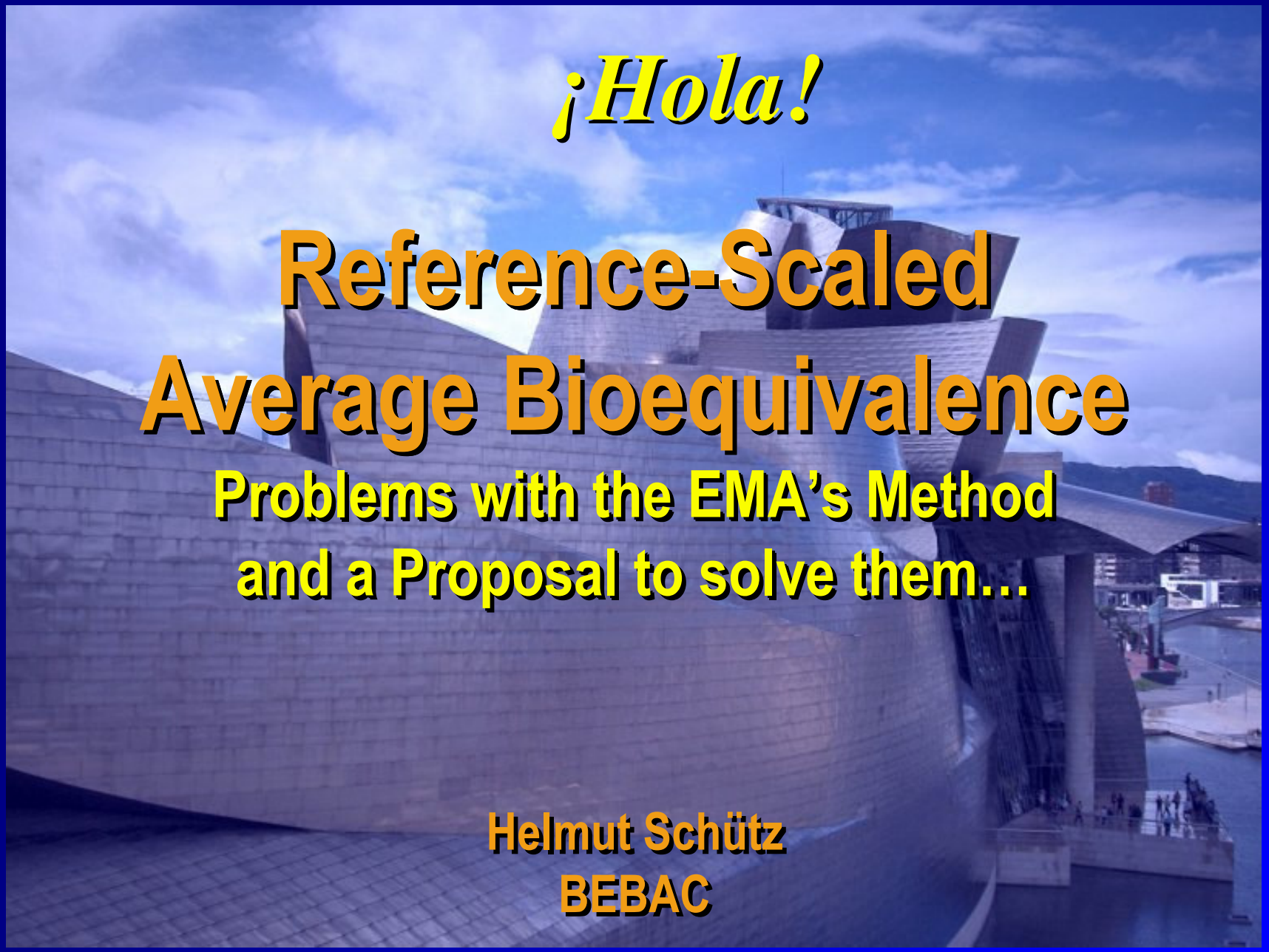


¡Hola!

Reference-Scaled Average Bioequivalence

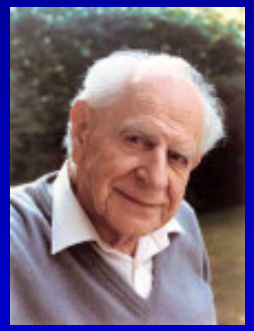
Problems with the EMA's Method and a Proposal to solve them…

Helmut Schütz
BEBAC

Wikimedia Commons • 2006 Georges Jansoone • CC BY-SA 3.0 Unported

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.

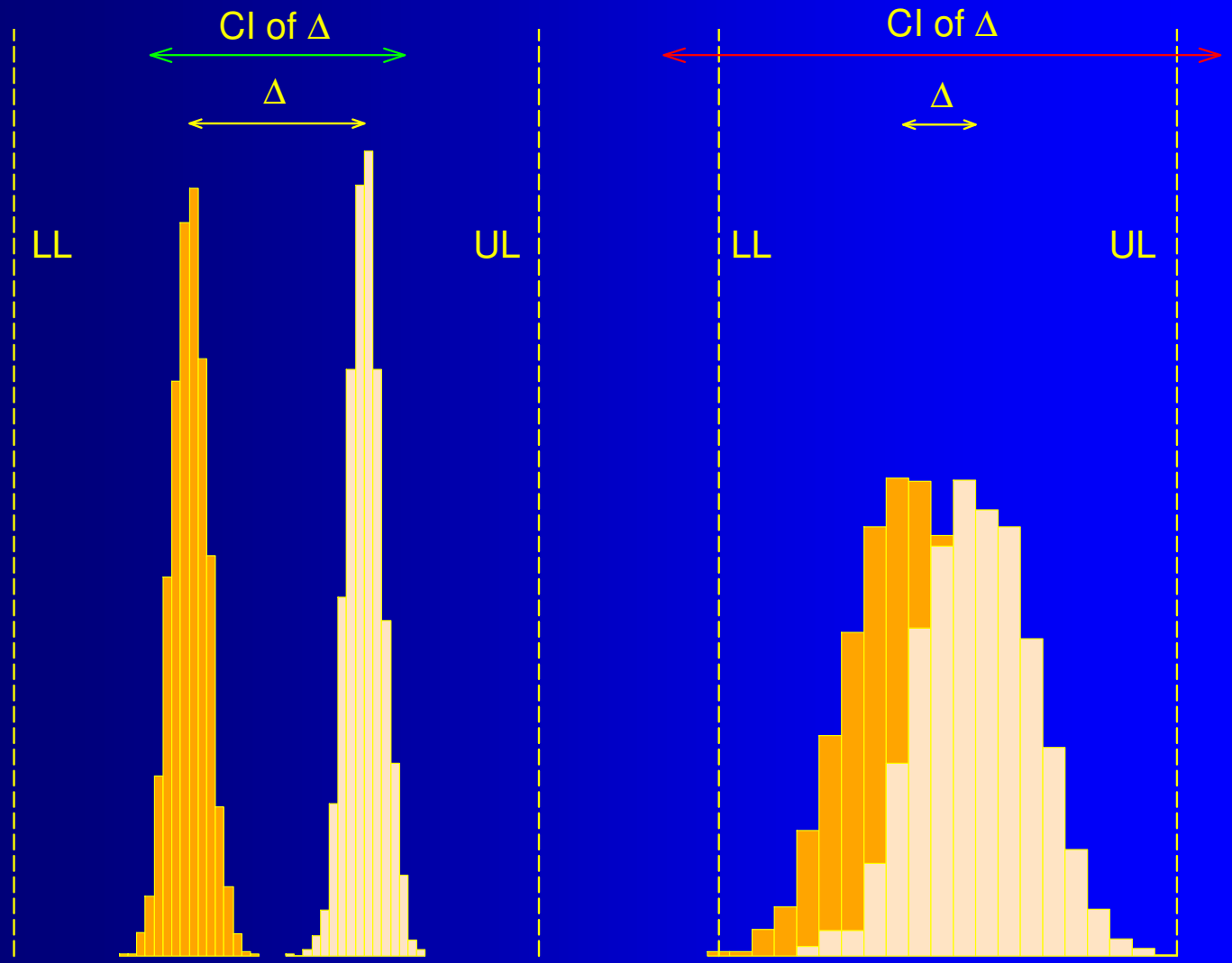


Karl R. Popper

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

Leslie Z. Benet

High variability



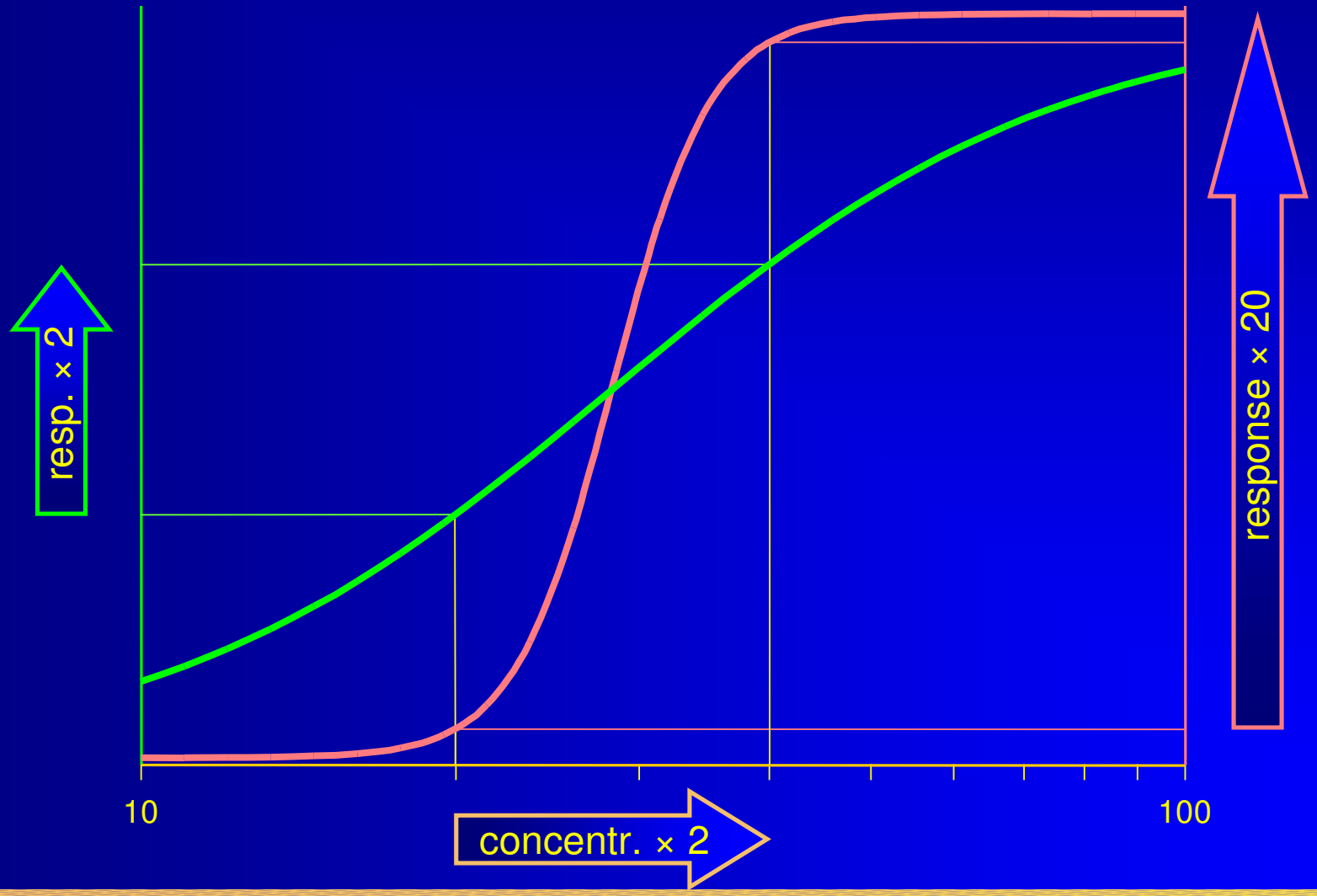
Modified from Fig. 1
Tothfálusi *et al.* (2009)

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.

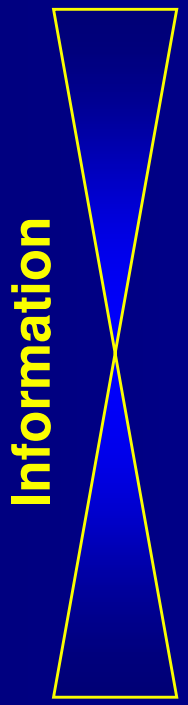
HVDs/HVDPs are safe

steep/flat PK/PD-curves



Hierarchy of Designs

- The more complex a design is, the more information can be obtained out of it.
  - Hierarchy of designs:



Full replicate (RTRT | TRTR or RTR | TRT),
Partial replicate (RRT | RTR | TRR)
Standard 2×2×2 cross-over (RT | TR)
Parallel (R | T)

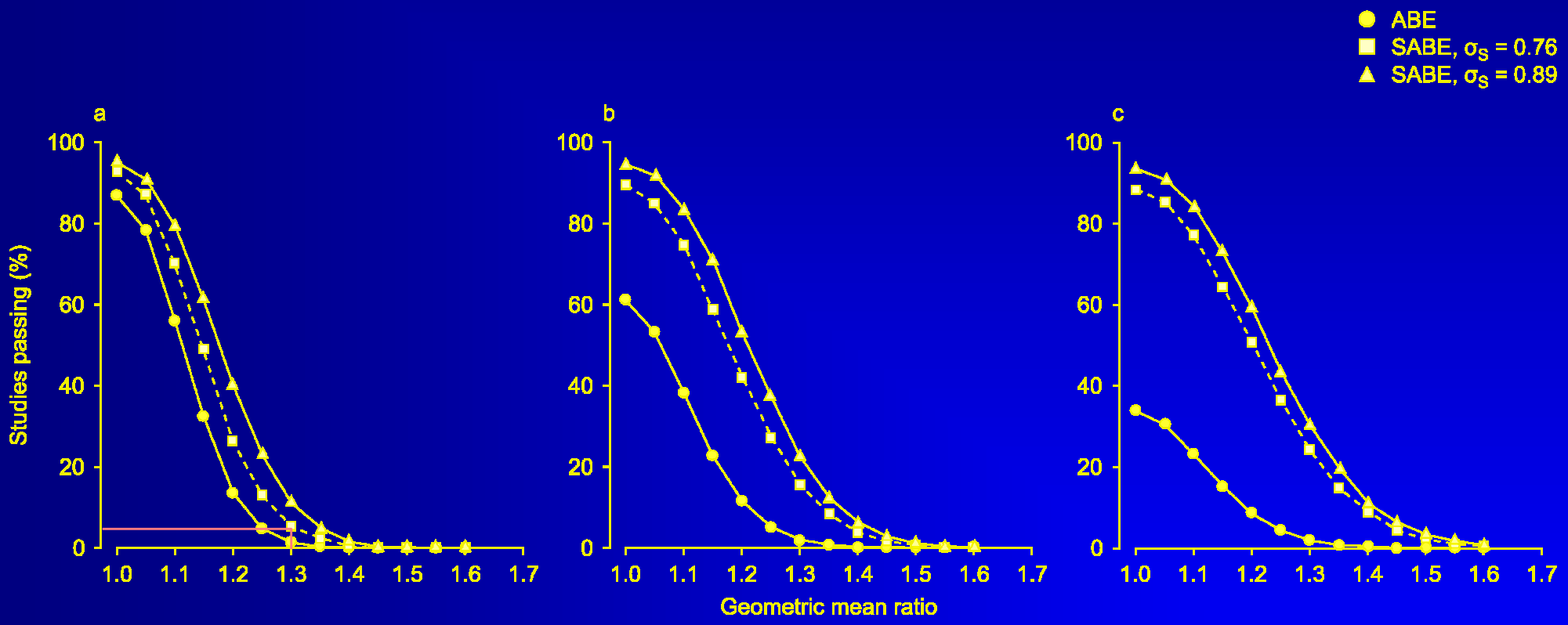
  - Variances which can be estimated:

Parallel:	total variance (between + within)
2×2×2 Xover:	+ between, within subjects
Partial replicate:	+ within subjects of reference
Full replicate:	+ within subjects of reference and test

Replicate Designs (Applications)

- Any replicate design can be evaluated by Average Bioequivalence (ABE)
  - Mandatory if scaling not allowed
    - AUC (generally …)
    - Other PK metrics if $CV_{wR} \leq 30\%$
  - Even if scaling is not intended, replicate designs give more information about formulations.
- Necessary for Scaled Average Bioequivalence (SABE)
  - C_{max}, $C_{ss,min}$, $C_{ss,\tau}$, partial AUCs of MR formulations if no clinical concerns (EMA) and $CV_{wR} > 30\%$.

Power (ABE *vs.* SABE)



Tothfálusi *et al.* (2009), Fig. 3

Simulated (n = 10,000) three-period full replicate design studies (RTR | TRT) in 36 subjects, GMR restriction 0.80 – 1.25.

(a) CV = 35%, (b) CV = 45%, (c) CV = 55%.

ABE: Average Bioequivalence, SABE: Scaled Average Bioequivalence.

σ_s 0.76: EMA criterion, σ_s 0.89: FDA criterion.

Regulatory models

- **Common to the EMA and the FDA**

  ABE model

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

  SABE model

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

  Regulatory regulatory switching condition θ_S is derived from the regulatory standardized variation σ_0 (proportionality between acceptance limits in ln-scale and σ_{wR} in the highly variable region).

EMA's Implementation of SABE

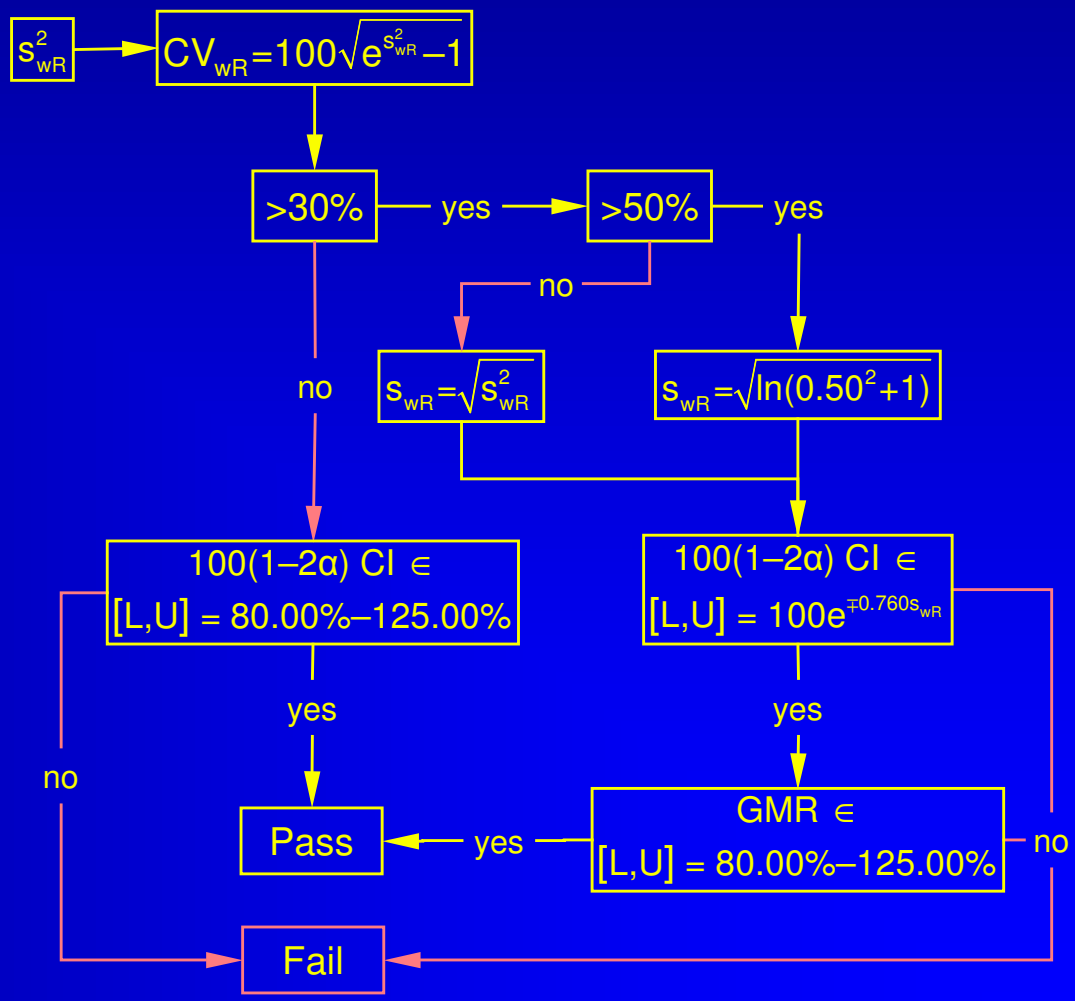
- Average Bioequivalence with Expanding Limits (ABEL)
  - All fixed effects model according to the EMA's Q&A-document preferred (*e.g.*, SAS `Proc GLM`, R `lm`).
  - Based on s_{wR} calculate the scaled acceptance range based on the regulatory constant k (θ_s = 0.760); limited at CV_{wR} 50%.
  - GMR within 0.80 – 1.25.
  - Justification that the widened acceptance range is clinically not relevant (important – different to FDA).
  - Demonstration that CV_{wR} >30% is not caused by outliers (box plots?).

$$[L-U] = e^{\mp k \cdot s_{wR}}$$

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

ABEL (EMA)

- Decision tree
  - The null hypothesis is modified* in the face of the data!
  - Acceptance limits themselves become random variables.
  - Type I error (consumer risk) might be inflated.



* In the strict sense the null hypothesis is undefined!

ABEL (EMA)

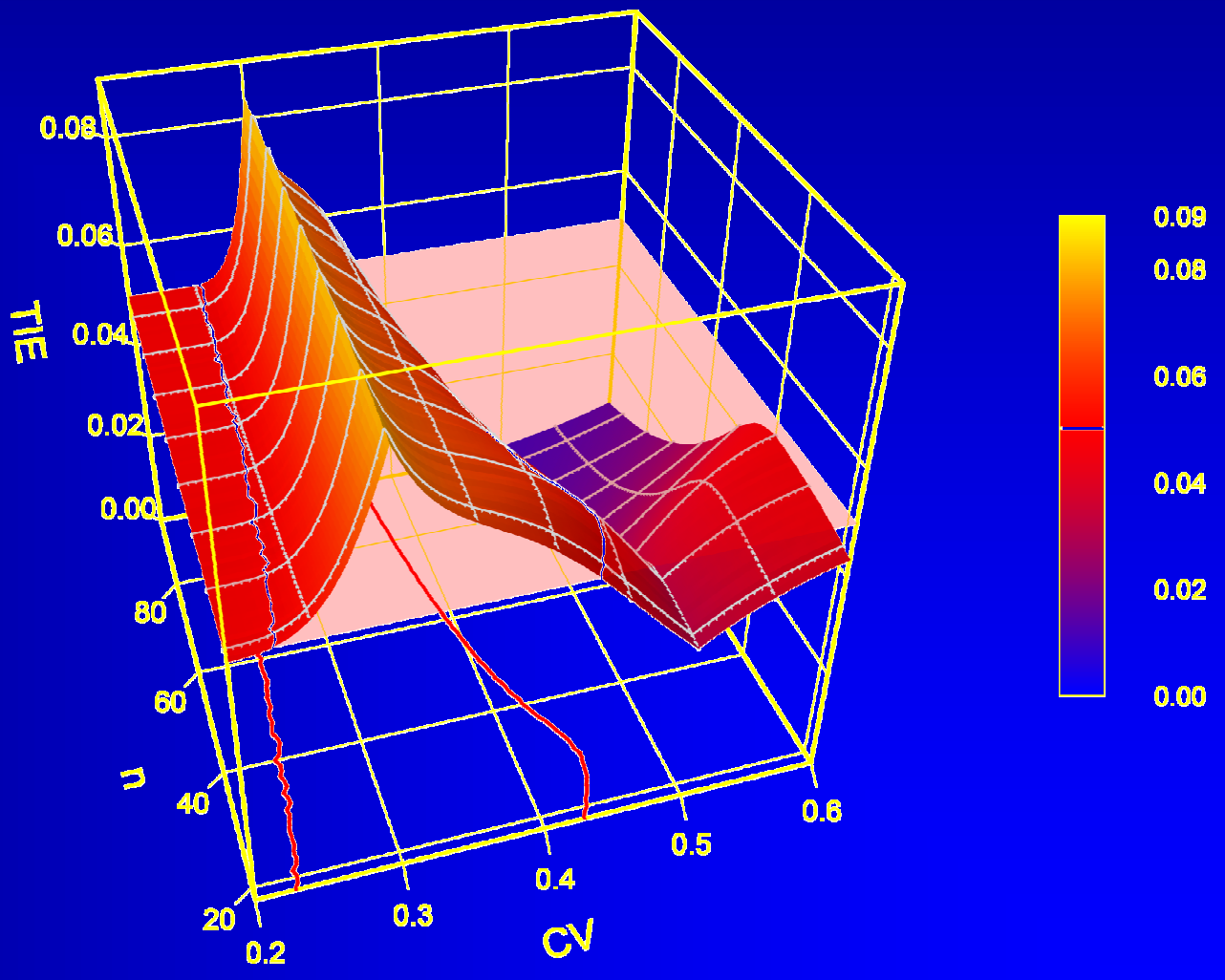
- Assessing the type I error (TIE)
  - TIE = falsely concluding BE at the limits of the acceptance range. In ABE the TIE is ≤0.05 at 0.8 and ≤0.05 at 1.25.
  - Due to the decision tree no direct estimation of the TIE at the scaled limits is possible. Extensive simulations are required (slow convergence: 1 mio BE studies mandatory).
  - Inflated TIE suspected
    (Tóthfalusi & Endrényi 2003, Chow & Liu 2009).
    Confirmed for ABEL
    (Labes@BEBA-Forum 2013, Wonnemann *et al.* 2015).

Labes D, Schütz H. Inflation of Type I Error in the Evaluation of Scaled Average Bioequivalence, and a Method for its Control. In preparation 2016.

ABEL (EMA)

- **Example:**
  RTRT | TRTR
  sample size 18 – 96
  CV_{wR} 20% – 60%

TIE_{max} 0.0837
(rel. increase of the consumer risk 67%)



ABEL (Problems)

- 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!

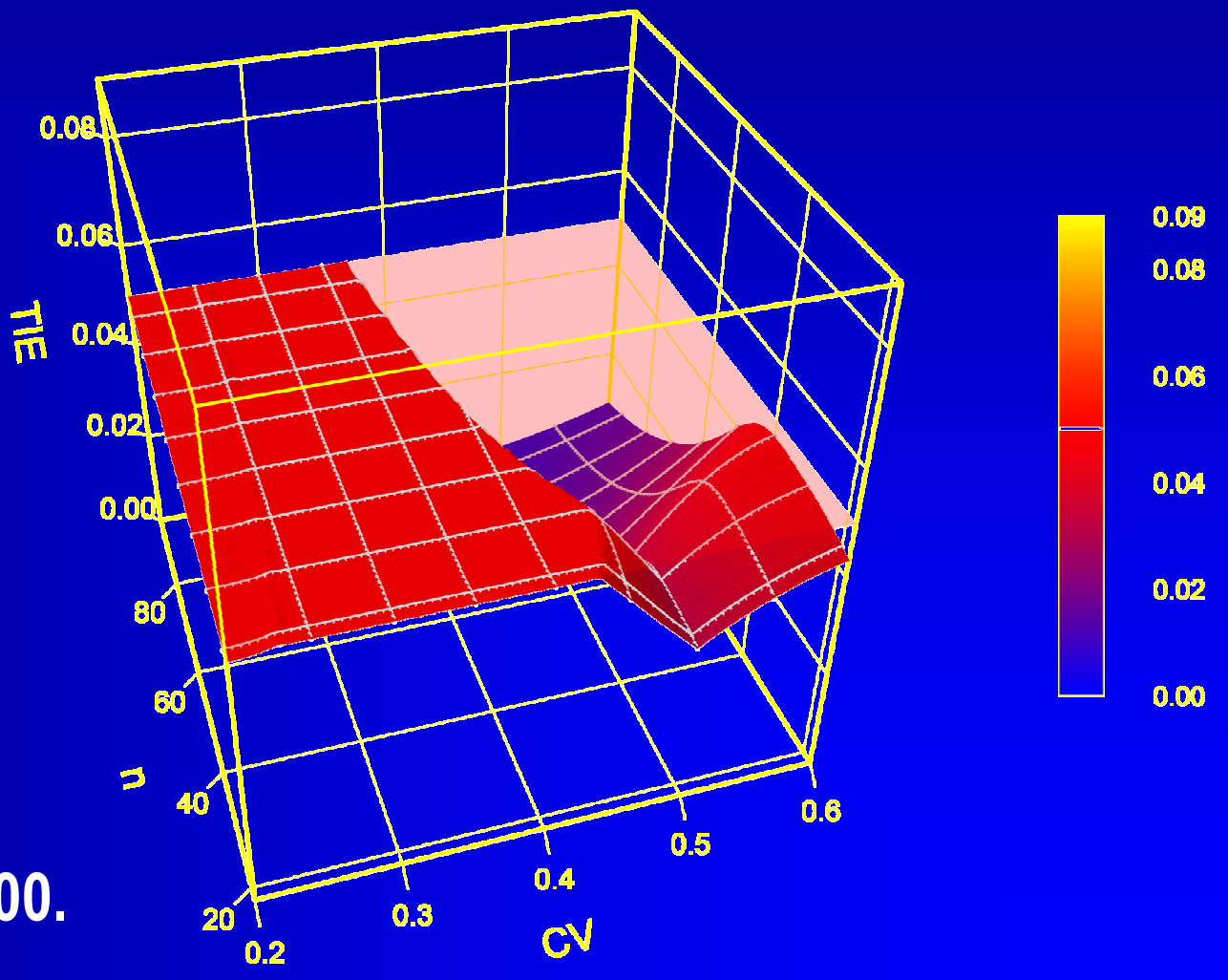
  - Only their *estimates* (GMR, s_{wR}) are accessible in the actual study.
  - At CV_{wR} 30% the decision to scale will be wrong in 50% of cases!
  - If moving away from 30% the chances of a wrong decision decrease and hence, the TIE.

ABEL (Solutions)

- What can we do?
  - Utopia: Agencies collect CV_{wR} from submitted studies. Pool them, adjust for designs / degrees of freedom. The EMA publishs a *fixed* acceptance range in the product-specific guidance. No need for replicate studies any more. 2×2×2 crossovers evaluated by ABE would be sufficient.
  - Halfbaked: Hope that Bonferroni preserves the consumer risk. Still apply ABEL, but with a 95% CI (α 0.025).
    But: Loss of power, substanial increase in sample sizes.
  - Proposal: Iteratively adjust α based on the study's CV_{wR} – in such a way that the consumer risk is preserved.

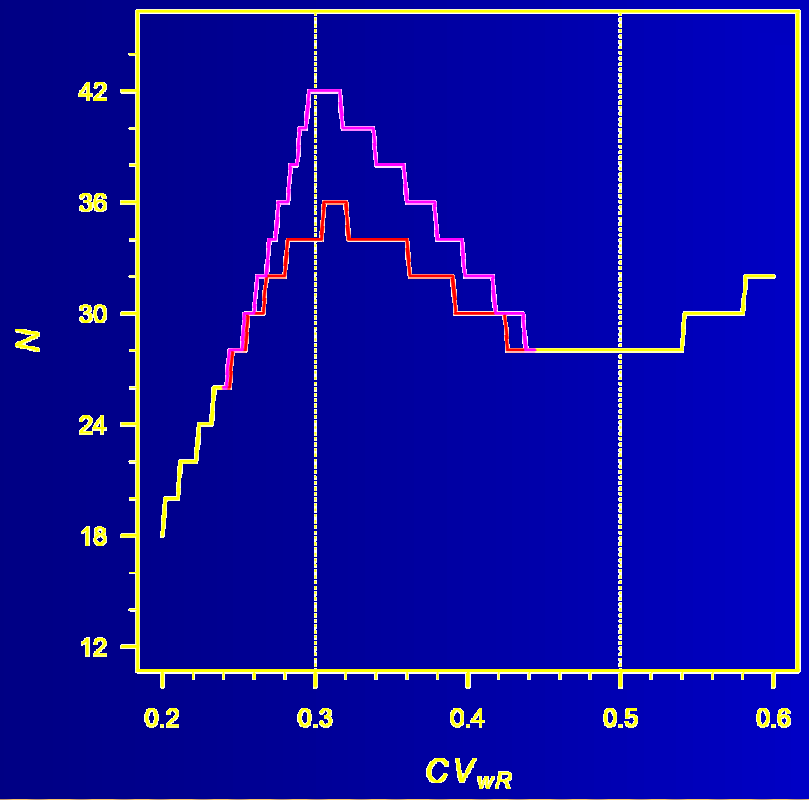
ABEL (iteratively adjusted α)

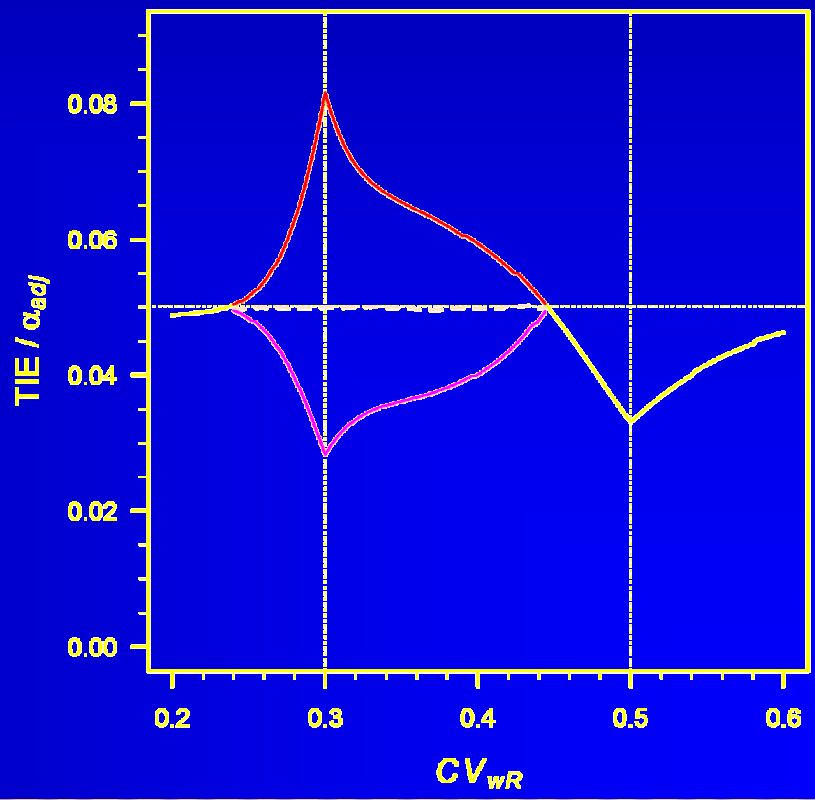
- Previous example
  - Assess the TIE for α 0.05.
  - If n.s. >0.05, stop.
  - Otherwise adjust α (downwards) until TIE 0.05.
  - At CV_{wR} 30% (dependent on the sample size) α_{adj} is 0.0273–0.0300.



ABEL (iteratively adjusted α)

- Potential impact on the sample size
  - Moderate in the critical region (—), none outside (—).





Conclusions

- EMA's ABEL evaluated with nominal α 0.05
  - Consumer risk unacceptably compromised in the critical region of CV_{wR} ~25% to ~45%.
  - However, no inflation of the TIE for any CV_{wR} >~45%.
- Prespecified α (*e.g.*, Bonferroni)
  - Adjusts even if not necessary. Hence, substantial impact on power.
- Iteratively adjusted α
  - Adjusts only if necessary while preserving the consumer risk. Always more powerful than Bonferroni.
  - Implemented in R `PowerTOST` function `scABEL.ad()`

¡Gracias!

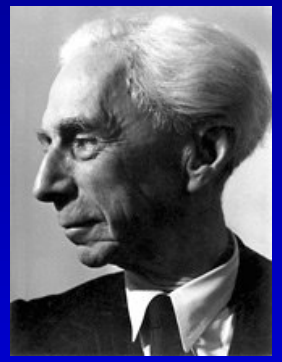
Reference-Scaled Average Bioequivalence

Open Questions?

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

To bear in Remembrance...

The fundamental cause of trouble in the world today is that the stupid are cocksure
while the intelligent are full of doubt. *Bertrand Russell*



100% of all disasters are failures of design, not analysis.

Ronald G. Marks

My definition of an expert in any field is a person who knows enough about what's really going on to be scared.

Phillip J. Auger