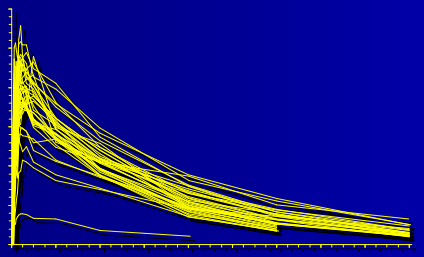
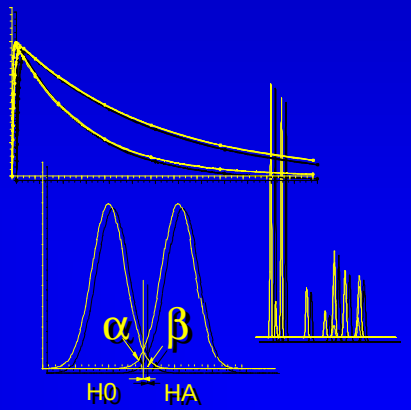
Statistical Design and Analysis I

Helmut Schütz
BEBAC

Assumptions

World *'Reality'*

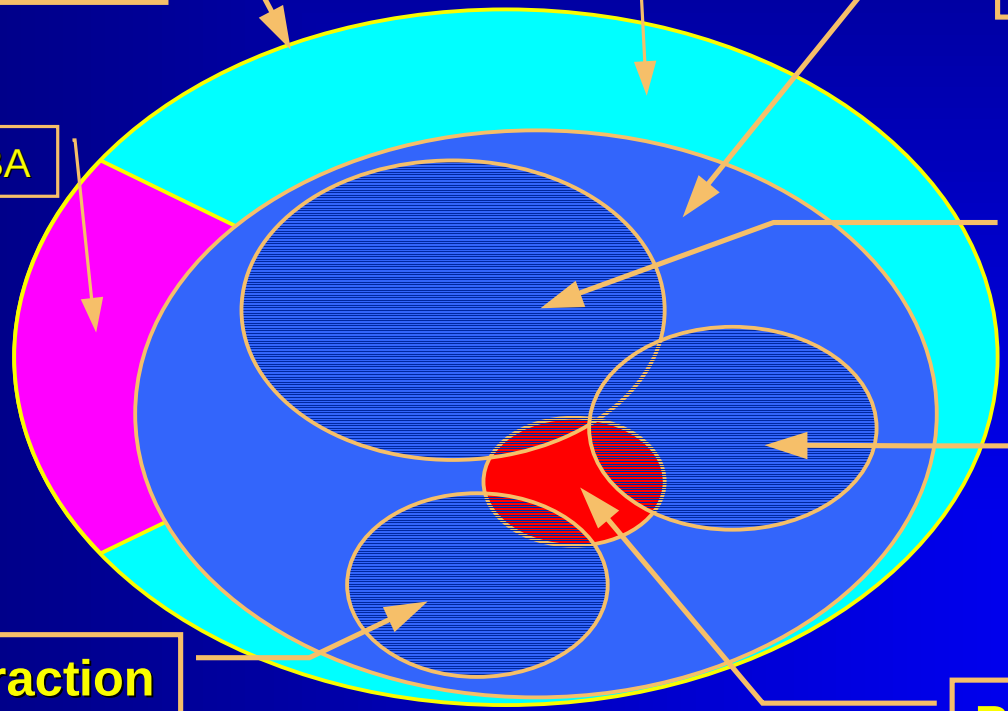




Model *'Data'*

Theory *'Truth'*

Terminology



Bioavailability

relative BA

Comparative BA

absolute BA

Bioequivalence

Food effect

PK interaction

Pilot study

Defining Study Objectives

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

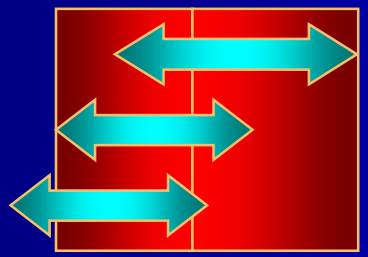
  'A bioequivalence study is basically a comparative bioavailability study designed to establish equivalence between test and reference products.'

  - 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.ema.europa.eu/pdfs/human/qwp/140198enfin.pdf

Defining Study Objectives

- Comparative BA
  - true experiment; no bibliographic comparison
- 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%]

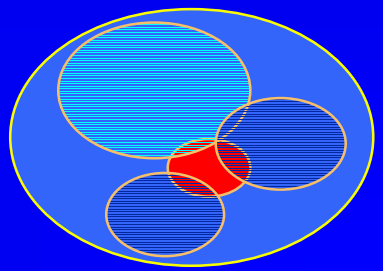
#2: BE [80%–110%]

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

Defining Study Objectives

- Definition of BE (EU GL, Section 1.1)

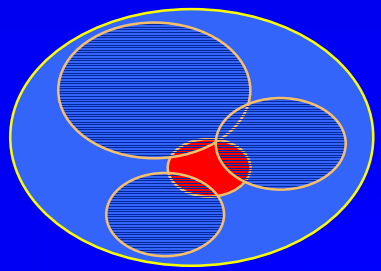
  'Two medicinal products containing the same active substance are considered bioequivalent if they are pharmaceutically equivalent or pharmaceutical alternatives and their bioavail-abilities (rate and extent) after administration in the same molar dose lie within acceptable predefined limits. These limits are set to ensure comparable in vivo performance, i.e. similarity in terms of safety and efficacy.'



Defining Study Objectives

- *In vivo* BE mandatory, if
  - Waiving (GL Section 4.2.2/Appendix III) 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.'

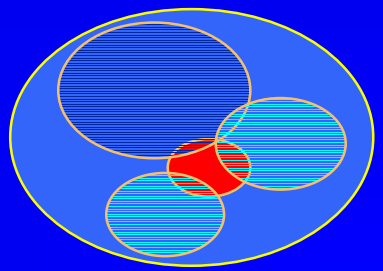


Defining Study Objectives

- 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 ≥5× $t_{½}$
  - Saturation phase long enough to reach steady-state: ≥5× $t_{½}$
  - Pre-dose samples (carry-over, compliance)

WHO (2006), EU GL (2010): For IR formulations sampling beyond 72 hours not required!

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)

WHO (2006), EU GL (2010): For IR formulations sampling beyond 72 hours not required!

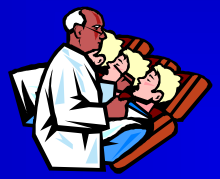
Defining Study Objectives

- 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, $t_{75\%}$

Human Guineapigs I

- BE studies as a surrogate for clinical efficacy / safety ('essential similarity')
  - We want to get unbiased estimates, *i.e.*, the point estimate from the study sample …

$$PE = \frac{\hat{X}_{Test}}{\hat{X}_{Reference}}$$



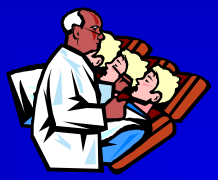
  - … should be representative for the population of patients.

$$F_{Pop} = \frac{\mu_{Test}}{\mu_{Reference}}$$

Human Guineapigs II

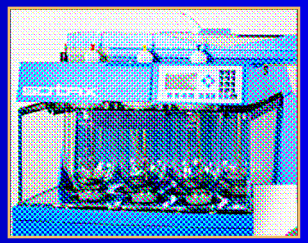
- BE studies as a special case of documented pharmaceutical quality
  - The *in vivo* release in the biostudy …

$$PE = \frac{\hat{X}_{Test}}{\hat{X}_{Reference}}$$



  - … should be representative for the *in vitro* performance.

$$f_2 = 50 \cdot \log\left[\frac{100}{\sqrt{1+\frac{\sum_{t=1}^{t=n}\left[\bar{R}(t)-\bar{T}(t)\right]^2}{n}}}\right]$$



Assumptions: Pharmacokinetics

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

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

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

Assumption 2: $CL_1 = CL_2$

Science → Regulations

- We can't compare bioavailabilities in the entire population of patients
  - Scientific Reductionism (based on assumptions)
    - 'Similar' concentrations in healthy subjects will lead to 'similar' effects in patients
    - Equal doses and inter-occasion clearances

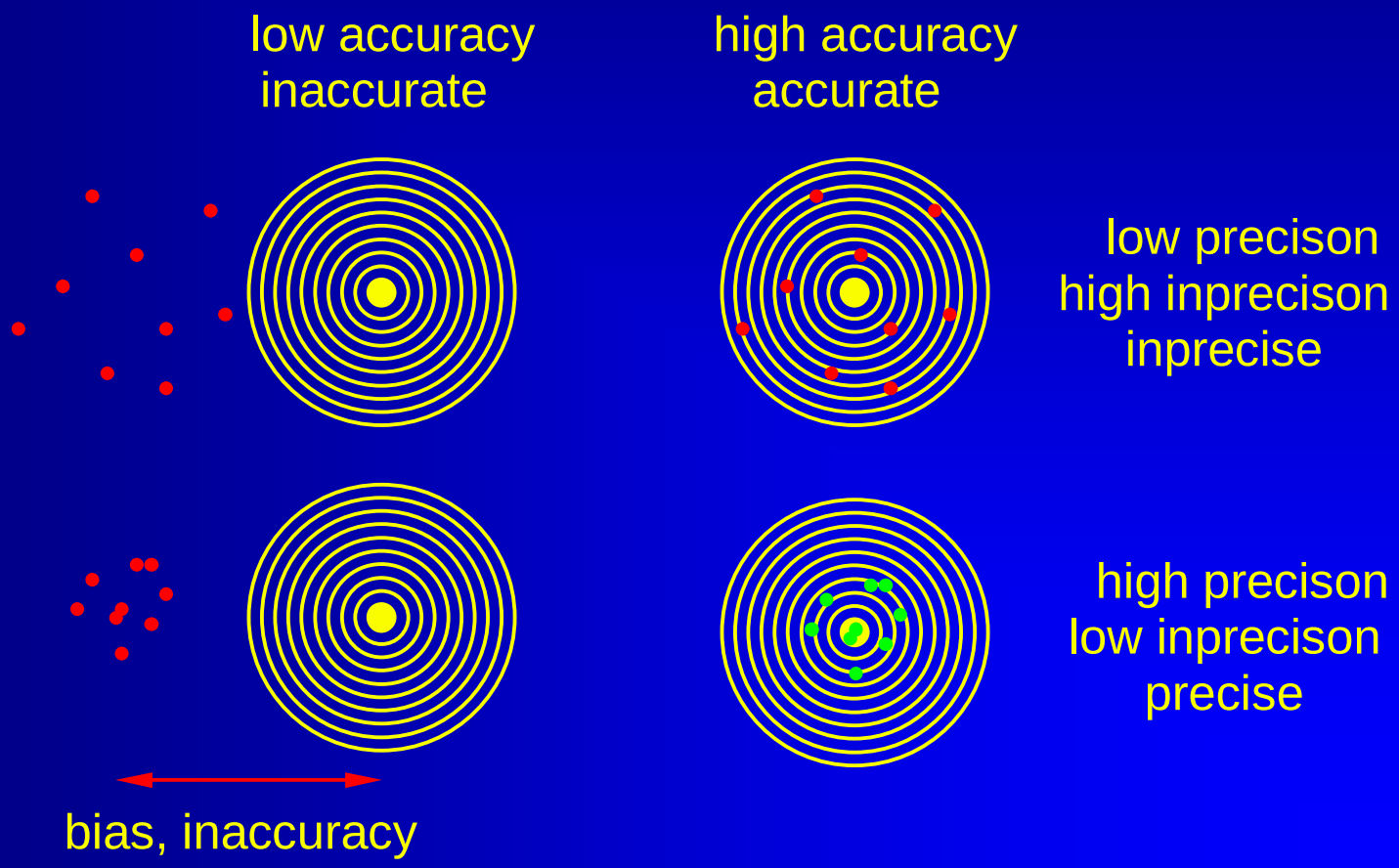
$$\frac{F_T \cdot AUC_T}{D_T \cdot CL_T}, \frac{F_R \cdot AUC_R}{D_R \cdot CL_R}$$

$$D_T = D_R, CL_T = CL_R$$

Highly Variable Drugs?

$$F_{rel}(BA) = \frac{AUC_T}{AUC_R}$$

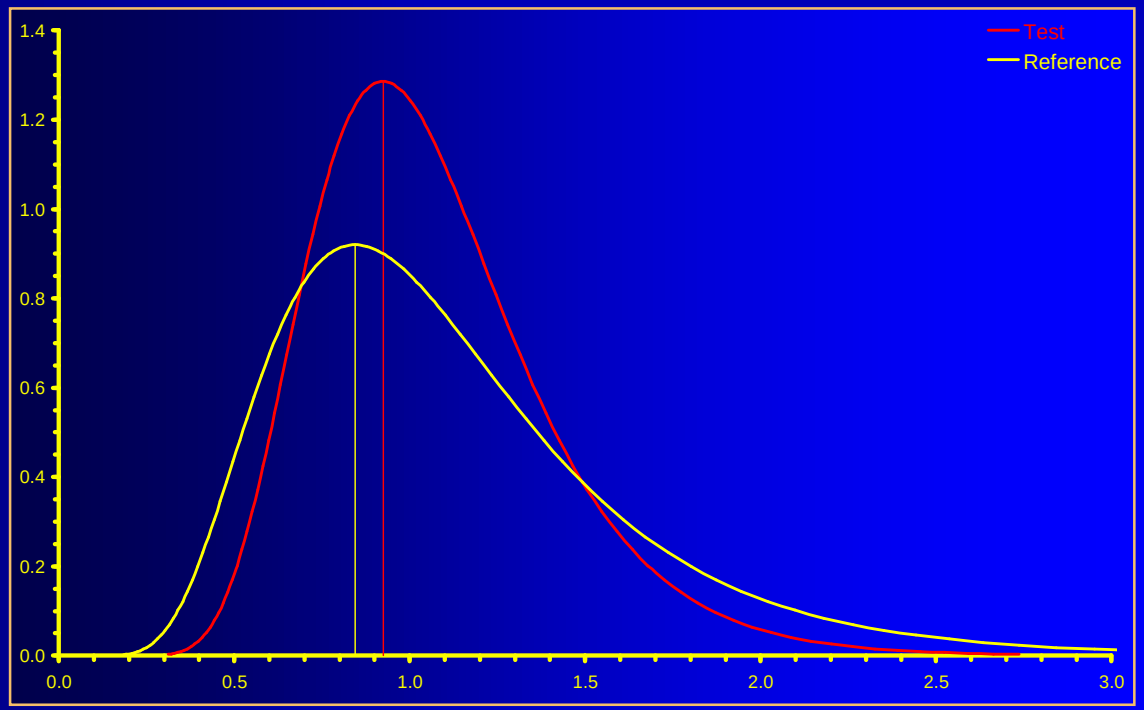
Assumptions: Statistics



Assumptions: Statistics

Distribution

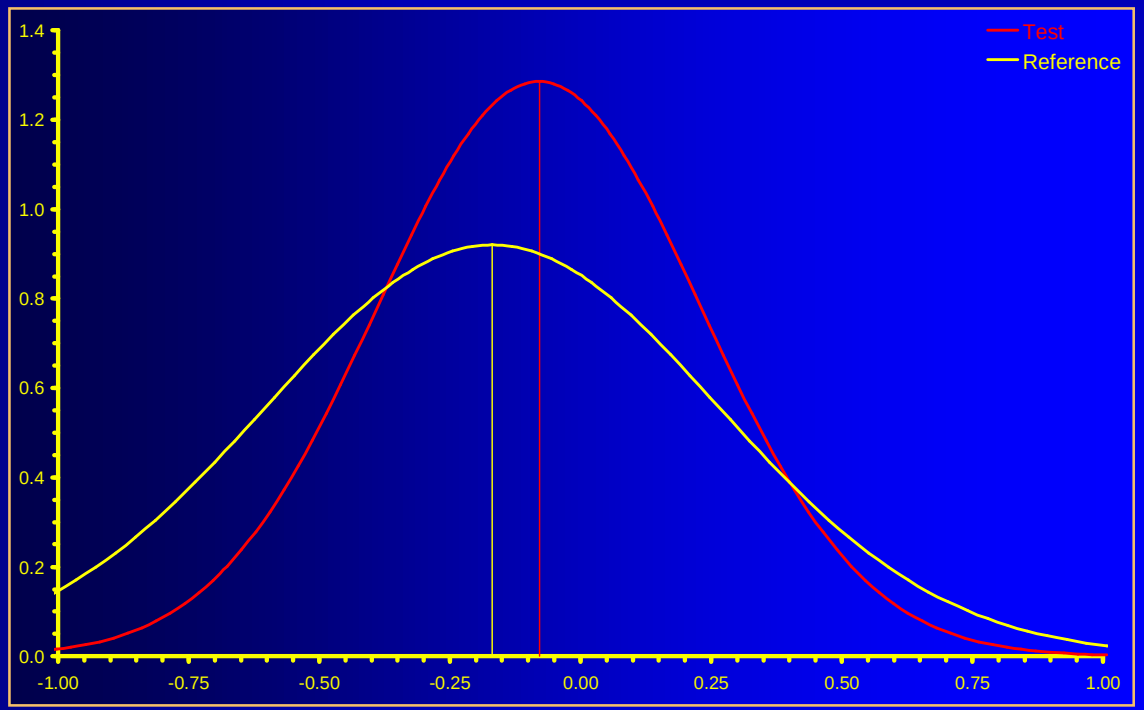
- IDD (Independent Identically Distribution)



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:

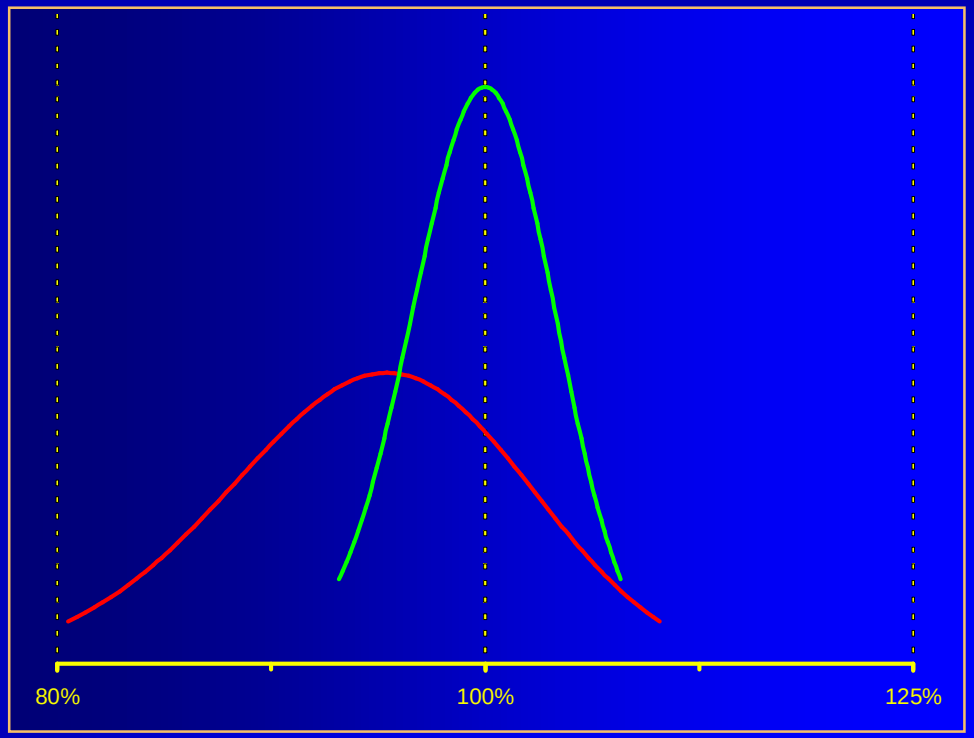
- All $ln\{s_{ik}\}$ and $ln\{e_{ijk}\}$ are independently and normally distributed about unity with variances σ^2_s and σ^2_e.
- All observations made on different subjects are independent.

Science → Regulations

- Scientific Reductionism (cont'd)
  - Independent Identically Distribution (IDD)

What if…

$$\sigma^2_{WT} \neq \sigma^2_{WR}$$



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 (11/2005), Saudia Arabia (05/2005), WHO (05/2006), Japan (11/2006),…

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

Deleted since 2008 draft

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/Corr.
Guideline on the Investigation of Bioequivalence (2010)

> *'Also interesting that they now say they will not accept non-parametric analyses. That seems a step backwards.'*
> (Walter Hauck, personal communication on the draft GL, Oct 2008)

Regulations = Science?

FDA 2010, EMA 2010

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.*, bootstrap)

ICH E9 1998, WHO 2006
Canada Draft 2009

Global Harmonization?

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

There can be only one!

Global Harmonization?

- Drugs with a narrow therapeutic range
  - USA, Japan: No difference to other drugs
  - WHO, EU, NZ, India: 90 % CI; Acceptance range may be tightened, *e.g.*, 0.9000–1.1111
  - RSA: 90 % CI within 0.80–1.25 (C_{max})
  - Brazil: 95 % CI within 0.80–1.25
  - Canada: No different procedure given in guideline, but considered in current draft
    - AUC: 90 % CI within 0.90–1.12
    - C_{max}: 90 % CI within 0.80–1.25

http://www.hc-sc.gc.ca/dhp-mps/alt_formats/pdf/consultation/drug-medic/draft_ebauche_cbs-eng.pdf (25 Jan 2010)

Basic Designs

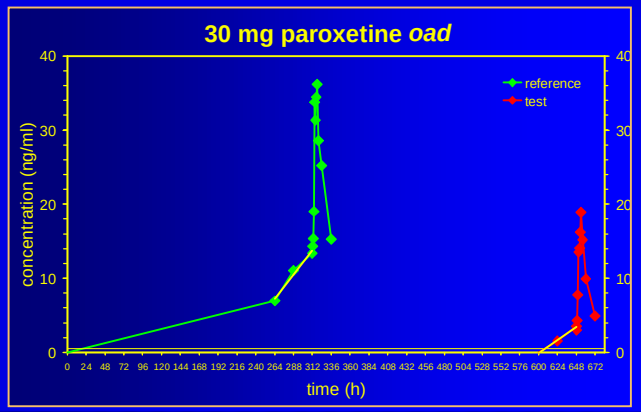
- Single Dose / Multiple Dose
  - Cross-over
    - Standard 2×2
    - Higher Order Designs (for more than two treatments)
      - Latin Squares
      - Variance Balanced Designs (Williams' Designs)
      - Incomplete Block Designs
    - Replicate designs
  - Parallel Groups

Single Dose / Multiple Dose

- Single Dose recommended in most GLs, 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 MD 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 should 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 the profile(s) should be excluded from the evaluation – if stated in protocol!



Single Dose / Multiple Dose

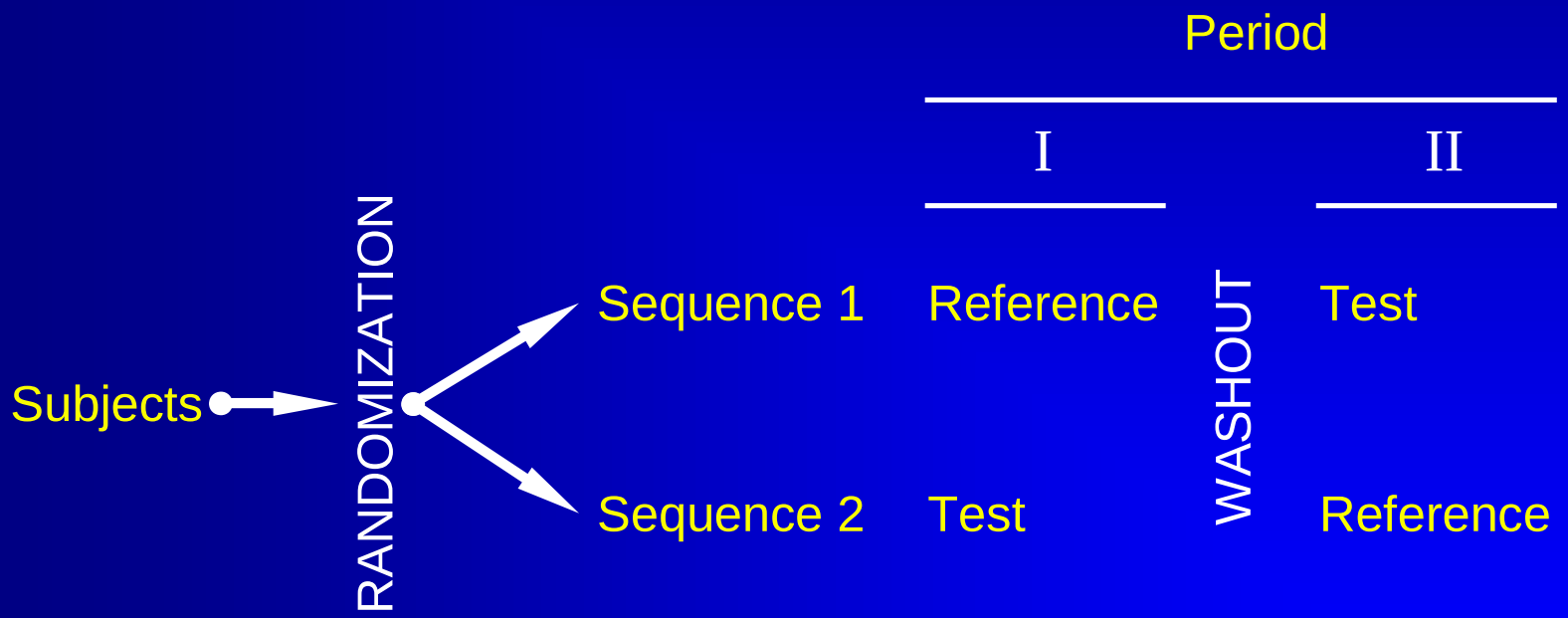
- Steady-state studies
  - Demonstration of steady-state (cont'd)
    - Multivariate method (simultaneous testing of all pre-dose values in all subjects)
      - *E.g.*, Hotellings T^2
      - Benefit: additional statement possible *when* steady-state was reached
      - Drawback: if significant result, 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

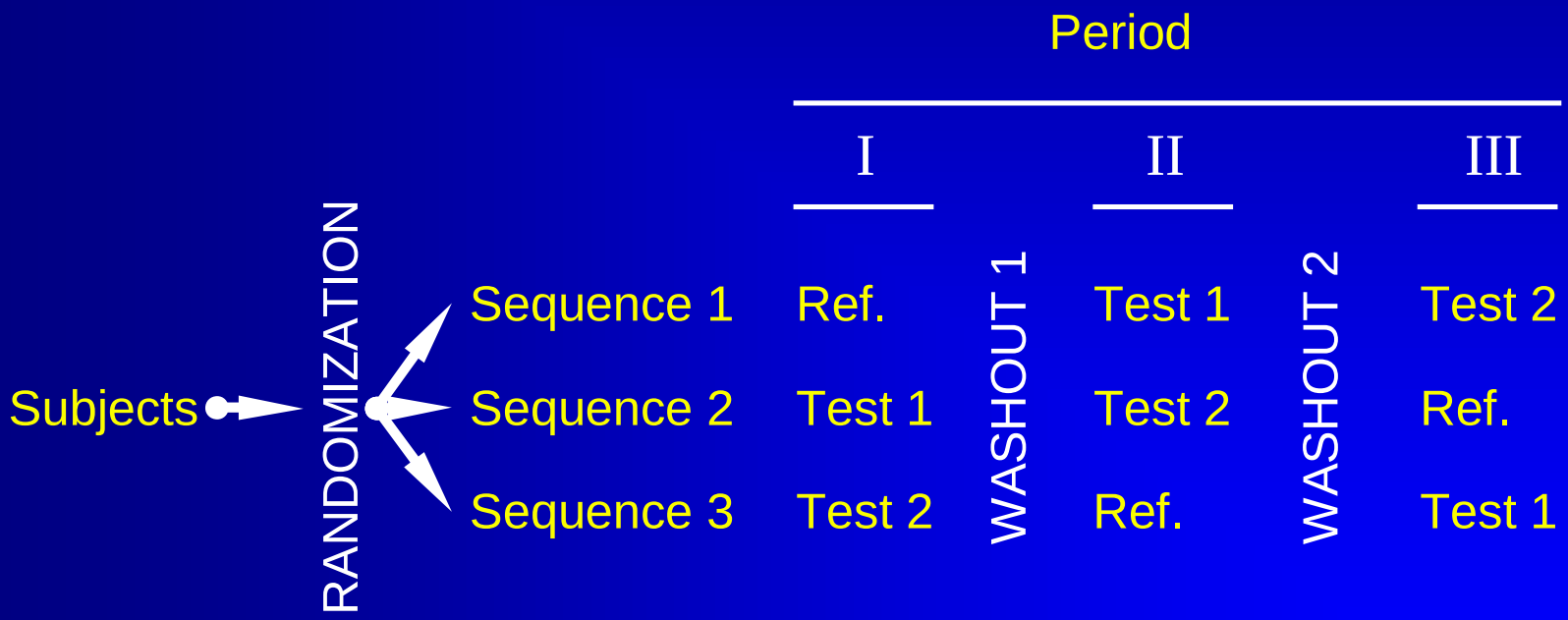
- Standard 2×2×2 design
  - Advantages
    - Globally applied standard protocol for BE
    - Straigthforward statistical analysis
  - Disadvantages
    - Not suitable for drugs with long half life (→ parallel groups)
    - Not optimal for studies in patients with instable diseases (→ parallel groups)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

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 pieces of software.
    - 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

imbalanced

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

  - Williams' design

balanced

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

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

- Higher Order Designs (cont'd)
  - Effect of α-adjustment on sample size (expected T/R 95%, CV_{intra} 20%, power 80%)

CV%	2×2 α 0.05	6×3 $\alpha_{adj.}$ 0.025	comp. 2×2	4×4 $\alpha_{adj.}$ 0.0167	comp. 2×2
10.0	8	12	+50%	16	+100%
12.5	10	12	+20%	16	+60%
15.0	12	18	+50%	16	+33%
17.5	16	24	+50%	24	+50%
20.0	20	24	+20%	28	+40%
22.5	24	30	+25%	36	+50%
25.0	28	36	+29%	40	+49%
27.5	34	42	+24%	48	+41%
30.0	40	54	+35%	56	+40%

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 an increased number of periods.
    - Same overall number of individual treatments!
    - Mandatory in the EU if scaled acceptance range for C_{max} is aimed at (CV_{WR}>30% must be demonstrated in the same study).

Cross-over designs

- Replicate designs
  - Advantages
    - Some experience from FDA's initiative on Population BE (PBE) and Individual BE (IBE).
    - Reference Scaled Average Bioequivalence (RSABE)
    - Handling of outliers (Subject-by-Formulation Interaction may be ruled out).
    - Mentioned in RSA GL; FDA's API GLs and EMA for C_{max}.
  - 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 (!)

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

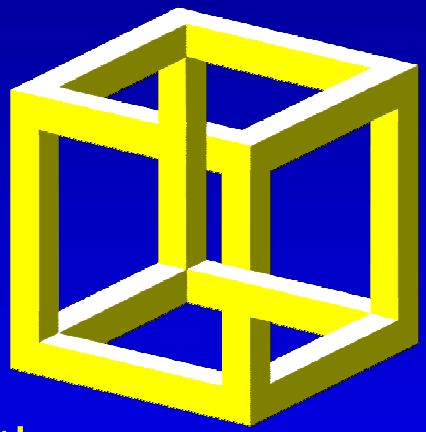
$$X_{ijkl} = \mu \cdot \pi_k \cdot \Phi_l \cdot s_{ij} \cdot e_{ijkl}$$

HVDs/HVDPs

- Highly Variable Drugs / Drug Products (intra-subject variability >30 %)
  - ✓USA Replicate Design recommended in product specific guidances: Minimum number of subjects (36?), restriction on GMR (0.8 – 1.25).
  - ± EU Widening of acceptance range (for C_{max} only: to maximum 69.84% – 143.19%), if CV_{WR} in the study >30%. Restriction on GMR (0.8 – 1.25).

HVDs/HVDPs

- Ways out?
  - Nonparametric methods
    A non-parametric analysis is **not acceptable**. (BE GL, Section 4.1.8)
  - Compartmental methods (Population PK)
    The use of compartmental methods for the estimation of parameters is **not acceptable**. (BE GL, Section 4.1.5)
  - Replicate designs could be considered e.g. for substances with highly variable pharmacokinetic characteristics. (EU BE GL, Section 4.1.1, 4.1.10)



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

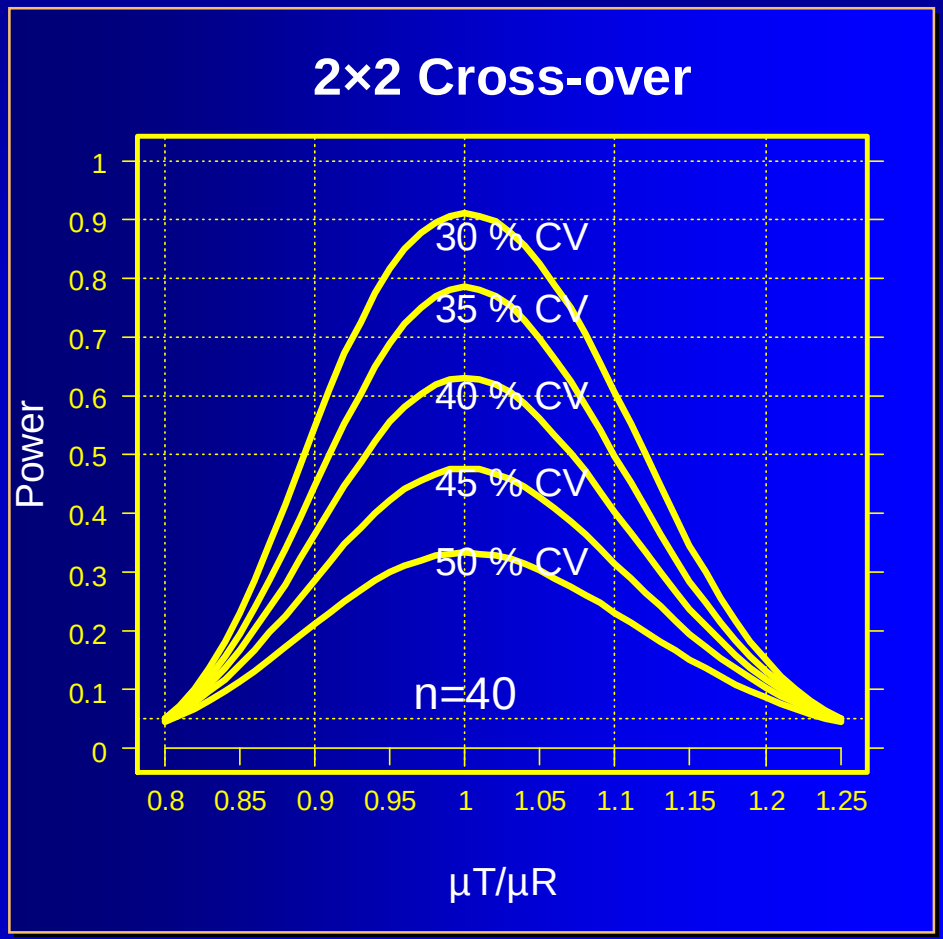
Davit BM, Conner DP, Fabian-Fritsch B, Haidar SH, Jiang X, Patel DT, Seo PR, Suh K, Thompson CL, and LX Yu
Highly variable drugs: observations from bioequivalence data submitted to the FDA for new generic drug applications
AAPS J 10(1): 148-56 (2008)

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 papers in 2008 (ref. in API-GLs)
  - Replicate study design [TRR–RTR–RRT]
  - Reference Scaled Average Bioequivalence (RSABE)
  - Minimum sample size 36 (?) 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
Haidar SH, Makhlouf F, Schuirmann DJ, Hyslop T, Davit B, Conner D, and LX Yu
Evaluation of a Scaling Approach for the Bioequivalence of Highly Variable Drugs
The AAPS Journal, 10/3, (2008) DOI: 10.1208/s12248-008-9053-4

HVDs/HVDPs

- Is suggested EU-method of any good?
  - Replicate designs *without scaling* (AUC)
    - 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 (!)

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

- EU GL 2001 (Section 4.1.8)
  - A statistical evaluation of t_{max} is not required. However, if rapid release is claimed to be clinically relevant and of importance for onset of action or is related to adverse events, there should be no apparent difference in median t_{max} and its variability between test and reference product.

How to assess that? Since 'A non-parametric analysis is not acceptable.'

Early Exposure (HVDP?)

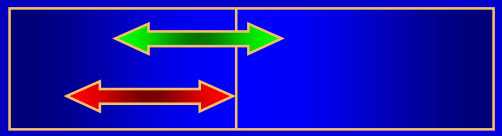
- Partial AUCs for Rapid Onset Drugs (cont'd)

Example	median t_{maxref}	PE	nonparametric CI		BE	FDA	parametric CI		BE	TGD	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%	yes
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%	no

  - 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 is unlikely to be able to demonstrate BE due to high variability of this metric. It is unclear how median t_{maxref} can be stated in the protocol (EMEA) – the innovator's SmPC (=label) mostly states only the *arithmetic mean*.

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.



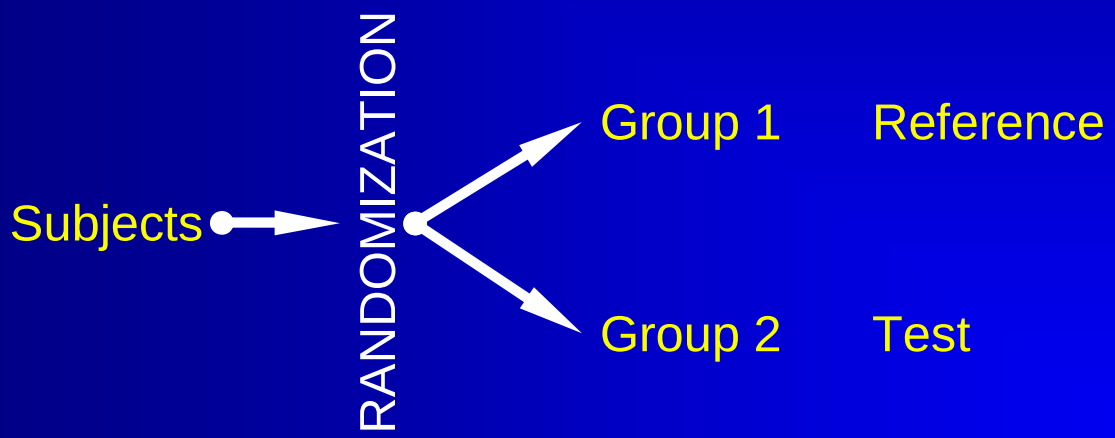
  - Denmark
    - DKMA considers that the 90% CI for the ratio test versus reference should include 100% […].
    - Deviations may be accepted if they can be adequately justified not to have impact on either the overall therapeutic effect or safety profile of the product.
      Danish Medicines Agency (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 *total (between+within)-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 *intra-subject* effects
    - One study of the major phenotype/genotype.
    - Two studies of the respective phenotype/genotype – only if requested!

Parallel Groups

- Design Issues
  - EMA GL on BE (2010)
    - 4.1.3 Subjects / Selction of Subjects
      ‘Phenotyping and/or genotyping of subjects may be considered for safety or pharmacokinetic reasons.
  - Wording has changed since old NfG, but content stayed the same!
  - Specifically not only for parallel designs!

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
    - 'Naïve 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

- 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 a correct answer?

Software / Method	equal variances	unequal variances
'manual' (Excel 2000)	63.51% – 110.19%	63.48% – 110.25%
R 2.10.1 (2009)	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 5.3 (2009)	63.51% – 110.20%	*not implemented!*
Phoenix/WNL 6.1 (2009)	63.51% – 110.20%	*not implemented!*
Kinetica 5.0.1 (2009)	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 2.10.1 (2009)	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 (naïve 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 BE testing (Phoenix/WinNonlin, Kinetica, EquivTest/PK)!

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

Thank You!
Statistical Design and Analysis I
Open Questions?

(References in the Handouts of Part II)

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