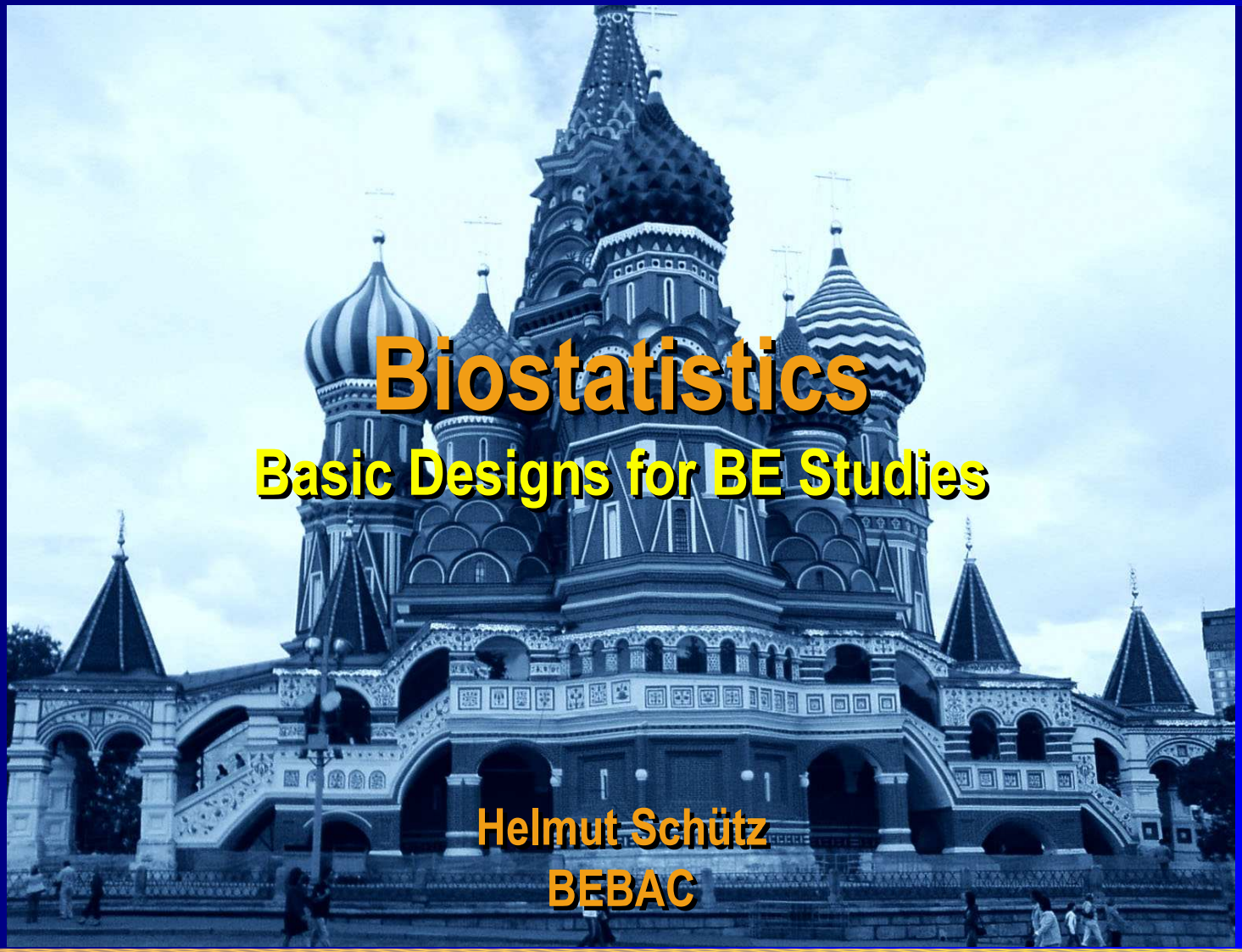


Biostatistics

Basic Designs for BE Studies

Helmut Schütz
BEBAC

To bear in Remembrance...

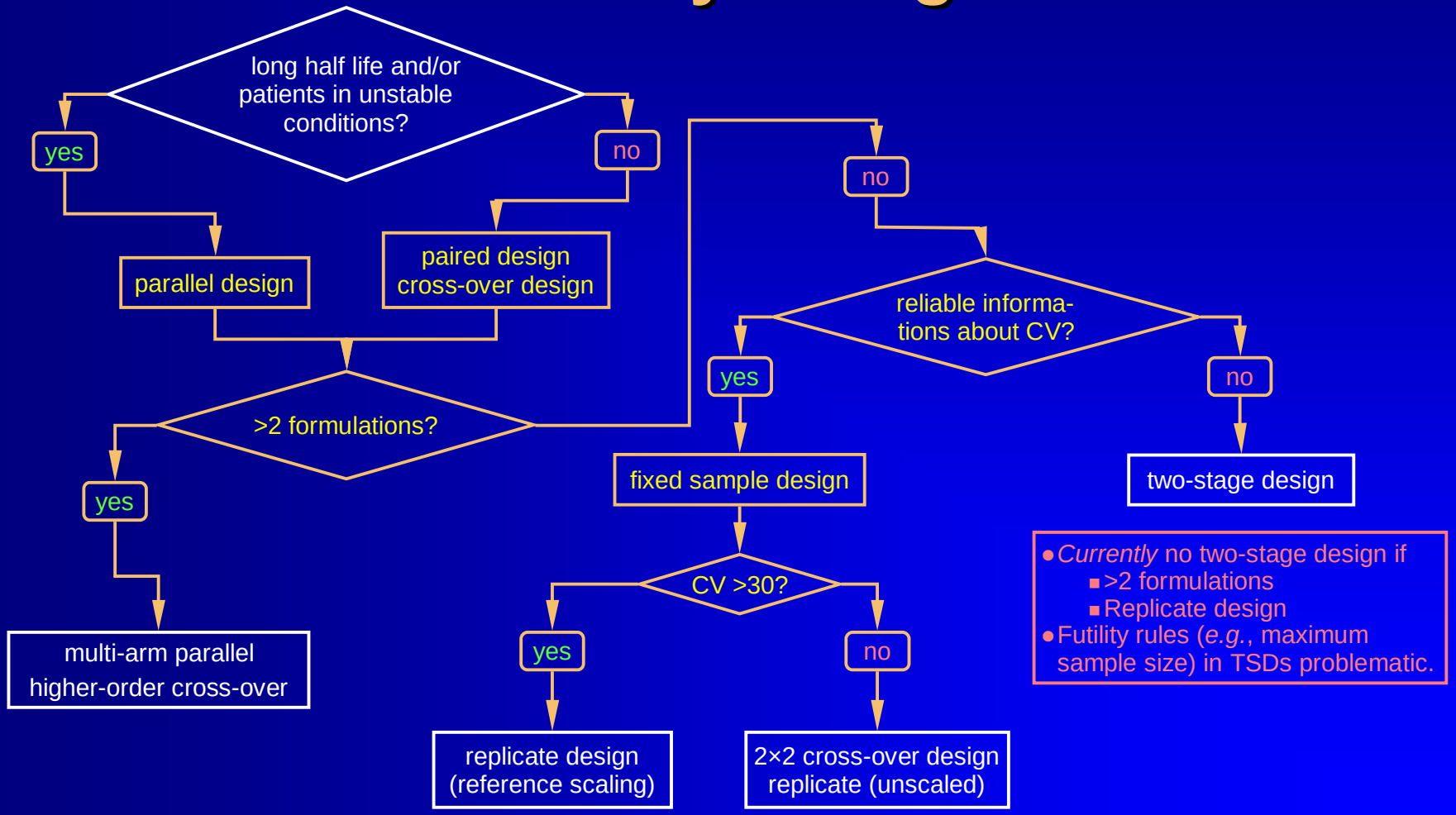
Whenever a theory appears to you as the only possible one, take this as a sign that you have neither understood the theory nor the problem which it was intended to solve.

Karl R. Popper

Even though it's *applied* science we're dealin' with, it still is – *science!*

Leslie Z. Benet

BE Study Designs



BE Study Designs

- The more ‘sophisticated’ a design is, the more information can be extracted
  - Hierarchy of designs:

    Full replicate (TRTR | RTRT or TRT | RTR),

    Partial replicate (TRR | RTR | RRT)

    Standard 2×2 cross-over (RT | RT)

    Parallel (R | T)

  - Variances which can be estimated:

    Parallel: total variance (between + within)

    2×2 Xover: + between, within subjects

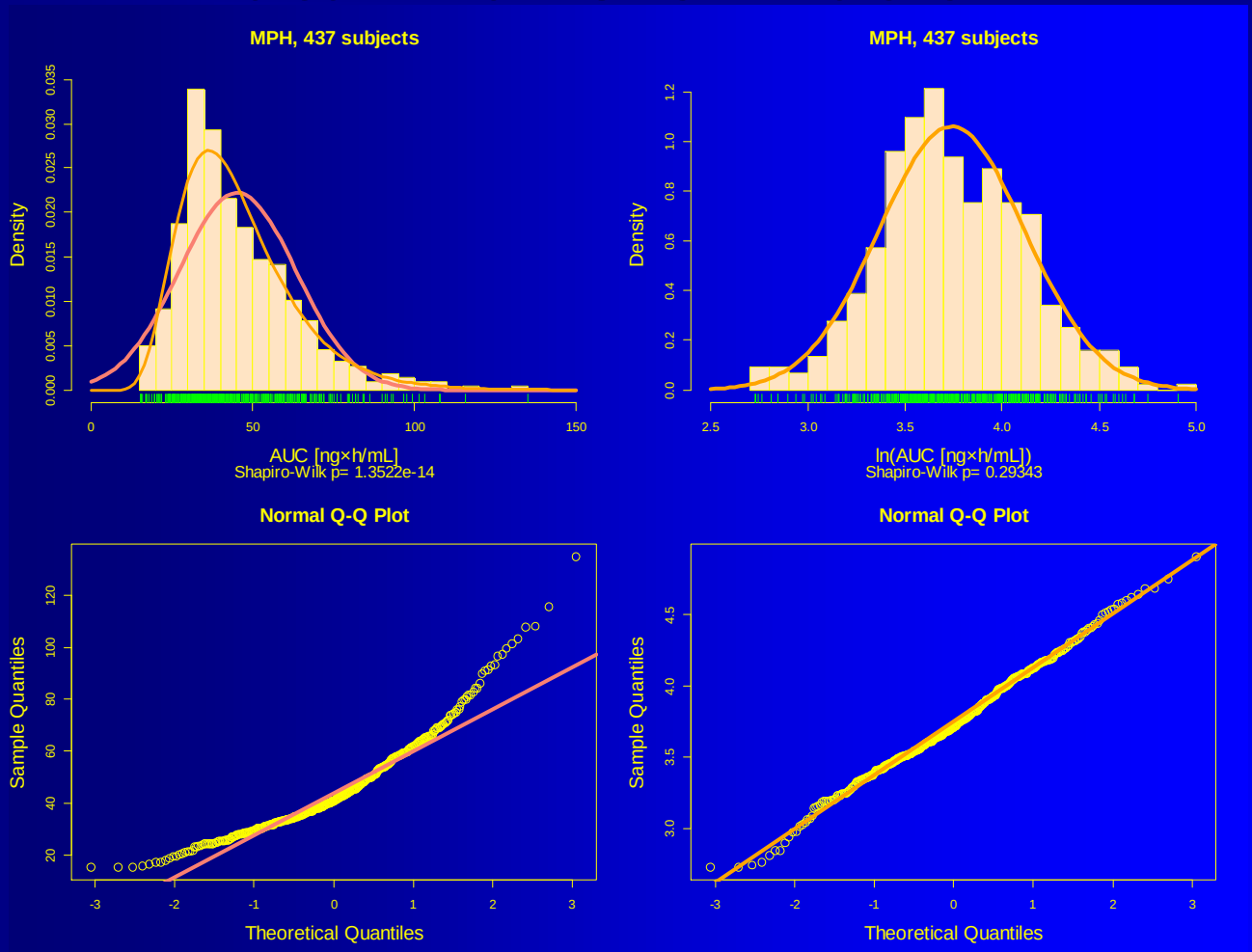
    Partial replicate: + within subjects (reference)

    Full replicate: + within subjects (reference, test)

Data Transformation?

- BE testing started in the early 1980s with an acceptance range of 80% – 120% of the reference based on the *normal* distribution
- Was questioned in the mid 1980s
  - Like many biological variables AUC and C_{max} *do not* follow a normal distribution
    - Negative values are impossible
    - The distribution is skewed to the right
    - Might follow a *lognormal* distribution
  - Serial dilutions in bioanalytics lead to multiplicative errors

Data Transformation?

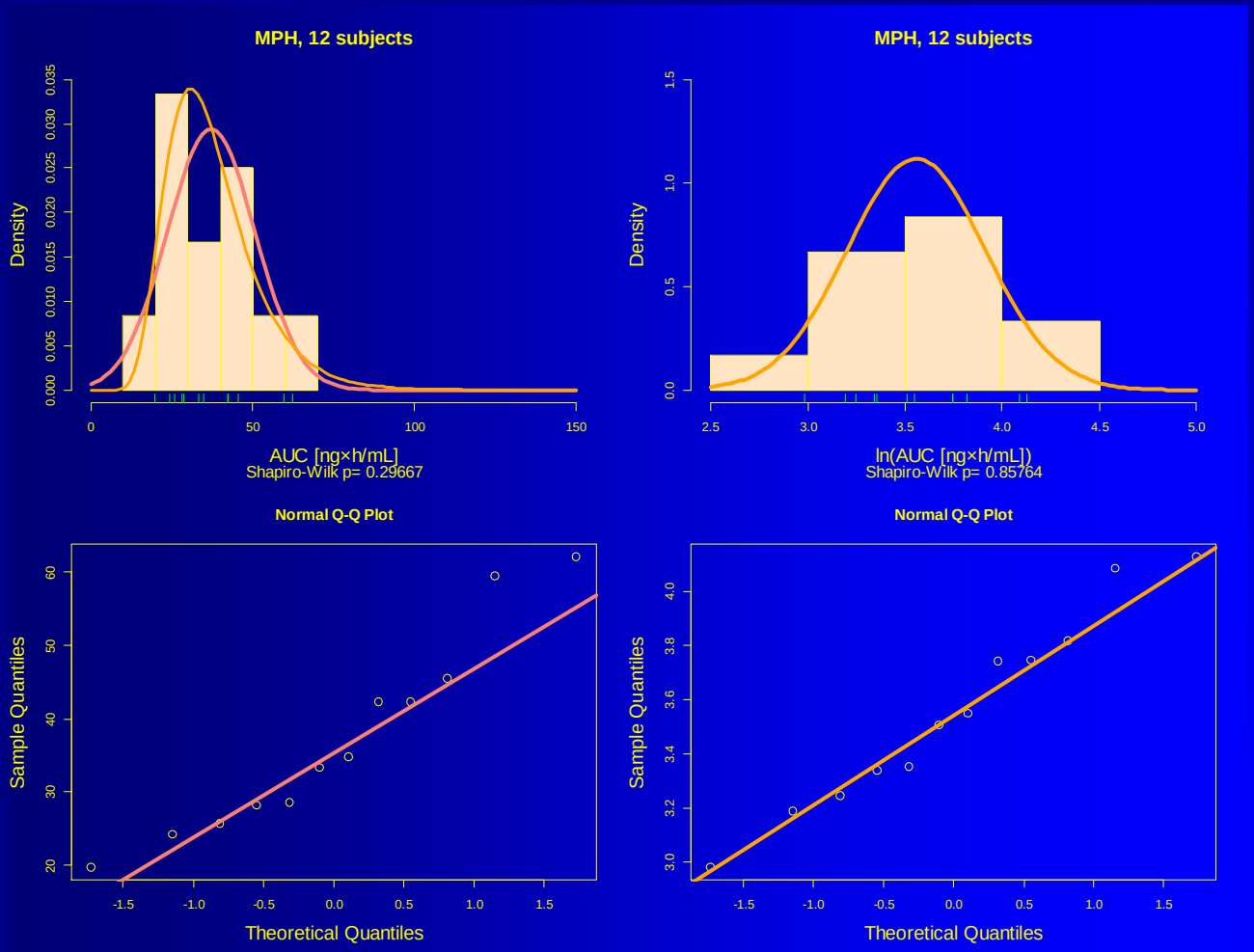


Pooled data from real studies.

Clearly in favor of a lognormal distribution.

Shapiro-Wilk test highly significant for normal distribution (assumption rejected).

Data Transformation!



Data of a real study.

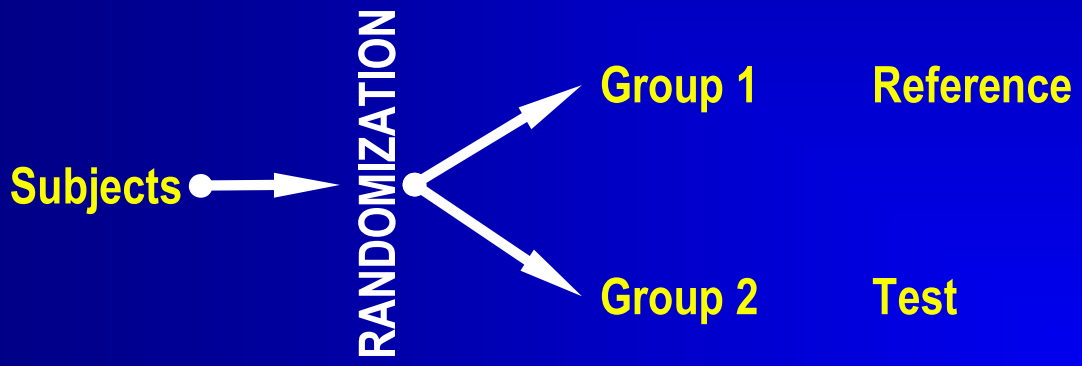
Both tests *not* significant (assumptions accepted).

Tests not acceptable according to GLs.

Transformation based on prior knowledge (PK)!

Parallel designs

- Two-Group Parallel Design

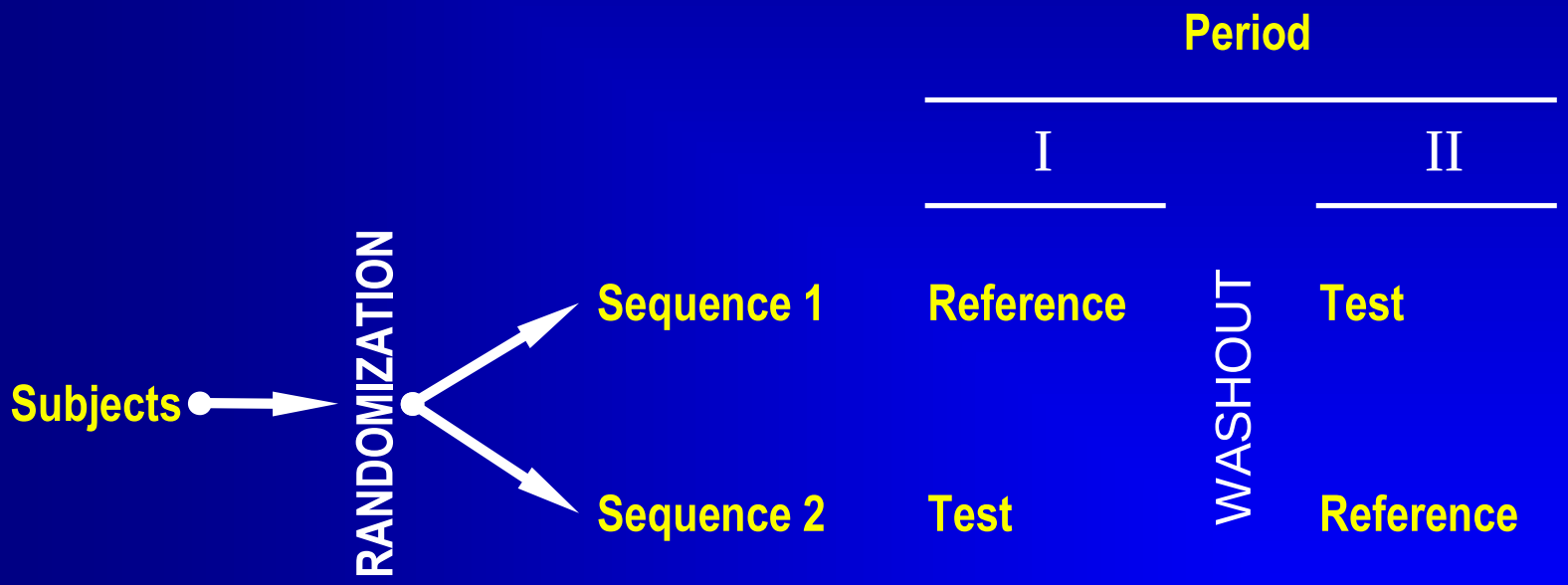


Parallel designs (cont'd)

- 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
    - Phenotyping mandatory for drugs showing polymorphism.

Cross-over designs

- Standard 2×2×2 Design



Cross-over designs (cont'd)

- Every subject is treated both with test and reference
- Subjects are randomized into two groups; one is receiving the formulations in the order RT and the other one in the order TR.
  These two orders are called 'sequences'.
- Whilst in a paired design we must rely on the assumption that no external influences affect the periods, a cross-over design will account for that.

Cross-over design: Model

Multiplicative Model (X-over without carryover)

$$\ln\left(X_{ijk}\right)=\ln\left(\mu\right)+\ln\left(\pi_k\right)+\ln\left(\Phi_l\right)+\ln\left(s_{ik}\right)+\ln\left(e_{ijk}\right)$$

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

X_{ijk}: 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 ($l=1,2$: $\Phi_1=\Phi_{test}$, $\Phi_2=\Phi_{ref.}$)

Cross-over design:
Assumptions

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 (cont'd)

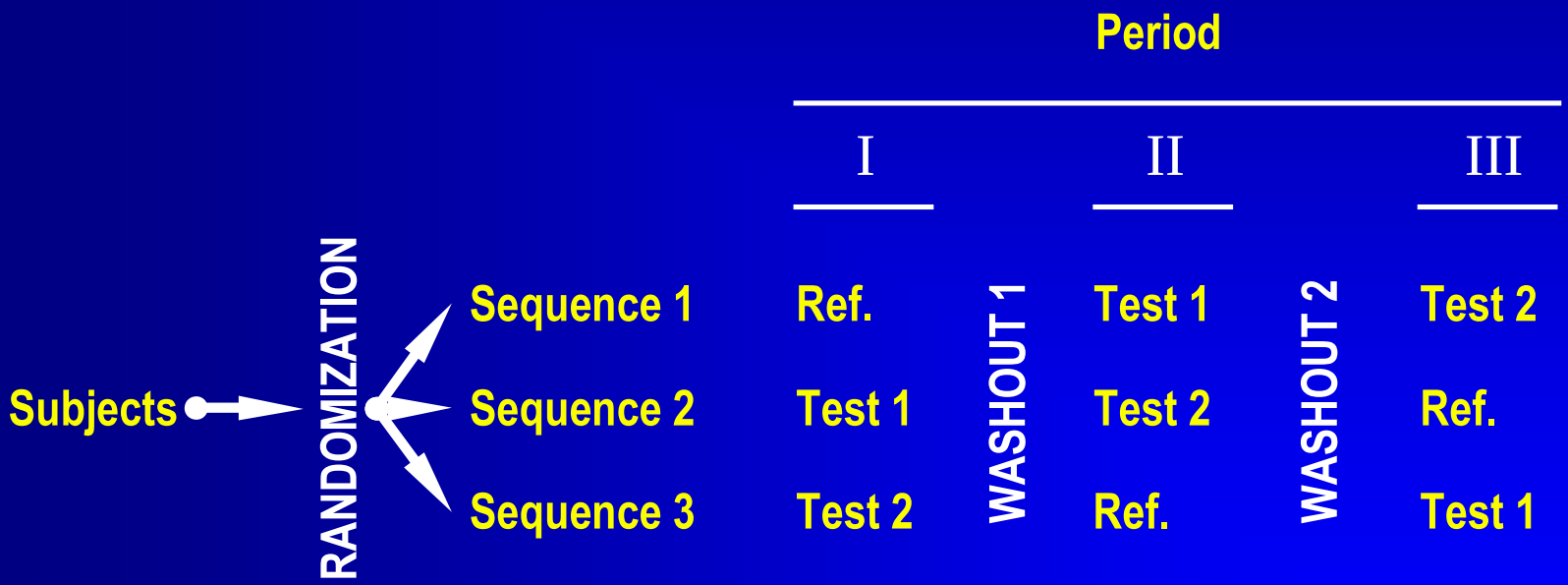
- Standard 2×2×2 design
  - Advantages
    - Globally applied standard protocol for bioequivalence, PK interaction, food studies
    - Straigthforward statistical analysis
  - Disadvantages
    - Not suitable for drugs with long half life
      → parallel design
    - Not optimal for studies in patients with instable diseases
      → parallel design
    - Not optimal for HVDs/HVDPs
      → replicate designs with reference-scaling

Cross-over designs (cont'd)

- 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 (cont'd)

- 3×3×3 Latin Square design



Cross-over designs (cont'd)

- 3×3×3 Latin Square design
  - Advantages
    - Allows to choose between two candidate test formulations or comparison of one 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 – not available in all software.
    - Pairwise comparisons are imbalanced.
    - May need measures against multiplicity (increasing the sample size).
    - Not mentioned in any guideline.

Cross-over designs (cont'd)

- 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 (cont'd)

- 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 (cont'd)

- 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 (cont'd)

- 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 are balanced.
    - Mentioned in Brazil's (ANVISA) and EMA guidelines.
  - 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 – not available in all software.
    - May need measures against multiplicity (increasing the sample size).

Cross-over designs (cont'd)

- Higher Order Designs (cont'd)
  - Bonferroni-correction needed (sample size!)
    - *If more than one formulation will be marketed* (for three simulta´-neous 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 (cont'd)

- 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%

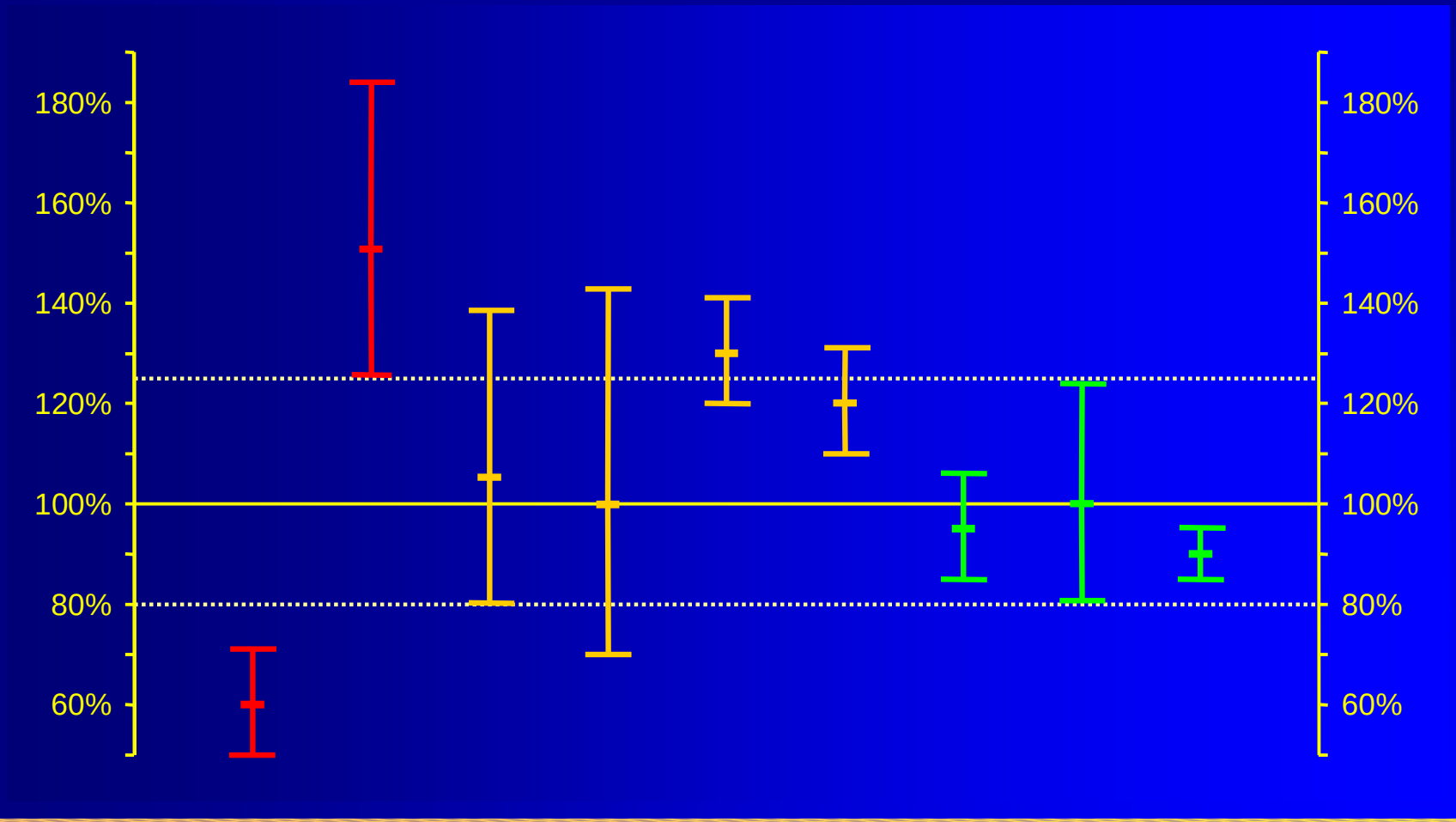
BE Evaluation

- Based on the design set up a statistical model.
- Calculate the test/reference ratio.
- Calculate a (generally 90%) confidence interval (CI) around the ratio.
- The *width* of the CI depends on the variability observed in the study.
- The *location* of the CI depends on the observed test/reference-ratio.

BE Assessment

- Decision based on the CI and the Acceptance Range (AR)
  - CI *entirely outside* the AR:
    Bioinequivalence proven
  - CI *overlaps* the AR (lies *not entirely within* the AR):
    Bioequivalence not proven – indecisive
  - CI lies *entirely within* the AR:
    Bioequivalence proven

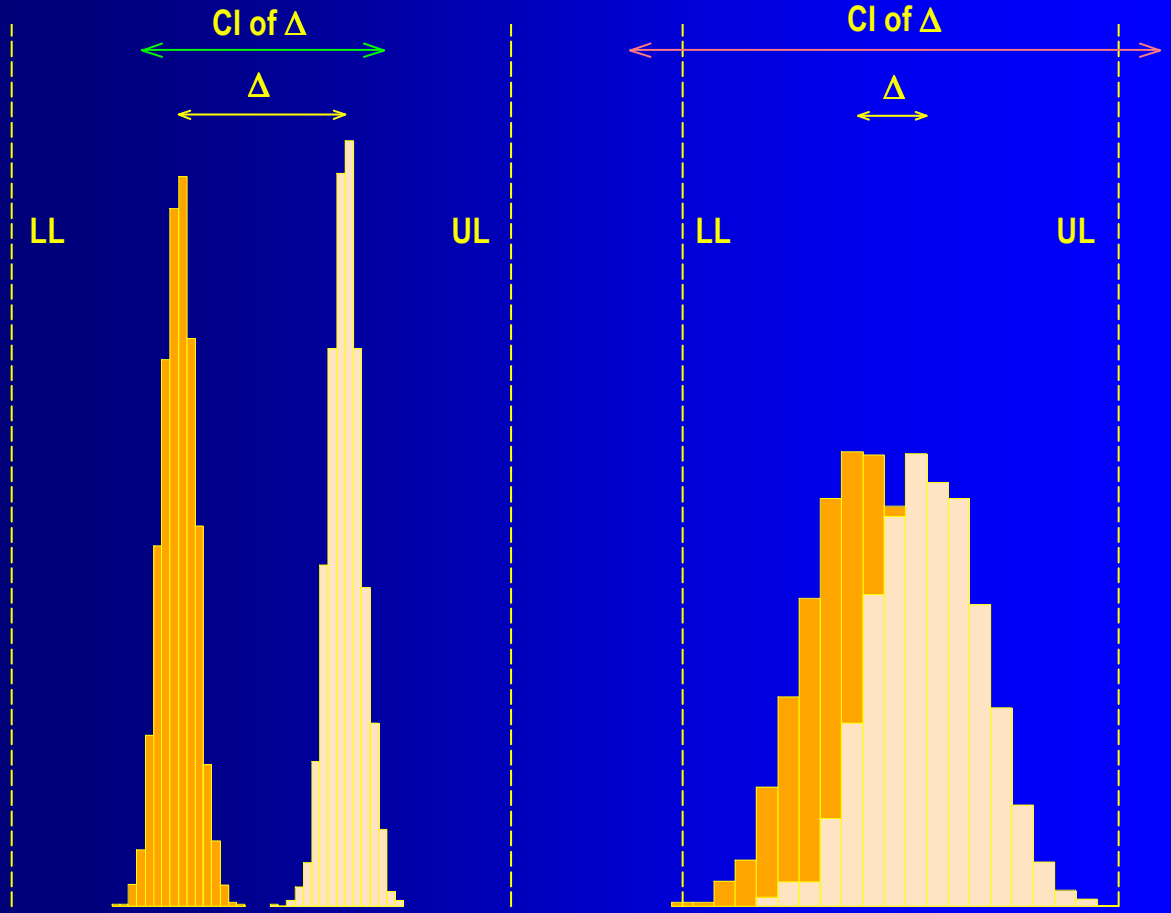
BE Assessment



Add-on / Two-Stage Designs

- Sometimes properly designed and executed studies fail due to
  - 'true' bioinequivalence,
  - poor study conduct (increasing variability),
  - pure chance (producer's risk hit),
  - false (over-optimistic) assumptions about variability and/or T/R-ratio.
- The patient's risk must be preserved
  - Already noticed at Bio-International Conferences (1989, 1992) and guidelines from the 1990s.

High variability…



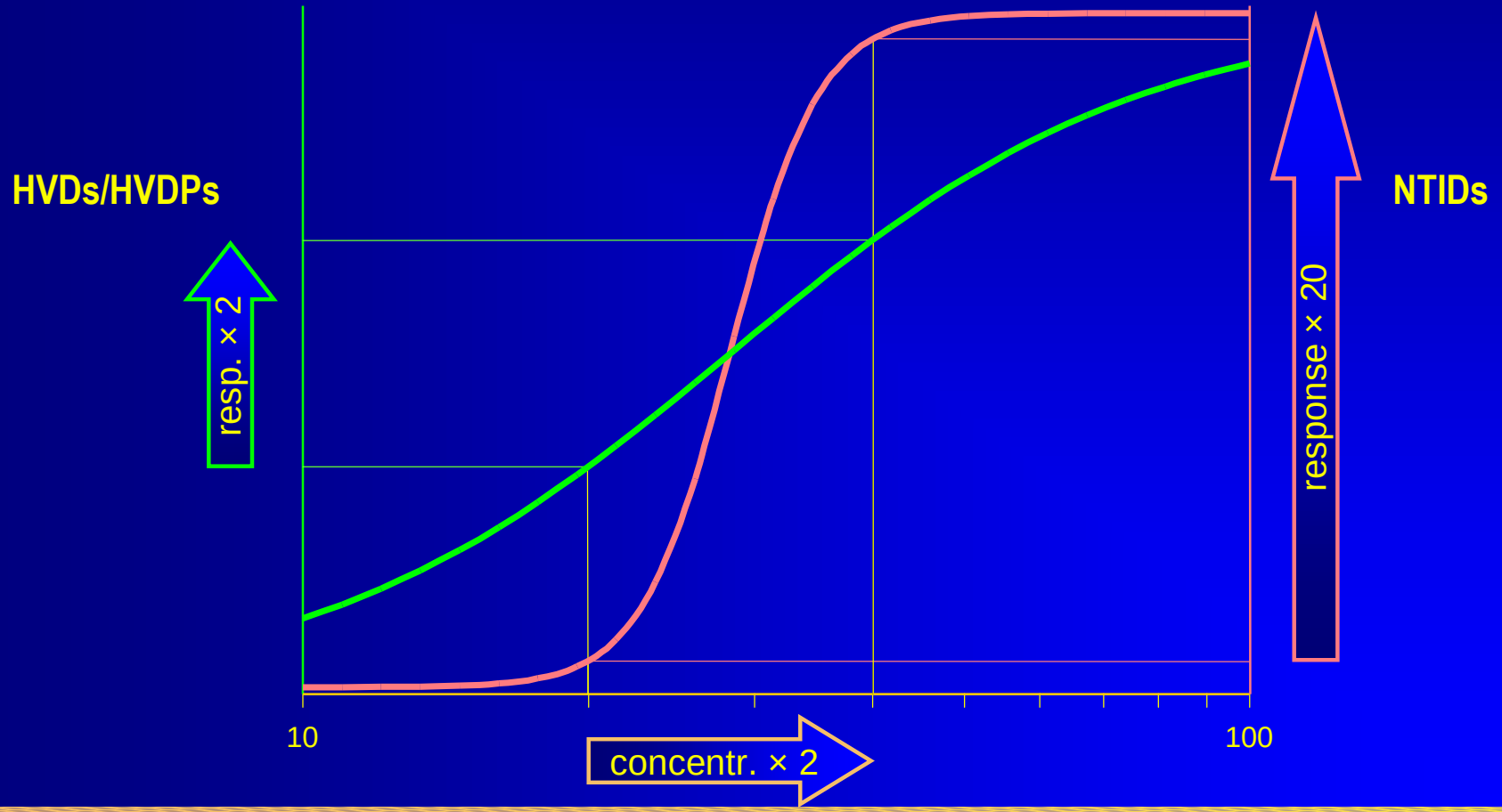
Modified from Fig. 1
Tothfálusi *et al.* (2009)

Counterintuitive concept of BE:

Two formulations with a large difference in means are declared bioequivalent if variances are low, but not bioequivalent – even if the difference is quite small – due to high variability.

HVDs/HVDPs are safe

flat & steep PK/PD-curves



High variability

- For Highly Variable Drugs / Drug Products (HVDs/HVDPs) it may be almost impossible to show BE with a reasonable sample size.
- The common 2×2 cross-over design over assumes Independent Identically Distributions (IID), which may not hold. If *e.g.*, the variability of the reference is higher than the one of the test, one obtains a high common (pooled) variance and the test will be penalized for the 'bad' reference.

Replicate designs

- Each subject is randomly assigned to sequences, where *at least one* of the treatments (generally the reference) 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 (biosamples to be analyzed)!

Replicate designs

- Any replicate design can be evaluated according to 'classical' (unscaled) Average Bioequivalence (ABE)
- ABE mandatory if scaling not allowed
  - FDA: s_{WR} <0.294 (CV_{WR} <30%); different models dependend on design (*i.e.*, SAS `Proc MIXED` for full replicate and `Proc GLM` for partial replicate).
  - EMA: CV_{WR} ≤30%; all fixed effects model according to 2011's Q&A-document preferred (*e.g.*, SAS `Proc GLM`).
  - Even if scaling is not intended or applicable, replicate designs give more information about formulation(s).

Application: HVDs/HVDPs

- CV_{WR} >30 %
  - ✓USA Recommended in API specific guidances. Scaling for AUC and/or C_{max} acceptable, GMR 0.80 – 1.25; ≥24 subjects enrolled.
  - ± EU Widening of acceptance range (only C_{max}) to maximum of 69.84 – 143.19%), GMR 0.80 – 1.25. Demonstration that CV_{WR} >30% is not caused by outliers. Justification that the widened acceptance range is clinically not relevant.

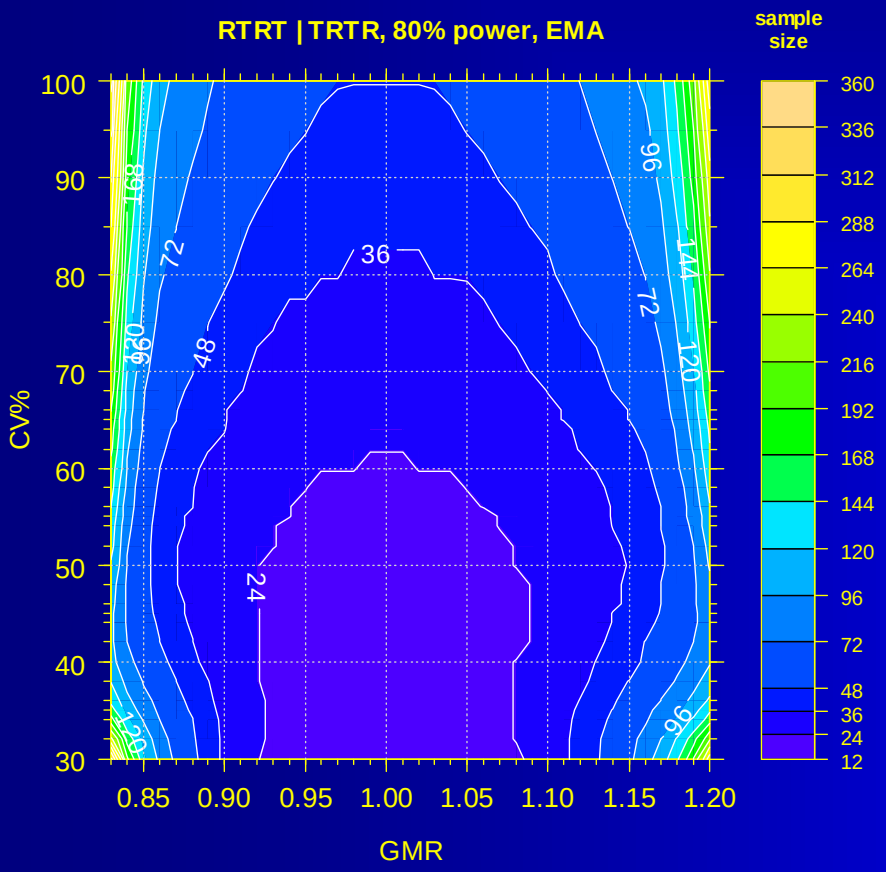
Replicate designs

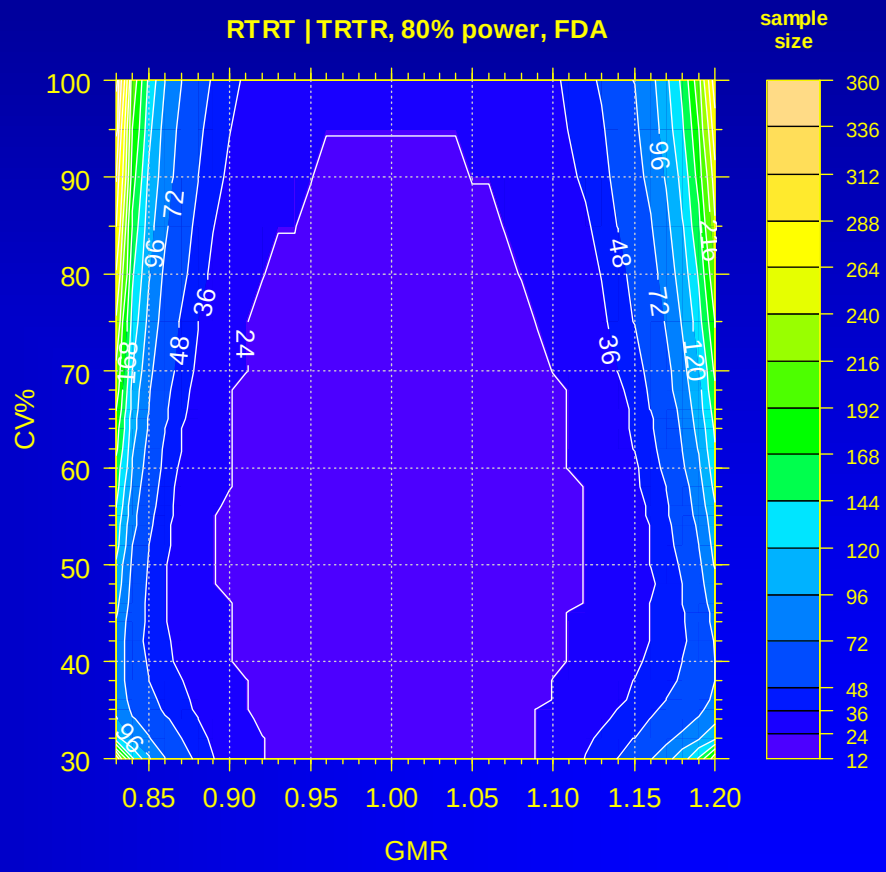
- Two-sequence three-period
  - T R T
  - R T R
- Two-sequence four-period
  - T R T R
  - R T R T
- and many others…

  (FDA: TRR | RTR | RRT, aka ‘partial replicate’)
- The statistical model is complicated and depends on the actual design!

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

HVDPs (EMA/FDA; sample sizes)





HVDPs (EMA)

- EU GL on BE (2010)
  - Average Bioequivalence (ABE) with Expanding Limits (ABEL)
    - Based on σ_{WR} (the *intra*-subject standard deviation of the reference formulation) calculate the scaled acceptance range based on the regulatory constant k (θ_s = 0.760); limited at CV_{WR} 50%.

$$[L-U] = e^{\mp k \cdot \sigma_{WR}}$$

CV_{WR}	$L-U$
≤30	80.00 – 125.00
35	77.23 – 129.48
40	74.62 – 143.02
45	72.15 – 138.59
≥50	69.84 – 143.19

HVDPs (EMA)

- **Q&A document (March 2011)**
  - **Two methods proposed (Method A preferred)**
    - **Method A: All effects fixed; assumes equal variances of test and reference, and no subject-by-formulation interaction; only a common within (*intra*-) subject variance is estimated.**
    - **Method B: Similar to A, but random effects for subjects. Common within (*intra*-) subject variance and between (*inter*-) subject variance are estimated.**
  - **Outliers: Boxplots (of model residuals?) suggested.**

Questions & Answers on the Revised EMA Bioequivalence Guideline
Summary of the discussions held at the 3rd EGA Symposium on Bioequivalence
June 2010, London
http://www.egagenerics.com/doc/EGA_BEQ_Q&A_WEB_QA_1_32.pdf

Example datasets (EMA)

- Q&A document (March 2011)
  - Data set I: Full replicate (RTRT | TRTR), 77 subjects, imbalanced, incomplete
    - FDA

      s_{WR} 0.446 ≥0.294 → apply RSABE (CV_{WR} 46.96%)
      a. critbound –0.0921 ≤0 and
      b. PE 115.46% ⊂ 80.00–125.00% ✓
    - EMA
      - CV_{WR} 46.96% → apply ABEL (> 30%)
      - Scaled Acceptance Range: 71.23–140.40%
      - Method A: 90% CI 107.11–124.89% ⊂ AR; PE 115.66% ✓
      - Method B: 90% CI 107.17–124.97% ⊂ AR; PE 115.73% ✓

Example datasets (EMA)

- Q&A document (March 2011)
  - Data set II: Partial replicate (TRR | RTR | RRT), 24 subjects, balanced, complete
    - FDA

      s_{WR} 0.114 <0.294 → apply ABE (CV_{WR} 11.43%)
      90% CI 97.05–107.76% ⊂ AR (CV_{intra} 11.55%) ✓
    - EMA
      - CV_{WR} 11.17% → apply ABE (≤30%)
      - Method A: 90% CI 97.32–107.46% ⊂ AR; PE 102.26% ✓
      - Method B: 90% CI 97.32–107.46% ⊂ AR; PE 102.26% ✓
      - A/B: CV_{intra} 11.86%

Thank You!

Basic Designs for BE Studies

Open Questions?

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

To bear in Remembrance...

To call the statistician after the experiment is done may be no more than asking him to perform a *post-mortem* examination: he may be able to say what the experiment died of.

Ronald A. Fisher

[The] impatience with ambiguity can be criticized in the phrase:
absence of evidence is not evidence of absence.

Carl Sagan

[…] our greatest mistake would be to forget that data is used for serious decisions in the very real world, and bad information causes suffering and death.

Ben Goldacre