SANOFI

BE BAC

Replicate Designs

Purpose
Regulatory Differences
Pitfalls

Hierarchy of Designs

- Information which can be extracted

  Design

  Full replicate (*e.g.*, TRTR | RTRT or TRT | RTR)

  Partial replicate (TRR | RTR | RRT)

  2×2×2 crossover (TR | RT)

  Parallel (T | R)

  Variances which can be estimated

Parallel	total (pooled)
2×2×2 crossover	+ between, within subjects
Partial replicate	+ within subjects of R
Full replicate	+ within subjects of T

Replicate Designs

- Four period full
  - TRTR | RTRT
  - TRRT | RTTR
  - TTRR | RRTT
  - TRTR | RTRT | TRRT | RTTR [1]
  - TRRT | RTTR | TTRR | RRTT [1]
- Three period full
  - TRT | RTR [2]
  - TRR | RTT [3]
- Three period partial
  - TRR | RTR | RRT [4]
  - TRR | RTR [5]
- Two period full
  - TR | RT | TT | RR [6]

1. Confounded effects, not recommended
2. ≥12 eligible subjects in sequence RTR (EMA)
3. ≥12 eligible subjects in sequence TRR (EMA)
4. Should be avoided if ever possible; convergence issues in mixed-effects ABE (FDA)
5. Extra-reference design; biased in the presence of period effects, not recommended
6. Balaam's design; not recommended due to its poor power characteristics

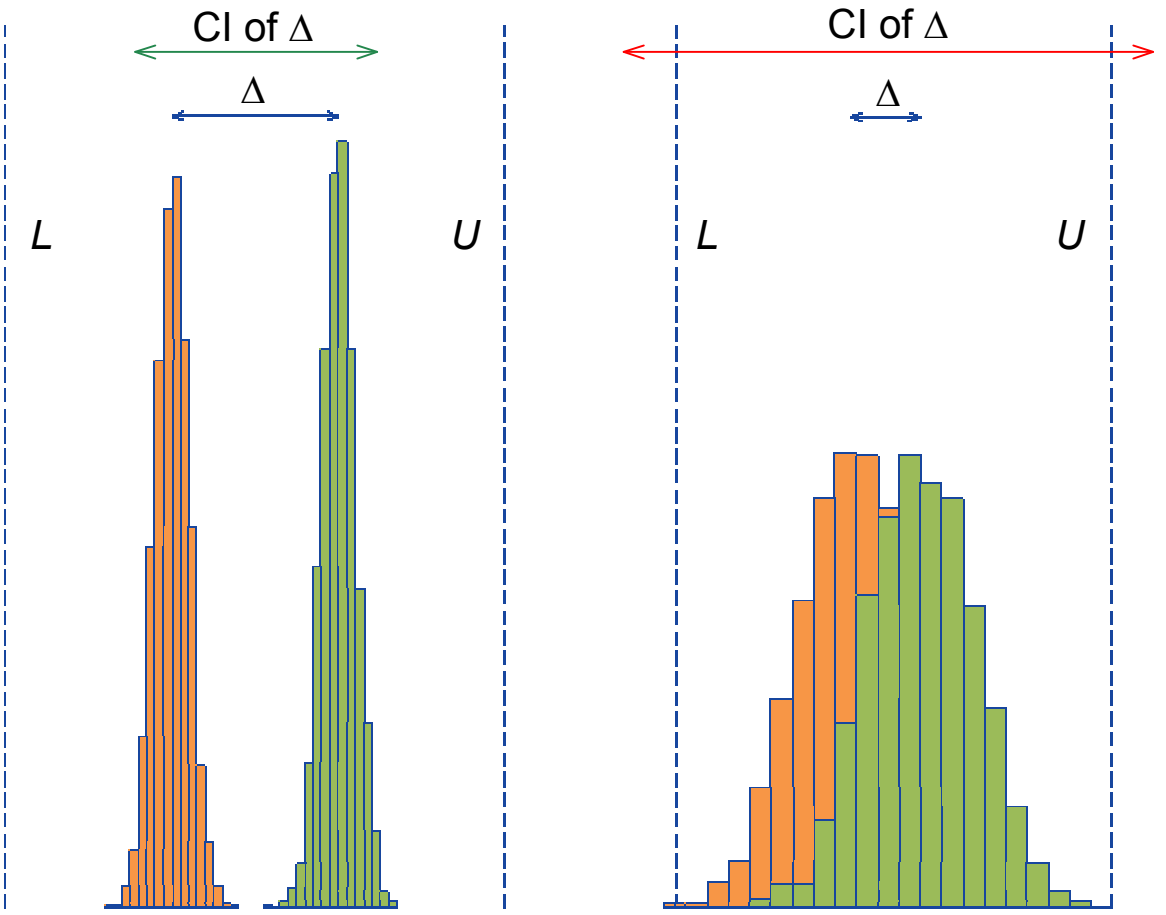
Highly Variable Drugs / Drug Products

- High (within-subject) variability can be
  - an intrinsic property of the drug itself (low absorption and/or inter-occasion clearance) and/or
  - attributed to the product's performance
    - Absorption, *i.e.*, rate of drug release and/or absorption window
    - Influence of excipients and/or food
      - on gastric motility and/or
      - on transporters
    - Physiology, *i.e*, enteric coated formulations and gastric emptying

HVD

HVDP

HVD(P)s



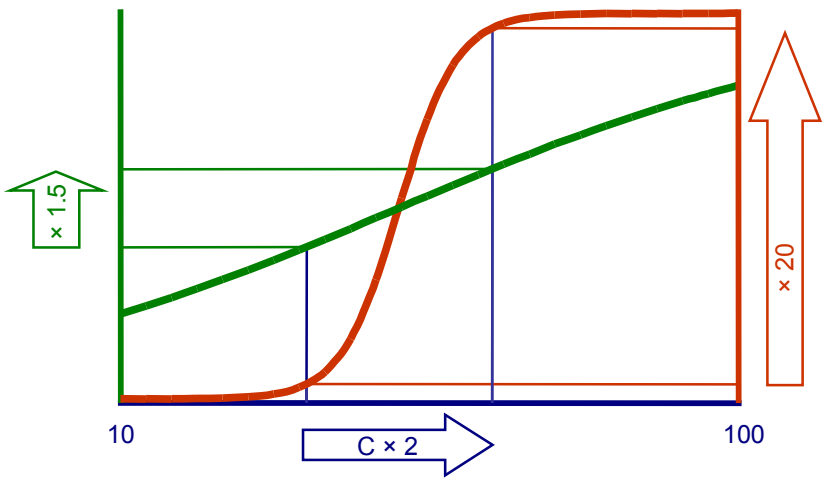
Counterintuitive concept of BE:

Two formulations with a large difference are declared bioequivalent if variabilities are small, but not BE – even if the difference is small – due to high variabilities

Tóthfalusi L, Endrényi L, García Arieta A. *Evaluation of Bioequivalence for Highly Variable Drugs with Scaled Average Bioequivalence*. Clin Pharmacokinet. 2009; 48(11): 725–43. doi:10.2165/11318040-000000000-00000.
Fig. 1 modified

HVD(P)s and NTIDs

- Clinically not relevant difference
  - Based on PK but extrapolated to similarity of safety and efficacy in the patient population (therapeutic equivalence)
    - Depends on the dose-response
      HVD(P)s – flat curve
      NTIDs – steep curve



  - Since HVD(P)s are considered to be safe and efficacious, in some jurisdictions a larger 'not clinically relevant' difference Δ is accepted

HVD(P)s

- It may be almost impossible to demonstrate BE with a reasonable sample size
  - Example: T/R ratio 0.90, *CV* 70%, power 80%, 2×2×2 design

```
library(PowerTOST)
sampleN.TOST(theta0 = 0.90, CV = 0.70, targetpower = 0.80, design = "2x2x2")

++++++++++++ Equivalence test - TOST ++++++++++++
            Sample size estimation
-----------------------------------------------
Study design:  2x2 crossover
log-transformed data (multiplicative model)
alpha = 0.05, target power = 0.8
BE margins = 0.8 ... 1.25
True ratio = 0.9,  CV = 0.7
Sample size (total)
  n     power
358   0.801175
```

HVD(P)s – Reference-scaling

- The BE limits are *scaled* (*i.e.*, expanded) based on the variability of the reference product
- Reference-scaling was introduced 2010 by the FDA and the EMA and many jurisdictions later
  - Requires a replicate design, where at least the reference product is administered twice *
    - Smaller sample sizes compared to the standard 2×2×2 design but outweighed by increased number of periods
    - Similar total number of individual treatments
  - Switching CV_{wR} 30% (estimate in the study used)
- Any replicate design can be evaluated for conventional (unscaled) Average Bioequivalence (ABE) as well

* Not necessarily to all subjects!

Reference-scaling (PK Metrics)

- Different ones acceptable in various jurisdictions
  - US FDA, China CDE — all PK metrics
  - Health Canada — AUC^{*}
  - WHO — C_{max} (*AUC* under certain conditions)
  - EMA and all others — C_{max}
  - EMA (MR products) — $C_{ss,min}$, $C_{ss,\tau}$, partial *AUC*s

* Only the T/R ratio of C_{max} has to lie within 80.0–125.0%. Hence, highly variable C_{max} is not an issue in Canada.

Models (in log-scale)

- ABE
  - Difference Δ of $\leq$20% is considered to be clinically not relevant
  - The limits [L, U] of the acceptance range are fixed at
    $\log(1-\Delta) = \log((1-\Delta)^{-1})$ or $L \sim -0.2231$ and $U \sim +0.2231$
  - The consumer risk α is fixed with 0.05 and BE is concluded if the 100(1 – 2α) confidence interval lies entirely within [L, U]
    $$-\theta_A \leq \mu_T - \mu_R \leq +\theta_A$$
- SABE
  - 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)
    $$-\theta_S \leq \frac{\mu_T - \mu_R}{\sigma_{wR}} \leq +\theta_S$$

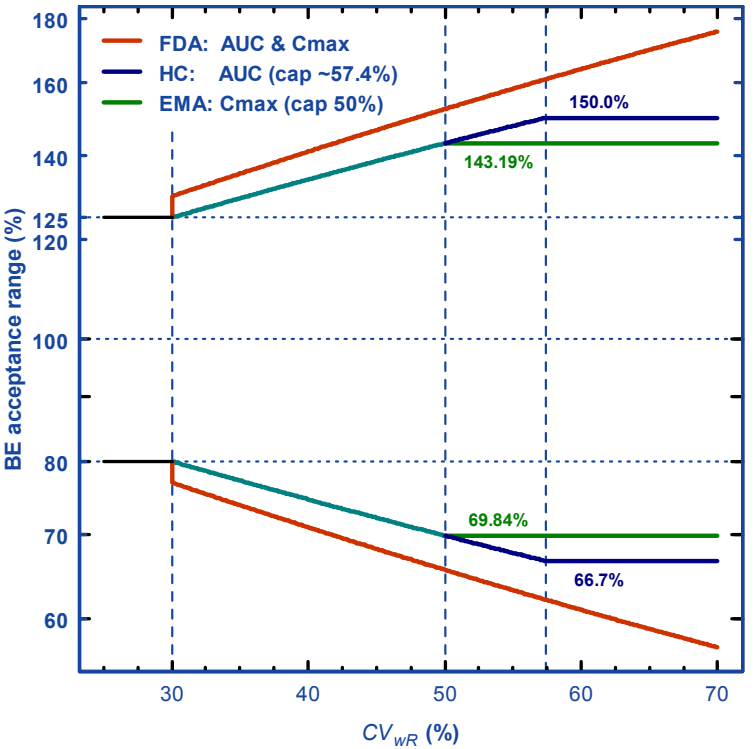
* Termed *k* in some guidelines.

Reference-scaling (Regulatory Approaches)

- Bioequivalence limits are derived from σ_0 and σ_{wR}

$$\theta_s = \frac{\log(1.25)}{\sigma_0},\ \left[L,U\right] = e^{\mp\theta_s \cdot \sigma_{wR}}$$

  - US FDA, China CDE
    - Scaling σ_{wR} 0.25 (θ_s 0.893) but applicable if CV_{wR} ≥30%
    - Discontinuity at CV_{wR} 30%
  - EMA and most others
    - Scaling σ_{wR} ~0.2936 (θ_s 0.760)
    - Upper cap at CV_{wR} 50%
  - Health Canada
    - Same scaling but upper cap at CV_{wR} ~57.4%
  - All: PE within 80.00–125.00%



Reference-scaling (Regulatory Approaches)

- Scaled limits based on CV_{wR}

FDA, CDE	
CV_{wR} (%)	$L - U$ (%)
<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
55	63.20 – 158.23
60	60.96 – 164.04
70	56.91 – 175.71
80	53.38 – 187.35
90.98	50.00 – 200.00

EMA, WHO, ANVISA	
CV_{wR} (%)	$L - U$ (%)
≤30	80.00 – 125.00
35	77.23 – 129.48
40	74.62 – 134.02
45	72.15 – 138.59
≥50	69.84 – 143.19

Health Canada	
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
55	67.66 – 147.80
≥57.4	66.67 – 150.00

- Due to the PE restriction in all jurisdictions power decreases beyond CV_{wR} ~50%

Reference-scaling (Regulatory Approaches)

- Example: T/R ratio 0.90, CV_{wR} 70%, power 80%

design	method	approach	regulator	Δ [1] (%)	CV_{wR} cap (%)	*L* – *U* (%)	*n*
2×2×2	ABE	un-scaled	any	20.00	–	80.00 – 125.00	358
2×2×4				20.00	–	80.00 – 125.00	180
2×2×4			GCC, EMEA [2]	25.00	–	75.00 – 133.33	76
2×2×4	ABEL	scaled	EMA, …	30.16	50.0	69.84 – 143.19	40
2×2×4			HC	33.33	57.4	66.67 – 150.00	32
2×2×4	RSABE		FDA, CDE	43.09	–	56.91 – 175.71	26

- Substantially smaller sample sizes compared to ABE
  - Smaller for Health Canada than for EMA due to higher CV_{wR} cap
  - Even smaller for the FDA than for the others due to unlimited scaling

1. Difference considered clinically not relevant
2. Q&A document (2006) to Note for Guidance (2001)

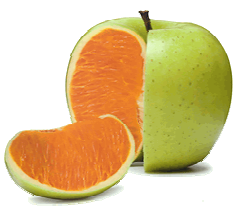
Reference-scaling (the EMA's Approach)

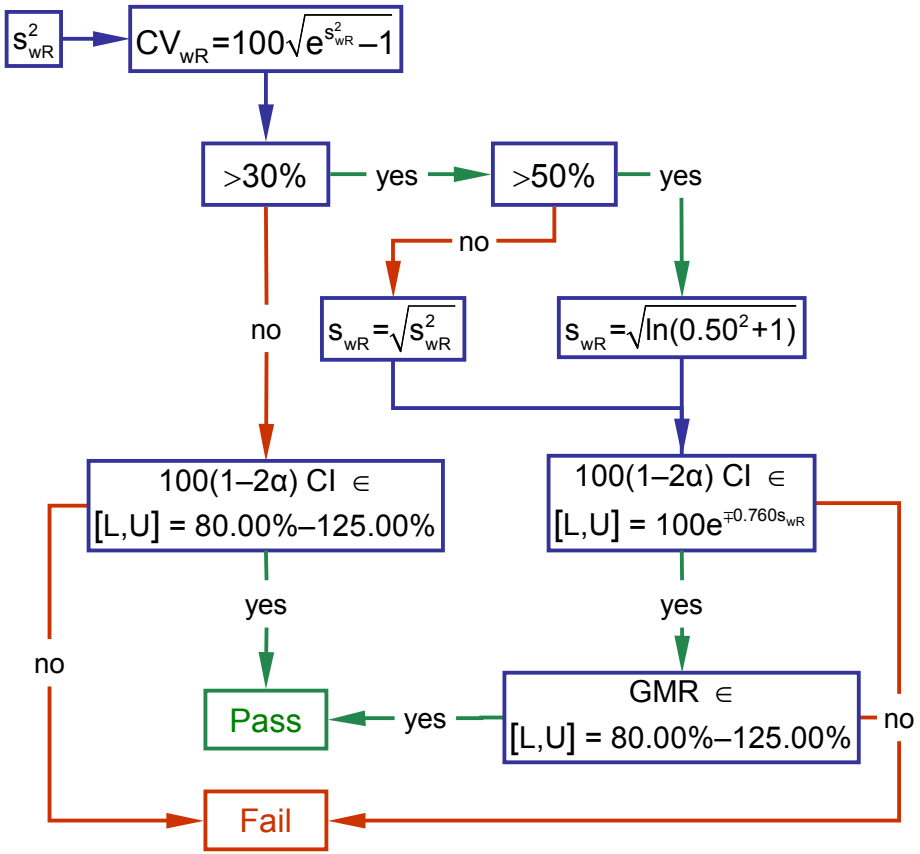
- Average Bioequivalence with Expanding Limits – ABEL (crippled from Endrényi and Tóthfalusi *)
  - Justification that the widened acceptance range is clinically not relevant
  - Assumes identical variances of T and R like in a 2×2×2 design
  - Fixed effects model according to the Q&A-document preferred
    - Mixed-effects model (allowing for unequival variances) is 'not compatible with CHMP guideline'…
  - Scaling capped at CV_{wR} 50%
  - PE within 80.00 – 125.00%
  - CV_{wR} >30% not caused by outliers (≠ ANVISA!)
  - At least 12 eligible subjects in sequence RTR of the 3-period full replicate design (Q&A document)

* Endrényi L, Tóthfalusi L. *Regulatory Conditions for the Determination of Bioequivalence of Highly Variable Drugs*. J Pharm Pharmaceut Sci. 2009; 12(1): 138–49. Open access.

Reference-scaling (the EMA's Approach)

- Decision Scheme
  - The Null Hypothesis is *generated* in the face of the data
  - Acceptance limits become *random* variables them-selves
  - Type I Error (consu-mer risk) might be inflated





Type I Error

- Falsely concluding BE at the expanded limits
  - Due to the decision scheme direct calculation of the TIE is impossible → extensive simulations required (10^6 BE studies mandatory)
  - Inflation of the TIE suspected [1–4]
  - Inflation of the TIE confirmed for the FDA's RSABE [5–8]

1. Chow S-C, Shao J, Wang H. *Individual bioequivalence testing under 2 × 3 designs*. Stat Med. 2002; 21(5): 629–48. doi:10.1002/sim.1056.
2. Willavize SA, Morgenthien EA. *Comparison of models for average bioequivalence in replicated crossover designs*. Pharm Stat. 2006; 5(3): 201–11. doi:10.1002/pst.212.
3. Chow S-C, J-p L. *Design and analysis of bioavailability and bioequivalence studies*. Boca Raton: Chapman & Hall/CRC Press; 2009. p. 596–8.
4. Patterson SD, Jones B. *Viewpoint: observations on scaled average bioequivalence*. Pharm Stat. 2012;11:1–7. doi:10.1002/pst.498.
5. Endrényi L, Tóthfalusi L. *Regulatory Conditions for the Determination of Bioequivalence of Highly Variable Drugs*. J Pharm Pharmaceut Sci. 2009; 12(1): 138–49. Open access.
6. Labes D. *RSABE/ABEL: 'alpha' of scaled ABE?* In: *Bioequivalence and Bioavailability Forum*. Vienna: BEBAC; 15 March 2013. Open access.
7. Muñoz J, Alcaide D, Ocaña J. *Consumer's risk in the EMA and FDA regulatory approaches for bioequivalence in highly variable drugs*. Stat Med. 2016;35(12):1933–43. doi:10.1002/sim.6834.
8. Deng Y, Zhou X-H. *Methods to control the empirical type I error rate in average bioequivalence tests for highly variable drugs*. Stat Meth Med Res. 2019. doi:10.1177/0962280219871589.

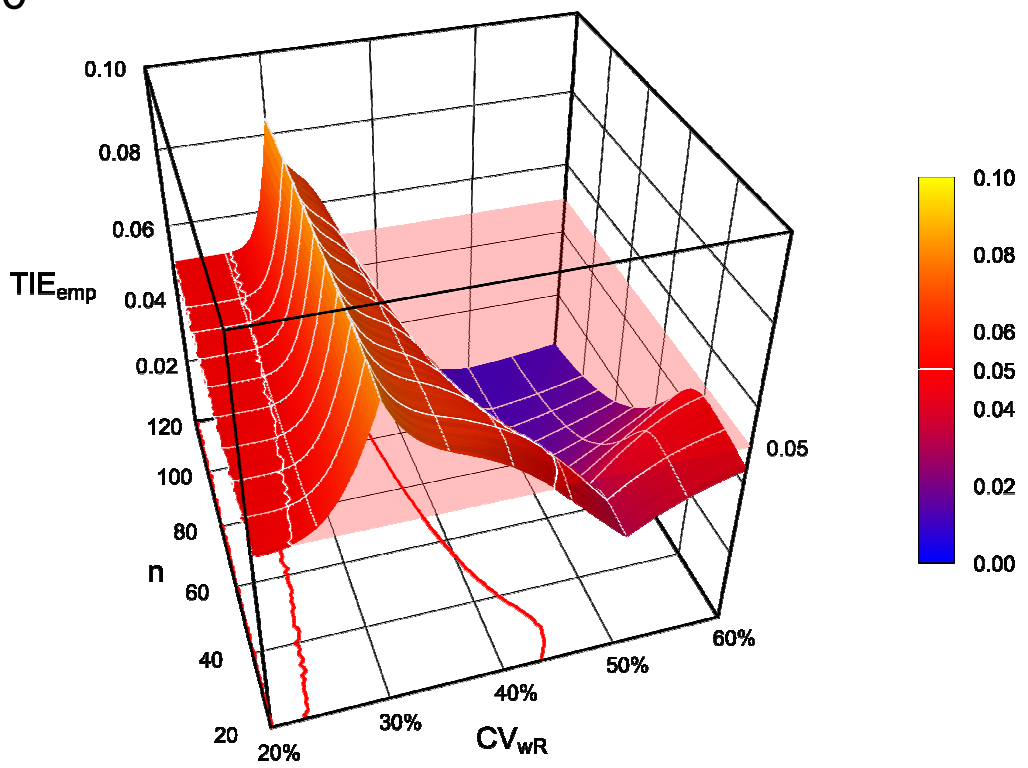
Type I Error

- Falsely concluding BE at the expanded limits
  - Inflation of the TIE confirmed for the EMA's ABEL [9–17]
- Agencies worry about a potential TIE of 0.0501 in TSDs but not about a much higher one in reference-scaling?

9. Endrényi L, Tóthfalusi L. *Regulatory Conditions for the Determination of Bioequivalence of Highly Variable Drugs*. J Pharm Pharmaceut Sci. 2009; 12(1): 138–49. Open access.
10. Labes D. *RSABE/ABEL: 'alpha' of scaled ABE?* In: *Bioequivalence and Bioavailability Forum*. Vienna: BEBAC; 15 March 2013. Open access.
11. Wonnemann M, Frömke C, Koch A. *Inflation of the Type I Error: Investigations on Regulatory Recommendations for Bioequivalence of Highly Variable Drugs*. Pharm Res. 2015; 32(1): 135–43. doi:10.1007/s11095-014-1450-z.
12. Muñoz J, Alcaide D, Ocaña J. *Consumer's risk in the EMA and FDA regulatory approaches for bioequivalence in highly variable drugs*. Stat Med. 2016;35(12):1933–43. doi:10.1002/sim.6834.
13. 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.
14. Tóthfalusi L, Endrényi L. *An Exact Procedure for the Evaluation of Reference-Scaled Average Bioequivalence*. AAPS J. 2016: 18(2); 476–89. doi:10.1208/s12248-016-9873-6.
15. Tóthfalusi L, Endrényi L. *Algorithms for Evaluating Reference Scaled Average Bioequivalence: Power, Bias, and Consumer Risk*. Stat Med. 2017: 36(27); 4378–90. doi:10.1002/sim.7440.
16. Molins E, Cobo E, Ocaña J. *Two-Stage Designs Versus European Scaled Average Designs in Bioequivalence Studies for Highly Variable Drugs: Which to Choose?* Stat Med. 2017: 36(30); 4777–88. doi:10.1002/sim.7452.
17. Deng Y, Zhou X-H. *Methods to control the empirical type I error rate in average bioequivalence tests for highly variable drugs*. Stat Meth Med Res. 2019. doi:10.1177/0962280219871589.

Type I Error

- Example (ABEL)
  - TRTR | RTRT
  - Sample sizes 20–120
  - CV_{wR} 20–60%
  - TIE_{max} 0.0838
  - Relative increase of the consumer risk 68%



Type I Error

- Explanation
  - SABE is stated in population *parameters* …

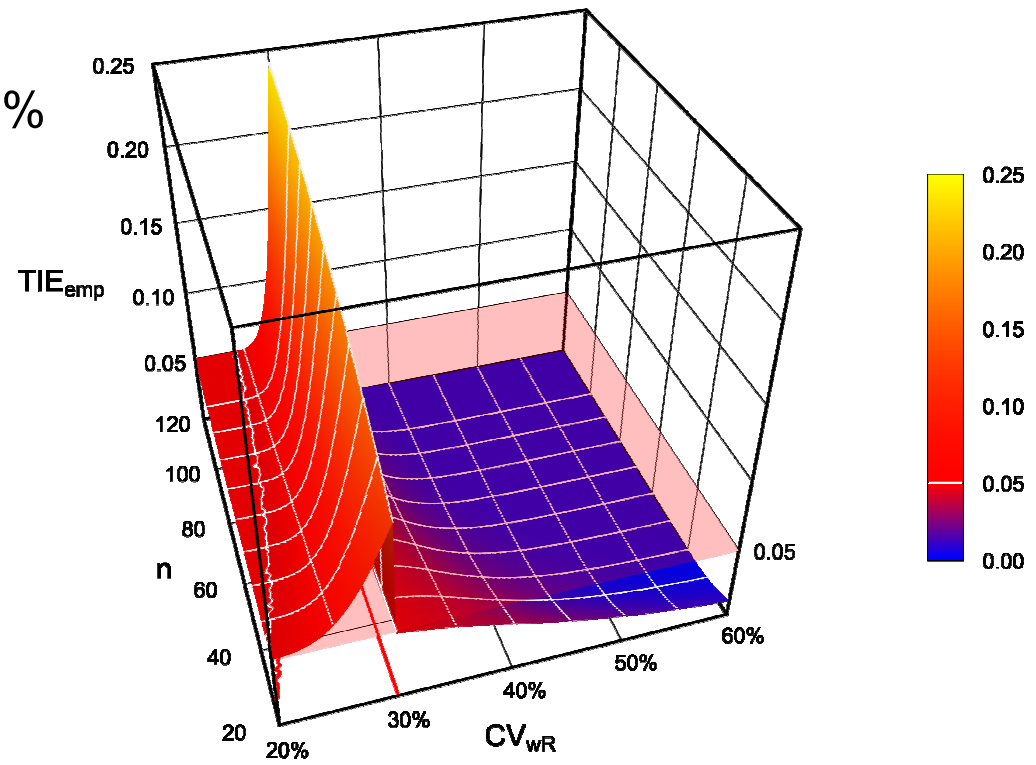
    $$-\theta_S \leq \frac{\mu_T - \mu_R}{\sigma_{wR}} \leq +\theta_S$$

    … which are *unknown*
  - Only their *estimates* (T/R ratio, 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% – in any direction – the chances of a wrong decision decrease and hence, the TIE
    - At high *CV*s (~43%) both the scaling cap and the PE-restriction help to maintain the TIE ≤0.05).

Type I Error

- Is the FDA's approach better?
  - At CV_{wR} >30%
    very conservative
  - Massive inflation of
    the TIE at CV_{wR} ≤30%
    - At CV_{wR} 30% de-
      pends strongly
      on *n*

24	0.1335
36	0.1536
48	0.1708
64	0.1916
120	0.2418



Preserving the Consumer Risk

- Utopia
  - Agencies collect CV_{wR} from submitted studies, pool them, adjust for designs / degrees of freedom
  - A *fixed* (expanded) acceptance range is published in product-specific guidances
    - Evaluation by ABE; TIE no issue any more *
    - All products follow the same rules and not different ones specific for each study – supporting switchability
- Half-baked
  - *Hope* that *e.g.*, Bonferroni preserves the consumer risk; still apply ABEL but with a 95% CI (α 0.025)
  - Drawback: Loss of power $\rightarrow$ increases required sample size

* In the 1990s for C_{max} a fixed acceptance range of 70–143% was applied in hundreds of European applications. Fixed limits of 75–133% are acceptable for members of the Gulf Cooperation Council (though a replicate design is required and CV_{wR} >30% has to be demonstrated).

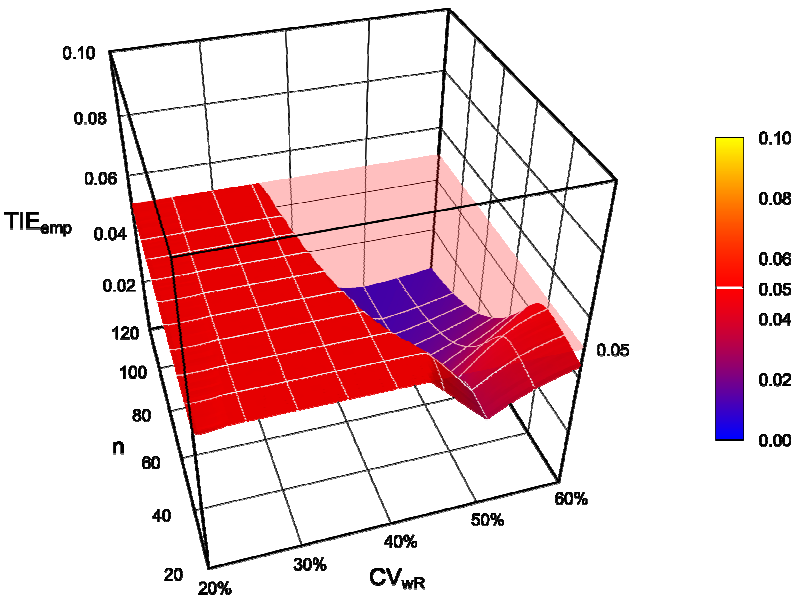
Preserving the Consumer Risk

- Proposals
  - Iteratively adjust α based on the study's conditions (design, sample size, CV_{wR}) – in such a way that the Type I Error is controlled [1]
  - Similar but iteratively adjust α for 'the worst possible' CV_{wR} 30% independent from the observed one [2]
  - 'Exact' procedure; regulatory acceptance unclear (modified model) [3,4]
  - All implemented in PowerTOST [5]

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. Molins E, Cobo E, Ocaña J. *Two-Stage Designs Versus European Scaled Average Designs in Bioequivalence Studies for Highly Variable Drugs: Which to Choose?* Stat Med. 2017: 36(30); 4777–88. doi:10.1002/sim.7452.
3. Tóthfalusi L, Endrényi L. *An Exact Procedure for the Evaluation of Reference-Scaled Average Bioequivalence*. AAPS J. 2016: 18(2); 476–89. doi:10.1208/s12248-016-9873-6.
4. Tóthfalusi L, Endrényi L. *Algorithms for evaluating reference scaled average bioequivalence: power, bias, and consumer risk*. Stat Med. 2017: 36(27); 4378–90. doi:10.1002/sim.7440.
5. Labes D, Schütz H, Lang B. *PowerTOST: Power and Sample Size for (Bio)Equivalence Studies*. 2019; R package version 1.4-9. https://cran.r-project.org/package=PowerTOST.

Iteratively adjusted α

- Example of slide 17
  - Algorithm
    - Estimate the empiric TIE for the nominal α 0.05
      - If the TIE ≤0.05 stop and use the 90% CI
      - Otherwise, adjust α down-wards until the TIE ≤0.05
      - At CV_{wR} 30% (dependent on the sample size) α_{adj} is 0.0273 – 0.0300
      - Evaluate the study with a 94.00 – 94.54% CI



Iteratively adjusted α

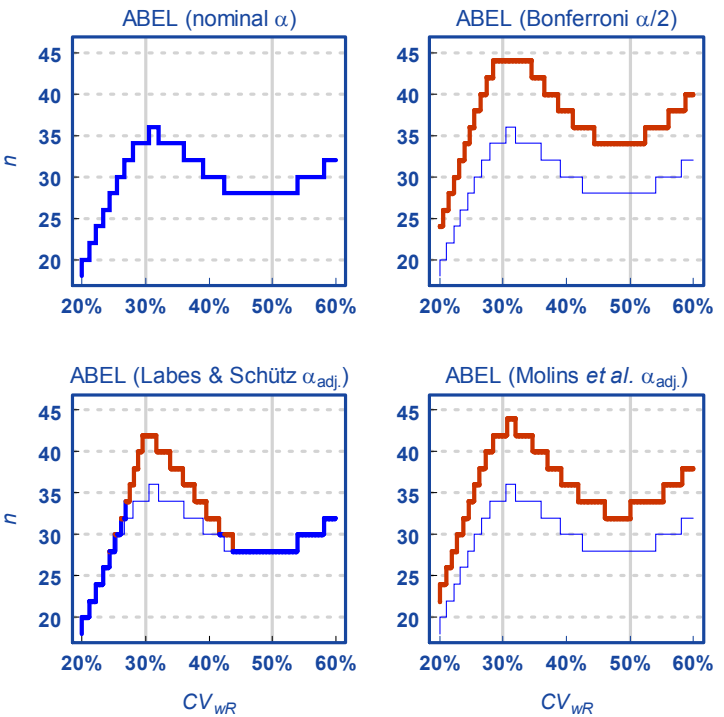
- Example: CV_{wR} 35%, 2×2×4 design → $n = 34$

```
library(PowerTOST)
CV                  <- 0.35
d                   <- "2x2x4" # using defaults: theta0 = 0.90, target power = 0.80
n                   <- sampleN.scABEL(CV = CV, design = d, details = FALSE,
                                      print = FALSE)[["Sample size"]]
expl                <- data.frame(Method = rep(NA, 4), alpha = 0.05, CI = NA,
                                  TIE = NA, power = NA)
tmp                 <- scABEL.ad(CV = CV, design = d, n = n, print = FALSE)
expl[1, 4:5]        <- tmp[c(12, 14)]
expl[2, c(2, 4:5)] <- tmp[15:17]
tmp                 <- scABEL.ad(CV = 0.30, design = d, n = n, print = FALSE)
expl[3, c(2, 4)]    <- tmp[15:16]
expl[3, 5]          <- power.scABEL(alpha = expl$alpha[3], CV = CV, design = d, n = n)
expl[4, 4]          <- power.RSABE2L.sds(CV = CV, design = d, n = n, nsims = 1e6,
                                         theta0 = (scABEL(CV = CV)[["upper"]]))
expl[4, 5]          <- power.RSABE2L.sds(CV = CV, design = d, n = n)
expl$CI             <- round(100 * (1 - 2 * expl$alpha), 2)
expl$power          <- round(100 * expl$power, 2)
expl[, 2:5]         <- signif(expl[, 2:5], 4)
expl[, 1]           <- c("EMA", "Labes and Schütz", "Molins et al.",
                         "Tóthfalusi and Endrényi")
print(expl, row.names = FALSE)
```

Method	alpha	CI	TIE	power
EMA (original)	0.05000	90.00	0.06557	81.18
Labes and Schütz	0.03630	92.74	0.05000	77.28
Molins et al.	0.02857	94.29	0.05000	74.05
Tóthfalusi and Endrényi	0.05000	90.00	0.04818	78.16

Iteratively adjusted α

- It comes with a price – loss in power
  - Can be counteracted by increasing the sample size based on the assumed CV_{wR} to maintain the target power
- Example
  - TRTR | RTRT
    T/R ratio 0.90, CV_{wR} 20–60%, power 80%
    - Bonferroni requires substantial increase
    - α_{adj} (Labes and Schütz) requires increase only in the area of inflated TIE
    - α_{adj} (Molins *et al.*) requires increase independent from CV_{wR}, although less than Bonferroni



Data Sets

- Different ones are required → state in the SAP
  1. For the calculation of the CI data of subjects who received both T and R at least once
  2. For the estimation of CV_{wR} data of subjects who received R twice
  3. For the estimation of CV_{wT} data of subjects who received T twice
  - Example: 16 subjects enrolled, dropouts •

#	seq	#	seq	#	seq	#	seq
1	RTR •	5	RTRT	9	TRTR	13	RTRT
2	RTRT	6	TR • •	10	TRTR	14	TRT •
3	RTRT	7	RTRT	11	RTRT	15	TRTR
4	TRTR	8	R • • •	12	TRTR	16	TRTR

data set	purpose	excluded	*n*
1	90% CI	8	15
2	CV_{wR}	6, 8, 14	13
3	CV_{wT}	1, 6, 8	13

  - Software does not necessarily handle that * automatically; check!

* Schütz H, Tomashevskiy M, Labes D. *replicateBE: Average Bioequivalence with Expanding Limits (ABEL)*. 2020; R package version 1.4.13. https://cran.r-project.org/package=replicateBE.

Misconceptions, Problems

- TRR|RTR|RRT is not the only * replicate design with three periods
- If you want only three periods (limited blood volume, higher number of droputs in a four period design expected), opt for one of the three period *full* replicates instead (slide 3)
  - Prevents problems with the FDA's implementation of ABE (if reference-scaling is not acceptable)
    - Its mixed-effects model is over-specified (T only administered once)
    - Problems in any software – if the software fails to converge, study done, money spent, no results …
  - Allows estimation of CV_{wT}

* Bad in the ANVISA's № 760.20, Section IV, Article 77: *A partially replicated (three-period) or fully replicated (four-period) crossover design should be used, and the comparator drug should be administered twice to each research participant.*

Thank You!

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