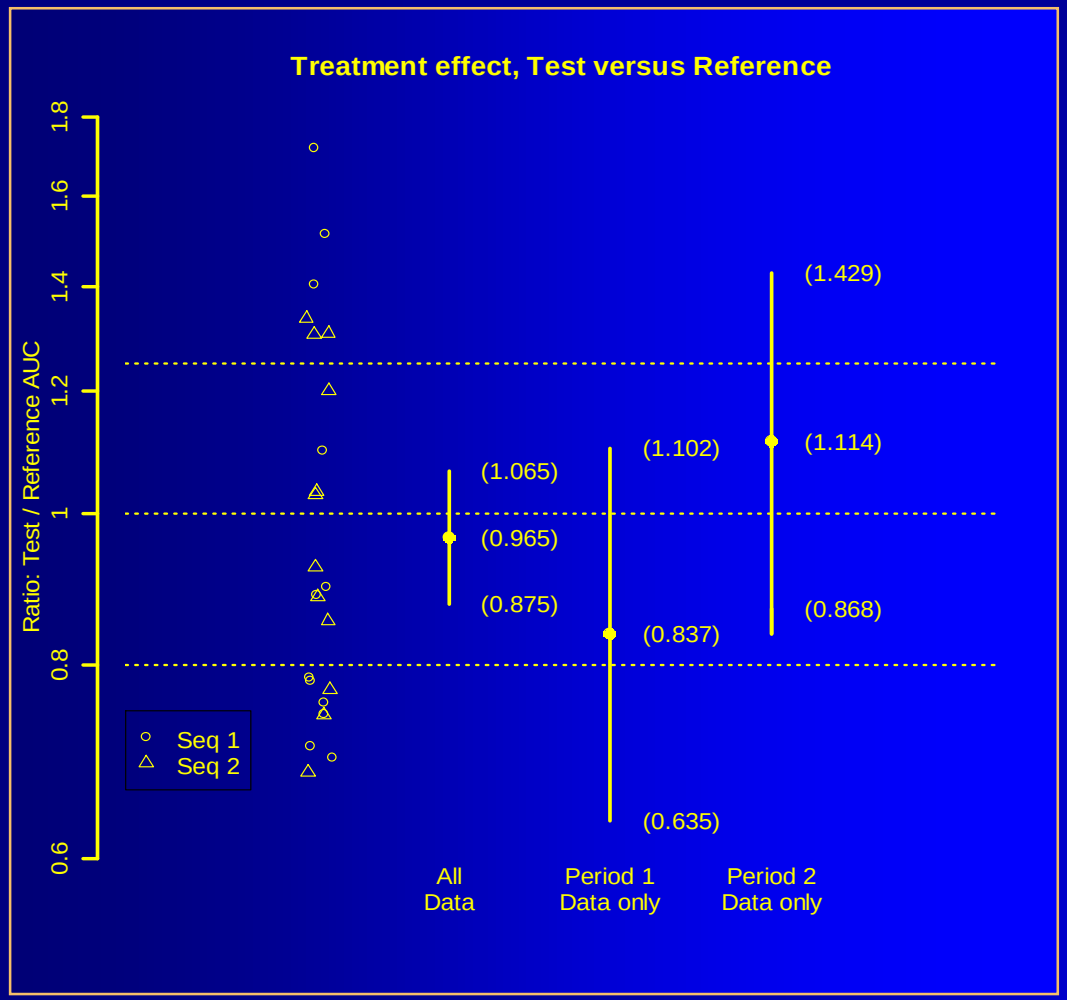


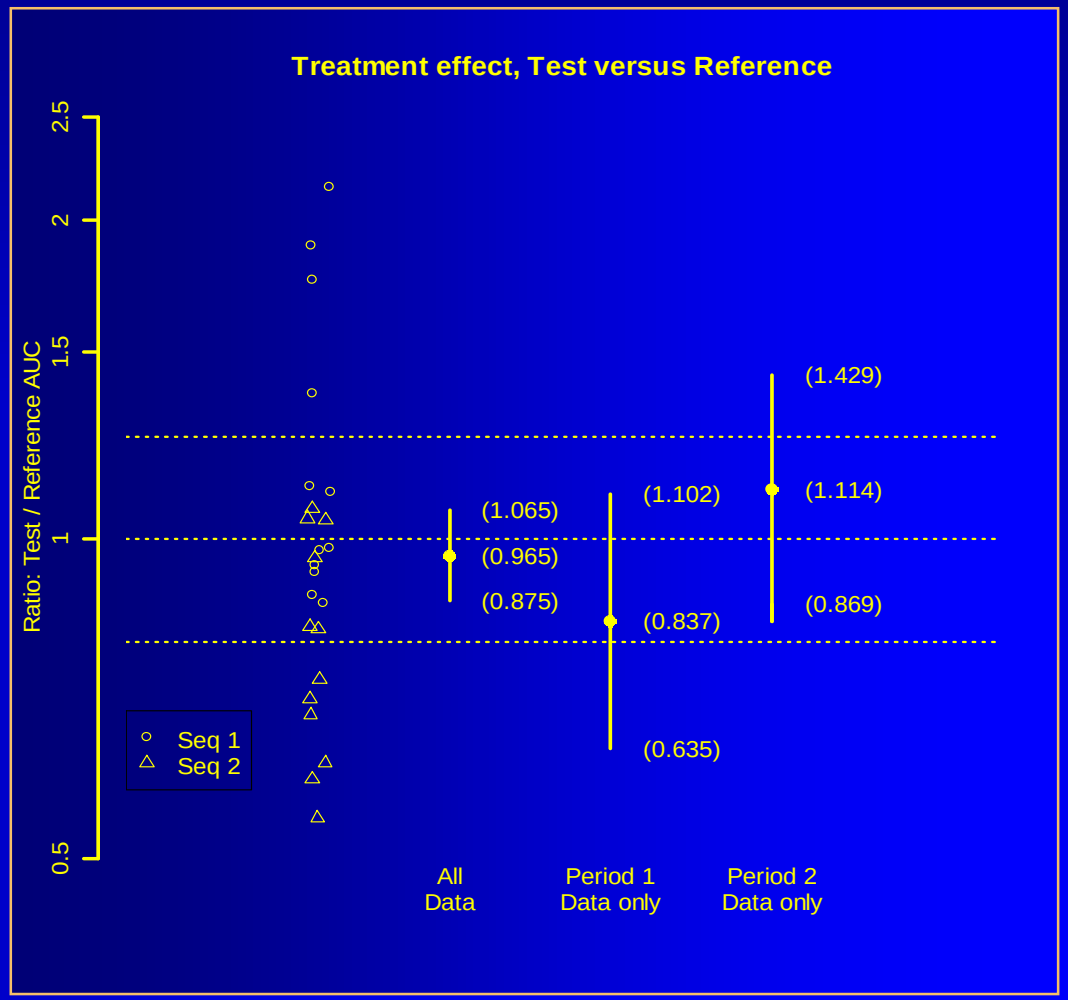


Nuisance: period effect

- Original data
  - AUC(p_2/p_1): 98.4%
  - Period: *p* 0.7856 (95% CI: 87.4% –110.8%)
  - Sequence: *p* 0.3239 (95% CI: 86.0% –154.8%)
  - GMR: 96.5% (90% CI: 87.5% –106.5%)
- Modified data (p_2 125% of original values)
  - AUC(p_2/p_1): 123.0%
  - Period: *p* 0.0015 (95% CI: 109.3% –138.5%)
  - Sequence: *p* 0.3239 (95% CI: 86.0% –154.8%)
  - GMR: 96.5% (90% CI: 87.5% –106.5%)

Nuisance: period effect





Nuisance: sequence effect

- In a ‘standard’ 2×2 cross-over design
  - the sequence effect is confounded with
    - the carry-over effect, and
    - the formulation-by-period interaction.
  - Therefore, a statistically significant sequence effect could indicate that there is
    - a true sequence effect,
    - a true carryover effect,
    - a true formulation by period interaction, or
    - a failure of randomization.

Nuisance: sequence effect

- 'Two-stage analysis'[1)] was – and regrettably still is – often applied.
  - Test for a significant sequence effect at α 0.10
  - If a significant sequence effect is found, evaluation of the first period as a parallel design
- This procedure was shown to be statistically flawed.[2)]

1) **JE Grizzle**
The two-period change over design and ist use in clinical trials
Biometrics 21: 467-480 (1965)

2) **P Freeman**
The performance of the two-stage analysis of two-treatment, two-period cross-over trials
Statistics in Medicine 8: 1421-1432 (1989)

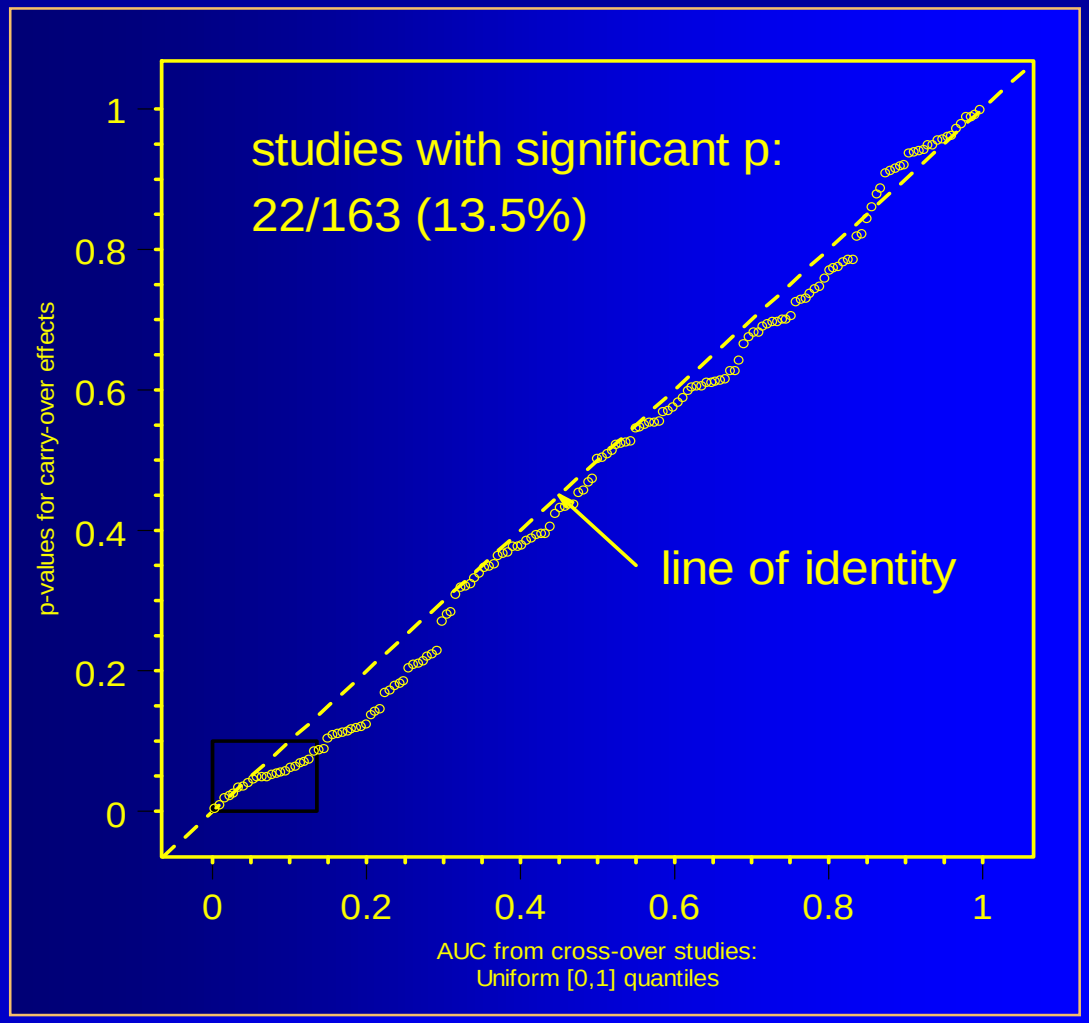
Nuisance: sequence effect

- In a large metastudy (n=420) significant sequence effects were found at ≈ α, both for AUC and C_{max}.*)
  - 2×2 studies (n=324)
    - AUC: 34/324 (10.5%) C_{max}: 37/324 (11.4%)
  - 6×3 studies (n=96)
    - AUC: 4/96 (4.2%) C_{max}: 4/96 (4.2%)
  - For both metrics the distribution of *p* values followed closely Uniform [0,1]

*) **D'Angelo G, Potvin D and J Turgeon**
Carry-over effects in bioequivalence studies
J Biopharm Stat 11: 35-43 (2001)

Nuisance: sequence effect

- These results could be confirmed (20 published studies, 143 studies from BEBAC's database; AUC):
  - Significant sequence effects in 22/163 studies (13.5%)
- Significant sequence effects in properly planned studies should be considered a statistical artefact (significant results are obtained in α of studies)



Nuisance: sequence effect

- Conclusions
  - No valid procedure exists to *correct* for a true sequence/carry-over effect
  - A true sequence/carry-over is *highly unlikely* in a BE study if
    - the study is performed in healthy subjects,
    - the drug is not an endogenous entity, and
    - an adequate washout period (no predose concentrations) was maintained.
  - Testing for a sequence effect is futile!

Nuisance: sequence effect

- Conclusions (cont'd)
  - EMEA Draft GL on BE (2008)
    - […] tests for difference and the respective confidence intervals for the treatment effect, the period effect, and the sequence effect should be reported for descriptive assessment. *A test for carry-over should not be performed and no decisions regarding the analysis (e.g. analysis of the first period, only) should be made on the basis of such a test.* The potential for carry-over can be directly addressed by examination of the pre-treatment plasma concentrations in period 2 (and beyond if applicable).

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.

Nuisance: group effect

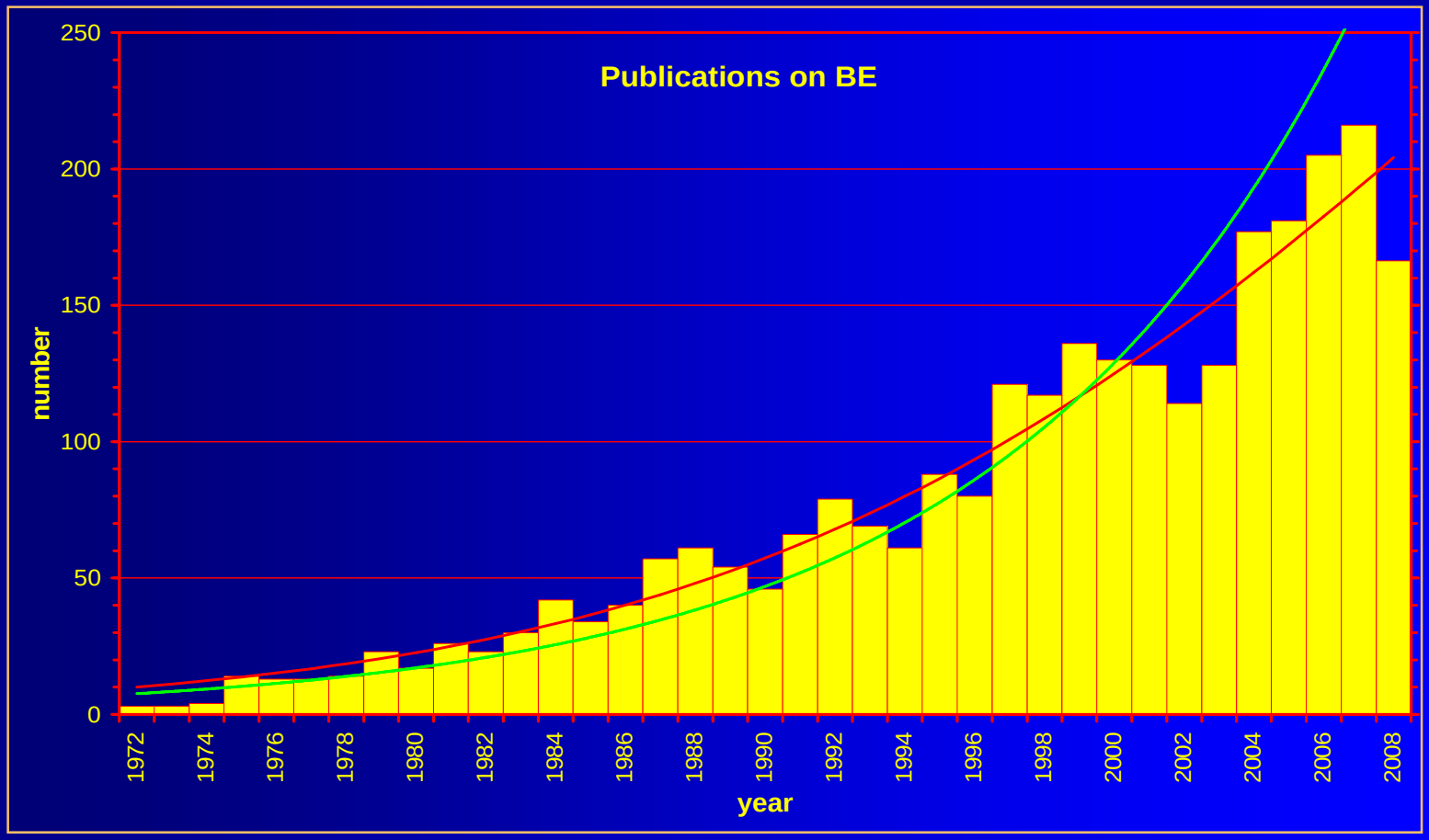
- Recently an increasing number of referrals (deficiency letters) from
  - Canada
  - Gulf States (Saudia Arabia, Emirates, Oman)
- Extended Statistical model (fixed effects in ANOVA)
  - Group
  - Group × Treatment Interaction
  - If both terms are not significant ($p>0.05$) pooling of groups is justified.

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 with the largest group only.

Are we making progress?

PubMed/MedLine: (bioequivalence) OR (comparative AND bioavailability),
Field: Title/Abstract, Limits: Humans, Publication Date



Are we making progress?

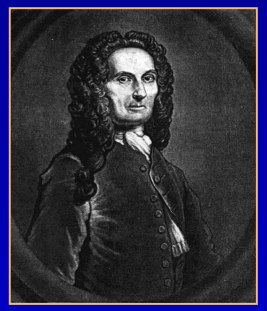
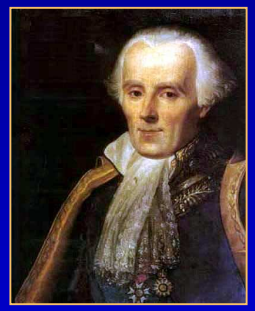
- About 3 000 – 10 000 BE studies / year are conducted worldwide; only ~ 1 – 5% of them are published.
- Although a standard for publishing data of BE studies was already suggested in 1992,[1)]
  - a review in 2002 found only 17 complete data sets on AUC and 12 on C_{max}.[2)]
  - Since no 'real world'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

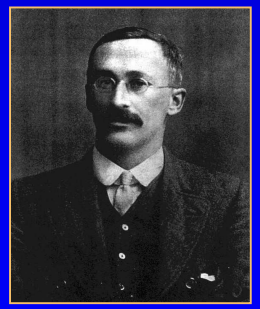
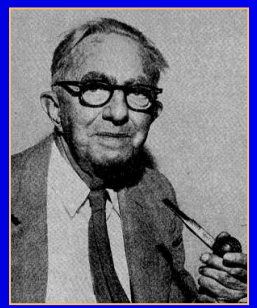
1) **Sauter R, Steinijans VW, Diletti E, Böhm E and H-U Schulz**
Int J Clin Pharm Ther Toxicol 30/Suppl.1, S7-S30 (1992)

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

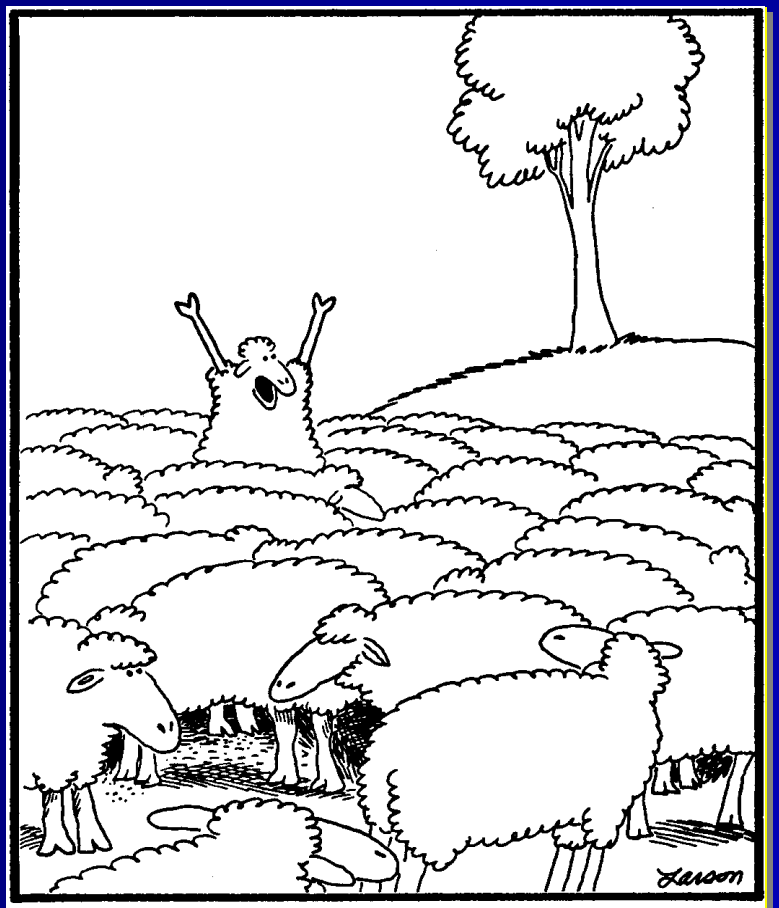
Bell curve (and beyond?)

- 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

Outlook



**“Wait! Wait! Listen to me! …
We don’t HAVE to be just sheep!”**

- David Bourne’s (Uni. Oklahoma) e-mail list
  - A rather active list (3200+ members, about 50 postings/week) covering almost any aspect of PK/PD/bio-analytics…
    - Subscription http://www.boomer.org/pkin/
    - Search page http://www.boomer.org/pkin/simple.html
- BA and BE Forum (BEBAC Vienna)
  - Specialized in BA/BE/bioanalytics.
    - No registration necessary to *read* posts. http://forum.bebac.at/
    - Registration (to post): http://forum.bebac.at/register.php

Thank You!

Statistical Design and Analysis

Open Questions?

(References in your handouts)

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

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)
- CDSCO
  Good Clinical Practices For Clinical Research In India (Schedule Y, Amended Version 2005)
- Indian Council of Medical Research
  Ethical Guidelines for Biomedical Research on Human Participants (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)

References

Center for Drug Evaluation and Research (CDER)
- ANDA Checklist for Completeness and Acceptability (2006)

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

Biostatistical Methodology in Clinical Trials (1993)

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)

Questions & Answers on the BA and BE Guideline (2006)

Draft Guideline on the Investigation of BE (2008)