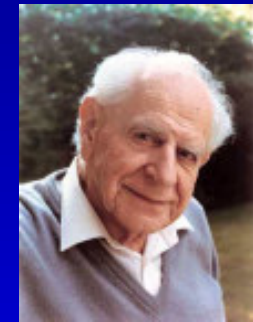BE
·BAC

Basic Designs for Bioequivalence 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.
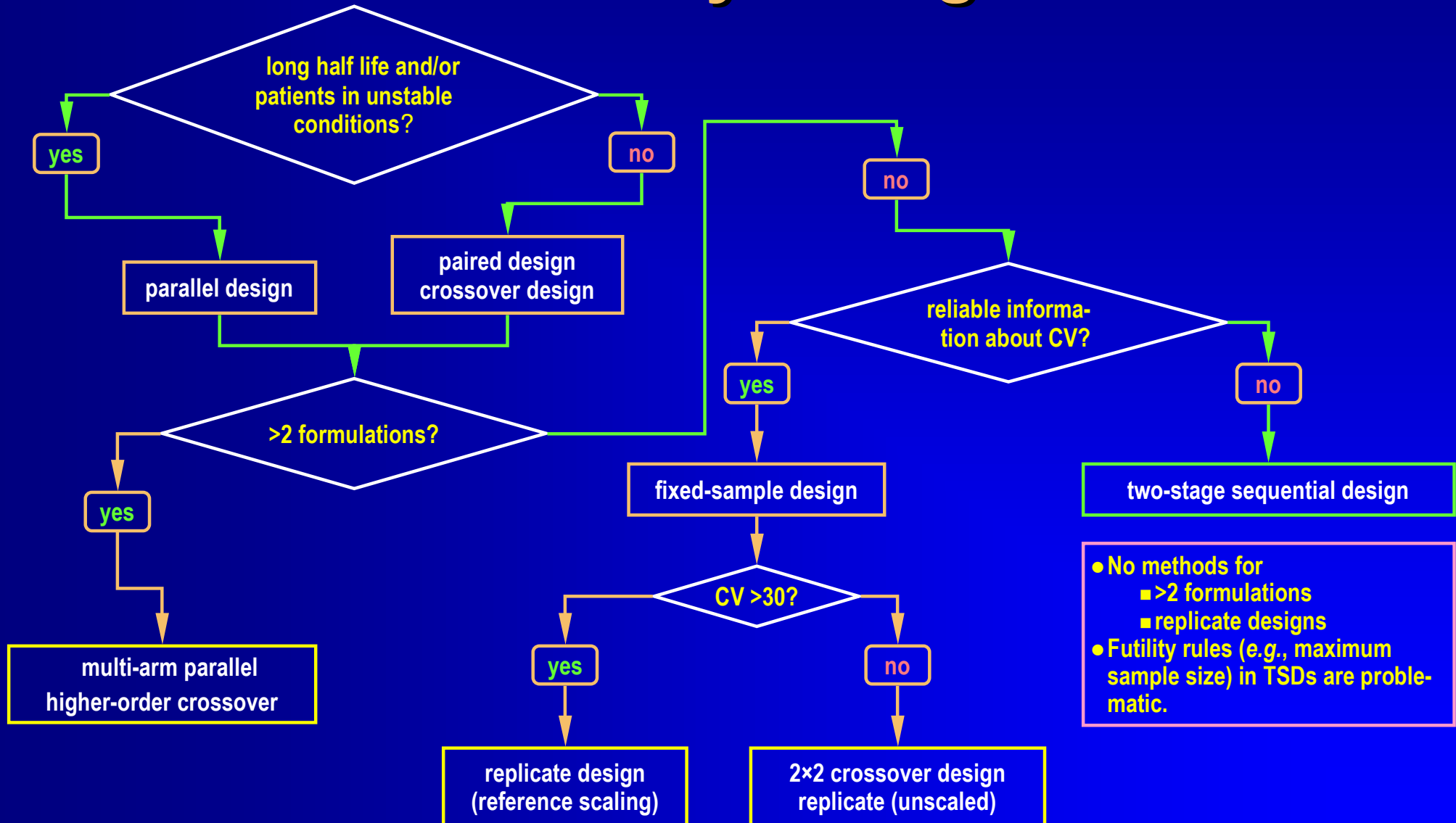

Karl R. Popper

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

Leslie Z. Benet

BE Study Designs

Flowchart:

- **long half life and/or patients in unstable conditions?**
  - **yes** → **parallel design**
  - **no** → **paired design / crossover design**

- **parallel design** / **paired design crossover design** → **>2 formulations?**
  - **yes** → **multi-arm parallel / higher-order crossover**
  - **no** → **reliable information about CV?**
    - **yes** → **fixed-sample design**
      - **CV >30?**
        - **yes** → **replicate design (reference scaling)**
        - **no** → **2×2 crossover design / replicate (unscaled)**
    - **no** → **two-stage sequential design**

- **No methods for**
  - **>2 formulations**
  - **replicate designs**
- **Futility rules (*e.g.*, maximum sample size) in TSDs are 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)
          2×2×2 crossover (TR | RT)
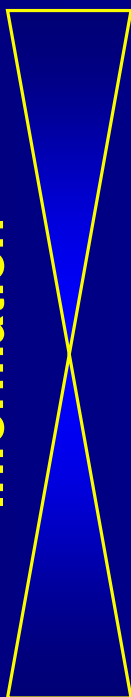   Parallel (R | T)

■ **Variances which can be estimated:**
          Parallel: total variance (between + within subjects)
      2×2×2 Xover: + between, within subjects
  Partial replicate: + within subjects (reference)
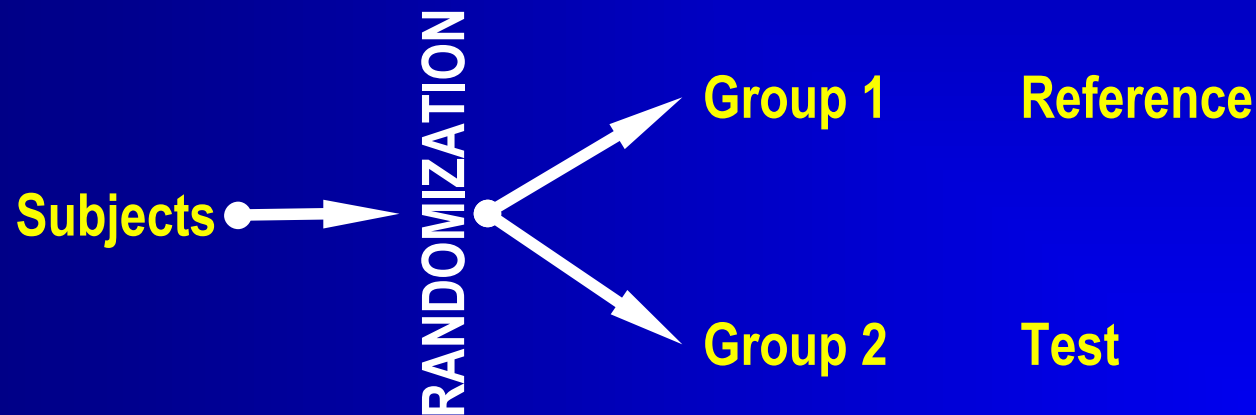Full replicate: + within subjects (reference, test)

Information

Parallel Designs

- **Two-Group Parallel Design**

Parallel Designs (cont'd)

- **Two-Group Parallel Design**
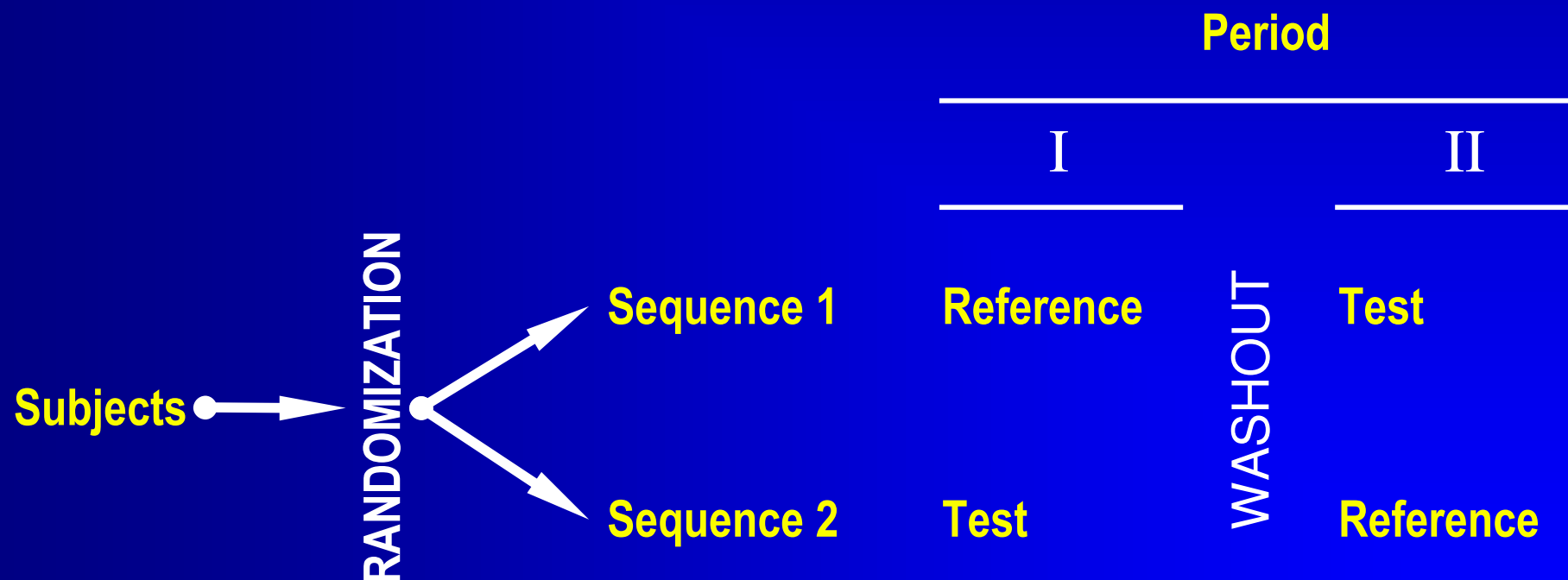 - **Advantages**
    - Clinical part – *sometimes* – faster than crossover.
    - Straigthforward statistical analysis.
    - Drugs with long half life.
    - Potentially toxic drugs or effect and/or AEs unacceptable in healthy subjects.
    - Studies in patients, where the condition of the disease irreversibly changes.
  - **Disadvantages**
    - Lower statistical power than crossover – high sample sizes.
    - Tight inclusion-/exclusion criteria to reduce between-subject variability.
    - Phenotyping mandatory for drugs showing polymorphism.

Crossover Designs

● **Standard 2×2×2 Design
(Two Treatments, Two Periods, Two Sequences)**

Period

	I		II

Subjects → **RANDOMIZATION**

Sequence 1 → **Reference** — WASHOUT → **Test**

Sequence 2 → **Test** — WASHOUT → **Reference**

Crossover Designs: Model

Multiplicative Model (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 (**$l=1,2$**:** $\Phi_1=\Phi_{test}$**,** $\Phi_2=\Phi_{ref.}$**)**

Crossover Designs: Assumptions

Multiplicative Model (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…

Crossover Designs (cont'd)

- **Standard 2×2×2 design**
  - **Advantages**
    - Globally applied standard protocol for bioequivalence, drug-drug interaction, food effect studies.
    - Straigthforward statistical analysis.
  - **Disadvantages**
    - Not suitable for studies in patients with instable diseases
      → parallel design
    - Not optimal for drugs with long half life
      → parallel design
    - Not optimal for highly variable drugs / drug products
      → replicate designs with reference-scaling

Crossover 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*.**

Crossover 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

Crossover Designs (cont'd)

- **Williams' Designs**
  - **Advantages**
    - Allows to choose between two candidate test formulations or comparison of a test formulation with two references.
    - Standard design for establishment of dose proportionality.
    - Paired comparisons are balanced.
    - Mentioned in EMA's and ANVISA's guidelines.
  - **Disadvantages**
    - More 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).

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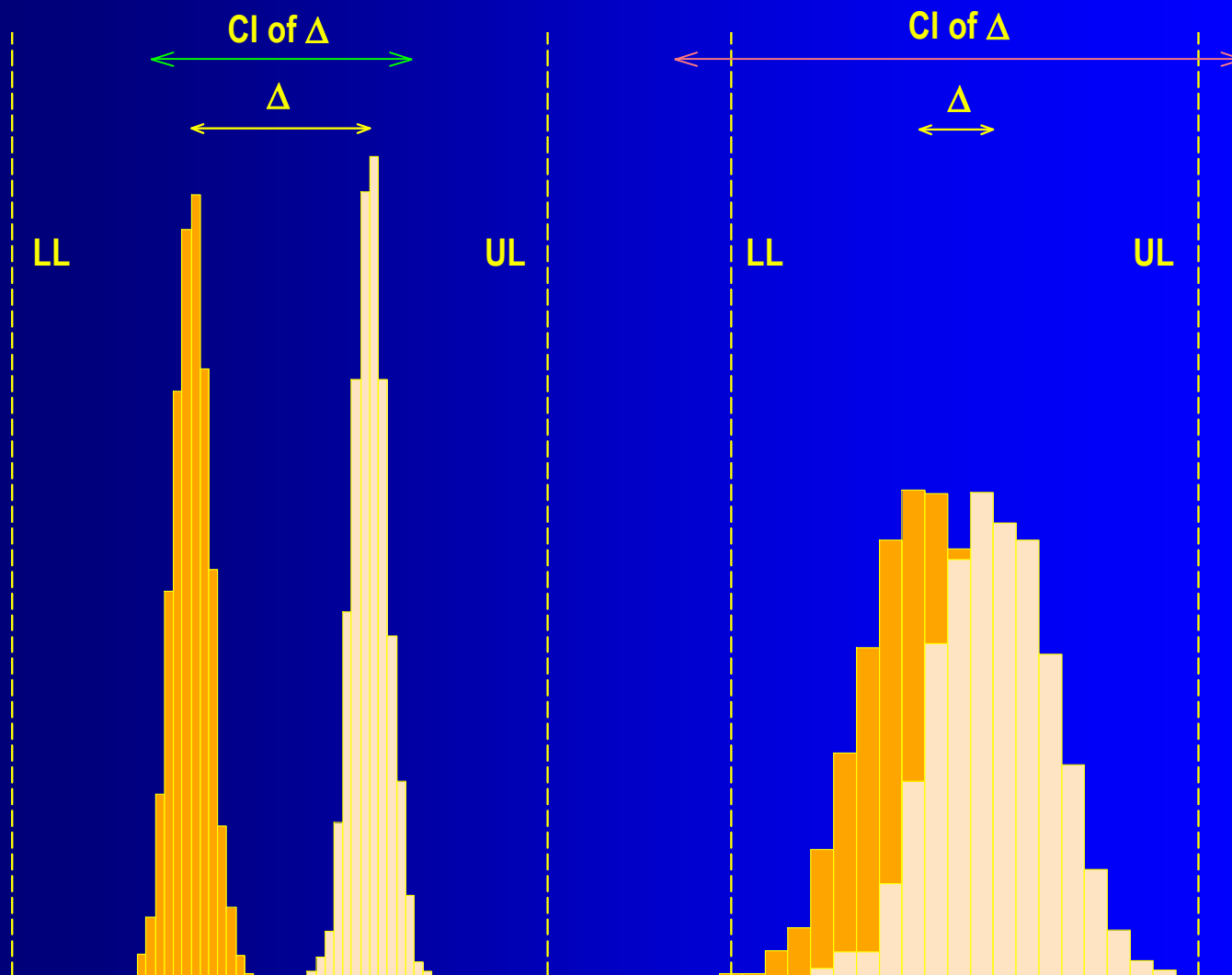
Modified from Fig. 1 Tothfálusi *et al*. (2009)

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.

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 2×2×2 crossover design assumes Independent Identically Distributions (IID) – which may not be correct.**

  - **If 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.**
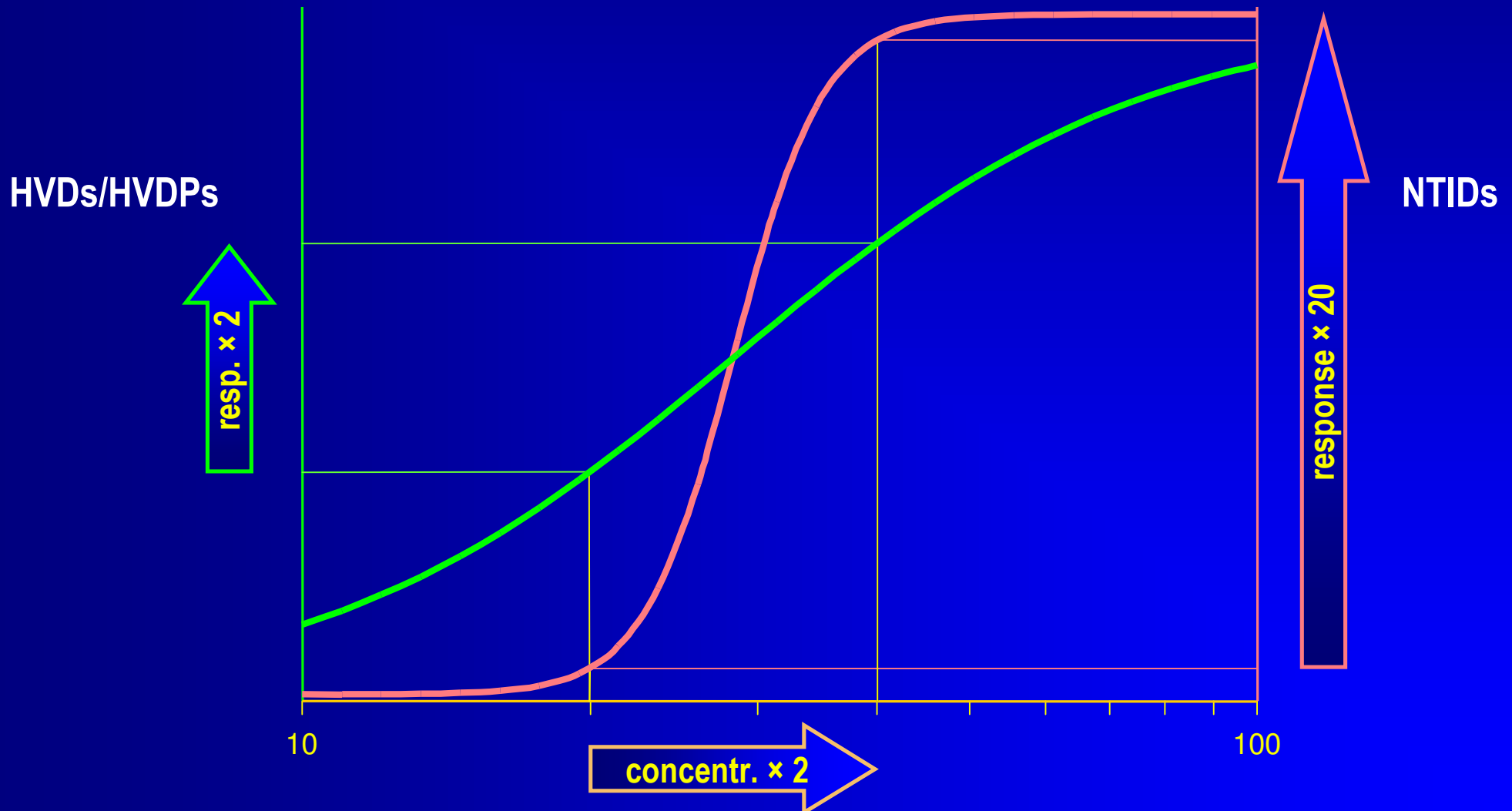
Replicate Designs

- **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 (study costs directly related to number of biosamples)!**

Replicate Designs

■ **Any replicate design can be evaluated for 'classical' (unscaled) Average Bioequivalence (ABE) as well.**

■ **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

● **Within-subject CV of the reference (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 a**
              maximum of 69.84 – 143.19%), GMR 0.80 – 1.25.
              Demonstration that CV_{WR} >30% is not caused
              by outliers (box plots).
              Justification that the widened acceptance range
              is clinically not relevant (safety, efficacy).
              Not less than 12 subjects in sequence RTR.

Replicate Designs

- **Two-sequence four-period**
  TRTR | RTRT

- **Two-sequence three-period**
  TRT | RTR

- **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)

- **EU GL on BE (2010)**

  - **Average Bioequivalence (AB) with Expanding Limits (EL) $\rightarrow$ "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%.**

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

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

Patients' Risk?

● **The Null-Hypothesis is modified 'in face of the data'**

  ■ **The acceptance range is not pre-specified (like in conventional ABE), but depends on the variability observed in the study.**

   ■ **Modifying H_0 generally requires adjustment of α in order to maintain the Type I Error $\leq$5%.**

   ■ **Inflation of the Type I Error known.[1,2]**

   ■ **Recommendation: Use an adjusted α of 0.025 (95% CI) for full replicate designs and α of 0.030 (94% CI) for the partial replicate design.**

1. **Endrényi L, Tóthfalusi (2009)**
   Regulatory Conditions for the Determination of Bioequivalence of Highly Variable Drugs
   J Pharm Pharmaceut Sci 12(1):138–49
2. **Wonnemann M, Frömke C, Koch A (2015)**
   Inflation of the Type I Error: Investigations on Regulatory Recommendations for Bioequivalence of Highly Variable Drugs
   Pharm Res 32(1):135–43 DOI: 10.1007/s11095-014-1450-z

Add-On / Two-Stage Designs

- Sometimes properly designed studies fail due to
  - 'true' bioinequivalence,
  - pure chance (producer's risk),
  - poor study conduct (increasing variability),
  - **false** (mainly over-optimistic) **assumptions about the** CV and/or T/R-ratio – leading to a too small sample size (insufficient power).
- Reminder:
  The chosen sample size is based on *assumptions*…

Add-On / Two-Stage Designs

- **Dealing with *inconclusive* BE studies (confidence interval not entirely with the acceptance range)**
  - **Repeat the study in a larger sample size.**
  - **Perform a meta-analysis of more than one study. Only acceptable if at least one study demonstrates BE.**
  - **Recruit additional subjects and pool data. Problematic!**
- **Discussed at Bio-International Conferences (1989, 1992) and guidelines of the 1990s.**
  - **The patient's risk must be preserved!**
  - **Among rivaling methods the one with with the highest power should be selected.**

Adaptive TS Sequential Designs

● **Two 'Types' of Two-Stage Sequential Designs**[1]

1. **The *same* adjusted α is applied in both stages (regardless whether a study stops already in the first stage or proceeds to the second stage).**
   - ■ **Based on Group Sequential Designs.**
   - ■ **In publications called 'Method B'.**

2. **An unadjusted α *may* be used in the first stage (dependent on interim power).**
   - ■ **Based on conventional BE testing + GSD.**
   - ■ **In publications called 'Method C, D, or C/D'.**

1. Schütz H (2015)
   Two-stage designs in bioequivalence trials
   Eur J Clin Pharmacol 71(3):271–81 DOI: 10.1007/s00228-015-1806-2

Adaptive TS Sequential Designs

- **The 94.12% CI (*i.e.*, an adjusted α of 0.0294) stated in the EMA's GL is *not* suitable to *all* designs.**

reference	Type	Method	T/R	target power	CV	α_{adj}	TIE_{max}	α_{adj}[1]	TIE_{max}[1]
Potvin *et al.*	1	B	0.95			0.0294	0.0485	0.0302	0.0501
	2	C					0.0510	0.0282	0.0501
Montague *et al.*	1	B	0.90	80%	10–100%	0.0280	0.0518	0.0270	0.0500
	2	D				–	–	0.0269	0.0502
Fuglsang	1	B	0.95			0.0284	0.0501	0.0286	0.0501
	2	C/D		90%	10–80%	0.0274	0.0503	0.0278	0.0503
	1	B	0.90			–	–	0.0286	0.0501
	2	C/D	0.90			0.0269	0.0501	0.0267	0.0500

1. **Schütz H, Labes D, Fuglsang A (in preparation 2015)**
   Modifications of 'Sequential design approaches for bioequivalence studies with crossover designs'

Evaluation

- **Design**

  - **The statistical model is defined.**
  - **The α which preserves the patient's risk ≤ 0.05 and the Acceptance Range (AR) for BE are specified.**

- **Evaluation**

  - **The test/reference ratio is calculated.**
  - **The $100(1 - 2\alpha)\%$ confidence interval (CI) around the ratio is calculated.**
    - **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.**

Assessment

● **Decision based on the CI and the pre-specified AR**

  ■ **Generally a 20% difference between formulations is considered *clinically not relevant*. This leads to**

  $$L = 100(1-\Delta), \; U = 100(1/L), \; \left[80-125\%\right]$$

  ■ **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

Assessment
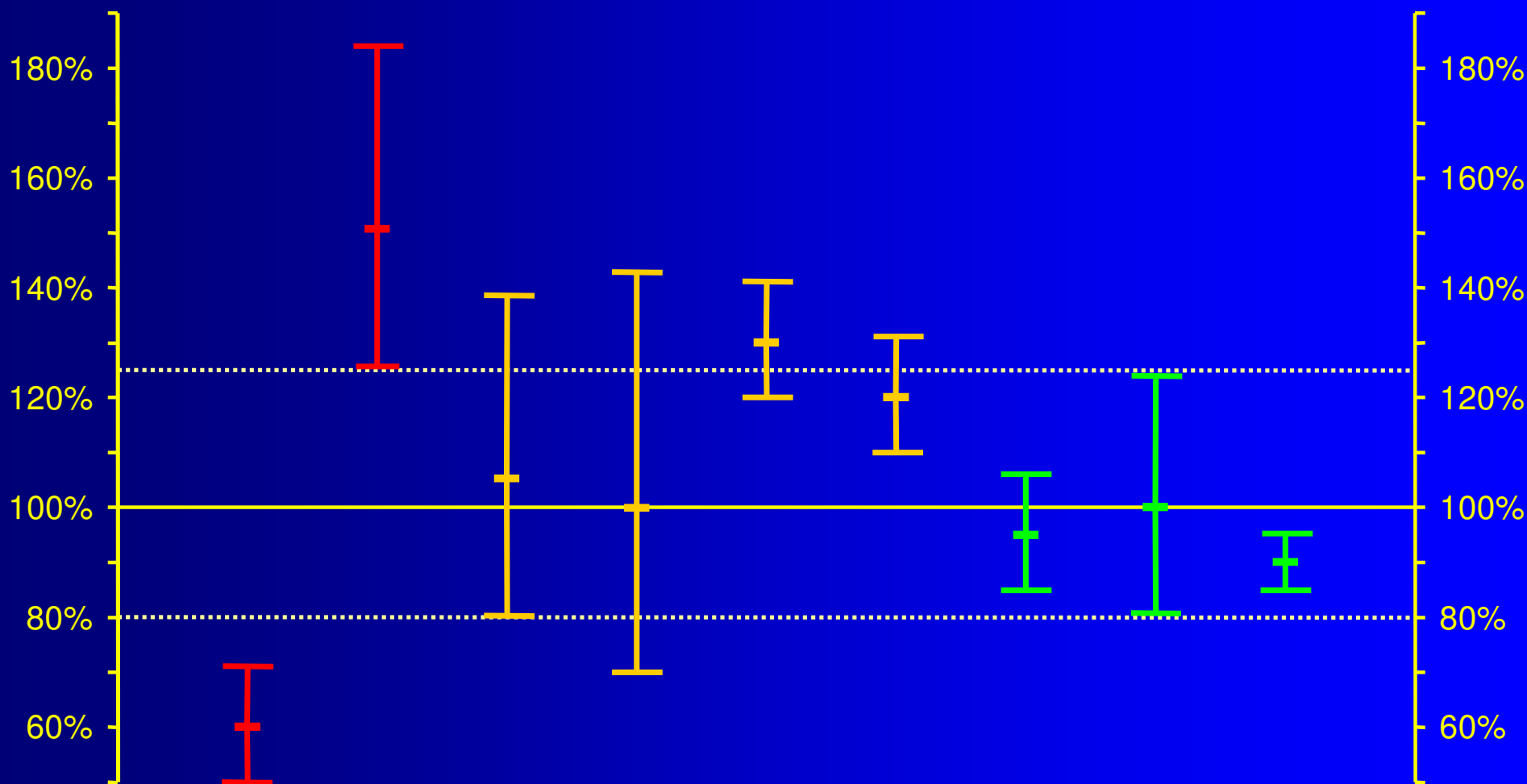
Thank You!

Basic Designs for BE Studies

Questions after the 2nd presentation, please.

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

In bioequivalence we must not forget the only important – *the patient*! He/she is living person, not just α 0.05.

Dirk Marteen Barends

It is a good morning exercise for a research scientist to discard a pet hypothesis every day before breakfast. It keeps him young.

Konrad Lorenz