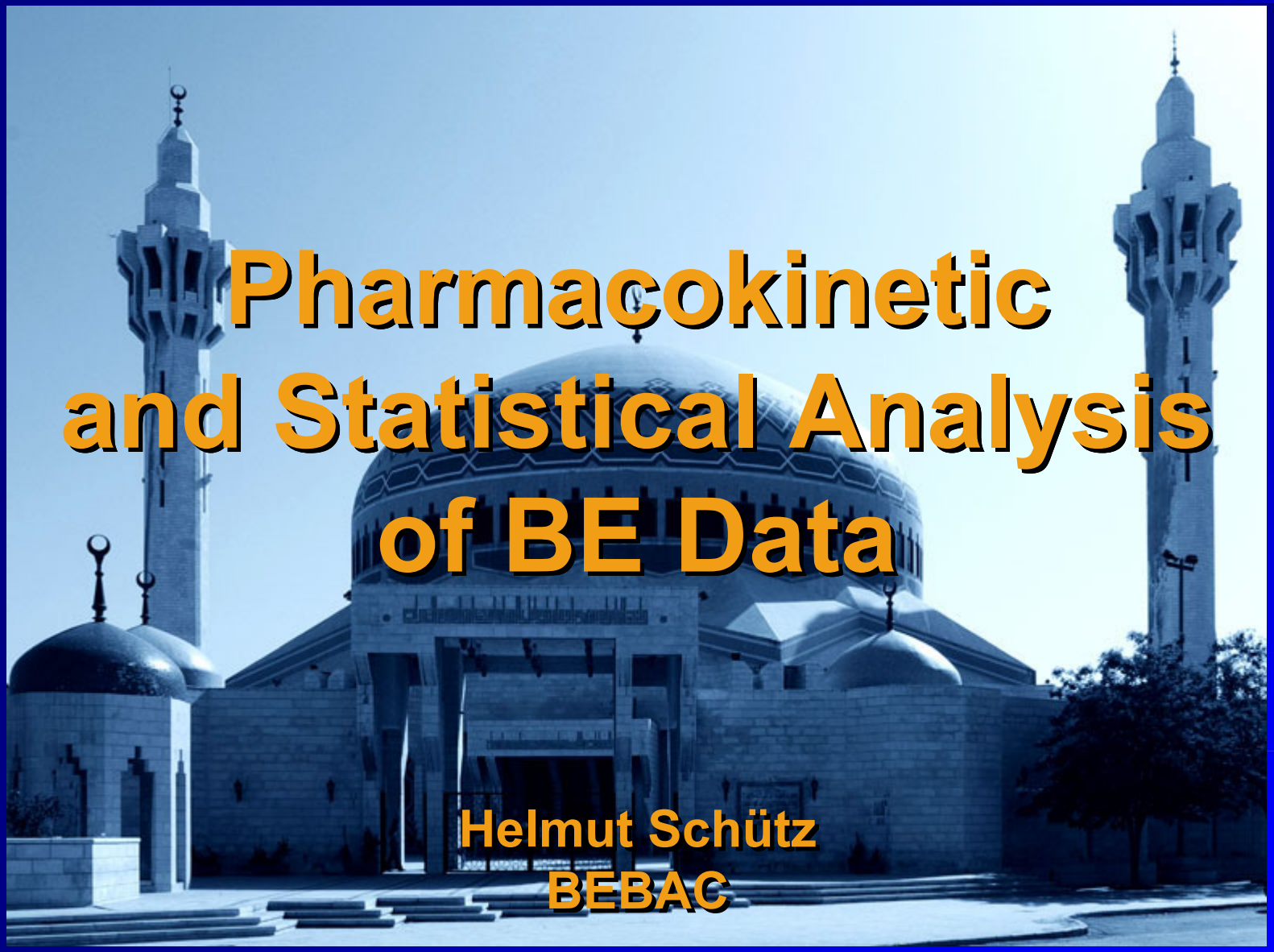
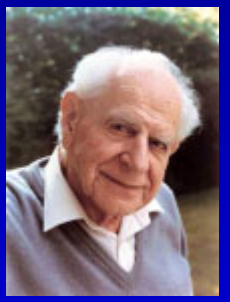


Pharmacokinetic and Statistical Analysis of BE Data

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

NCA *vs.* PK Modeling

- Pharmacokinetic models
  - Useful for understanding the drug/formulation
    - Study design of BA/BE, *e.g.*,
      washout, accumulation / saturation to steady state
  - Drawbacks
    - Almost impossible to validate (fine-tuning of side conditions, weighting schemes, software, …)
    - Still a mixture of art and science
    - Impossible to recalculate any given dataset using different software – sometimes even different versions of the same software!
    - Not acceptable for *evaluation* of BE studies!

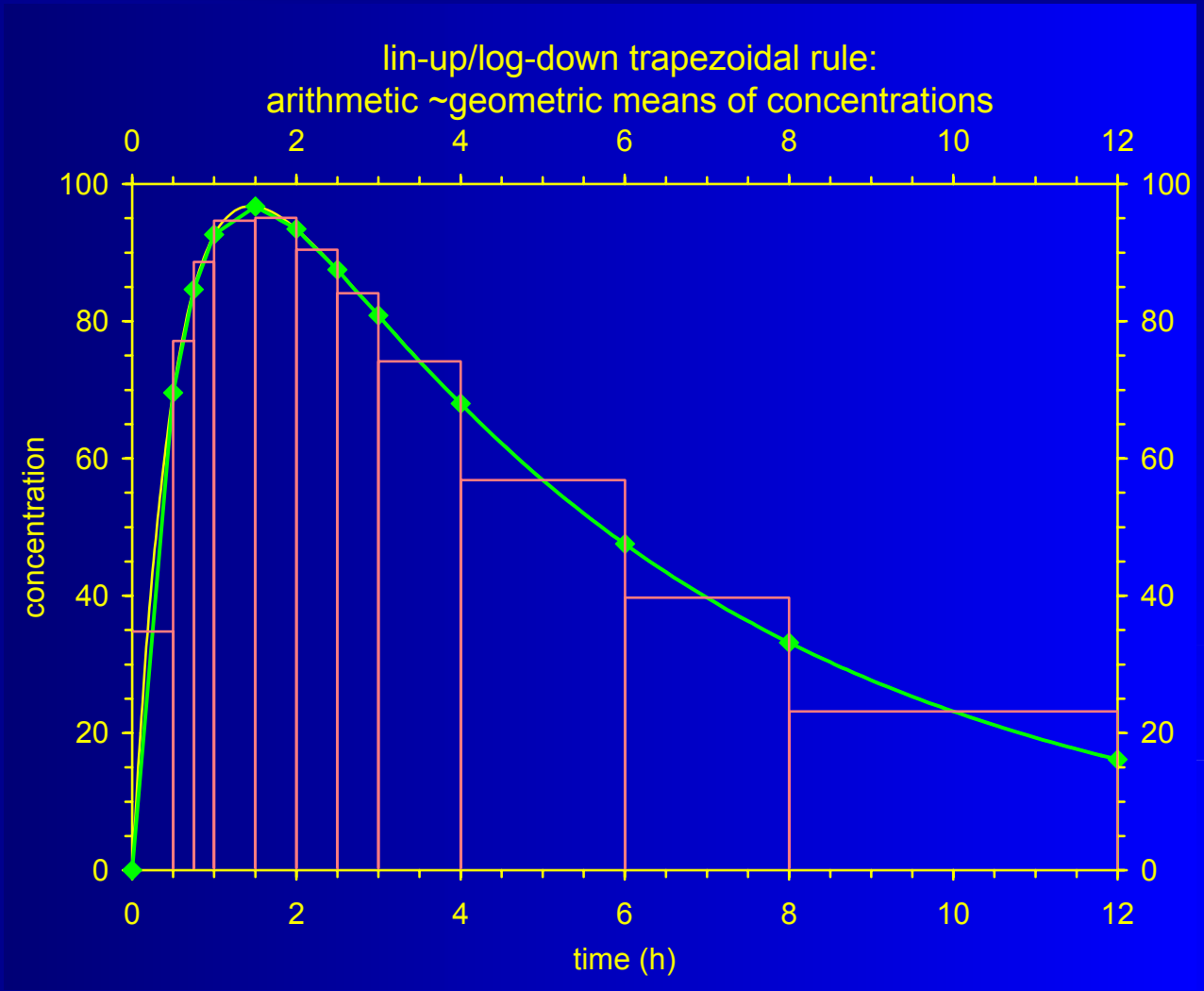
NCA: Single Dose

- Noncompartmental methods do not rely on a PK (=compartmental) model
- Also known as SHAM (Shape, Height, Area, Moments)
  - Metrics (plasma, single dose)
    - Extent of absorption (EU…), total exposure (US): AUC (Area Under the Curve)
    - Rate of absorption (EU…), peak exposure (US): C_{max}
    - t_{max} (EU…)
    - Early exposure (US, CAN): $pAUC_{tmax}$; AUC truncated at population's (CAN: subject's) t_{max} of the reference
    - Others: C_{min}, Fluctuation, MRT, Occupancy time, t_{lag},…

NCA: AUC

- Recommended: lin-up/log-down trapezoidal rule
  - Hybrid of linear and log-linear
  - Sections with *increasing or equal* concentrations ($C_{i+1} \geq C_i$) calculated by linear trapezoidal rule
  - Sections with *decreasing* concentrations ($C_{i+1} < C_i$) calculated by log-linear trapezoidal rule
  - Avoids bias in both absorption and distribution/ elimination phases
  - Suitable for IV and EV
  - Suitable for multiphasic profiles

NCA: AUC


lin-up/log-down trapezoidal rule:
arithmetic ~geometric means of concentrations
concentration
time (h)
0
20
40
60
80
100
0
2
4
6
8
10
12

NCA: AUC Extrapolation

- $AUC_{0-\infty}$
  - Unweighted log-linear regression of ≥3 data points in the elimination phase
  - Don't rely on softwares' automatic methods; visual inspection of the fit mandatory
  - Extrapolation from AUC_{0-t} (regardless the method)

$$AUC_{\infty} = AUC_t + \frac{C_t}{\hat{\lambda}_z} \quad \text{or better} \quad AUC_{\infty} = AUC_t + \frac{\hat{C}_t}{\hat{\lambda}_z}$$

NCA: other PK Metrics

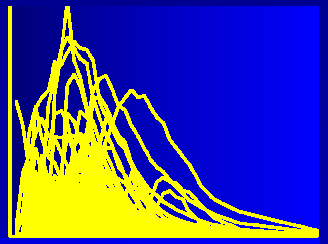
- Single dose
  - C_{max} and t_{max} directly from profile
  - Metrics describing the shape of the profile
    - Early exposure (US, CAN): AUC_{tmax} = $pAUC$ truncated at population (CAN: subject's) t_{max} of the reference
    - Biphasic MR formulations: $pAUCs$ truncated at a prespecified cut-off time point
      - FDA: Product specific guidances (methylphenidate, zolpidem)
      - EMA: All products

        Questions & Answers: Positions on specific questions addressed to the pharmacokinetics working party
        EMA/618604/2008 Rev. 7 (13 February 2013)
        http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500002963.pdf

NCA: other PK Metrics

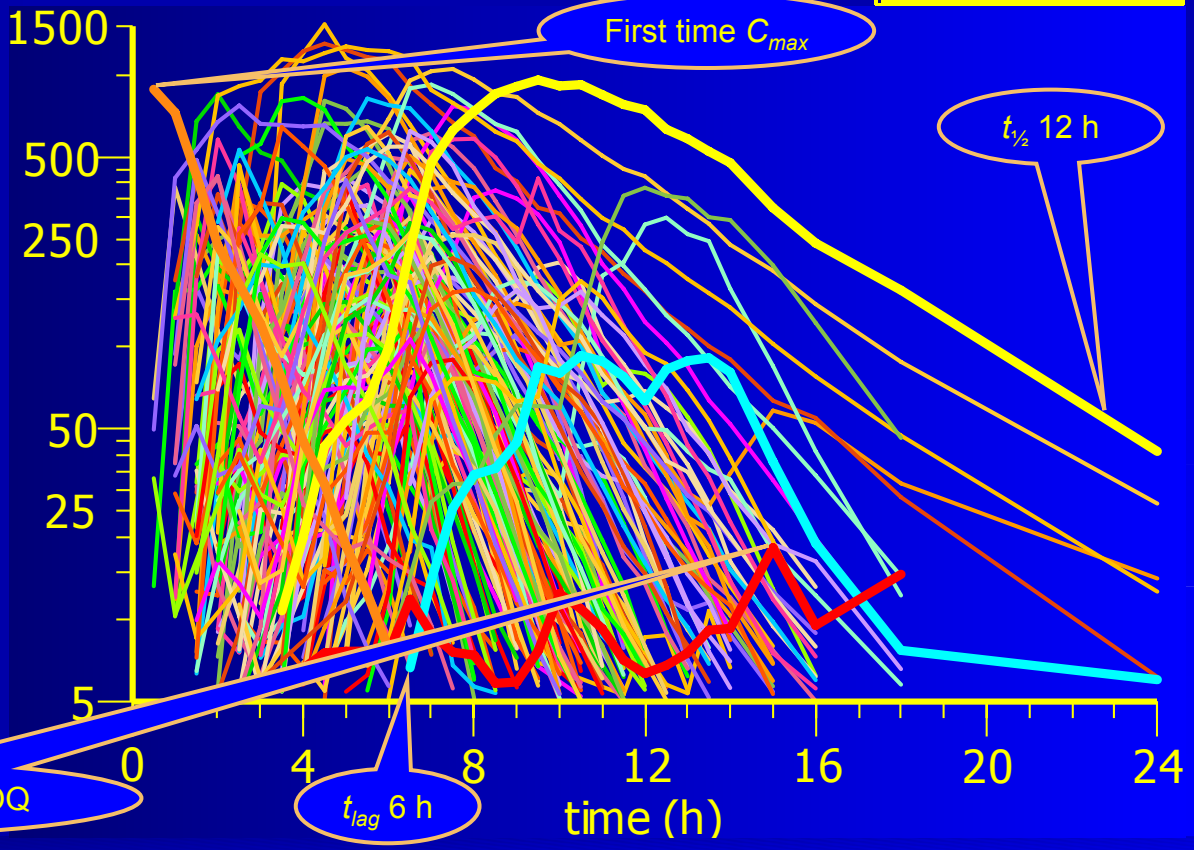
- Single dose
  - Metrics describing the shape of the profile
    - C_{max}/AUC
    - $t_{75\%}$ = $POT\text{-}75$ (Plateau time, Peak-Occupancy-Time 75: time interval where $C(t) \geq 75\%$ of C_{max})
    - HVD = $POT\text{-}50$ (Half Value Duration, Peak-Occupancy-Time 50: time interval where $C(t) \geq 50\%$ of C_{max})
    - Occupancy time, $t \geq MIC$ (time interval where $C(t)$ is above some limiting concentration)

Case Study (PPI)



- Attempt to deal with high variability

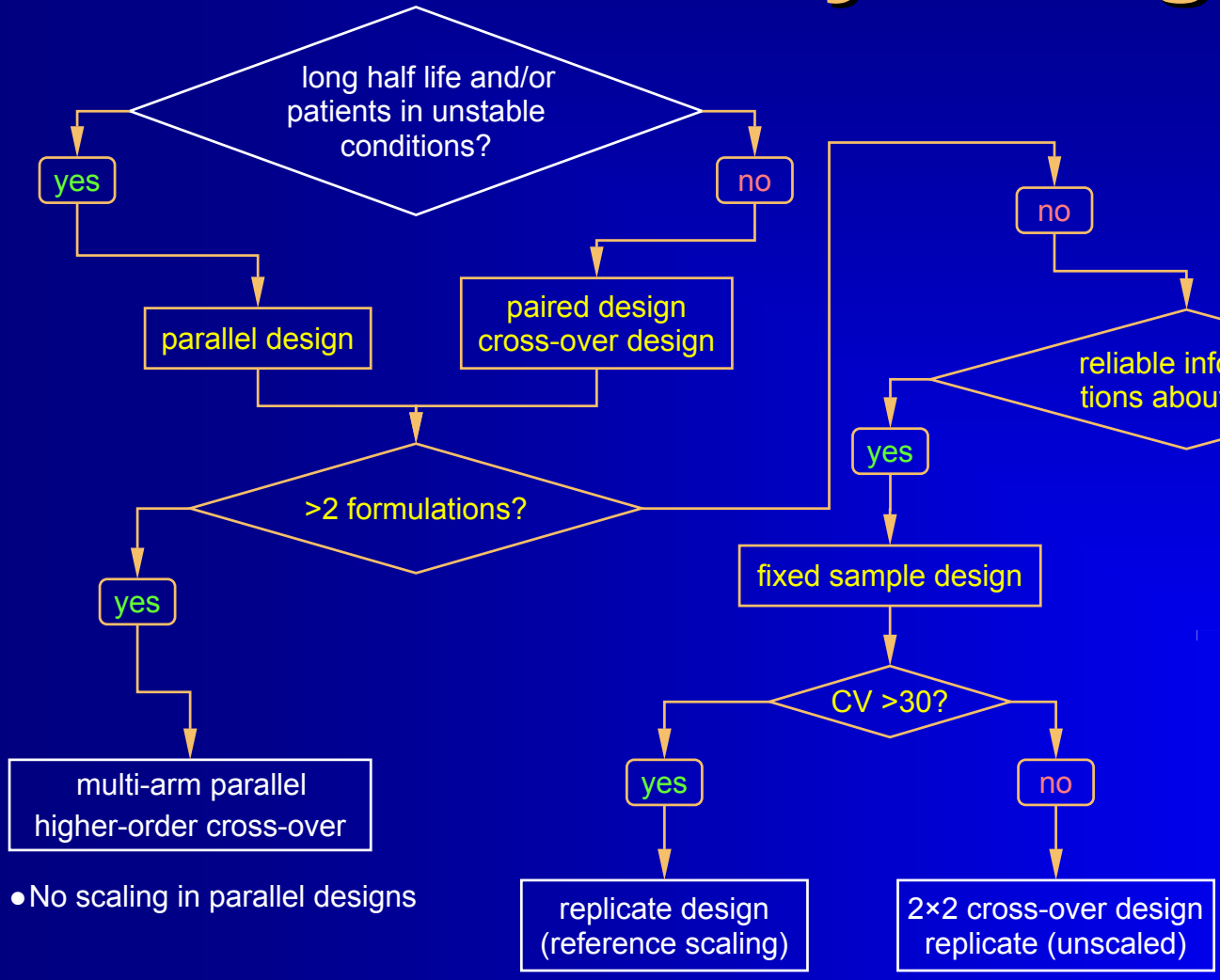
Powered to 90% according to CV from previous studies; 140 (!) subjects and to 80% for expect-ed dropout rate. Sampling every 30 min up to 14 hours (7,785 total)



NCA: Multiple Dose

- AUC_{τ} (dosage interval τ) or $AUC_{ss,24h}$ (if more than *o.a.d.* and chronopharmacological variation)
- No extrapolation!
- $C_{ss,max}$ and $C_{ss,min}$ directly from profile
- Peak-Trough-Fluctuation: $(C_{ss,max} - C_{ss,min}) / C_{ss,av}$, where $C_{ss,av} = AUC_{\tau} / \tau$
- Swing: $(C_{ss,max} - C_{ss,min}) / C_{ss,min}$

BE Study Designs



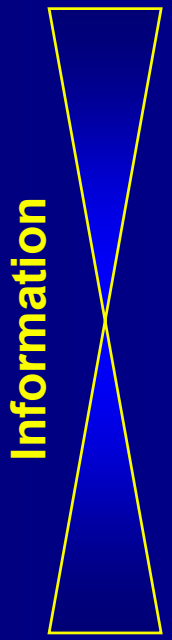
- No scaling in parallel designs

- *Currently* no two-stage design if
  - Parallel design
  - >2 formulations
  - Replicate design
- Futility rules (*i.e.*, maximum sample size) in two-stage designs problematic

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)



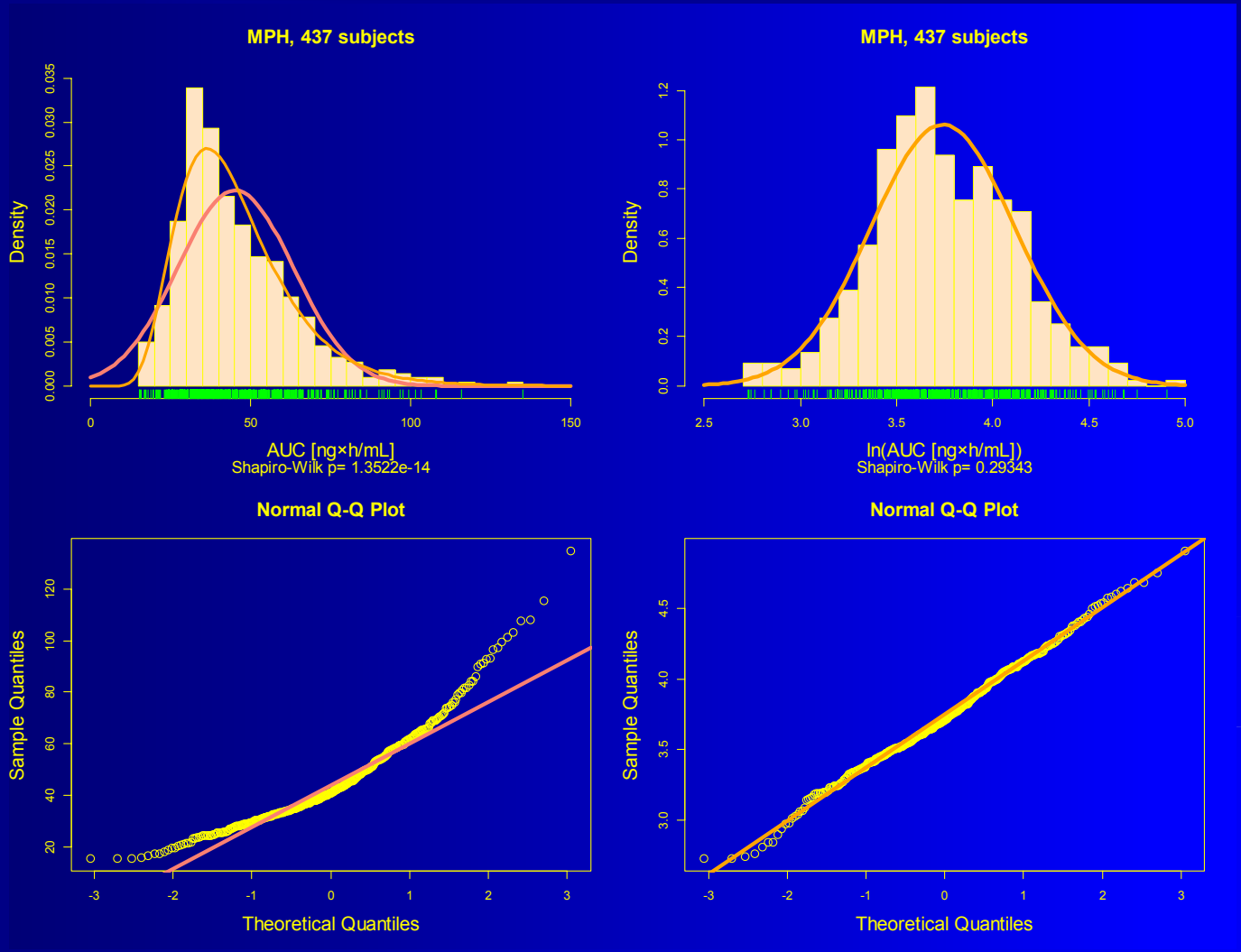
  - Variances which can be estimated:

    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?

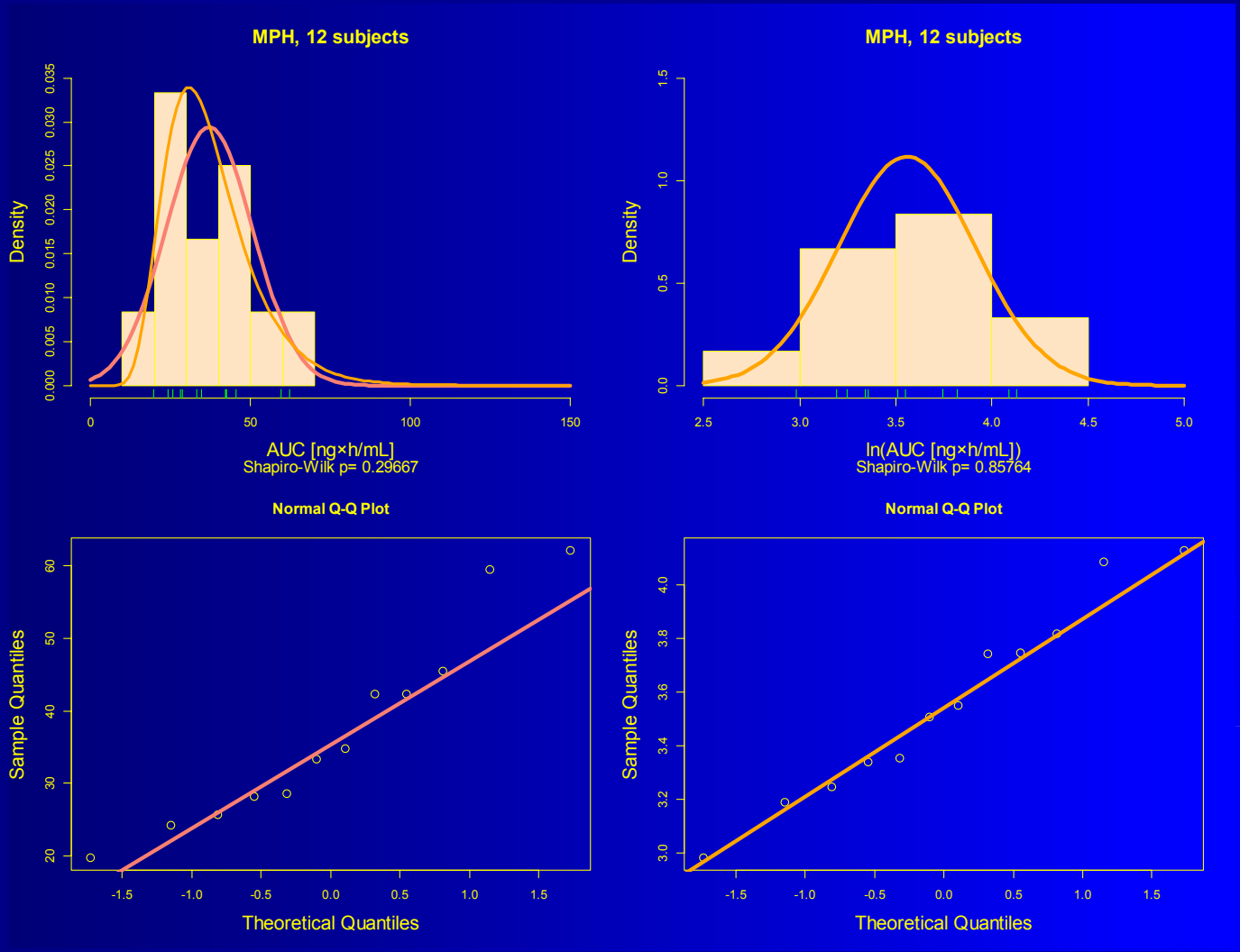


Pooled data from real studies.

Clearly in favor of a lognormal distribution.

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

Data Transformation!



Data of a real study.

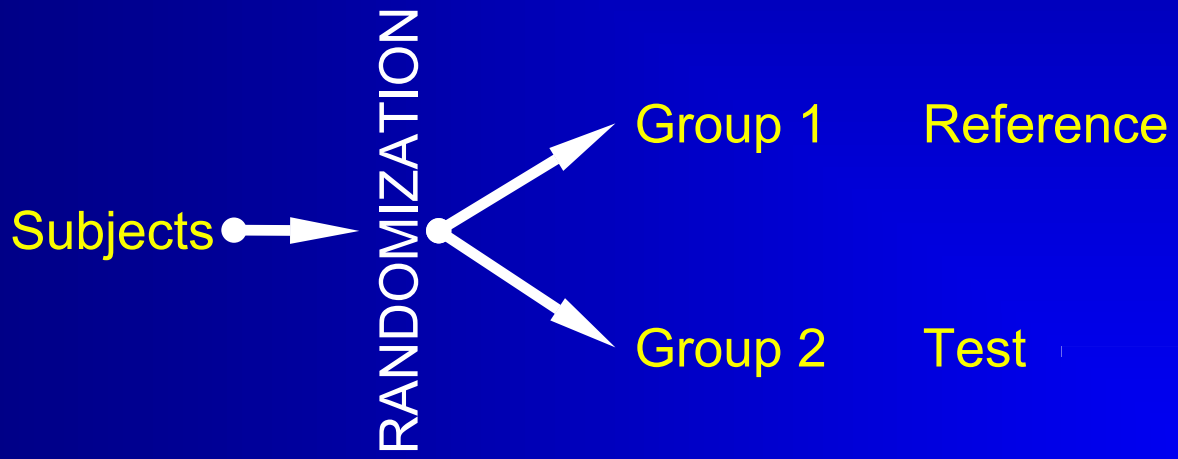
Both tests *not* significant (assumptions accepted).

Tests not acceptable according to GLs.

Transformation based on prior knowledge (PK)!

Parallel designs

- Two-Group Parallel Design

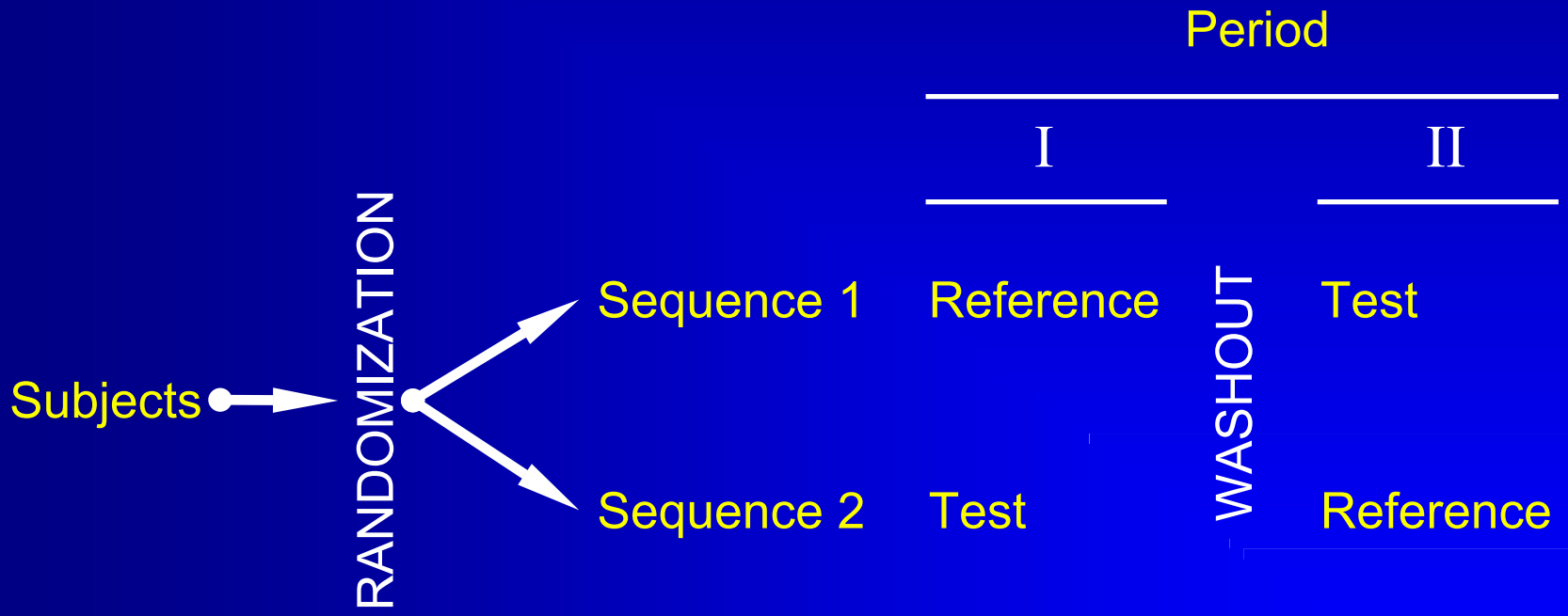


Parallel designs (cont'd)

- 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 (*rule of thumb:* sample size should at least be doubled).
    - 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 groups)
    - Not optimal for studies in patients with instable diseases (→ parallel groups)
    - Not optimal for HVDs/HVDPs (→ Replicate Designs)

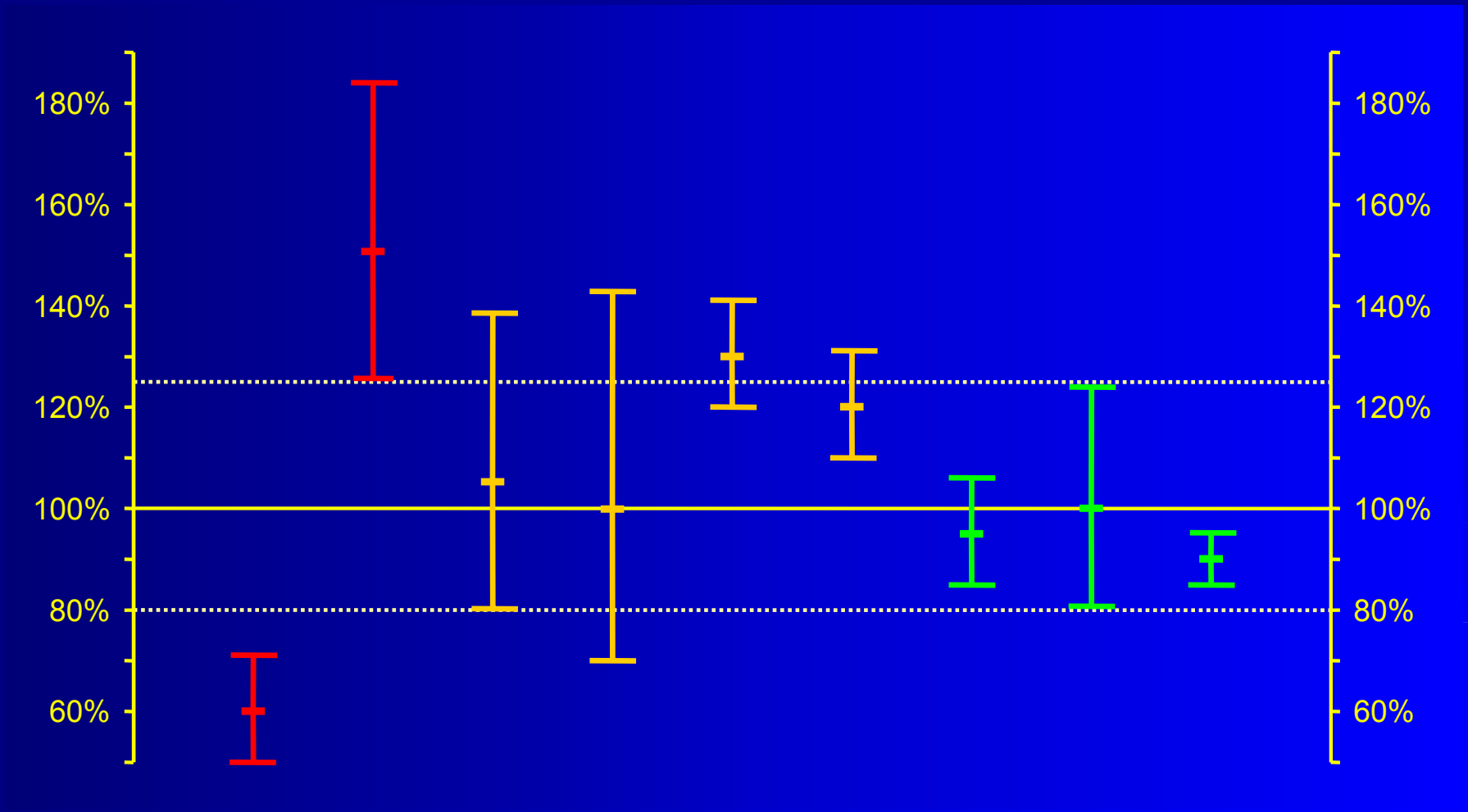
BE Evaluation

- Based on the design set up a statistical model.
- Calculate the test/reference ratio.
- Calculate the 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 rules 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.

Sequential Designs

- Have a long and accepted tradition in clinical research (mainly phase III)
  - Based on work by Armitage *et al.* (1969), McPherson (1974), Pocock (1977), O'Brien and Fleming (1979), Lan & DeMets (1983), …
    - First proposal by Gould (1995) in the area of BE did not get regulatory acceptance in Europe, but
    - new methods stated in recent guidelines.

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

Sequential Designs

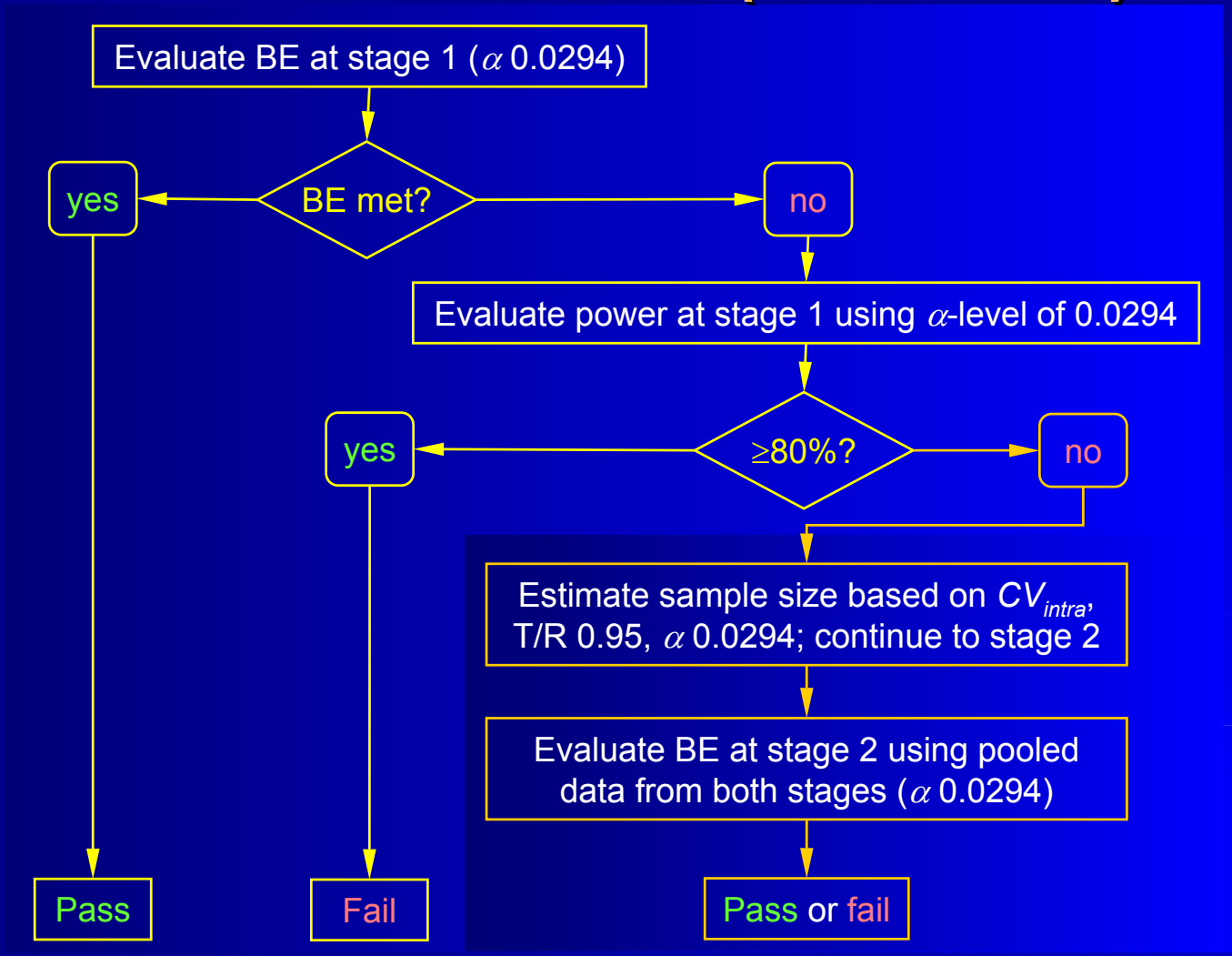
- Methods by Potvin *et al.* (2008) first validated framework in the context of BE
  - Supported by the 'Product Quality Research Institute' (members: FDA/CDER, Health Canada, USP, AAPS, PhRMA…)
    - Three of BEBAC's protocols accepted by German BfArM, one product approved in 06/2011.

Potvin D, Diliberti CE, Hauck WW, Parr AF, Schuirmann DJ, and RA Smith
Sequential design approaches for bioequivalence studies with crossover designs
Pharmaceut Statist 7(4), 245–62 (2008) DOI: 10.1002/pst.294

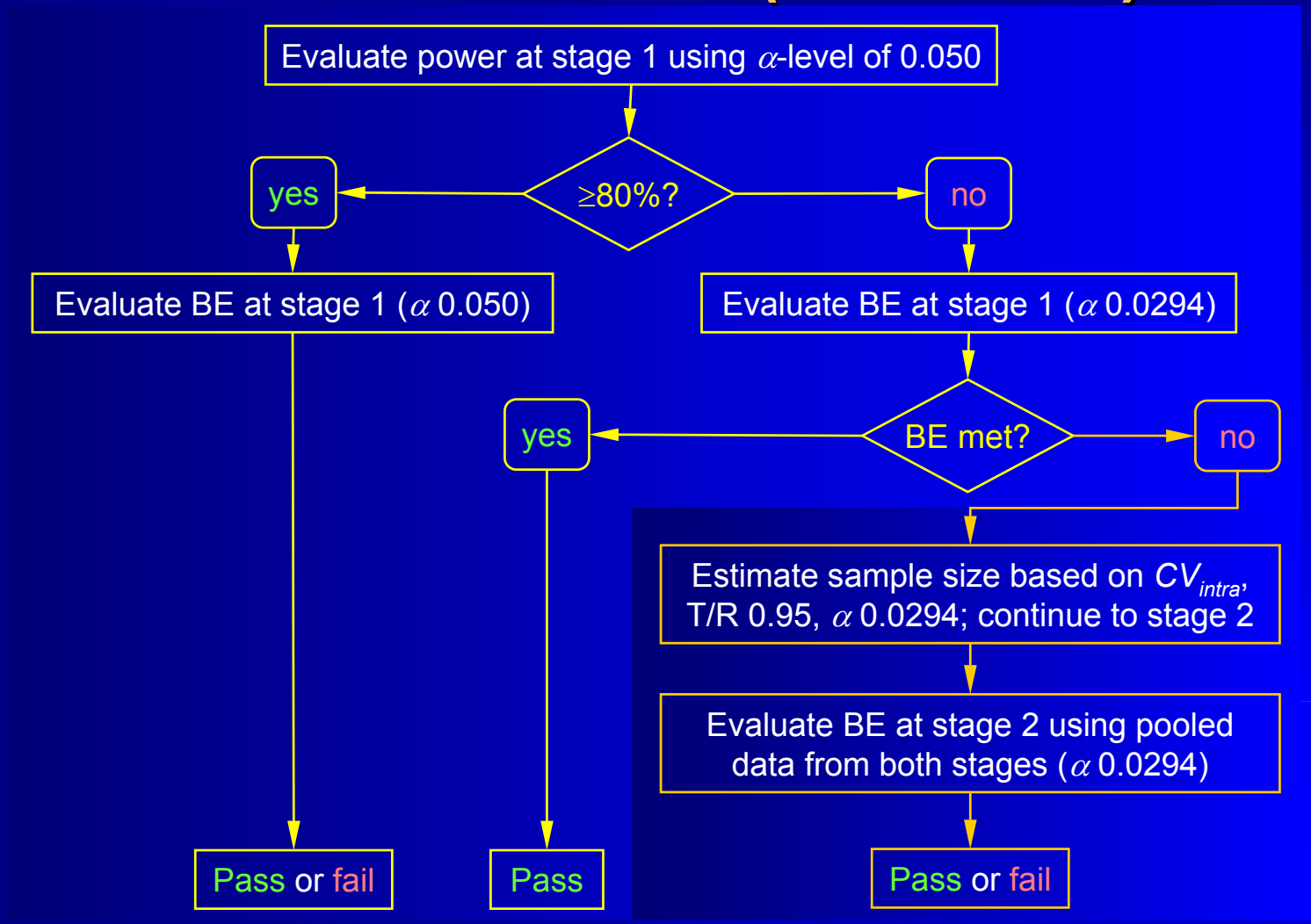
Review of Guidelines

- EMA (Jan 2010)
  Acceptable; Potvin *et al.* Method B preferred (?)
- Russia (Draft 2011)
  Acceptable (Methods B and C)
- Canada (May 2012)
  Potvin *et al.* Method C recommended
- FDA (Jun 2012)
  Potvin *et al.* Method C recommended
  API specific guidances: Loteprednol, Dexamethasone / Tobramycin

Potvin *et al.* (Method B)



Potvin *et al.* (Method C)



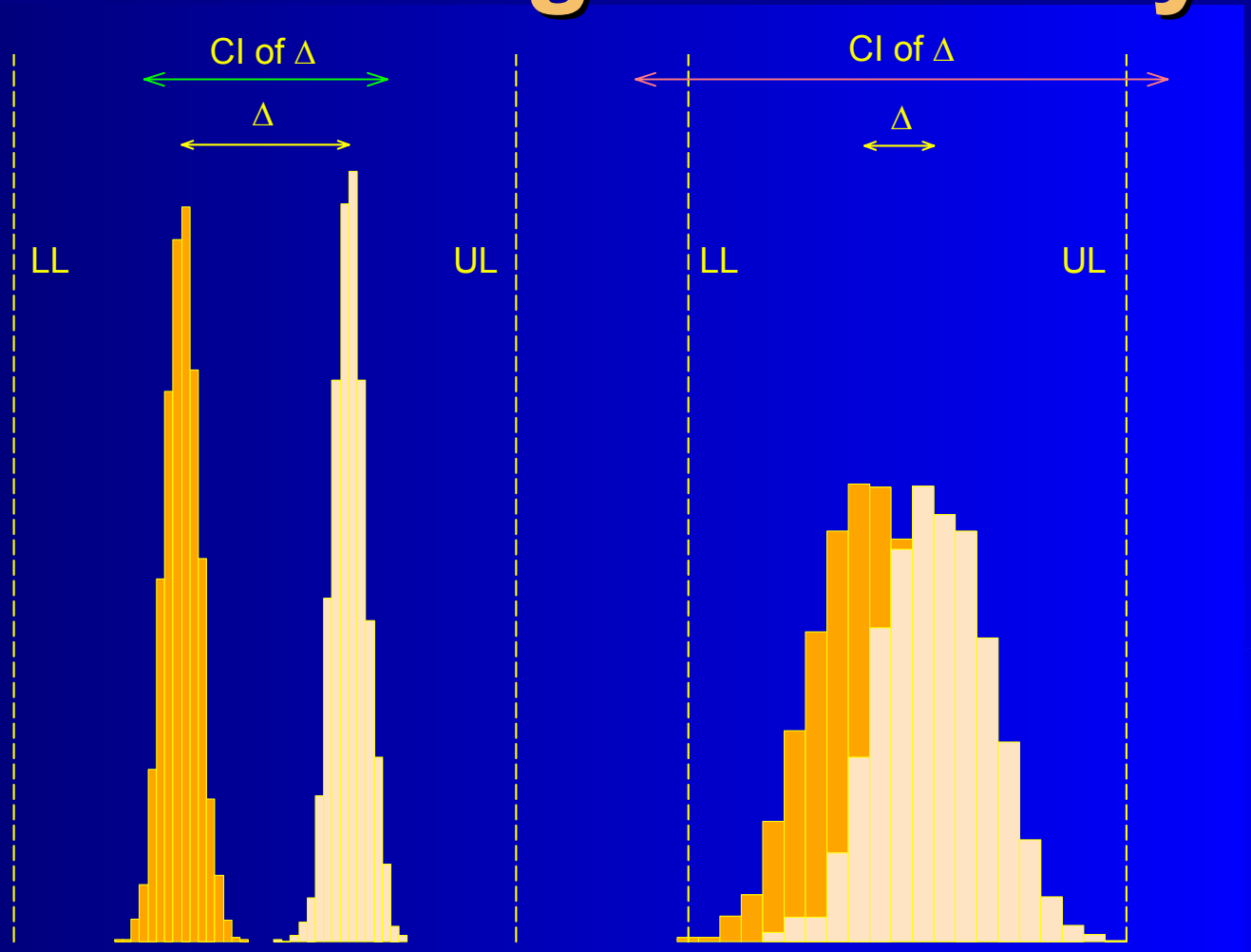
TSDs: Alternatives

- Methods by Potvin *et al.* (2008) limited to T/R of 0.95 and 80% power
  - Follow-up papers (T/R 0.95…0.90, 80…90% power)

reference	method	T/R	target power	*CV*	$\alpha_{adj.}$	max.$\alpha_{emp.}$
Potvin *et al.*	B	0.95	80%	10–100%	0.0294	0.0485
	C	0.95				0.0510
Montague *et al.*	D	0.90			0.0280	0.0518
Fuglsang	B	0.95	90%	10–80%	0.0284	0.0501
	D				0.0274	0.0503
	D	0.90			0.0269	0.0501

Montague TH, Potvin D, DiLiberti CE, Hauck WW, Parr AF, and DJ Schuirmann
Additional results for 'Sequential design approaches for bioequivalence studies with crossover designs'
Pharmaceut Statist 11(1), 8–13 (2011) DOI: 10.1002/pst.483
A Fuglsang
Sequential Bioequivalence Trial Designs with Increased Power and Controlled Type I Error Rates
AAPS J 15(3), 659–61 (2013) DOI: 10.1208/s12248-013-9475-5

High variability



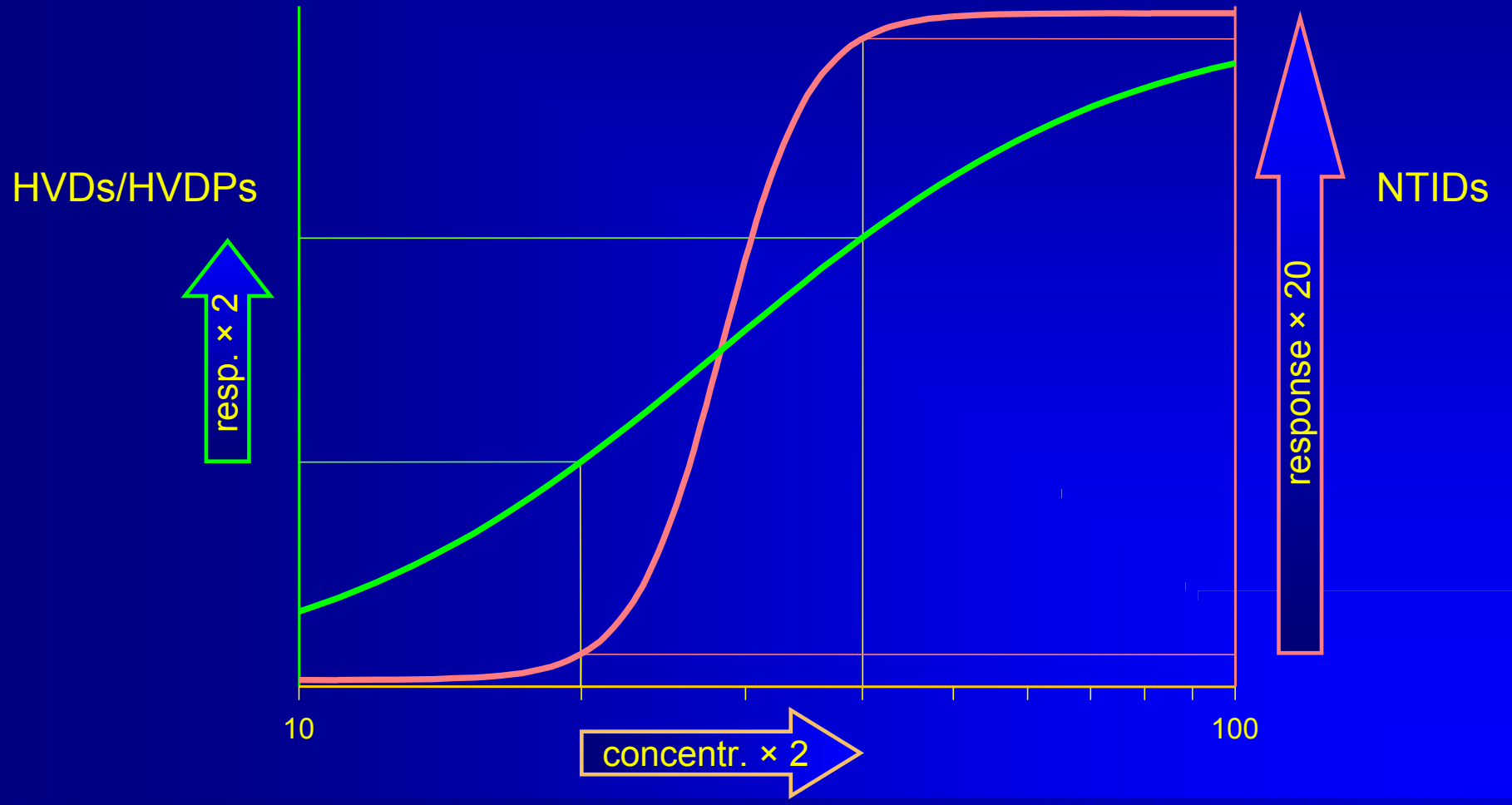
Modified from Fig. 1
Tothfálusi *et al.* (2009)

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 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. Note: Same overall number of individual treatments!

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 depend on design (*e.g.*, SAS **`Proc MIXED`** for full replicate and SAS **`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, replicate design give more informations about formulation(s)

Application: HVDs/HVDPs

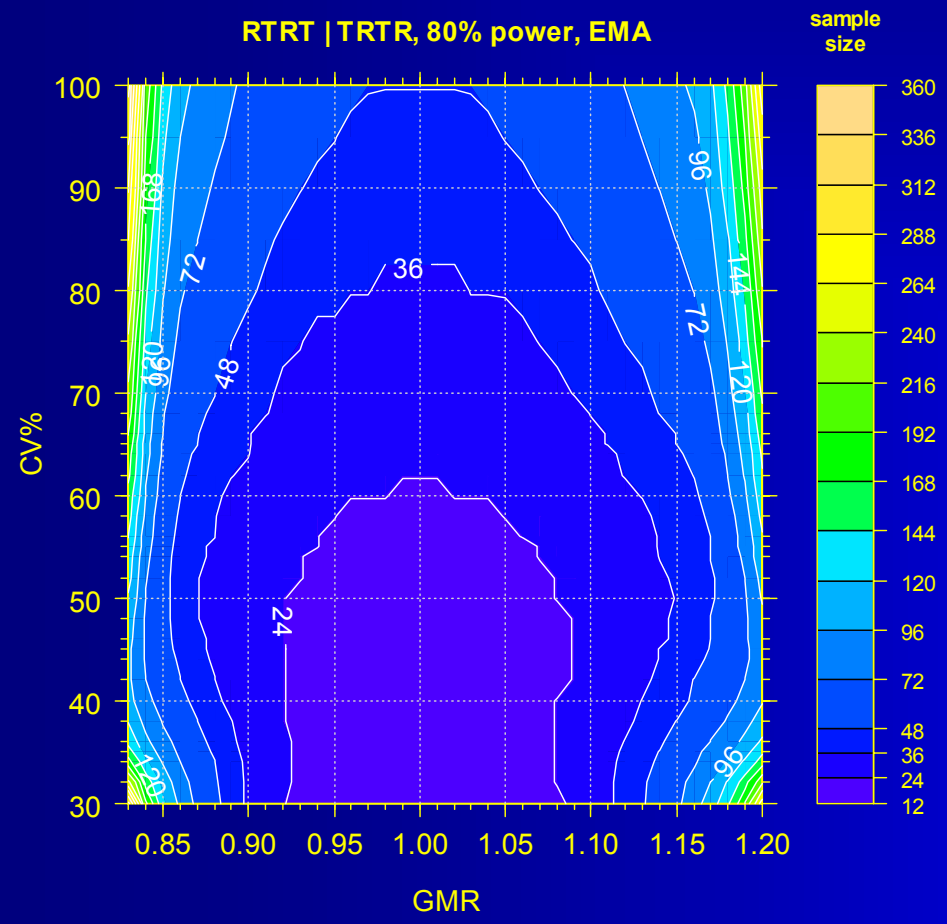
- CV_{WR} >30 %
  - ✓USA Recommended in API specific guidances. Scaling for AUC and/or C_{max} acceptable, GMR 0.80 – 1.25; ≥24 subjects enrolled.
  - ± EU Widening of acceptance range (only C_{max}) to maximum of 69.84% – 143.19%), GMR 0.80 – 1.25.
    Demonstration that CV_{WR} >30% is not caused by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

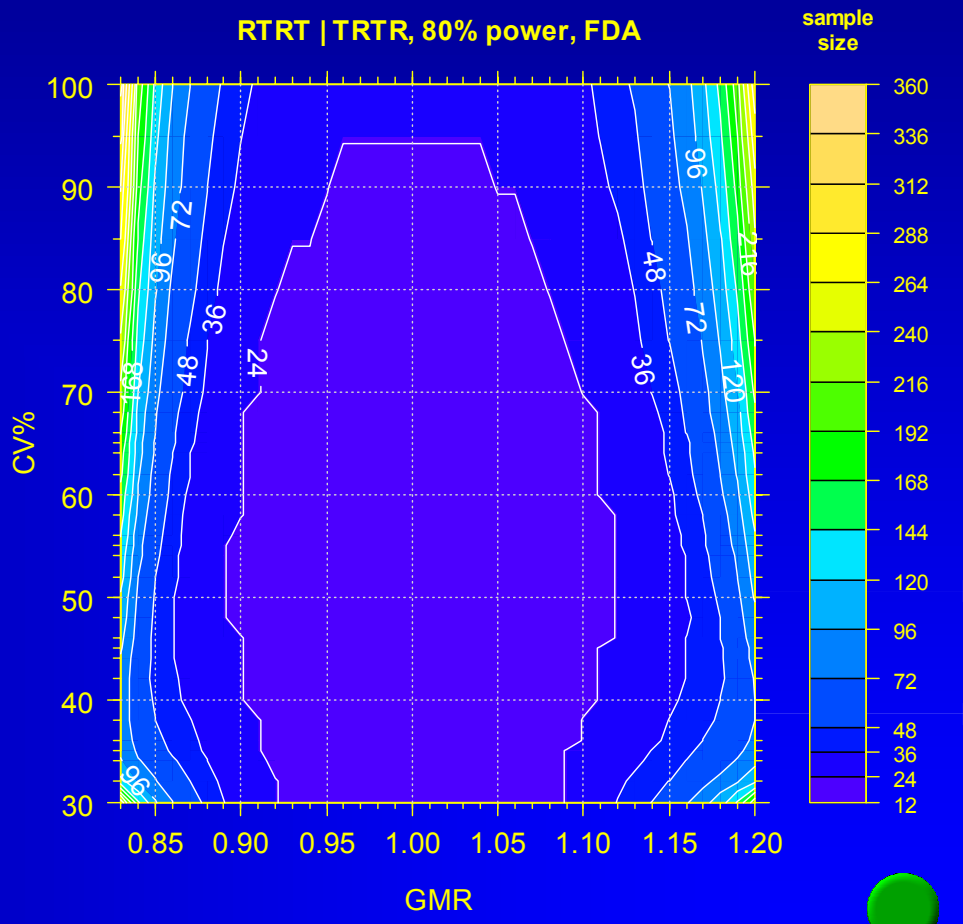
Replicate designs

- 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
    RTRT | TRTR full replicate, 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
    TRR | RTR | RRT partial replicate, 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%

Outliers (EMA)

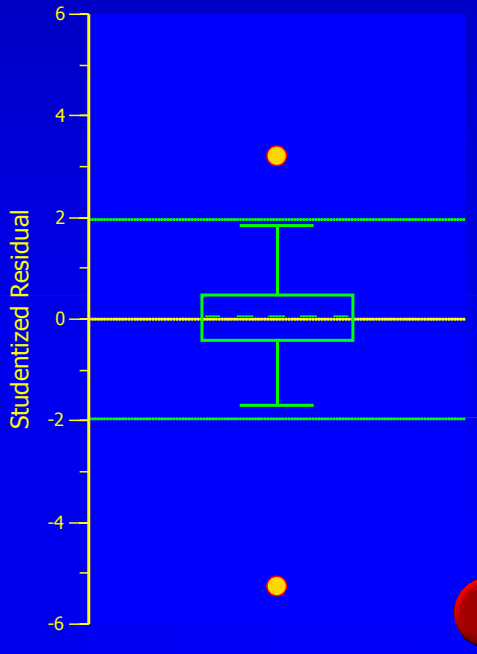
- EMA GL on BE (2010), Section 4.1.10
  - The applicant should justify that the calculated intra-subject variability is a reliable estimate and that it is not the result of outliers.
- EGA/EMA Q&A (2010)
  - Question:
    How should a company proceed if outlier values are observed for the reference product in a replicate design study for a Highly Variable Drug Product (HVDP)?

Outliers (EMA)

- EGA/EMA Q&A (2010)
  - Answer:
    The outlier cannot be removed from evaluation […] but should not be taken into account for calculation of within-subject variability and extension of the acceptance range.
    An outlier test is not an expectation of the medicines agencies but outliers could be shown by a box plot. This would allow the medicines agencies to compare the data between them.

Outliers (EMA)

- Data set I (full replicate)
  - CV_{WR} 46.96%
    EL 71.23–140.40%
    Method A: 107.11–124.89%
    Method B: 107.17–124.97%
  - But there *are* two outliers!
    By excluding subjects 45 and 52
    CV_{WR} drops to 32.16%.
    EL 78.79–126.93%
    Almost no more gain compared to conventional limits…



شكرا لك!

Pharmacokinetic and Statistical Analysis of BE Data

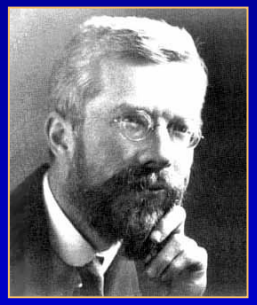
Open Questions?

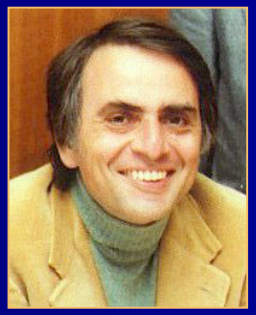
Helmut Schütz
BEBAC
Consultancy Services for
Bioequivalence and Bioavailability Studies
1070 Vienna, Austria
helmut.schuetz@bebac.at

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