medicines for europe
better access. better health.

Statistical challenges and opportunities in ICH M13C

Helmut Schütz
Center for Medical Data Science
of the Medical University of Vienna
BEBAC, Vienna

3RD BIOEQUIVALENCE CONFERENCE

HILTON AMSTERDAM AIRPORT SCHIPHOL

26 FEBRUARY 2025

Scaled Average Bioequivalence (SABE) for HVD(P)s

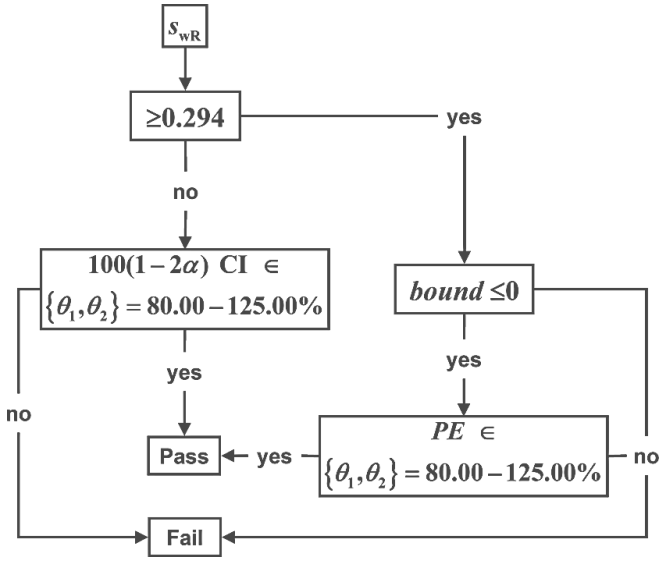
- HVD(P)s show large within-subject variability
  - Safe and efficacious *despite* their high variability
  - Large sample sizes required for Average BE based on the clinical relevant difference Δ = 20% (80 – 125%)
- Δ > 20% discussed at BioInternational conferences (1989 – 2005) and meetings of the FDA Advisory Committee for Pharmaceutical Science (1997 – 2006) if $CV_{wR} \geq 30\%$
  - *Fixed* limits based on Δ = 25% (75 – 133%) or Δ = 30% (70 – 143%)
  - *Scaled* limits based on the observed CV_{wR} in a (at least reference-) replicated design study

Implemented methods

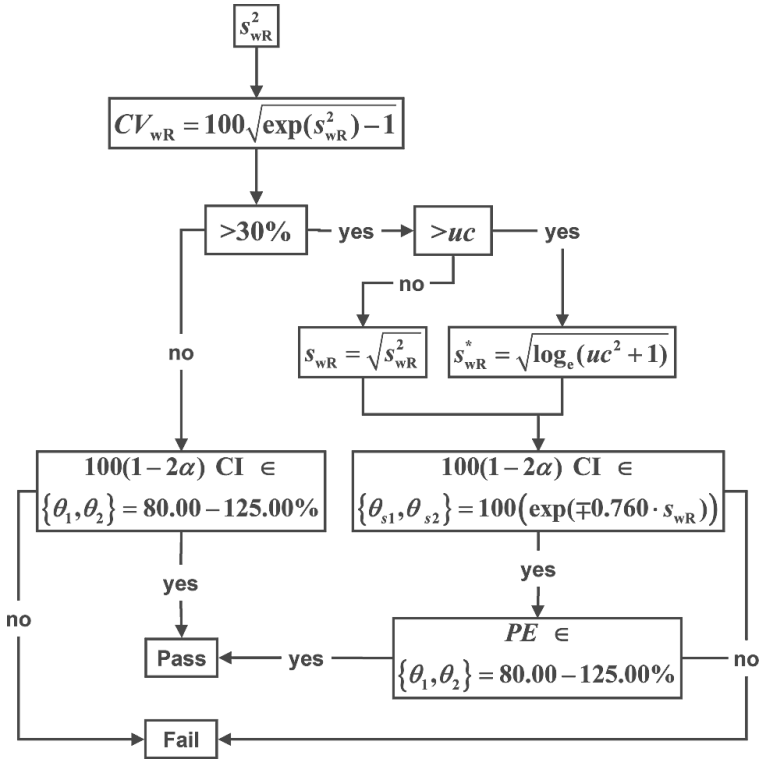
- Reference-Scaled Average Bioequivalence (RSABE)
  - US FDA and China CDE
  - If $s_{wR} \geq 0.294$ ($CV_{wR} \geq 30.05\%$); otherwise, by ABE
- Average Bioequivalence with Expanding Limits (ABEL)
  - Recommended in all [*sic*] other jurisdictions accepting SABE
  - If $CV_{wR} > 30\%$; otherwise, by ABE
  - Upper cap of expansion
    - 50% → max. 69.84 – 143.19%
    - ≈57.4% → max. 66.7 – 150.0% (Health Canada)
- In both: Point estimate constraint (80.00 – 125.00%)

The Type I Error can be inflated (increased patient's risk)

- Implemented methods of SABE are frameworks
  - BE limits are random variables dependent on the reference's variance
  - Δ unknown beforehand
  - Drugs *will* be misclassified if <u>observed</u> $s^2_{wR} \neq$ <u>true</u> σ^2_{wR}

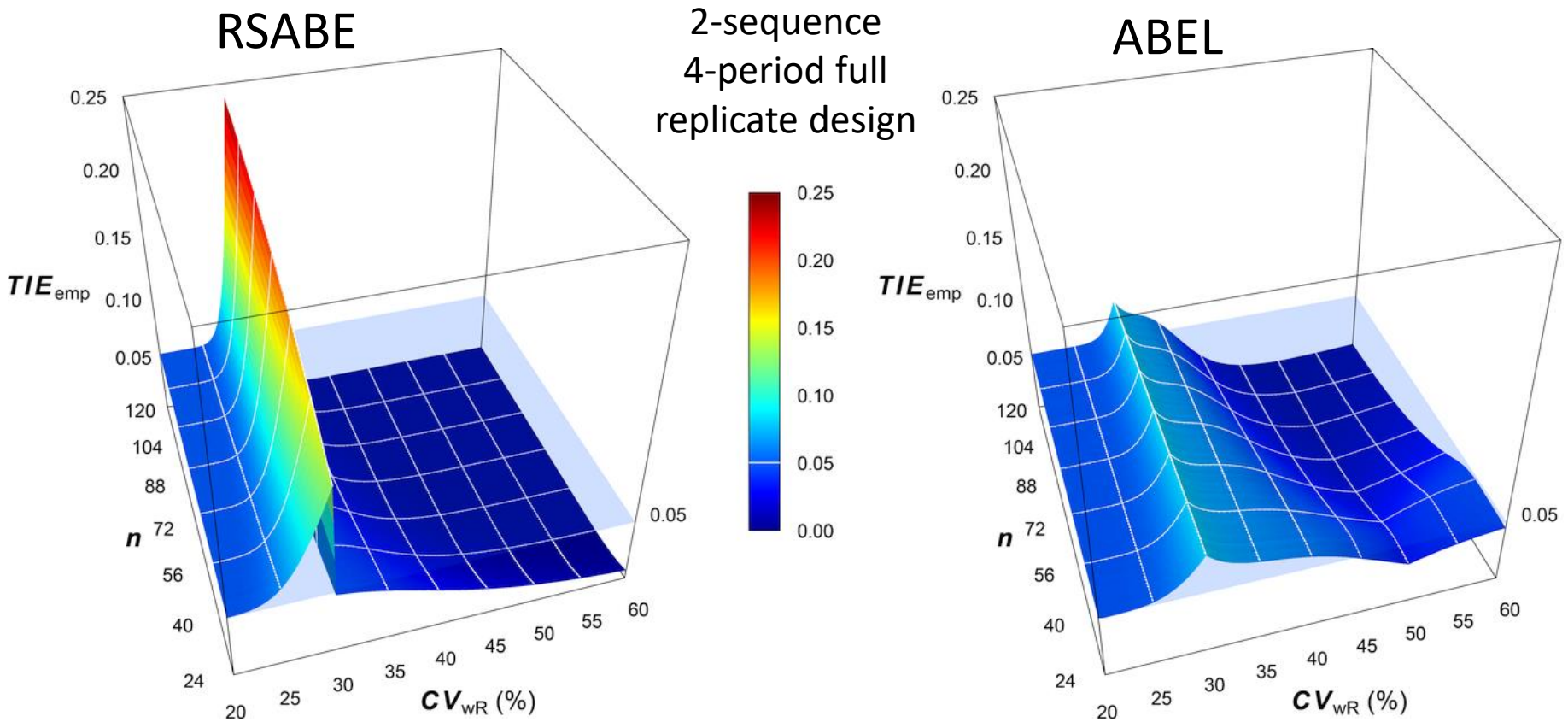


RSABE



ABEL

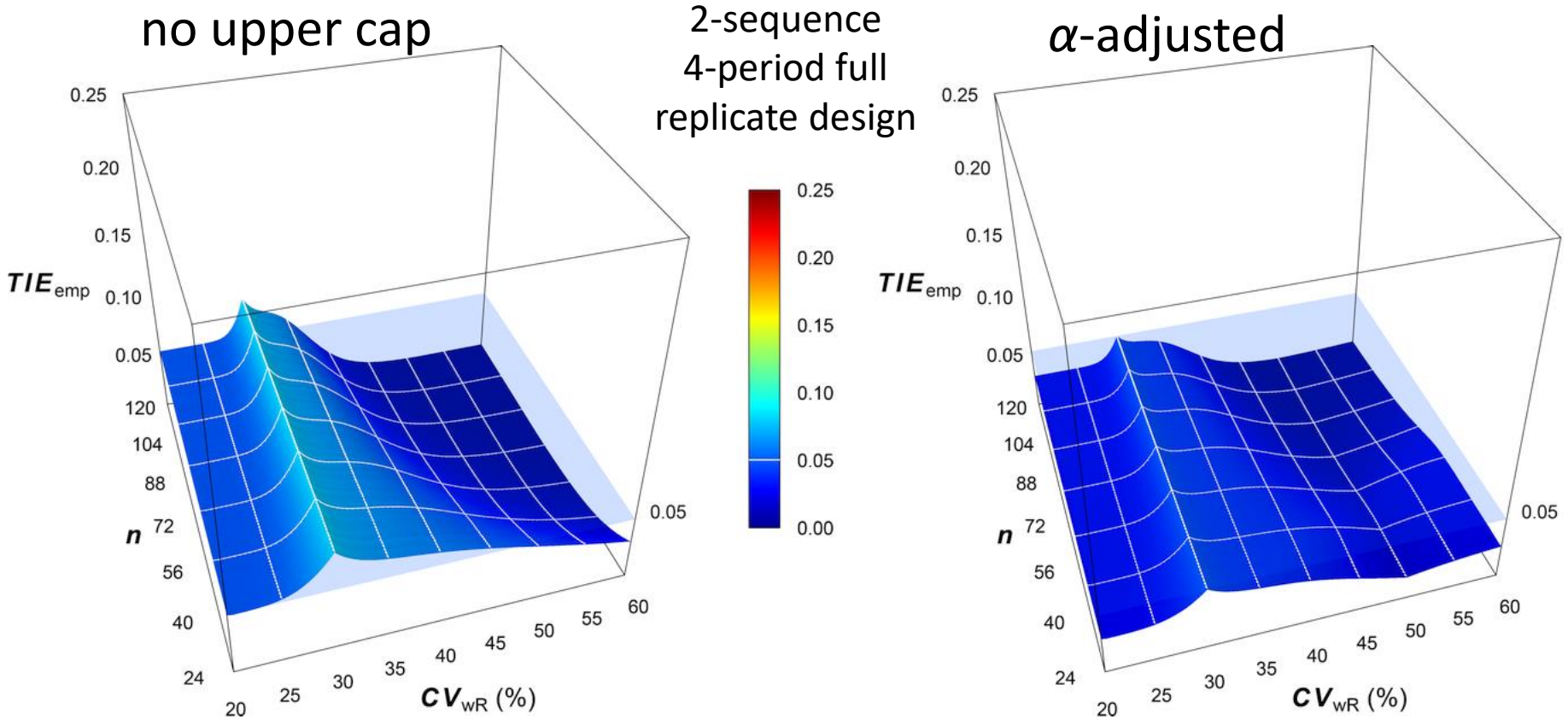
Empiric Type I Error in the implemented methods



TIE_{emp} at CV_{wR} 30%; *n* 24: 0.1335, *n* 120: 0.2418

TIE_{emp} at CV_{wR} 30%; *n* 24: 0.0804, *n* 120: 0.0838

Empiric Type I Error (ABEL, modifications)



TIE_{emp} at CV_{wR} 30%; n 24: 0.0804, n 120: 0.0838

TIE_{emp} at CV_{wR} 30%; n 24 to 120: 0.05

Conclusions (RSABE)

- The implemented method is beyond repair
  - Its maximum Type I Error is much larger than by ABEL
  - Assessing the TIE via the ‘desired consumer risk model’ (Davit *et al*. 2012) is a mere magician’s trick; I do not agree that it »maintains an acceptable Type I Error rate« (6.63% with 24 subjects in a full replicate design)
  - The decision of equivalence (*i.e.*, whether the upper bound of the linearized criterion is non-negative) is incomprehensible for physicians
  - If $s_{wR} < 0.294$ in a partial replicate design, the model is over-specified and may not converge

Conclusions (ABEL)

- The upper cap of expansion lacks a scientific rationale
  - 50% introduced in most jurisdictions due to reservations of <u>one</u> European member state
  - HC's ≈57.4% likely to give a 'nice' max. expansion of 67.7 – 150.0%
  - No issues with the Type I Error due to the inherent conservatism of the TOST procedure and the PE-constraint;
    lower sample sizes for large CV_{wR}
- α-adjusted methods
  - Control the Type I Error
  - Compromise power → large sample sizes required if true CV_{wR} > 50%

Suggestions for ICH M13C

- ABEL with modifications should be considered
  - Should be acceptable for all PK metrics
  - The upper cap should be removed
  - Biased-corrected Howe-LO and iteratively adjusted α are promising – control the Type I Error with less loss in power than other methods
- Heretical utopia (utopian heresy?)
  - Full replicate studies mandatory for the originator; alternatively agencies could collect and exchange CV_{wR} of studies → PSGs
  - *Fixed* limits ($\Delta > 20\%$): replicate designs no more needed and the Type I Error is *always* controlled

Group-Sequential (GS), Adaptive Two-Stage (TS) Designs

- In the conventional approach of pilot / pivotal studies
  - Only part of the information (T/R ratio, *CV*) of the former is used to design the latter
  - The individual data of the pilot – or a failed pivotal – are not used (they are only supportive information in the application)
- Since the T/R-ratio and the *CV* are estimates or assumptions, even a properly powered pivotal study of a bioequivalent product may fail in a fixed-sample design (probability = 1 – power)
- GSDs and TSDs allow decisions in an ‘interim analysis’

Group-Sequential Designs

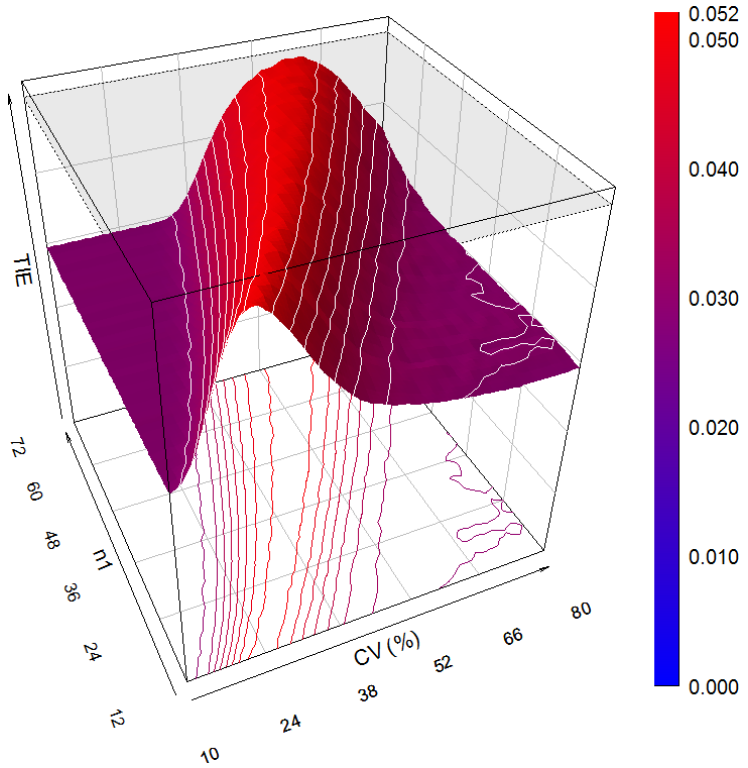
- The total sample size *n* is estimated as in a fixed-sample design
  - Analyses (interim and final) are performed with adjusted alphas (< 0.05), which must not be the same
- An interim analysis is performed at *n*/2
  - If the study passes BE already → stop for success, otherwise 2nd group is administered
  - (If the result looks promising → 2nd group is administered, otherwise stop for futility)
- Even if the study fails in the interim only by a slight margin, still *n*/2 have to be administered

Adaptive Two-Stage Designs

- If a pivotal study of a bioequivalent product fails in a fixed-sample design (probability = 1 – power) and is repeated in a larger sample size, the data of the first is not used
- In the interim analysis
  - Whether or not an adjusted α has to be used depends on the method
  - If the study passes BE already → stop for success, otherwise the total sample size is re-estimated and the 2nd group dosed
  - (If the result looks promising → 2nd group is dosed, otherwise stop for futility)

Operating Characteristics (simulation-based TSD)

Type I Error



Maximum TIE at *CV* 24% and n_1 12: 0.04895

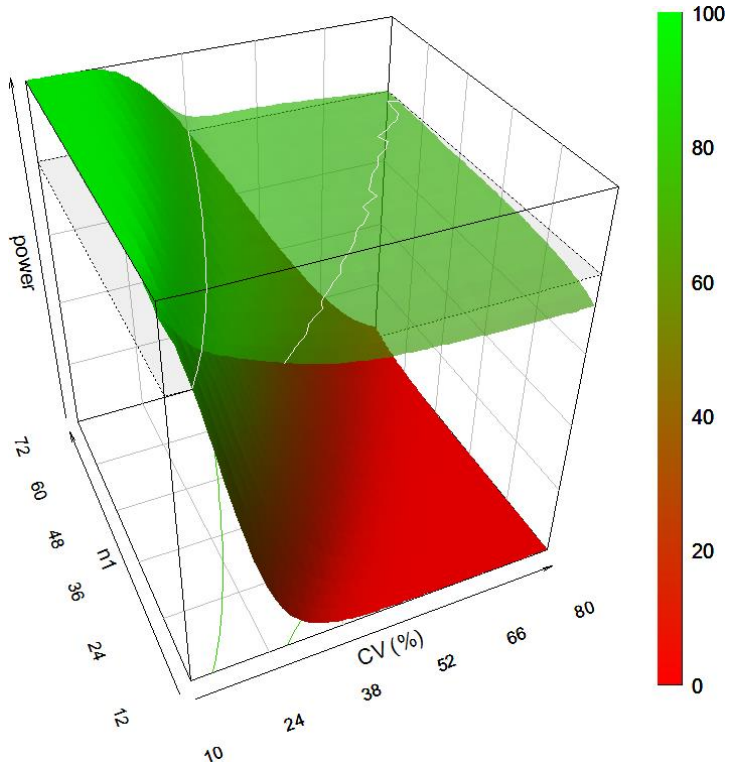
'Type 1' TSD
(Potvin *et al.* B)

Conditions:

α_{adj}	0.0294
GMR	0.95
power	80%
CV	10 – 80% (step 2%)
n_1	12 – 72 (step 2)

1 mio simulations in all combinations; significance limit for the Type I Error (TIE) 0.05036

Power



Minimum final power at *CV* 80% and n_1 12: 72.24%

Remarks

- GSDs
  - If interim is not at $n/2$ and/or final not at n → further adjustment of α
- TSDs
  - Exact methods only for 2×2×2 crossover design
  - Futility rules
    - Reduce the Type I Error
    - Negative impact on power → simulations recommended
  - Small first stage not recommended → large sample size penalty
  - 0.0294 is not a ‘natural constant’ (different conditions → different α_{adj})

Suggestions for ICH M13C

- Inverse Normal Method for 2×2×2 crossover should be recommended (based on a mathematical proof)
  - Maximum combination test or
  - Standard combination test
- Simulation-based methods should be acceptable if maximum empiric Type I Error for the entire grid of *CV* / n_1-combinations with 1 mio simulations ≤ 0.05036
  - Published methods
  - Alternatively simulations of the applicant with exhaustive documentation

Statistical challenges and opportunities in ICH M13C

Thank You!

Helmut Schütz

Center for Medical Data Science

MEDICAL UNIVERSITY OF VIENNA

1090 Vienna, Austria

helmut.schuetz@muv.ac.at

BEBAC

1070 Vienna, Austria

helmut.schuetz@bebac.at

References

Scaled Average Bioequivalence

1. Tóthfalusi L, Endrényi L, García-Arieta A. *Evaluation of bioequivalence for highly variable drugs with scaled average bioequivalence*. 2009. doi:10.2165/11318040-000000000-00000
2. Haidar SH, Makhlouf F, Schuirmann DJ, Hyslop T, Davit B, Conner D, Yu LX. *Evaluation of a Scaling Approach for the Bioequivalence of Highly Variable Drugs*. 2008. doi:10.1208/s12248-008-9053-4
3. Endrényi L, Tóthfalusi L. *Regulatory and Study Conditions for the Determination of Bioequivalence of Highly Variable Drugs*. 2009. doi:10.18433/j3zw2c
4. Karalis V, Symillides M, Macheras P. *On the leveling-off properties of the new bioequivalence limits for highly variable drugs of the EMA guideline*. 2011. doi:10.1016/j.ejps.2011.09.008
5. Davit BM, Chen ML, Conner DP, Haidar SH, Kim S, Lee CH, Lionberger RA, Makhlouf FT, Nwakama PE, Patel DT, Schuirmann DJ, Yu LX. *Implementation of a Reference-Scaled Average Bioequivalence Approach for Highly Variable Generic Drug Products by the US Food and Drug Administration.* 2012. doi:10.1208/s12248-012-9406-x
6. Wonnemann M, Frömke C, Koch A. *Inflation of the Type I Error: Investigations on Regulatory Recommendations for Bioequivalence of Highly Variable Drugs*. 2015. doi:10.1007/s11095-014-1450-z
7. Muñoz J, Alcaide D, Ocaña J. *Consumer's risk in the EMA and FDA regulatory approaches for bioequivalence in highly variable drugs.* 2016. doi:10.1002/sim.6834
8. Labes D, Schütz H. *Inflation of Type I Error in the Evaluation of Scaled Average Bioequivalence, and a Method for its Control*. 2016. doi:10.1007/s11095-016-2006-1
9. Tóthfalusi L, Endrényi L. *An Exact Procedure for the Evaluation of Reference-Scaled Average Bioequivalence*. AAPS J. 2016. doi:10.1208/s12248-016-9873-6
10. Tóthfalusi L, Endrényi L. *Algorithms for Evaluating Reference Scaled Average Bioequivalence: Power, Bias, and Consumer Risk*. 2017. doi:10.1002/sim.7440
11. Molins E, Cobo E, Ocaña J. *Two-Stage Designs Versus European Scaled Average Designs in Bioequivalence Studies for Highly Variable Drugs: Which to Choose?* 2017. doi:10.1002/sim.7452
12. Endrényi L, Tóthfalusi L. *Bioequivalence for highly variable drugs: regulatory agreements, disagreements, and harmonization*. 2019. doi:10.1007/s10928-019-09623-w
13. Deng Y, Zhou XH. *Methods to control the empirical type I error rate in average bioequivalence tests for highly variable drugs*. 2019. doi:10.1177/0962280219871589
14. Ocaña J, Muñoz J. *Controlling type I error in the reference-scaled bioequivalence evaluation of highly variable drugs*. 2019. doi:10.1002/pst.1950
15. Schütz H, Labes D, Wolfsegger MJ. *Critical Remarks on Reference-Scaled Average Bioequivalence*. 2022. doi:10.18433/jpps32892
16. Muñoz J, Ocaña J, Suárez R, Millapán C. *Scaled average bioequivalence methods for highly variable drugs: Leveling-off soft limits and the EMA's 2010 guideline (some ways to improve its type I error control).* 2024. doi:10.1002/sim.10021
17. Schütz H, Tomashevskiy M, Labes D. *replicateBE: Average Bioequivalence with Expanding Limits (ABEL).* 2022. doi:10.32614/CRAN.package.replicateBE
18. Labes D, Schütz H, Lang B. *PowerTOST: Power and Sample Size Based on Two One-Sided t-Tests (TOST) for (Bio)Equivalence Studies*. 2024. doi:10.32614/CRAN.package.PowerTOST

Group-Sequential Designs

1. Gould AL. *Group Sequential Extension of a Standard Bioequivalence Testing Procedure*. 1995. doi:10.1007/BF02353786
2. Hauck WW, Preston PE, Bois FY. *A Group Sequential Approach to Crossover Trials for Average Bioequivalence*. 1997. doi:10543409708835171 4
3. Jennison C, Turnbull BW. *Group sequential methods with applications to clinical trials*. Boca Raton: Chapman & Hall/CRC; 1999.

References

Adaptive Two-Stage Designs, Open-Source Software

1. Potvin D, DiLiberti CE, Hauck WW, Parr AF, Schuirmann DJ, Smith RA. *Sequential design approaches for bioequivalence studies with crossover designs.* 2008. doi:10.1002/pst.294
2. Montague TH, Potvin D, DiLiberti CE, Hauck WW, Parr AF, Schuirmann DJ. *Additional results for 'Sequential design approaches for bioequivalence studies with crossover designs'*. 2011. doi:10.1002/pst.483
3. Fuglsang A. *Sequential Bioequivalence Trial Designs with Increased Power and Controlled Type I Error Rates*. 2013. doi:10.1208/s12248-013-9475-5
4. Karalis V, Macheras P. *An Insight into the Properties of a Two-Stage Design in Bioequivalence Studies.* 2013. doi:10.1007/s11095-013-1026-3
5. Karalis V. *The role of the upper sample size limit in two-stage bioequivalence designs*. 2013. doi:10.1016/j.ijpharm.2013.08.013
6. Fuglsang A. *Futility Rules in Bioequivalence Trials with Sequential Designs*. 2014. doi:10.1208/s12248-013-9540-0
7. Fuglsang A. *Sequential Bioequivalence Approaches for Parallel Designs*. 2014. doi:10.1208/s12248-014-9571-1
8. Golkowski D, Friede T, Kieser M. *Blinded sample size reestimation in crossover bioequivalence trials.* 2014. doi:10.1002/pst.1617
9. Karalis V, Macheras P. *On the Statistical Model of the Two-Stage Designs in Bioequivalence Assessment*. 2014. doi:10.1111/jphp.12164
10. Jones B, Kenward MG. *Design and analysis of crossover trials*. Boca Raton: Chapman & Hall/CRC; 3rd ed. 2015.
11. Schütz H. *Two-stage designs in bioequivalence trials*. 2015. doi:10.1007/s00228-015-1806-2
12. Kieser M, Rauch G. *Two-stage designs for cross-over bioequivalence trials*. 2015. doi:10.1002/sim.6487
13. Zheng Ch, Zhao L, Wang J. *Modifications of sequential designs in bioequivalence trials*. 2015. doi:10.1002/pst.1672
14. Xu J, Audet C, DiLiberti CE, Hauck WW, Montague TH, Parr AF, Potvin D, Schuirmann DJ. *Optimal adaptive sequential designs for crossover bioequivalence studies*. 2016. doi:10.1002/pst.1721
15. Molins E, Cobo E, Ocaña J. *Two-stage designs versus European scaled average designs in bioequivalence studies for highly variable drugs: Which to choose?* 2017. doi:10.1002/sim.7452
16. Molins E, Labes D, Schütz H, Cobo E, Ocaña J. *An iterative method to protect the type I error rate in bioequivalence studies under two-stage adaptive 2×2 crossover designs*. 2021. doi:10.1002/bimj.201900388
17. König F, Wolfsegger M, Jaki T, Schütz H, Wassmer G. *Adaptive two-stage bioequivalence trials with early stopping and sample size re-estimation*. 2015. doi:10.1186/1745-6215-16-S2-P218
18. Maurer W, Jones B, Chen Y. *Controlling the type 1 error rate in two-stage sequential designs when testing for average bioequivalence*. 2018. doi:10.1002/sim.7614
19. Schütz H. *Two-Stage Designs. Dealing with Uncertainty*. Presentation at Network for Scientific Excellence. Campinas. February 2020. https://bebac.at/lectures/Campinas2020-3.pdf
20. Kaza M, Sokolvskyi A, Rudzki PJ. *10th Anniversary of a Two-Stage Design in Bioequivalence. Why Has it Still Not Been Implemented?* 2020. doi:10.1007/s11095-020-02871-3
21. 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*. 2020. doi:10.1002/cpt.2050
22. Molins E, Labes D, Schütz H, Ocaña J. *betsd: Adjusting significance levels in two-stage adaptive 2×2 cross-over designs*. 2019. https://github.com/eduard-molins/betsd
23. Labes D, Lang B. Schütz H. *Power2Stage: Power and Sample-Size Distribution of 2-Stage Bioequivalence Studies*. 2021. doi:10.32614/CRAN.package.Power2Stage