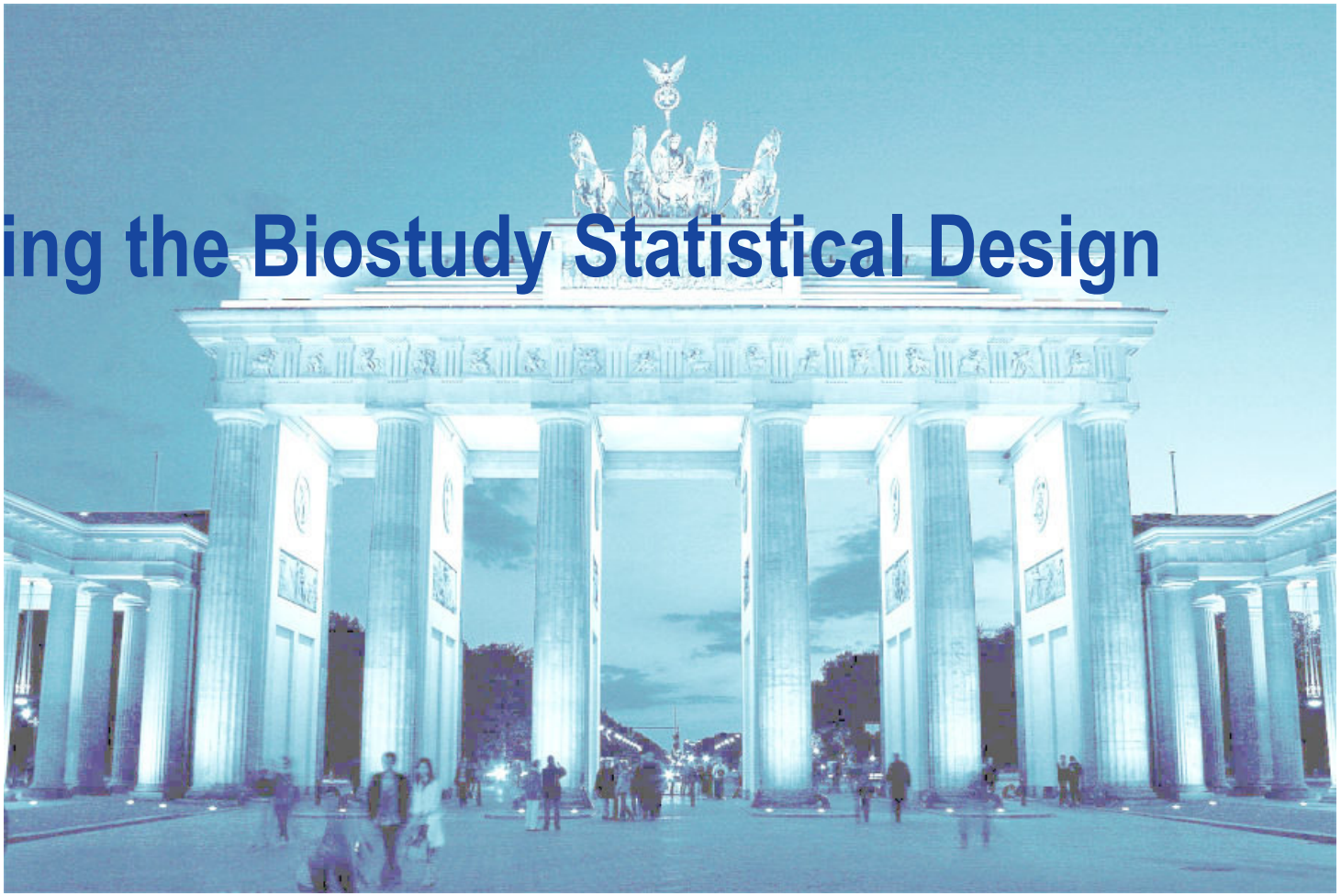
Establishing the Biostudy Statistical Design

Helmut Schütz



Study Designs

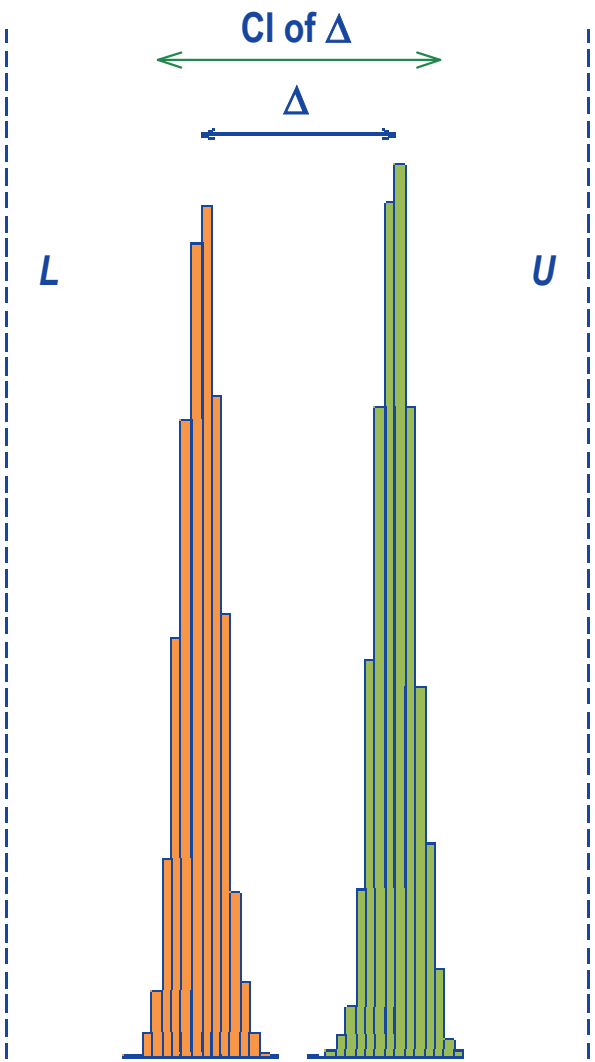
The more ‘sophisticated’ a design is, the more information can be extracted.

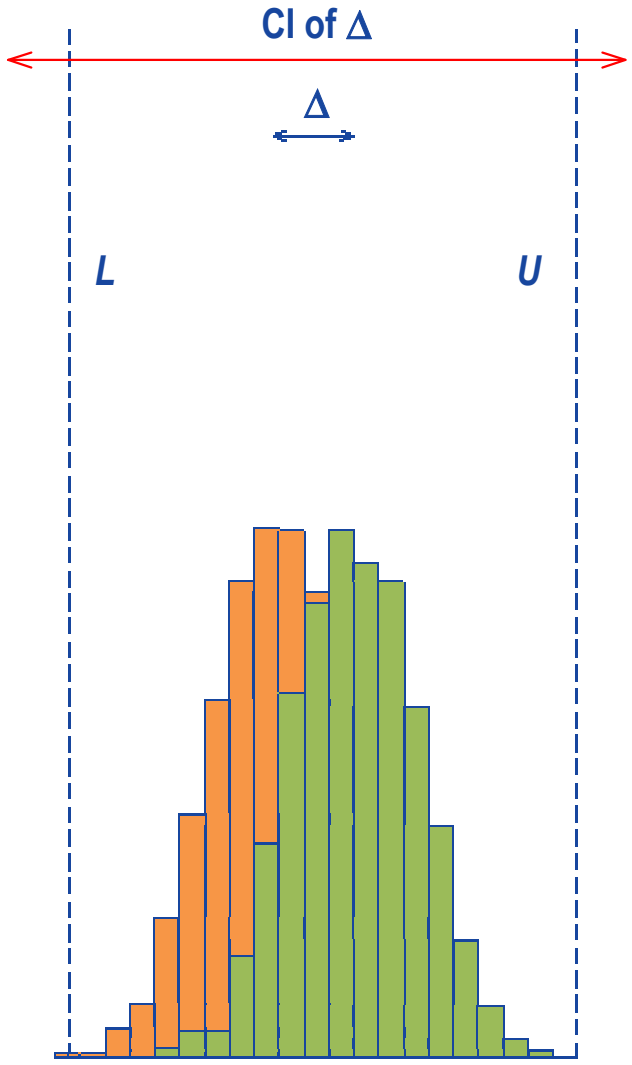
- **Hierarchy of designs:**
  - Full replicate (RTRT | TRTR or RTR | TRT)
    - Partial replicate (RRT | RTR | TRR)
      - 2×2×2 cross-over (RT | TR)
        - Parallel (R | T)

- **Variances which can be estimated:**

Parallel:	*total* variance (*pooled* of between + within subjects)
2×2×2 cross-over:	+ *between*, *within* subjects
Partial replicate:	+ within subjects (*of* R)
Full replicate:	+ within subjects (of R *and* T)

Highly Variable Drugs / Drug Products





Counterintuitive concept of BE:

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

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

HVD(P)s – Reference-scaling

It may be almost impossible to demonstrate ABE with a reasonable sample size.

- Reference-scaling (*i.e.*, widening the acceptance range based of the variability of the reference) in 2010 introduced by the FDA and EMA and in 2016 by Health Canada.
  - Requires a replicate design, where at least the reference product is administered twice.
  - Smaller sample sizes compared to a standard 2×2×2 design but outweighed by increased number of periods.
  - Similar total number of individual treatments.
  - Any replicate design can be evaluated for 'classical' (unscaled) Average Bioequivalence (ABE) as well. Switching CV_{wR} 30%:
    - FDA: *AUC* and C_{max}
    - EMA: C_{max}; MR products additionally: $C_{ss,τ}$, $C_{ss,τ}$, partial *AUC*s
    - Health Canada: *AUC*

HVD(P)s – Reference-scaling

Models (in log-scale).

- **ABE Model:**
  - A difference Δ of ≤20% is considered to be clinically not relevant.
  - The limits [*L*, *U*] of the acceptance range are fixed to $\log(1-\Delta) = \log((1-\Delta)^{-1})$ or $L \sim -0.2231$ and $U \sim +0.2231$.
  - The consumer risk (α) is fixed with 0.05. BE is concluded if the 100(1 – 2α) confidence interval lies entirely within the acceptance range.

$$-\theta_A \leq \mu_T - \mu_R \leq +\theta_A$$

- **SABEL Model:**
  - Switching condition θ_S is derived from the regulatory standardized variation σ_0 (proportionality between acceptance limits in log-scale and σ_{wR} in the highly variable region).

$$-\theta_S \leq \frac{\mu_T - \mu_R}{\sigma_{wR}} \leq +\theta_S$$

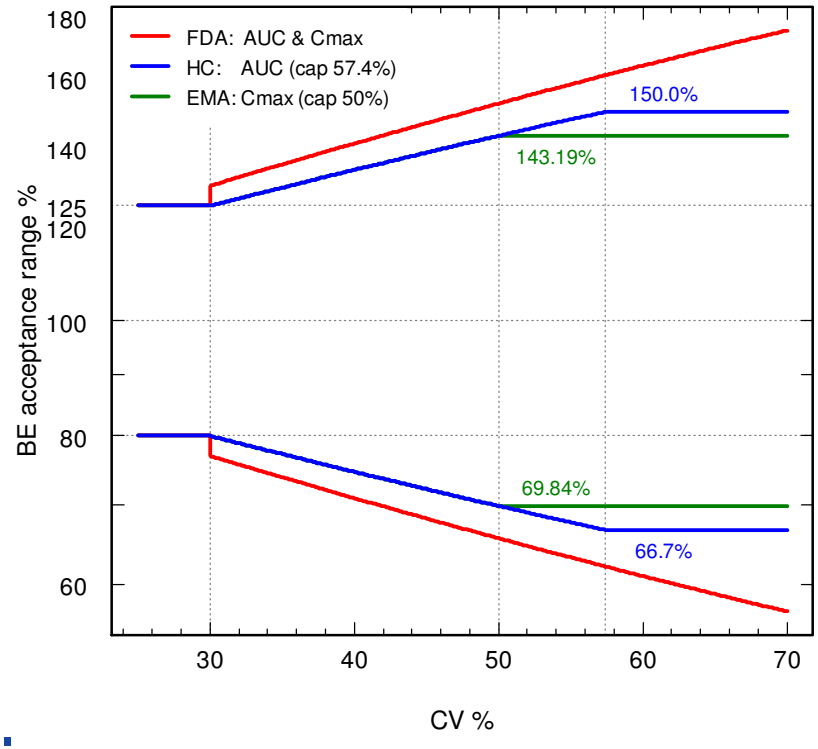
HVD(P)s – Reference-scaling

Regulatory Approaches.

- Bioequivalence limits derived from σ_0 and σ_{wR}

$$\theta_s = \frac{\log(1.25)}{\sigma_0}, \quad [L,U] = e^{\pm\theta_S \cdot \sigma_{wR}}$$

- FDA
  - Scaling σ_{wR} 0.25 (θ_S 0.893) but applicable at CV_{wR} ≥30%.
  - Discontinuity at CV_{wR} 30%.
- EMA
  - Scaling σ_0 0.2936 (θ_S 0.760).
  - Upper cap at CV_{wR} 50%.
- Health Canada
  - Like EMA but upper cap at CV_{wR} 57.4%.



HVD(P)s – Reference-scaling

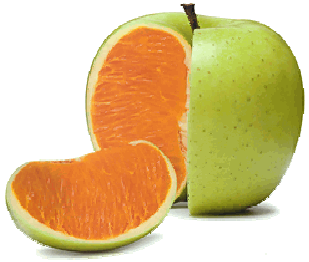
The EMA's Approach.

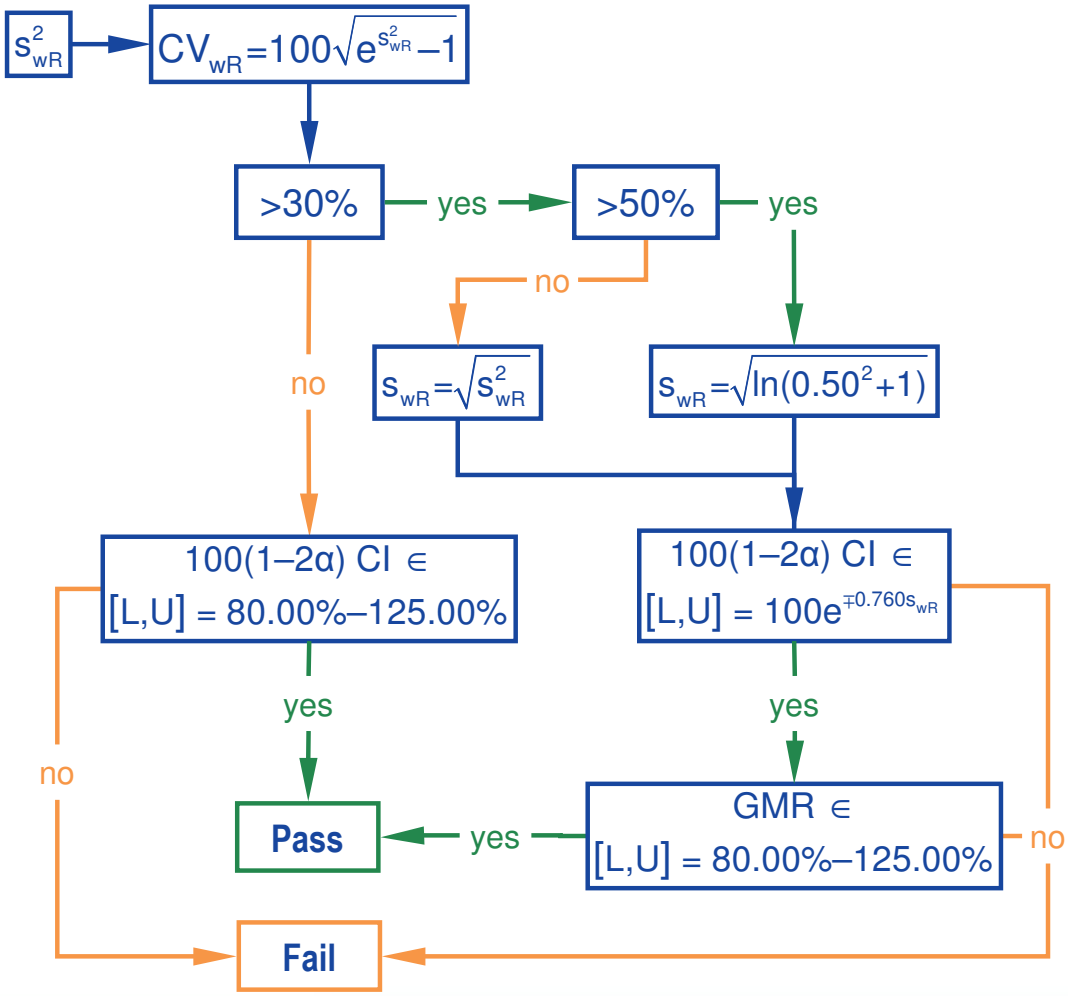
- Average Bioequivalence with Expanding Limits – ABEL (crippled from Endrényi and Tóthfalusi 2009).
  - Justification that the widened acceptance range is clinically not relevant (important – different to the FDA).
  - Assumes identical variances of *T* and *R* [*sic*] like in a 2×2×2.
  - All fixed effects model according to the Q&A-document preferred.
  - Mixed-effects model (allowing for unequival variances) is 'not compatible with CHMP guideline'…
  - Scaling limited at a maximum of CV_{wR} 50% (*i.e.*, to 69.84 – 143.19%).
  - *GMR* within 0.8000 – 1.2500.
  - Demonstration that CV_{wR} >30% is not caused by outliers (box plots of studentized intra-subject residuals?)…
  - ≥12 subjects in sequence RTR of the 3-period full replicate design.

HVD(P)s – Reference-scaling

The EMA's Approach.

- **Decision Scheme.**
  - The Null Hypothesis is *specified* in the face of the data.
  - Acceptance limits themselves become random variables.
  - Type I Error (consumer risk) might be inflated.





HVD(P)s – Reference-scaling

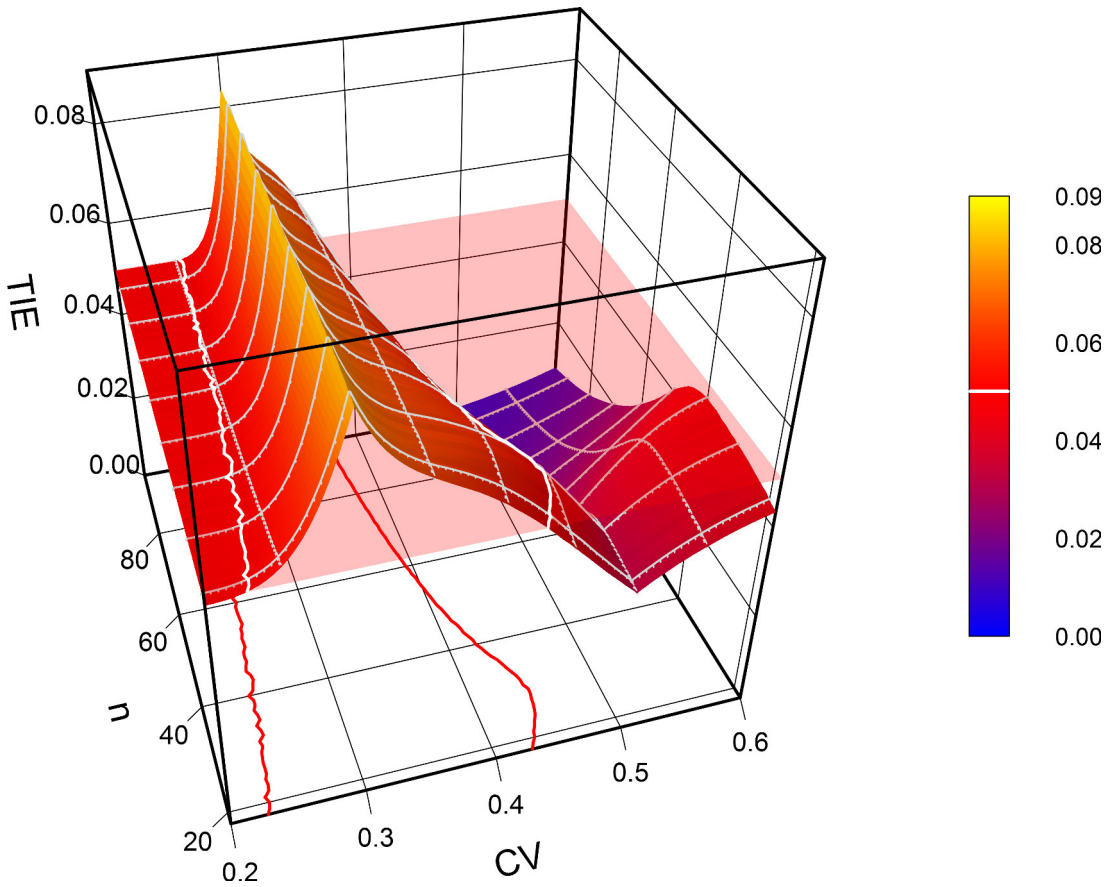
Assessing the Type I Error (TIE).

- TIE = falsely concluding BE at the limits of the acceptance range.
- Due to the decision scheme direct calculation of the TIE at the scaled limits is not possible;
  → extensive simulations required (10^6 BE studies mandatory).
- Inflation of the TIE suspected.
  (Chow *et al*. 2002, Willavazie and Morgenthien 2006, Chow and Liu 2009, Patterson and Jones 2012).
- Confirmed.
  - EMA's ABEL
    (Tóthfalusi and Endrényi 2009, BEBA-Forum 2013, Wonnemann *et al*. 2015, Muñoz *et al*. 2016, Labes and Schütz 2016).
  - FDA's RSABE
    (Tóthfalusi and Endrényi 2009, BEBA-Forum 2013, Muñoz *et al*. 2016).

HVD(P)s – Reference-scaling

Example for ABEL

- RTRT | TRTR
  sample size 18 – 96
  CV_{wR} 20% – 60%
  - TIE_{max} 0.0837.
  - Relative increase of the consumer risk 67%!



HVD(P)s – Reference-scaling

What is going on here?

- SABE is stated in model *parameters* …

$$-\theta_S \leq \frac{\mu_T - \mu_R}{\sigma_{wR}} \leq +\theta_S$$

… which are *unknown*.

  - Only their *estimates* (*GMR*, s_{wR}) are accessible in the actual study.
  - At CV_{wR} 30% the decision to scale will be wrong in ~50% of cases.
  - If moving away from 30% the chances of a wrong decision decrease and hence, the TIE.
  - At high *CV*s (>43%) both the scaling cap and the *GMR*-restriction help to maintain the TIE <0.05).

HVD(P)s – Reference-scaling

Outlook.

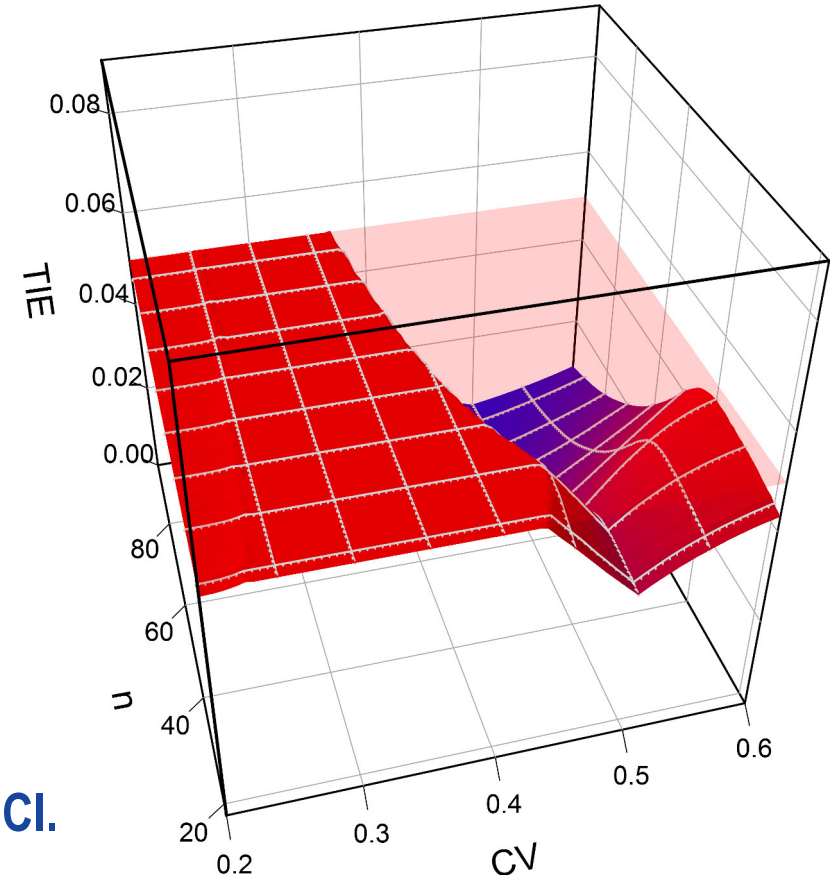
- Utopia
  - Agencies collect CV_{wR} from submitted studies. Pool them, adjust for designs / degrees of freedom. The EMA publishes a fixed acceptance range in the product-specific guidance. No need for replicate studies any more. 2×2×2 cross.overs evaluated by ABE would be sufficient.
- Halfbaked
  - Hope [*sic*] that *e.g.*, Bonferroni preserves the consumer risk. Still apply ABEL, but with a 95% CI (α 0.025).
  - Drawback: Loss of power, substantial increase in sample sizes.
- Proposal
  - Iteratively adjust α based on the study's CV_{wR} and sample size – in such a way that the consumer risk is preserved (Labes and Schütz 2016).

Labes D, Schütz H. *Inflation of Type I Error in the Evaluation of Scaled Average Bioequivalence, and a Method for its Control.* Pharm Res. 2016; 33(11): 2805–14. DOI 10.1007/s11095-016-2006-1

ABEL (iteratively adjusted α)

Previous example

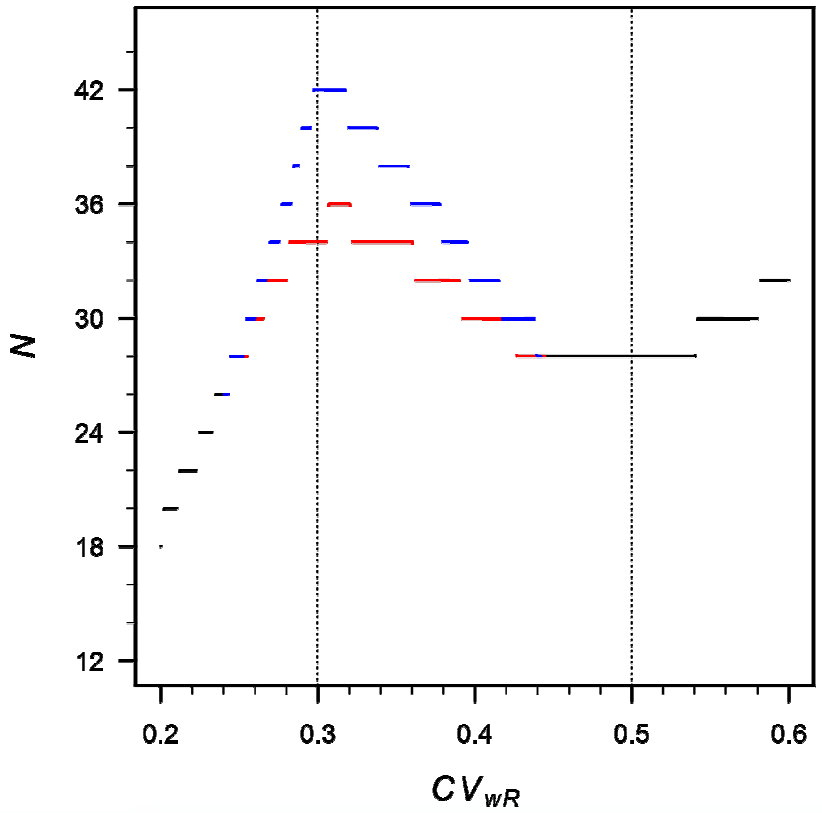
- Algorithm
  - Assess the TIE for the nominal α 0.05.
  - If the TIE ≤ 0.05, stop.
  - Otherwise adjust α (downwards) until the TIE ≅ 0.05.
  - At CV_{wR} 30% (dependent on the sample size) α_{adj} is 0.0273 – 0.0300;
    → use a 94.00 – 94.54% CI.



ABEL (iteratively adjusted α)

Potential impact on the sample size.

- Example: RTRT | TRTR, θ_0 0.90, target power 0.80.
  - Moderate in the critical region (— —).
    - CV_{wR} 30%: 36 → 42 (+17%);
    - CV_{wR} 35%: 34 → 38 (+12%);
    - CV_{wR} 40%: 30 → 32 (+7%).
  - None outside (—).



ABEL (iteratively adjusted α)

Example (RTRT | TRTR, expected CV_{wR} 35%, θ_0 0.90, target power 0.80); R package PowerTOST (≥1.3-3).

- **Estimate the sample size.**

```
sampleN.scABEL(CV=0.35, theta0=0.90, targetpower=0.80, design="2x2x4",
               details=FALSE, print=FALSE)[["Sample size"]]
[1] 34
```

- **Estimate the empiric TIE for this study.**

```
UL <- scABEL(CV=0.35)[["upper"]] # scaled limit (1.2948 for CVwR 0.35)
power.scABEL(CV=0.35, theta0=UL, n=34, design="2x2x4", nsims=1e6)
[1] 0.065566
```

- **Iteratively adjust α.**

```
scABEL.ad(CV=0.35, n=34, design="2x2x4")
++++++++++++ scaled (widened) ABEL ++++++++++++
         iteratively adjusted alpha
-----------------------------------------------
CVwR 0.35, n(i) 17|17 (N 34)
Nominal alpha                  : 0.05
Null (true) ratio              : 0.9000
Regulatory settings            : EMA (ABEL)
Empiric TIE for alpha 0.0500   : 0.06557
Power for theta0 0.900         : 0.812
Iteratively adjusted alpha     : 0.03630
Empiric TIE for adjusted alpha: 0.05000
Power for theta0 0.900         : 0.773
```

ABEL (iteratively adjusted α)

- **Optionally compensate for the loss in power (0.812 → 0.773) by increasing the sample size:**

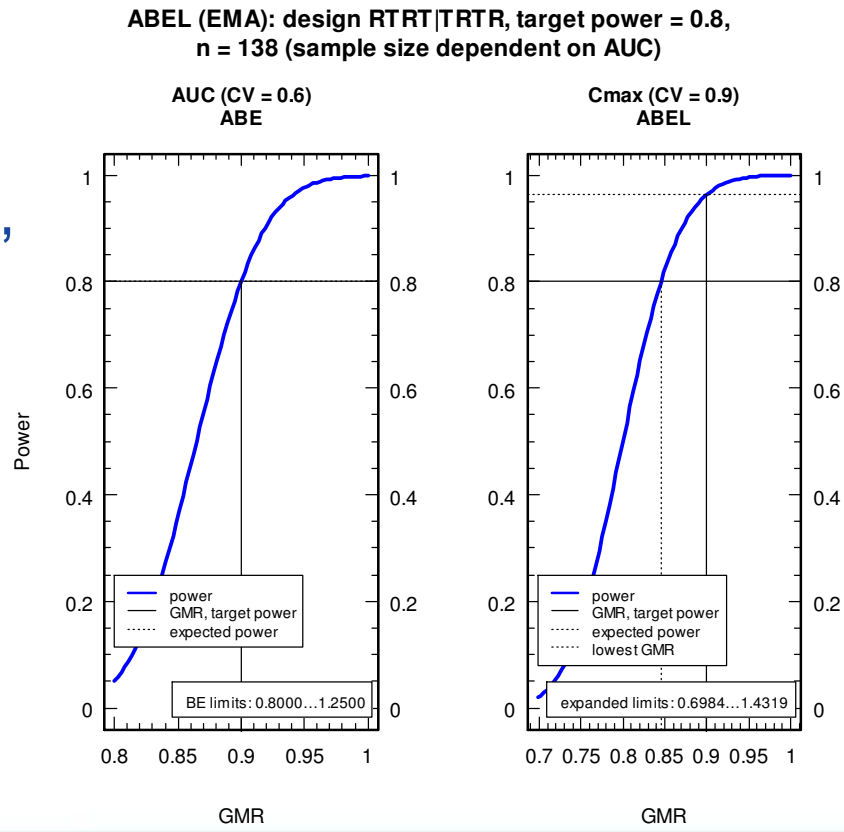
```
sampleN.scABEL.ad(CV=0.35, theta0=0.90, targetpower=0.80, design="2x2x4")
++++++++++++ scaled (widened) ABEL ++++++++++++
             Sample size estimation
         for iteratively adjusted alpha
--------------------------------------------
Study design: 2x2x4 (RTRT|TRTR)
Expected CVwR 0.35
Nominal alpha      : 0.05
Null (true) ratio  : 0.9000
Target power       : 0.8
Regulatory settings: EMA (ABEL)
Switching CVwR     : 30%
Regulatory constant: 0.760
Expanded limits    : 0.7723...1.2948
Upper scaling cap  : CVwR 0.5
PE constraints     : 0.8000...1.2500
n  38,   adj. alpha: 0.03610 (power 0.8100), TIE: 0.05000
```

– ***n*** **34 → 38 (+12%), power 0.773 → 0.810, α_{adj} 0.0363 → 0.0361.**

Side Effect

Allowing ABEL only for C_{max}.

- Some drugs show high variability in *AUC* as well.
  - Since in such a case the sample size is mandated by *AUC*, products with high deviations in C_{max} will be approved.
  - Example: CV_{wR} 90% (C_{max}), 60% (*AUC*), θ_0 0.90, target power 80% ⟶ the study is 'overpowered' for C_{max}; C_{max}-*GMR*s of [0.846 – 1.183] will pass BE. Really desirable?
  - With the FDA's RSABE the study could be performed in only 34 subjects…



NTIDs – tighter BE limits

EMA (2010)

- In *specific cases* of products with a narrow therapeutic range, the acceptance interval *may* need to be tightened.
  - The acceptance interval for *AUC* should be tightened to 90.00 – 111.11%.
  - Where C_{max} is of particular importance for safety, efficacy or drug level monitoring the 90.00 – 111.11% acceptance interval should also be applied for this parameter.
  - It is not possible to define a set of criteria to categorise drugs as narrow therapeutic index drugs (NTIDs) and it must be decided *case by case* if an active substance is an NTID *based on clinical considerations*.

EMA (Product-specific guidance 2013 – 2016)

- Sirolimus: 80.00 – 125.00% for C_{max}, 90.00 – 111.11% for AUC_{0-t}.
- Tacrolimus: 80.00 – 125.00% for C_{max}, 90.00 – 111.11% for AUC_{0-72h}.

NTIDs – tighter BE limits

Impact of tighter BE limits on sample size

- Example: *CV* 15%, *GMR* 0.975, target power 90%, 2×2×2 design.
  - Conventional 80.00 – 125.00%

```
library(PowerTOST)
sampleN.TOST(CV=0.15, theta0=0.975, targetpower=0.90,
             theta1=0.8000, theta2=1.2500,
             print=FALSE)[["Sample size"]]
[1] 14
```

  - 90.00 – 111.11%

```
library(PowerTOST)
sampleN.TOST(CV=0.15, theta0=0.975, targetpower=0.90,
             theta1=0.9000, theta2=1.1111,
             print=FALSE)[["Sample size"]]
[1] 62
```

NTIDs – reference scaling

FDA

- First recommended in the guidance for warfarin (2012).
  - Scale bioequivalence limits to the variability of the reference product.
  - Compare test and reference product within-subject variability.
  - A fully replicated 4-period study (RTRT | TRTR) is mandatory.
- Scaling approach similar to the FDA's for HVD(P)s.
  - σ_0 0.10 (*CV* ≈10.02505%)
  - Δ 1.11111
- Must demonstrate:
  - BE with the scaled approach.
  - BE with the conventional limits.
  - Variance of T not higher than of R (upper 90% CI of $\sigma_{wT}/\sigma_{wR} \leq 2.5$)

NTIDs – sample sizes

GMR 0.975, $CV_{wT} = CV_{wR}$, target power 90%, 2×2×4 design.

	90.00 – 111.11% EMA	RSABE FDA
CV_{wR}	*n*	
5.0	12*	44
7.5	12*	26
10.0	14	22
12.5	22	20
15.0	32	20
20.0	54	18

As above; $CV_{wT} = 1.5 \times CV_{wR}$, FDA's RSABE

CV_{wR}	CV_{wT}	*n*
5.0	7.50	48
7.5	11.25	36
10.0	15.00	32
12.5	18.75	30
15.0	22.50	30
20.0	30.00	28

Establishing the Biostudy Statistical Design

Thank You!

Open Questions?

Helmut Schütz

BEBAC

Consultancy Services for
Bioequivalence and Bioavailability Studies
1070 Vienna, Austria

helmut.schuetz@bebac.at