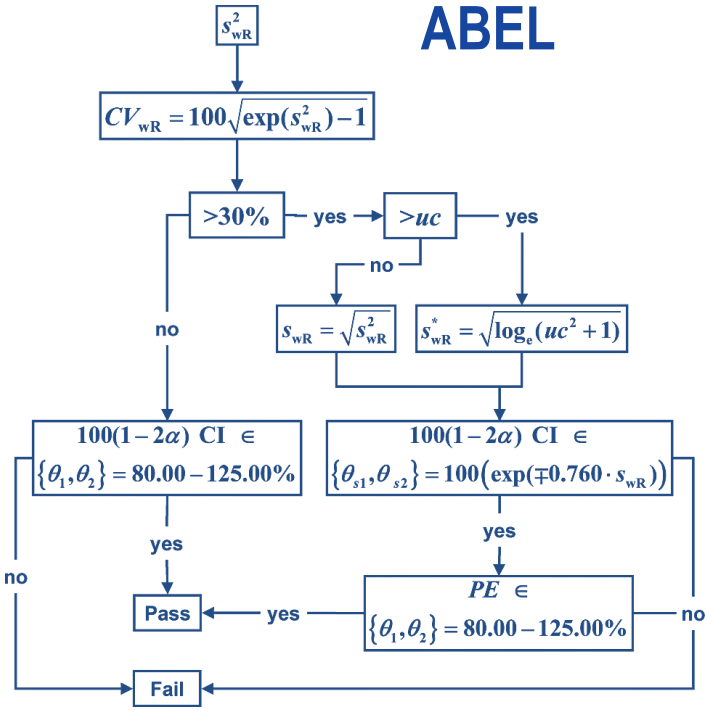
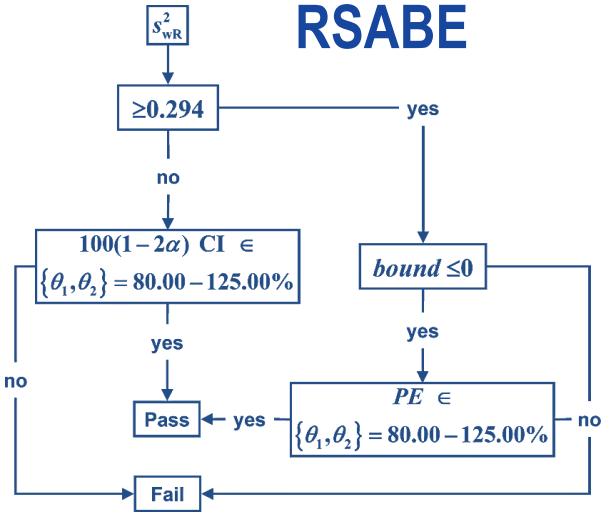
# Highly Variable Drugs and Type I Error


## Helmut Schütz

Center for Medical Data Science of the Medical University of Vienna
Faculty of Pharmacy, Universidade de Lisboa, Lisbon
BEBAC, Vienna

# Why can the Type I Error (TIE) be inflated?

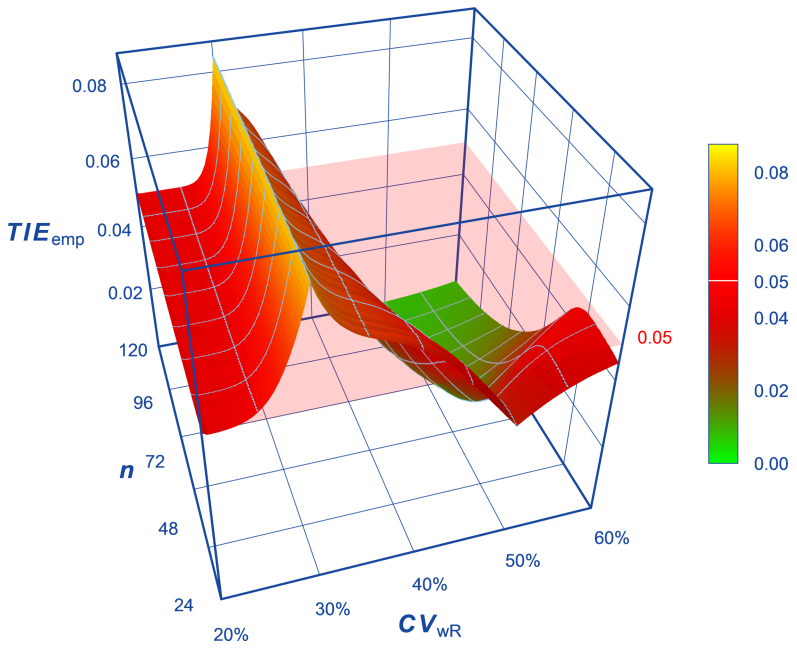



- Implemented Scaled Average Bioequivalence are frameworks
  - Limits are random variables dependent on the reference's variance
  - Drugs will be misclassified if the <u>observed</u> $CV_{wR} \neq$ <u>true</u> $CV_{wR}$

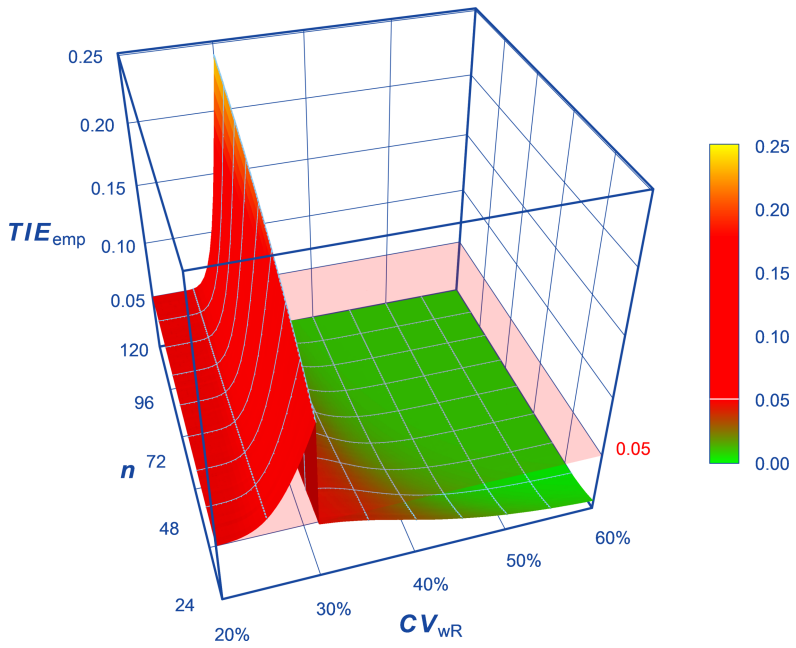
# TIE in SABE as implemented

## ABEL (EMA and others)




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

## RSABE (FDA 'implied limits')




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

**2-sequence 4-period full replicate design**

# The FDA's 'desired consumer risk model'

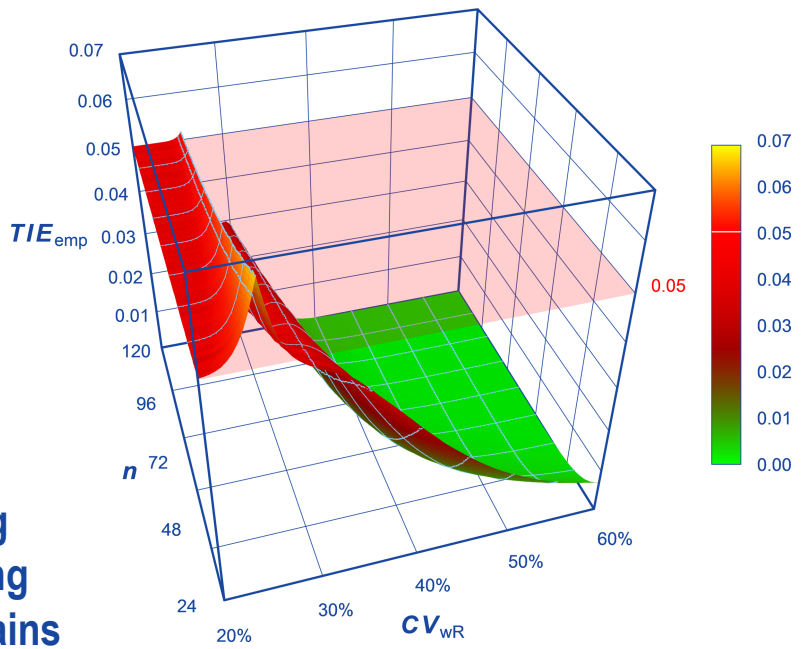
## Type I Error assessed at

- 0.8000 or 1.2500 if $s_{wR} \leq 0.25$
- $\exp(\pm k \cdot s_{wR})$ if $s_{wR} > 0.25$

Davit *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. https://doi.org/10.1208/s12248-012-9406-x

## Section 'Controversies'

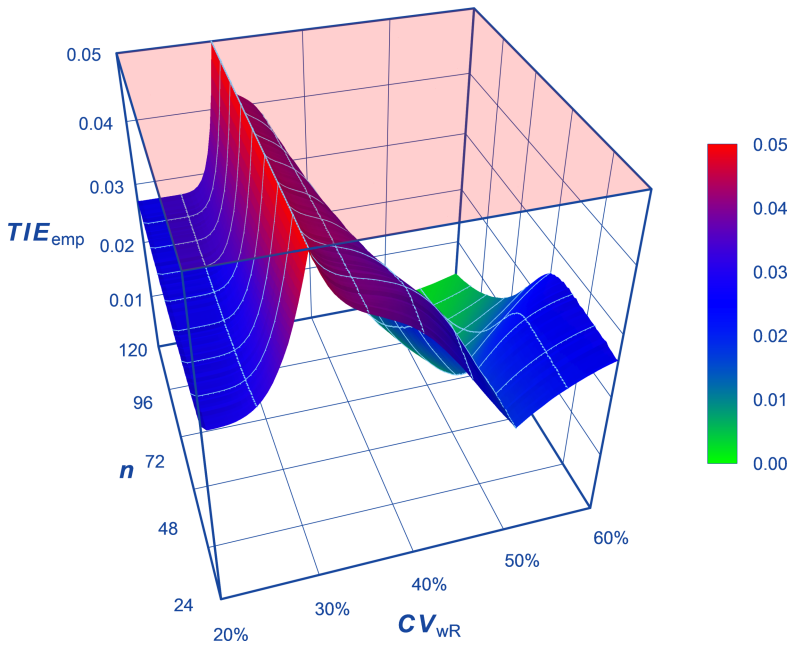
» Results of simulations conducted by members of the HV Drug Working Group support the position that using a cutoff value of 0.294 for $s_{wR}$ maintains an acceptable [*sic*] type I error rate relative to FDA's desired consumer risk model. «




$TIE_{emp}$ at $CV_{wR} \approx 25.396\%$ ($s_{wR}$ 0.25); *n* 24: 0.0663, *n* 120: 0.0501

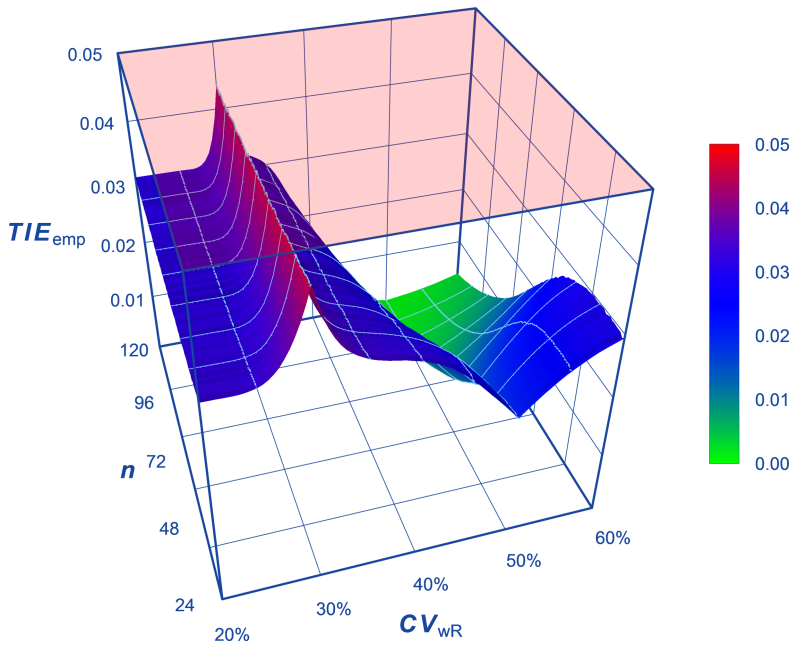
# Alternatives for ABEL: Iteratively adjusted $\alpha$[7]

## Molins *et al.*[10]




$TIE_{emp}$ at $CV_{wR}$ 30%: 0.0500 ✓

## Ocaña *et al.*[13]

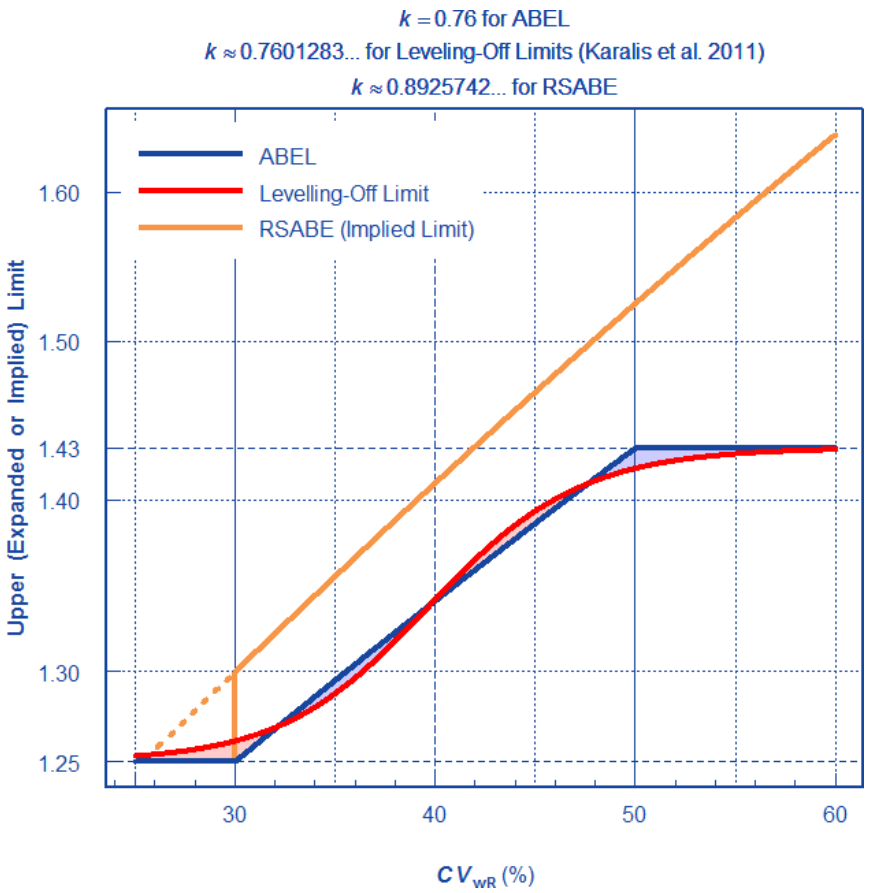



$TIE_{emp}$ at $CV_{wR}$ 30%; *n* 24: 0.0430, *n* 120: 0.0456 ✓

**2-sequence 4-period full replicate design**

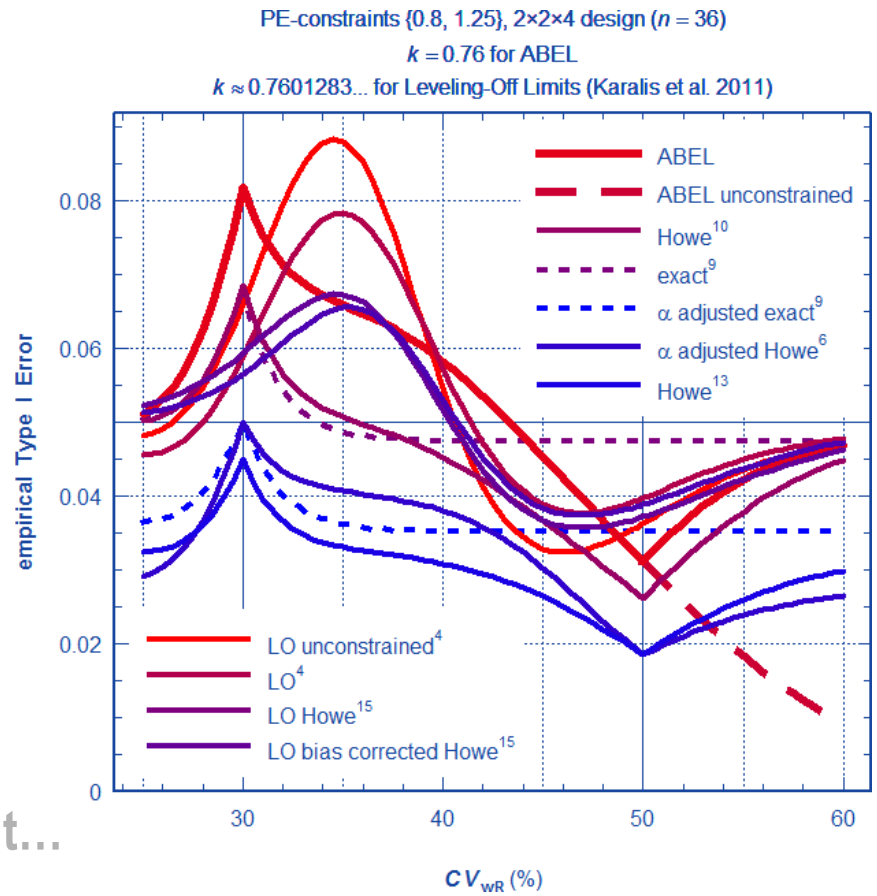
# RSABE, modified ABEL: Leveling-off Limits[4]

- Unconstrained scaling in RSABE
  - Discontinuity $s_{wR}$ 0.294
- For ABEL a sigmoidal function with inflection at 40% was proposed[4]
  - Esthetically more appealing…
  - Higher inflation of the TIE in the red areas and lower in the blue ones



# Desperate Attempts…

- First paper showing an inflated TIE published already in 2009[1] (before [*sic*] SABE was implemented by agencies)
  - Proposals to control the TIE[6 –13,15]
  - All but three[6,9,13] failed to resolve the problem completely
  - The exact method[9] drops the PE constraint…



# Conclusions (ABEL)

- The upper cap of expansion lacks a scientific rationale
  - 50% introduced due to reservations of one member state
  - Health Canada's ≈57.38% likely only to give a 'nice' maximum expansion of 67.7 – 150.0%
  - If removed, no issues with the TIE (like in RSABE); controlled by the inherent conservatism of the TOST and PE-constraint
- $\alpha$-adjusted Howe-ABEL[6] and Howe-ABEL[13] control the TIE
  - Compromise power $\rightarrow$ larger sample sizes required
- Leveling-Off approaches[4,15] are problematic
  - With the original even more inflation of the TIE than in ABEL
  - At low $CV_{wR}$ always inflated TIE

# Conclusions (RSABE)

- The FDA's RSABE is beyond repair
  - The (correct) TIE is more than twice as large as with ABEL
  - Correlation of the TIE with the sample size
  - Assessing the TIE via the 'desired consumer risk model' is a mere magician's trick
    I do not agree with Davit *et al*. (2012) that it
    »maintains an acceptable TIE 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 or not) is incomprehensible for physicians
  - If $s_{wR} < 0.294$ (ABE-branch) in a partial replicate design, the model is over-specified and may not converge

# Suggestion for Harmoni$\frac{s}{z}$ation

- ABEL
  - Should be acceptable for <u>all</u> PK metrics in <u>all</u> jurisdictions
  - The upper cap of expansion should be removed
  - Biased-corrected Howe-LO[15] and iteratively adjusted $\alpha$[6,7] are promising – control the Type I Error with less loss in power than other methods
- RSABE should be abandoned in favor of a variant of ABEL controlling the Type I Error
- Heresy
  - Full replicate studies mandatory for the originator (Les Benet, Bio-International, Munich 1994); alternatively regulators could collect and exchange $CV_{wR}$ of studies → PSGs
  - Fixed limits, replicate designs not needed, TIE always controlled[14]

Reach for the stars, even if you have to stand on a cactus.
– Susan Longacre

# Highly Variable Drugs and Type I Error

MEDICAL UNIVERSITY OF VIENNA

BEBAC

## Thank You!


**Helmut Schütz**
Center for Medical Data Science, Medical University of Vienna
1090 Vienna, Austria
helmut.schuetz@muv.ac.at
Faculty of Pharmacy, Universidade de Lisboa, Lisbon
1649-003 Lisbon, Portugal
helmut@edu.ulisboa.pt
**BEBAC**
1070 Vienna, Austria
helmut.schuetz@bebac.at

# Backup

- We can calculate a 1 – $\alpha$ confidence interval of the $CV_{wR}$ based on its associated variance and the $\chi^2$-distribution with $n$ – 2 degrees of freedom
  - Full replicate design, 95% CI of $CV_{wR}$ = 30%

| *n* | 95% CI | |
|---|---|---|
| 24 | 21.28 | 49.53 |
| 26 | 21.55 | 48.24 |
| 28 | 21.79 | 47.15 |
| 30 | 22.01 | 46.22 |
| 32 | 22.21 | 45.42 |
| 34 | 22.40 | 44.71 |
| 36 | 22.57 | 44.08 |

- If a study is performed with 24 subjects, we can expect with 95% probability in ABEL the *entire range* between no scaling
  21.28% → {80.00 – 125.00%}
  to almost maximum expansion
  49.53% → {70.13 – 142.59%}
- Not rocket science

# Assessment of the consumer risk in SABE

## Empirical Type I Error simulated under the Null, i.e., at $\exp(\pm k \cdot s_{wR})$


1. 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. https://doi.org/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*. AAPS J. 2008; 10(3): 450–4. https://doi.org/10.1208%2Fs12248-008-9053-4
3. Endrényi L, Tóthfalusi L. *Regulatory and Study Conditions for the Determination of Bioequivalence of Highly Variable Drugs*. J Pharm Pharmaceut Sci. 2009; 12(1): 138–49. https://doi.org/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*. Europ J Pharm Sci. 2011; 44(4): 497–505. https://doi.org/10.1016/j.ejps.2011.09.008
5. 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. https://doi.org/10.1007/s11095-014-1450-z
6. 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. https://doi.org/10.1002/sim.6834
7. 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. https://doi.org/10.1007/s11095-016-2006-1
8. Tóthfalusi L, Endrényi L. *An Exact Procedure for the Evaluation of Reference-Scaled Average Bioequivalence.* AAPS J. 2016; 18(2): 476–89. https://doi.org/10.1208/s12248-016-9873-6
9. Tóthfalusi L, Endrényi L. *Algorithms for Evaluating Reference Scaled Average Bioequivalence: Power, Bias, and Consumer Risk*. Stat Med. 2017; 36(27): 4378–90. https://doi.org/10.1002/sim.7440
10. 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. https://doi.org/10.1002/sim.7452
11. Endrényi L, Tóthfalusi L. *Bioequivalence for highly variable drugs: regulatory agreements, disagreements, and harmonization*. J Pharmacokin Pharmacodyn. 2019; 46(2): 117–26. https://doi.org/10.1007/s10928-019-09623-w
12. Deng Y, Zhou XH. *Methods to control the empirical type I error rate in average bioequivalence tests for highly variable drugs*. Stat Methods Med Res. 2019; 29(6): 1650–67. https://doi.org/10.1177/0962280219871589
13. Ocaña J, Muñoz J. *Controlling type I error in the reference-scaled bioequivalence evaluation of highly variable drugs*. Pharm Stat. 2019; 18(5): 583–99. https://doi.org/10.1002/pst.1950
14. Schütz H, Labes D, Wolfsegger MJ. *Critical Remarks on Reference-Scaled Average Bioequivalence*. J Pharm Pharmaceut Sci. 2022; 25: 285–96. https://doi.org/10.18433/jpps32892
15. 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).* Stat Med. 2024; 43(7): 1475–88. https://doi.org/10.1002/sim.10021