Bioequivalence

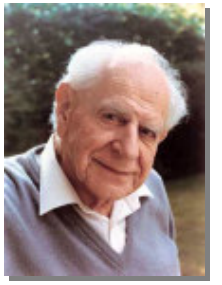
An Old Area with some Uncharted Territories

Biometric Colloquium | Vienna, 15 December 2021

Helmut Schütz

To Keep in Memory…

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

Pharmacokinetics (PK)

- φαρμακός (drug) + κινητικός (putting in motion)
  - Coined in 1953
    - Friedrich H. Dost
      Der Blutspiegel.
      Kinetik der Konzentrationsabläufe in der Kreislaufflüssigkeit (1953)
  - Pharmacokinetics may be simply defined as what the body does to the drug, as opposed to pharmacodynamics which may be defined as what the drug does to the body.
    - Leslie Z. Benet
      Pharmacokinetics.
      Basic Principles and Its Use as a Tool in Drug Metabolism (1984)

PK

- (L)ADME

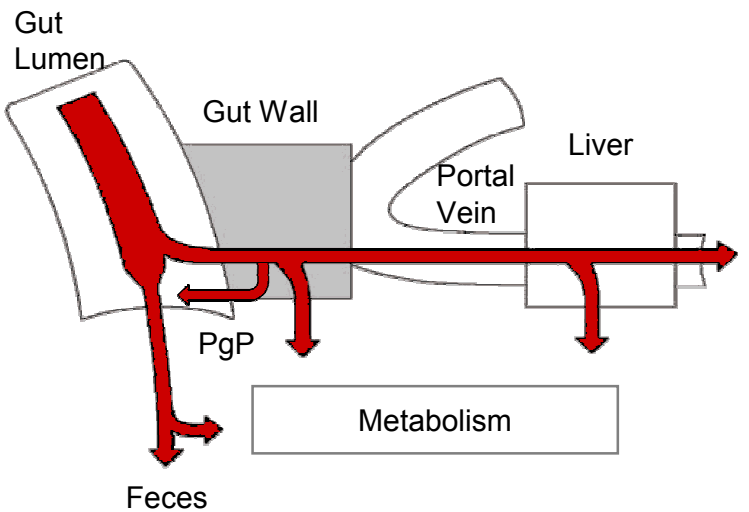
Formulation specific

Biopharmaceutical phase
- Disintegration
- Release
- Dissolution

} **L**iberation

Drug specific

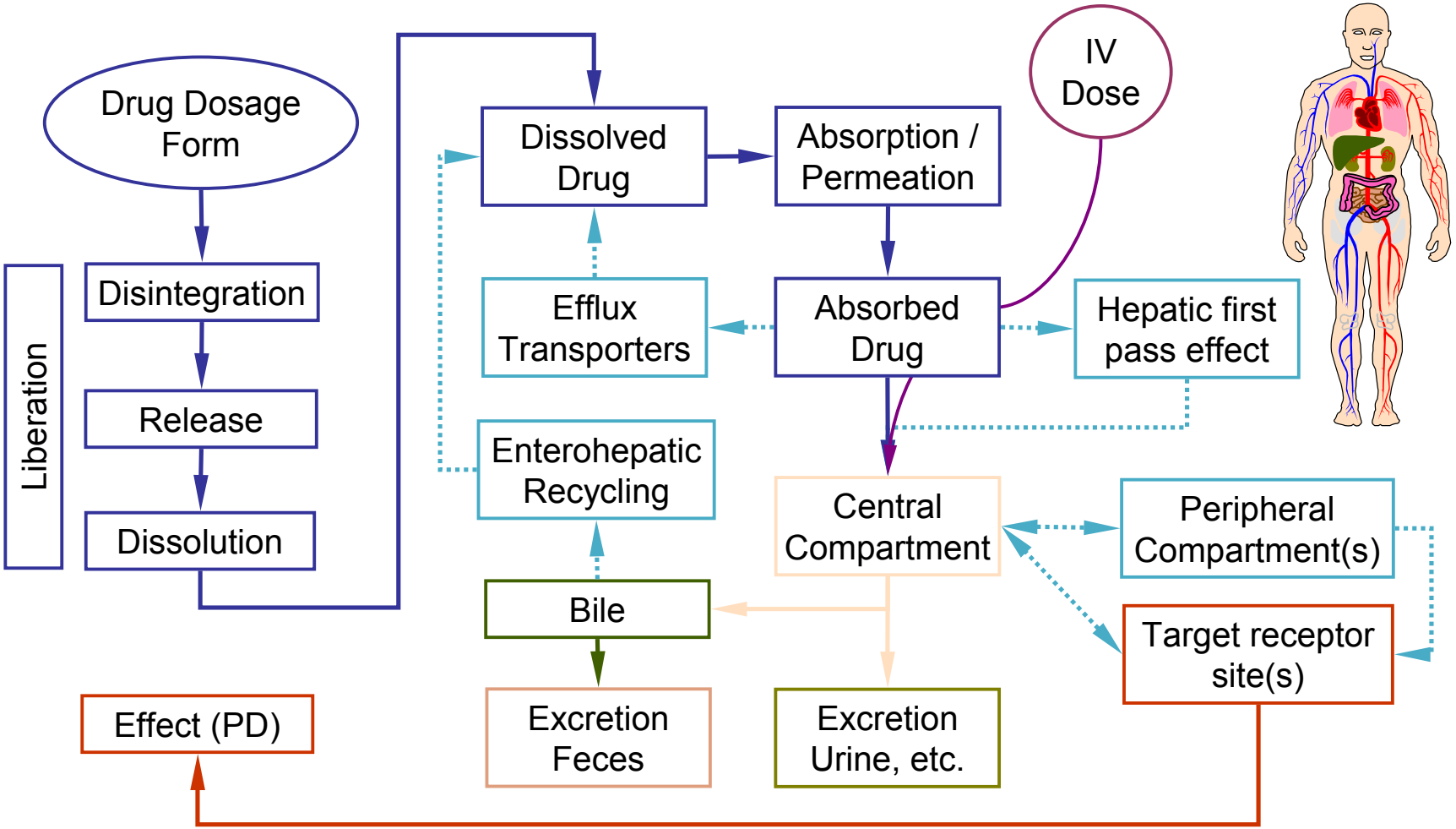
Pharmacokinetic phase
- **A**bsorption / Permeation
  - Passive diffusion
  - Active transport
- **D**istribution
- **M**etabolism
  - Intestinal first pass
  - Membrane first pass
  - Hepatic first pass
- **E**xcretion



Elimination = **M** + **E**

Rowland M, Tozer TN. *Clinical PK and PD*. Philadelphia: Wolters Kluwer; 2011.

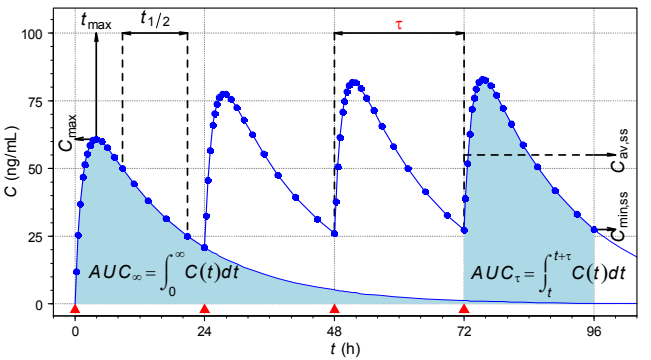
PK *vs* Pharmacodynamics (PD)



Bioavailability (BA)

- Portmanteau of ‘biological’ and ‘availability’
  - Given by the ‘Area Under the [Concentration-time] Curve’

$$AUC = \int_{t=0}^{t=\infty} C(t)\,\mathrm{d}t$$



  - Variants of the PK’s basic equation

$$AUC = \frac{f \cdot D}{CL} = \frac{f \cdot D}{V_d \cdot k_e}, \text{ where}$$

- f is the fraction absorbed (≤1), ⇑ f → ⇑ AUC
- D the administered Dose, ⇑ D → ⇑ AUC
- CL the total body Clearance, ⇑ CL → ⇓ AUC
- V_d the Volume of Distribution, and ⇑ V_d → ⇓ AUC
- k_e the elimination rate constant ($k_e = \log_e 2 / t_{½}$) ⇑ $t_{½}$ → ⇑ AUC

  - AUC is independent from the absorption rate constant k_a

Types of BA

- Absolute BA
  - An extravasal dose compared to an intravenous one

$$f_{abs} = \frac{AUC_{EV,0-\infty}}{AUC_{IV,0-\infty}}, \text{ or } F_{abs} = 100\frac{AUC_{EV,0-\infty}}{AUC_{IV,0-\infty}}$$

- Relative BA
  - A solution compared to an IV dose

$$f_{rel} = \frac{AUC_{PO,0-\infty}}{AUC_{IV,0-\infty}}, \text{ or } F_{rel} = 100\frac{AUC_{PO,0-\infty}}{AUC_{IV,0-\infty}}$$

Comparative BA

- Influence of the formulation (Liberation)
  - A drug (Test) compared to a solution

$$F_{rel} = 100 \frac{AUC_{T,0-\infty}}{AUC_{PO,0-\infty}}$$

- Comparison of formulations (Test *vs* Reference)

$$F_{rel} = 100 \frac{AUC_{T,0-\infty}}{AUC_{R,0-\infty}}$$

- Others
  - Assessment of linear PK (multiple *vs* single dose)
  - Types of formulations (controlled release *vs* immediate release)
  - Food effects (fed *vs* fasting, different types of food)
  - Drug-Drug Interactions

Bioequivalence (BE) – Regulatory Definitions

- EMA (BE GL, 2010)

  Two medicinal products containing the same active substance are considered bio-equivalent if they are pharmaceutically equivalent or pharmaceutical alternatives and their bioavailabilities (rate and extent) after administration in the same molar dose lie within acceptable predefined limits. These limits are set to ensure compar-able *in vivo* performance, *i.e.* similarity in terms of safety and efficacy.

- WHO (TRS 992, Annex 6, 2017)

  Two pharmaceutical products are bioequivalent if they are pharmaceutically equiva-lent or pharmaceutical alternatives, and their bioavailabilities, in terms of rate (C_{max} and t_{max}) and extent of ab-sorption (area under the curve (AUC)), after administra-tion of the same molar dose under the same conditions, are similar to such a degree that their effects can be expected to be essentially the same.

- FDA (CFR 21–320.23(b)(1), 2021)

  Two drug products will be considered bioequivalent drug pro-ducts if they are phar-maceutical equivalents or pharmaceutical alternatives whose rate and extent of ab-sorption do not show a significant difference when administered at the same molar dose of the active moiety under similar experimental conditions, either single dose or multiple dose.

BE – Regulatory Approaches

- Comparison of BA of Test with Reference (rate and extent of absorption by Noncompartmental Analysis, NCA [1]), where
  - Rate = C_{max} maximum observed concentration
  - Extent = AUC all jurisdictions: $AUC_{0-t_{last}}$ [2]
    FDA (additionally): $AUC_{0-\infty}$
  - Other PK metrics required dependent on the design (*i.e.*, multiple dose) and/or the type of formulation

1. According to all guidelines pharmacokinetic modeling is not acceptable in BE.
2. t_{last} = time of the last quantifiable concentration

BE – Assumptions

- BE as a surrogate for Therapeutic Equivalance
  - Studies in healthy volunteers in order to minimize variability (*i.e.*, lower sample sizes than in patients)
  - Emphasis on *in vivo* release ('human dissolution apparatus')
- Concentrations in the sample matrix reflect concentrations at the target receptor site
  - In the strict sense only valid in steady state
  - *In vivo* similarity in healthy volunteers can be extrapolated to the patient population
- $AUC_T = \frac{f_T \cdot D_T}{CL}$, $AUC_R = \frac{f_R \cdot D_R}{CL}$; assuming $D_T = D_R$ and $CL = const$

  $\rightarrow \frac{F_T}{F_R} \approx 100 \frac{AUC_T}{AUC_R}$

BE – Background

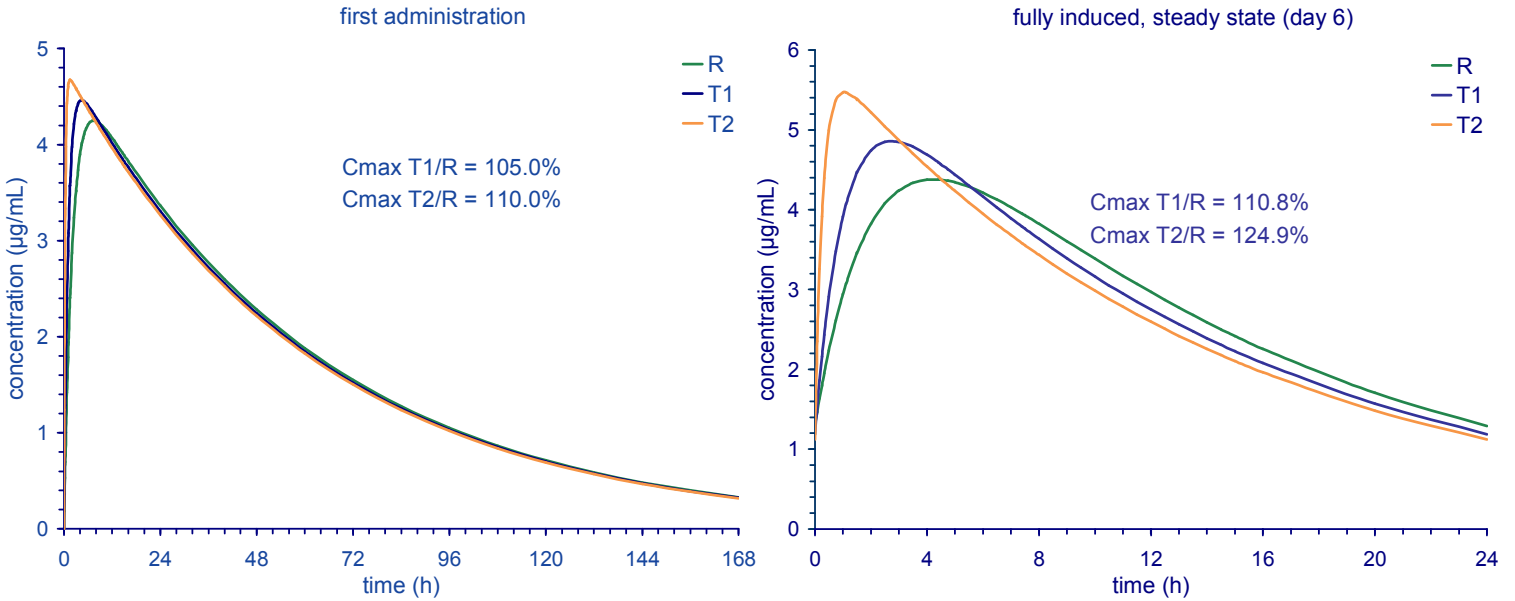
- BE = (desired) result of a comparative BA study
  - Generally only for extravascular routes. Exceptions for IV:
    - Excipients which may interact with the API (complex formation)
  - Case-by-case: Liposomal formulations, emulsions
  - Focus on the ‘core’ API
    - Different salts, esters, isomers, complexes are considered the same active substance
  - Same molar dose
  - Same conditions (*e.g.*, fasting/fed, single/multiple dose)
  - Clinically not relevant difference $\Delta = 20\%$, except for
    - Narrow Therapeutic Index Drugs – NTIDs $< 20\%$
    - Highly Variable Drugs / Drug Products – HVDP(s) $> 20\%$
  - $100(1 - 2\alpha)$ confidence interval of T/R within $\{1 - \Delta, (1 - \Delta)^{-1}\}$ (*e.g.*, for Δ 20%: 80.00 – 125.00%)

BE – Background (cont')

- Design should allow accurate (unbiased) assessment of the treatment effect
  - Generally healthy volunteers (lower variability); except:
    - Not ethical due to effects or AEs → study in patients
  - Cross-over design preferred
    - Each subject serves as its own 'reference'
      - Hence, the comparison is performed *within* subjects
      - More powerful than parallel design
  - Parallel design as an alternative
    - Studies in patients were the disease state is not stable
    - Studies of drugs with (very) long half lives
    - Comparison is performed *between* subjects
      - Requires high degree of standardization
      - Less powerful than cross-over

BE – Background (cont')

- Assessment of the treatment effect (cont'd)
  - Carbamazepine ($k_{a(R)}$ 0.472 h^{-1}, $k_{a(T_1)}$ 0.94 h^{-1}, $k_{a(T_2)}$ 3.6 h^{-1})
    - $t_{½}$ after first administration 43 h → 10 h after full auto-induction
    - A – rare – example where a multiple dose study is more sensitive to detect differences in the rate of absorption than a single dose study



BE – Background (cont')

- Assessment of the treatment effect (cont'd)
  - Parent compound *vs.* metabolite(s)
    - Absorption of parent expected to be the best measure of Liberation and Absorption (formulation dependent)
    - Parent may be difficult to measure (*e.g.*, pro-drugs: low concentrations together with fast elimination)
      - Metabolite as alternative (irrelevant whether active or not)
      - If possible the *first* metabolite in the chain should be measured; the further 'downstream' a metabolite is, the less sensitive it is to detect differences in absorption of the parent
  - Fasting *vs*. fed state
    - Fasting state generally considered the most sensitive, except
      - Intake with food required according to the reference's label / SmPC
      - Studies in fasting *and* fed state for prolonged release* products (EMA and some of the FDA's product-specific guidances)

* a.k.a. controlled release (CR), extended release (ER/XR), long-acting (LA)

BE – Background (cont')

- Assessment of the treatment effect (cont'd)
  - Dose strength
    - The one which is considered to be most sensitive
    - If linear PK *
      - Generally highest strength
      - If highly soluble, a lower strength is acceptable
      - A lower strenght is also acceptable if safety/tolerability issues in healthy subjects (requires justification)
    - If nonlinear PK
      - Higher than proportional increase in *AUC* over the dose range
        - » Generally highest strength; similar exceptions as for linear PK
      - Lower than proportional increase in *AUC* over the dose range
        - » Lowest *and* highest strength
        - » Under certain conditions testing only the lowest strength can be justified

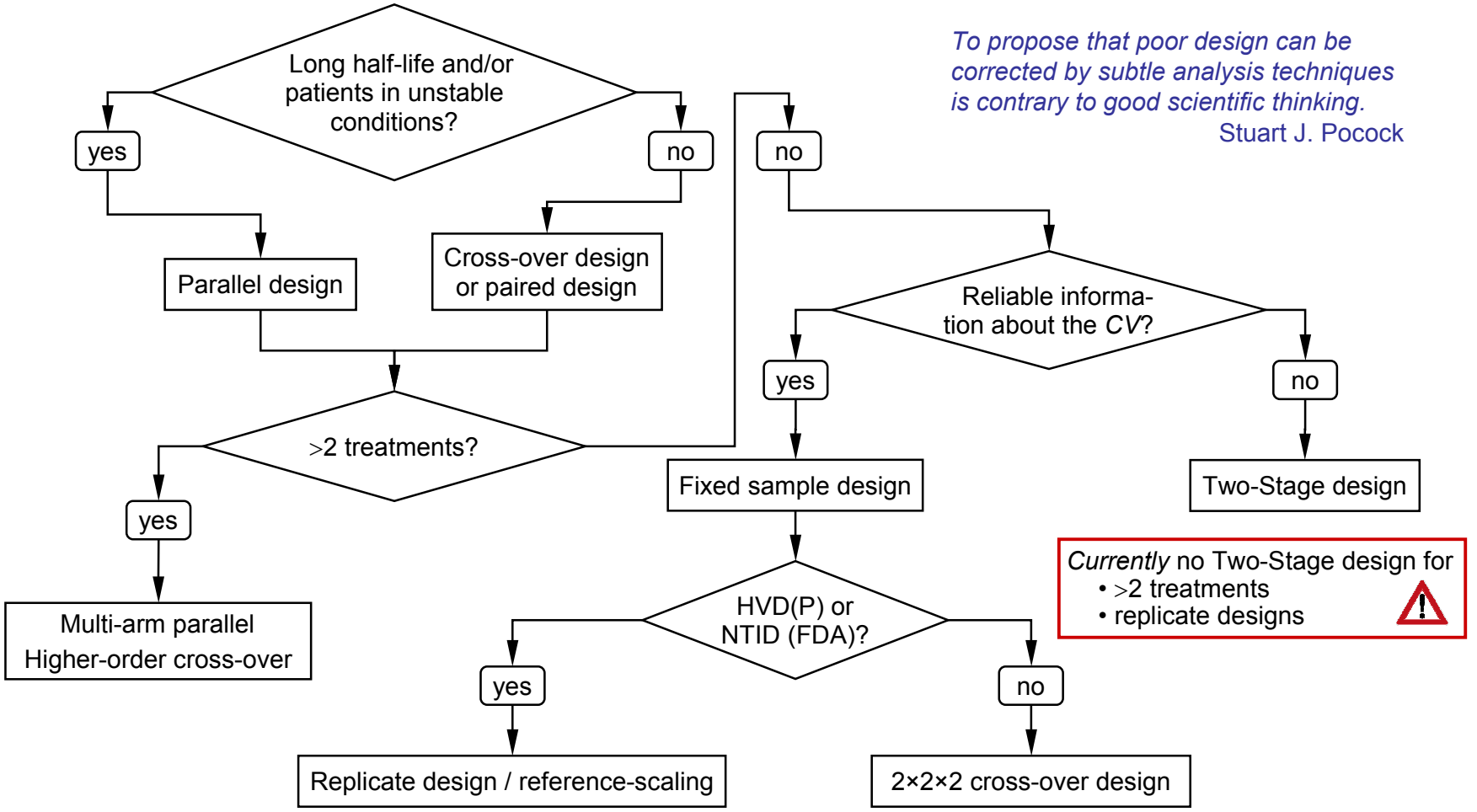
* Bauer A, Kagedal M, Wolfsegger MJ. *Assessment of PK linearity after repeated drug administration using the superposition principle*. Manuscript submitted 2021.

BE – Background (cont')

- Assessment of the treatment effect (cont'd)
  - If multiple strengths, biostudies can be waived ('Proportionality Biowaiver') given certain conditions
    - One *in vivo* BE study performed
    - Similarity of other strengths demonstrated *in vitro* (dissolution in p_H 1.2, 4.5, and 6.8): bootstrapped lower confidence limit of $f_2 \geq 50$
  - If nonlinear PK for range of strengths, waiving still possible within strenghts which show linear PK *
- Special case Immediate Release formulations
  - For Biopharmaceutical Classification System 'class I' and – some – 'class III' drugs *in vitro* similarity is sufficient ('BCS-based Biowaiver') – except for NTIDs

* Wolfsegger MJ, Bauer A, Labes D, Schütz H, Vonk R, Lang B, Lehr S, Jaki TF, Engel W, Hale MD. *Assessing goodness-of-fit for evaluation of dose-proportionality*. Pharm Stat. 2021; 20(2): 272–81. doi:10.1002/pst.2074.

Selecting an Appropriate Design



To propose that poor design can be corrected by subtle analysis techniques is contrary to good scientific thinking.
Stuart J. Pocock

Currently no Two-Stage design for
- >2 treatments
- replicate designs

Designs – Background

- 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 cross-over (TR | RT) ⇦

Parallel (T | R)

Information

- Variances which can be estimated

Parallel $s^2_{t(p)} = s^2_b + s^2_w$

2×2×2 cross-over $s^2_b,\ s^2_w$

Partial replicate $s^2_T = s^2_{bT} + s^2_{wT},\ s^2_{bR},\ s^2_{wR}$

Full replicate $s^2_{bT},\ s^2_{wT},\ s^2_{bR},\ s^2_{wR}$

Standard Designs in Comparative BA

- Parallel design (T | R)
  - Standard design for studies
    - in patients with an instable disease or if AEs in healthy volunteers
    - in healthy volunteers for drugs with long half lives
    - Assumes lacking difference in groups
    - Similar anthropometric properties (sex, age, BMI, …)
    - If the drug is subjected to genetic polymorphism, [1] geno- / phenotyping is strongly recommended
  - Equal variances must not be assumed [2]
    - the *t*-test is liberal in case of
      - heterogenicity and/or
      - unequal group sizes

1. Subpopulations of extensive (fast) and poor (slow) metabolizers.
2. FDA, CDER. *Guidance for Industry. Statistical Approaches to Establishing Bioequivalence*. January 2001.

Standard Designs in Comparative BA (cont'd)

- Cross-over design (TR | RT)
  - Globally applied standard protocol for bioequivalence, drug-drug interaction, and food effect studies
  - Assumes that the treatment effect is independent from the sequence [1] of administration
    (if not, the estimate would be biased)
  - Sufficiently long washout between periods
    - No residual concentrations in higher period(s) [2]
    - No remaining effect which may influence ADME [3]
    - Patients: Stable disease
  - Period effects are not relevant (accounted for in the model)
  - Assumes homoscedasticity

1. Actually assuming equal carryover.
2. According to all guidelines a subject with a pre-dose concentration of > 5% of C_{max} should be excluded from the comparison.
3. Not stated in any guideline. However, especially important for auto-inducers and auto-inhibitors.

Standard Designs in Comparative BA (cont'd)

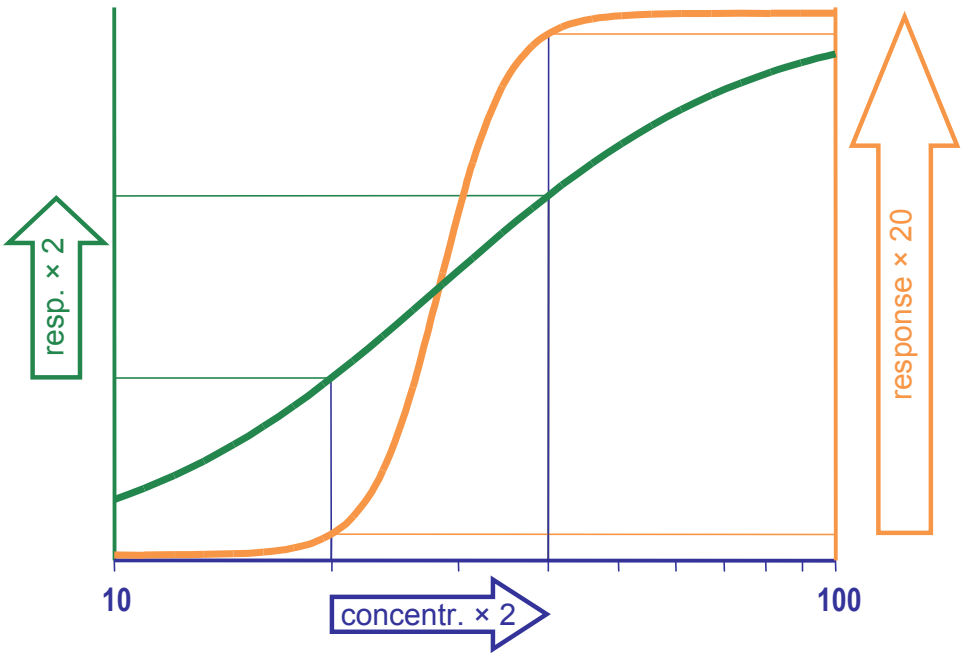
- Higher-order cross-over designs
  - Same assumptions and requirements like simple cross-over
  - Applications
    - Product performance in fasting and fed state (*i.e.*, $T_{fasting}$ *vs* $R_{fasting}$ and T_{fed} *vs* R_{fed})
    - Assessing BE of one Test vs R from two regions (*e.g.*, T *vs* the RLD and T *vs* a European reference)
    - Assessing dose-proportionality
    - Selection between candidate formulations in a pilot study (*e.g.*, T_1 *vs* R and T_2 *vs* R and T_3 *vs* R)
    - Establishing an *in vitro in vivo* correlation (*e.g.*, CR with fast – target – slow release characteristics)

Standard Designs in Comparative BA (cont'd)

- Replicate designs
  - Same assumptions and requirements like cross-over
  - Mandatory for Scaled Average Bioequivalence (SABE) of HVD(P)s and NTIDs (FDA)
  - At least R is administered twice
    - 'Partial' aka 'semireplicate' designs
      - TRR | RTR | RRT (most popular)
      - TRR | RTR (extra-reference design; not recommended)
    - Full replicate designs
      - Three periods
        TRT | RTR or TRR | RTT
      - Four periods, two sequences
        TRTR | RTRT, TRRT | RTTR, TTRR | RRTT
      - Four periods, four sequences
        TRTR | RTRT | TRRT | RTTR, TRRT | RTTR | TTRR | RRTT

Recent Developments (1) NTIDs and HVD(P)s

- Clinically not relevant difference Δ?
  - Based on PK but extrapolated to similarity of safety and efficacy in the patient population
    - Δ depends on the dose-response curves: NTID (steep), HVD (flat)



Recent Development (1a) NTIDs

- $\Delta < 20\%$
  - EMA * 10% → fixed limits of 90.00 – 111.11%
  - FDA Limits scaled based on the reference's variability CV_{wR}

CV_{wR}	limits (%)	Δ_r
5.00	94.87 – 105.41	5.13
7.50	92.41 – 108.21	7.59
10.00	90.02 – 111.08	9.98
10.03	90.00 – 111.11	10.00
15.00	85.46 – 117.02	14.54
20.00	81.17 – 123.20	18.83
21.50	80.00 – 125.00	20.00

$$100\exp(\mp 1.053605 \cdot s_{wR})$$

Additionally

- must pass 80.00 – 125.00%
- $\sigma_T / \sigma_R \leq 2.5$

* And the WHO, Health Canada, any many, many others…

Recent Development (1b) HVD(P)s

- $\Delta > 20\%$
  - GCC 25% → fixed limits of 75.00 – 133.33% (C_{max} only)
  - EMA Scaled based on CV_{wR} (C_{max} only)
  - WHO Like EMA (*AUC* also)
  - HC Like EMA (*AUC* only)
  - FDA Scaled based on CV_{wR} (*AUC* and C_{max})

EMA, WHO	
CV_{wR}	limits (%)
≤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

Health Canada	
CV_{wR}	limits (%)
≤30	80.0 – 125.0
35	77.2 – 129.5
40	74.6 – 143.0
45	72.2 – 138.6
50	69.8 – 143.2
≥57.4	66.7 – 150.0

FDA	
CV_{wR}	limits (%)
≤30	80.00 – 125.00
35	73.83 – 135.45
40	70.90 – 141.04
45	68.16 – 146.71
50	65.60 – 152.45
60	60.96 – 164.04

$100\exp(\mp 0.760 \cdot s_{wR})$ $100\exp(\mp 0.8925742 \cdot s_{wR})$

Interlude – Hypotheses in BE

- Average Bioequivalence (ABE)

  $H_0 : \frac{\mu_T}{\mu_R} \ni \{\theta_1, \theta_2\}$ *vs* $H_1 : \theta_1 < \frac{\mu_T}{\mu_R} < \theta_2$,

  where the – fixed – limits $\{\theta_1, \theta_2\}$ of the acceptance range depend on the clinically not relevant difference Δ by

  $\theta_1 = 1 - \Delta,\ \theta_2 = (1 - \Delta)^{-1}$

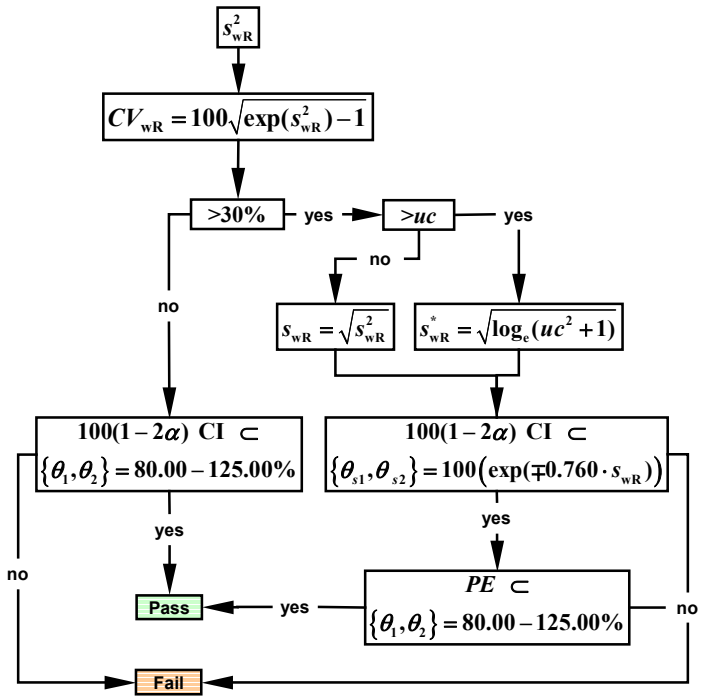
- Scaled Average Bioequivalence (SABE)

  $H_0 : \frac{\mu_T}{\mu_R} \Big/ \sigma_{wR} \ni \{\theta_{s_1}, \theta_{s_2}\}$ *vs* $H_1 : \theta_{s_1} < \frac{\mu_T}{\mu_R} \Big/ \sigma_{wR} < \theta_{s_2}$,

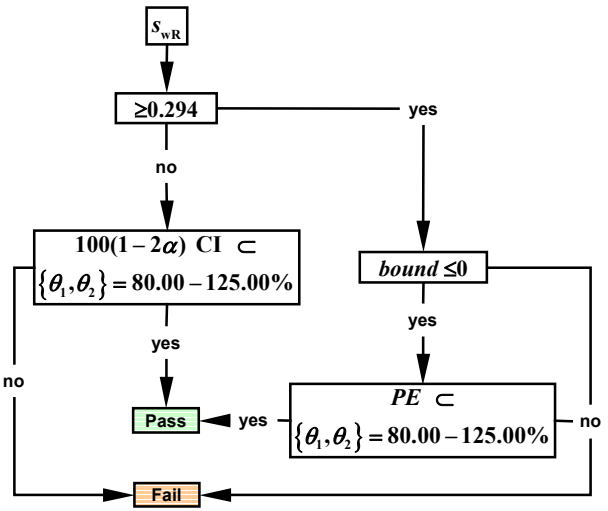
  where σ_{wR} is the standard deviation of the reference and the scaled limits $\{\theta_{s_1}, \theta_{s_2}\}$ of the acceptance range depend on conditions given by the agency

SABE – Frameworks

- Implemented



ABEL [1]
(EMA, …)



RSABE [2]
(FDA, CDE)

1. Average Bioequivalence with Expanding Limits
2. Reference-scaled Average Bioequivalence

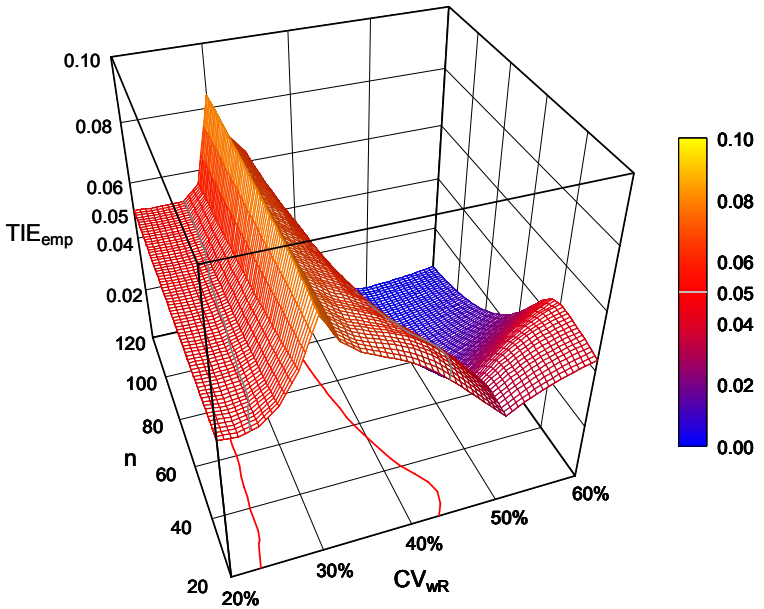
SABE – Problems

- SABE as implemented …
  - RSABE (Reference-Scaled Average Bioequivalene)
  - ABEL (Average Bioequivalence with Expanding Limits)
- … are frameworks, where the acceptance limits are random variables depending on the realized variability
  - Strictly speaking, Δ is not defined beforehand
  - The *model* is based on the true (but unknown) parameter σ_{wR}, whereas the *study* is assessed based on the realized s_{wR}
  - This may lead to a misclassification, *i.e.*,
    - the limits are scaled (because $CV_{wR} > 30\%$), although the drug is *not* highly variable and hence,
    - the chance to pass increases, compromising the patient's risk [1,2]

1. 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.
2. Schütz H, Labes D. *Critical remarks on reference-scaled average bioequivalence*. Manuscript submitted 2021.

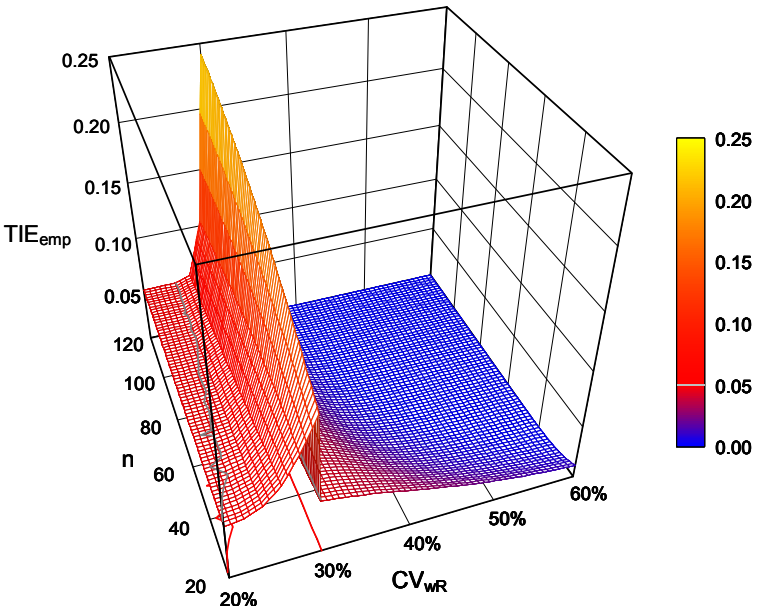
SABE – Inflation of the Type I Error

- Example TRTR | RTRT



ABEL

Inflated *TIE* with CV_{wR} ≈24 – ≈42%
low dependency on sample size
($n = 20$: 0.0800, $n = 120$: 0.0838)



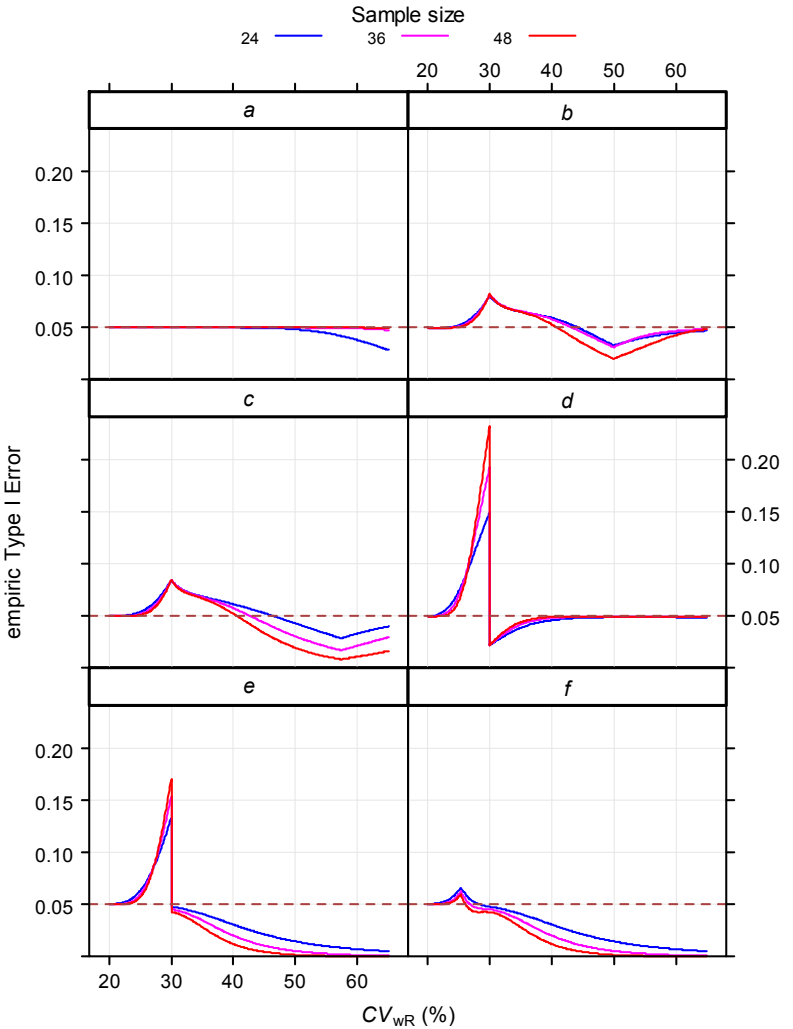
RSABE

Inflated *TIE* with CV_{wR} <30%
high dependency on sample size
($n = 20$: 0.1251, $n = 120$: 0.2421)

Maximum empiric *TIE* at true $CV_{wR} = 30\%$

SABE – Inflation of the Type I Error

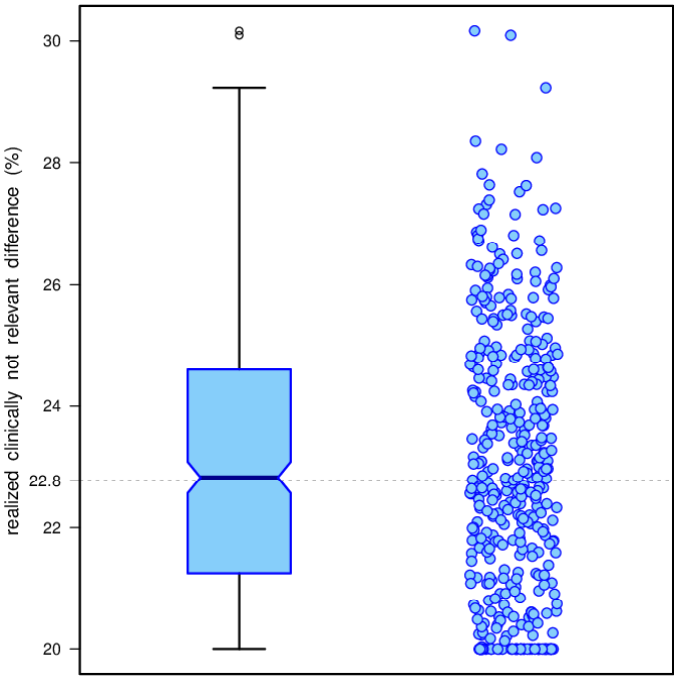
- Example TRTR | RTRT
  - 10^6 simulated [1] studies
  - $n = 24, 36, 48$
  - True $CV_{wR} = 20 – 65\%$
  - True $\theta_0 = \theta_{s2}$
    - *a* ABE
    - *b* ABEL (EMA and others)
    - *c* ABEL (Health Canada)
    - *d* ABEL (GCC)
    - *e* RSABE (implied limits) [2]
    - *f* RSABE (desired consumer risk model) [2]



1. No exact method exists for power and hence, the *TIE* in the implemented regulatory frameworks. Therefore, extensive simulations under the Null are required.
2. Davit BM *et al*. *Implementation of a Reference-Scaled Average Bioequivalence Approach for Highly Variable Generic Drug Products by the US Food and Drug Administration*. AAPS J 2012; 14(4): 915–24. doi:10.1208/s12248-012-9406-x.

SABE – Realized Δ_r

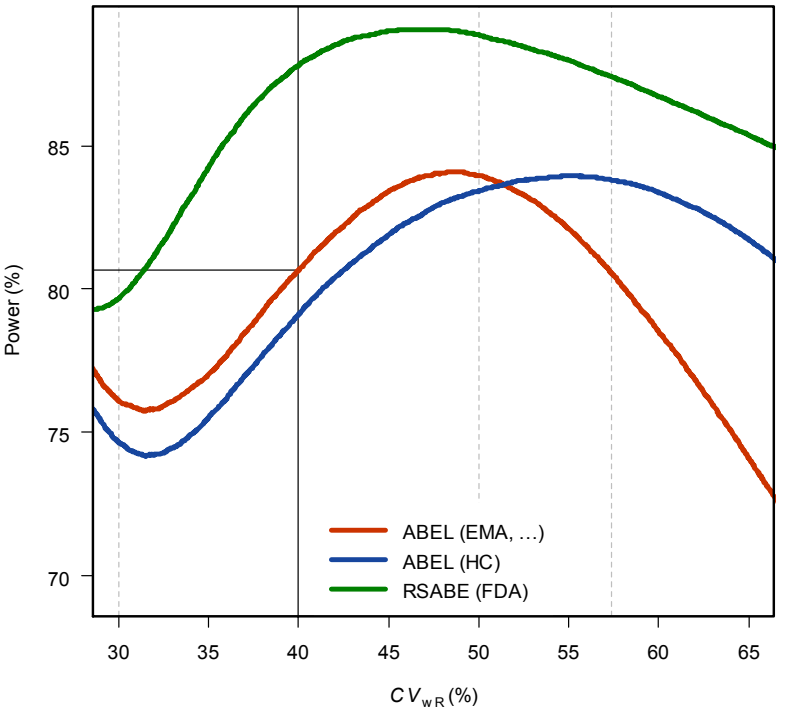
- Example TRTR | RTRT
  - 500 simulated studies for ABEL
  - $n = 34$ (81.2% power)
  - True $CV_{wR} = 35\%$ ($\Delta = 22.77\%$)
  - True $\theta_0 = 0.90$
    - 417 studies passed (83.4%)
    - Realized CV_{wR} 22.30 – 51.25% (Δ_r 20.00 – 30.16%)



- Every study sets its own rules, awarding ones with high CV_{wR}
  - Without access to the study report, Δ_r is unknown to physicians, pharmacists, and patients alike
  - This is an unsatisfactory situation – we put the cart before the horse

SABE – Lack of Harmonization

- Example TRTR | RTRT
  - Designed for ABEL (EMA)
    - Assumed $CV_{wR} = 40\%$
    - Assumed $\theta_0 = 0.90$
    - Target power $\geq 80\%$
  - $n = 30$ (80.7% power)
- For any given n RSABE is more powerful than the ABEL variants
  - Always requires a smaller sample size for target power
  - Hypothetical situation
    - The *same* study is submitted to different agencies
    - Might *pass* for one and *fail* for another



RSABE – Partial replicate Design

- If in RSABE $s_{wR} < 0.294$, the study has to be assessed for ABE [1,2]
  - The recommended [1,2,3] mixed-effects model is over-specified, since T is not replicated
  - The software *might* fail to converge – terrible consequence: *no* result at all
- Remedy
  - No issues in full replicate designs (the model is appropriate)
  - If poor bioanalytical method (large blood volume required) or concerns about dropouts, opt for one of the three period full replicate designs (TRT | RTR or TRR | RTT)

1. FDA, OGD. *Draft Guidance on Progesterone*. Recommended Apr 2010; Revised Feb 2011.
2. FDA, CDER. *Draft Guidance for Industry. Bioequivalence Studies with Pharmacokinetic Endpoints for Drugs Submitted Under an ANDA*. August 2021.
3. FDA, CDER. *Guidance for Industry. Statistical Approaches to Establishing Bioequivalence*. January 2001.

RSABE – Partial replicate Design

- The FDA's model decomposes the covariance matrix as $V = ZGZ^t + R,$ where
  - Z is the design matrix for the random effects,
  - G contains the between-subject variance components, and
  - R contains the within-subject variance components
- Contary to fully replicated designs, in the partial replicate design the between- and within-subject variances s^2_{bT}, s^2_{wT} cannot be uniquely estimated, only the *total* variance $s^2_{\mathrm{T}} = s^2_{\mathrm{bT}} + s^2_{\mathrm{wT}}$
  - In the 'best case' the estimated s^2_{bT}, s^2_{wT} is plain nonsense and differs between software…
  - In the worst case the optimizer fails to converge

Recent Development (2) Two-Stage Designs

- Regulatory guidelines
  - Partly ambiguous and/or even flawed statements…

    - 2010 European Economic Area
    - 2011 Australia (EMA GL adopted)
    - 2012 Canada
    - 2013 USA,[1,2] Russian Federation
    - 2015 New Zealand (Australian GL adopted)
    - 2016 Eurasian Economic Union, Gulf Cooperation Council
    - 2017 WHO, Egypt
    - 2019 Brazil (under public consultation)

1. Davit B, Braddy AC, Conner DP, Yu LX. *International Guidelines for Bioequivalence of Systemically Available Orally Administered Generic Drug Products: A Survey of Similarities and Differences*. AAPS J. 2013; 15(4): 974–90. doi:10.1208/s12248-013-9499-x.
2. Lee J, Feng K, Xu M, Gong X, Sun W, Kim J, Zhang Z, Wang M, Fang L, Zhao L. *Applications of Adaptive Designs in Generic Drug Development*. Clin Pharmacol Ther. 2021; 110(1): 32–5. doi:10.1002/cpt.2050.

Recent Development (2) Two-Stage Designs

- Alternative to pilot / pivotal studies
  - If no reliable information about the *CV* and – in some methods – the T/R-ratio is available
    - Due to multiplicity ($\rightarrow$ inflated Type I Error), the level of the test has to be adjusted

- In a first stage a fixed number of subjects (n_1) is treated and an interim analysis performed
  - If BE is demonstrated, the study stops (pass)
  - If BE is unlikely even with a large additional number of subjects, the study stops for futility (fail)
  - If BE is achievable with a given power, the sample size is re-estimated (N) and a second stage initiated with $n_2 = N - n_1$ subjects
- In the final analysis BE is assessed (pass | fail)

Recent Development (2) Two-Stage Designs

- Futility criteria for stopping in the IA can be pre-specified
  - If n_2 higher than economically reasonable and/or
  - if the PE or its CI is outside certain limits, making success in the final analysis unlikely
- Minimum stage 2 sample size ($n_{2,\min}$) can be pre-specified
  - Some methods require $n_2 = 2$ or $n_2 = 4$; a larger one protects against loss in power due to dropouts
- Maximum total sample size ($N_{\max}$) can be pre-specified
  - If $N > N_{\max}$, the second stage is performed in $N_{\max} - n_1$ subjects
  - Although it compromises power, it might still be sufficient; can be explored in simulations
- Fully adaptive methods allow not only taking the *CV* but also the PE into account

Simulation-based TSDs – Problems (?)

- Most published Two-Stage Designs [1] are based on extensive simulations of
  - power (10^5 sim's with fixed θ_0) and
  - the Type I Error (10^6 sim's under H_0)
- α_{adj} is selected in such a way that the *TIE* is controlled
  - α_{adj} depends on the desired conditions (θ_0, target power)
  - Valid only for the assessed simulation grid (n_1 and *CV*) under specific conditions (θ_0, target power)
- European regulators expressed reservations against simulation-based methods
  - Exact method [2,3] (strict *TIE* control) *only* for 2×2×2 cross-over

1. Schütz H. *Two-stage designs in bioequivalence trials*. Eur J Clin Pharm. 2015; 71(3): 271–81. doi:10.1007/s00228-015-1806-2.
2. König F, Wolfsegger M, Jaki T, Schütz H, Wassmer G. *Adaptive two-stage bioequivalence trials with early stopping and sample size re-estimation*. Trials. 2015; 16(Suppl 2): P218. doi:10.1186/1745-6215-16-S2-P218.
3. Maurer W, Jones B, Chen Y. *Controlling the type 1 error rate in two-stage sequential designs when testing for average bioequivalence*. Stat Med. 2018; 37(10): 1–21. doi:10.1002/sim.7614.

Summary

- BE is a regulatory requirement
  - Rate and extent of absorption are a regulatory ‘invention’
  - Not contained in the PK toolbox
- Traditional BE is well established and highly regulated
  - Not a scientific theory in the Popperian sense but an *ad hoc* solution to a pressing problem in the 1970s *
  - Limited space for new approaches
  - Apart from C_{max} and *AUC*, clinical relevance of more recent PK metrics (*e.g.*, partial *AUC*, C_{τ}) unclear
- SABE and Two-Stage Designs are complex
  - Inflation of the Type I Error in the former is not resolved yet (only suggestions published); approaches are not harmonized
  - In the latter complying with guidelines – which include ambiguous statements and/or are partly flawed – can be demanding

* However, we have decades of empiric evidence that the concept is sufficient in practice. Apart from occasional anecdotal reports, no problems are evident switching between bioequivalent drugs in terms of lack of efficacy or compromised safety.

Thank You!

Helmut Schütz

BEBAC

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