Critical Remarks on Reference-Scaled Average Bioequivalence

Helmut Schütz[1,2], Detlew Labes[3], Martin J Wolfsegger[4]

[1]BEBAC. Neubaugasse 36/11, 1070 Vienna, Austria; [2]Institute of Medical Statistics of the Medical University of Vienna, Austria; [3]Consultant. Lössauer Straße 18, 13055 Berlin, Germany; [4]Baxalta Innovations GmbH, a Takeda company. Donau-City-Straße 7, 1220 Vienna, Austria

Corresponding author: Helmut Schütz, BEBAC, Neubaugasse 36/11, 1070 Vienna, Austria; TEL: (+43) 1 2311746; email: helmut.schuetz@bebac.at; helmut.schuetz@meduniwien.ac.at

Received, June 6, 2022; Revised, July 7, 2022; Accepted, August 24, 2022; Published, September 15, 2022

ABSTRACT — Purpose: More than a decade ago the option to assess highly variable drugs / drug products by reference-scaled average bioequivalence was introduced in regulatory practice. Recommended approaches differ between jurisdictions and may lead to different conclusions even for the same data set. According to our knowledge, implemented methods have not been directly compared for their operating characteristics (Type I Error and power). **Methods:** We performed Monte Carlo simulations to assess the consumer risk and the clinically relevant difference for the recommended regulatory settings. **Results:** In all methods for reference-scaled average bioequivalence the Type I Error can be inflated with a consequently compromised consumer risk. Furthermore, the clinically relevant difference could vary between studies performed with the same reference product. **Conclusions:** Only average bioequivalence with fixed – widened – limits would both maintain the consumer risk and offer an unambiguously defined clinically not relevant difference. As long as such an approach is not implemented in regulatory practice, we recommend adjusting the level of the test α.

INTRODUCTION

Comparative bioavailability studies are used in the regulatory approval process of generics, supporting formulation changes, developing line extensions to approved products, and for testing of drug-drug interactions or food effects. Regulatory acceptance depends on whether the two-sided 1–2α (α commonly set to 0.05) confidence intervals (CI) for the ratios of geometric least squares means for specific pharmacokinetic (PK) metrics of two treatments fall completely within specific limits representing the clinically not relevant difference (Δ). The two treatments under investigation are often termed the Test (T) and Reference (R) treatments, where R is the comparator. Crossover studies are designed based on a fixed nominal α, assumed T/R-ratio and intra-subject coefficient of variation (CV_w) to obtain a given target power (commonly between 80 – 90%). It should be noted that from a regulatory perspective the outcome of a comparative bioavailability study is dichotomous (pass|fail). In some jurisdictions for highly variable drugs (HVDs) / drug products (HVDPs), i.e., with an intra-subject coefficient of variation of the reference product (CV_{wR}) of >30%, it is possible to *expand* the conventional acceptance range of 80.00 – 125.00% based on the observed CV_{wR} [the EMA (1-3) the WHO (4,5), Australia (6) the East African Community (7), ASEAN states (8) the Eurasian Economic Union (9), Egypt (10), New Zealand (11), Chile (12), Brazil (13), the Russian Federation (14), Belarus (14), Canada (15)], directly *widen* the limits (member states of the Gulf Cooperation Council (16)], or *scale* the limits (the FDA (17), China (18)]. In most jurisdictions (1,2,4,6-14,16) enlarging the limits focuses on the maximum concentration (C_{max}). Recommended statistical models differ substantially between jurisdictions and consequently also Δ and the consumer risks.

METHODS

In the following Greek letters denote (unknown) population parameters and italicised Latin letters sample estimates.

Models and hypotheses

In the next sections the concept of average bioequivalence (ABE), scaled average bioequivalence (SABE), average bioequivalence with expanding limits (ABEL), and reference-scaled

average bioequivalence (RSABE) are introduced, where ABEL and RSABE approaches require replicated crossover designs (i.e., at least the reference must be administered twice) and ABE can be analysed from a conventional 2×2×2 (i.e., two-treatment, two-sequence, two-period) crossover design.

For all SABE-approaches it is necessary to estimate the within-subject variability of the reference treatment (expressed as CV_{wR} or s_{wR}). For ABEL this is performed by using the actual study data via applying an analysis of variance (ANOVA) of $\log_e$-transformed PK metrics for the subset of the reference treatment with effects sequence, subject (sequence), period. CV_{wR} is then estimated from the corresponding ANOVA's residual means squares error (*MSE*) by

$$CV_{wR} = 100\sqrt{\exp(MSE) - 1} \qquad \text{(Eq. 1)}$$

For RSABE s_{wR} is estimated by an ANOVA via an intra-subject contrast of the reference treatment's $\log_e$-transformed PK metrics and sequence as factor.

$$s_{wR} = \sqrt{MSE / 2} \qquad \text{(Eq. 2)}$$

It should be noted that by the FDA's mixed-effects model (17) the within-subject variance of both treatments could be directly estimated.

Average Bioequivalence (ABE)

In ABE regulatory goalposts are defined based on Δ. This difference is commonly set to 20% but may be narrower [e.g., 10% for narrow therapeutic index drugs (1,15)] or wider [e.g., 25% (19,20)] or more for highly variable drugs / drug products). Based on a multiplicative model of original data, i.e., analysis of $\log_e$-transformed data in an additive model, fixed lower (θ_1) and upper (θ_2) limits of the acceptance range are given by

$$\theta_1 = 1 - \Delta,\ \theta_2 = (1 - \Delta)^{-1} \qquad \text{(Eq. 3)}$$

Bioequivalence must be assessed by the confidence interval inclusion approach

$$H_0 : \frac{\mu_T}{\mu_R} \not\subset \{\theta_1,\ \theta_2\} \quad vs. \quad H_1 : \theta_1 < \frac{\mu_T}{\mu_R} < \theta_2 \qquad \text{(Eq. 4)}$$

where μ_T and μ_R are the geometric least squares means of T and R, respectively. It should be noted that the confidence interval inclusion approach is operationally identical to the Two One-Sided Tests procedure [TOST (21)].

The conclusion of bioequivalence is reached if the null hypothesis H_0 in Eq. 4 is rejected at level α (commonly fixed at 0.05) for each of the PK metrics required in the corresponding jurisdiction.

Scaled Average Bioequivalence (SABE)

For SABE the ratio given in Eq. 4 is scaled by the standard deviation of the reference σ_{wR} and the criterion becomes

$$H_0 : \frac{\mu_T}{\mu_R} \Big/ \sigma_{wR} \not\subset \{\theta_{s1},\ \theta_{s2}\} \ vs.$$
$$H_1 : \theta_{s1} < \frac{\mu_T}{\mu_R} \Big/ \sigma_{wR} < \theta_{s2} \qquad \text{(Eq. 5)}$$

where the scaled limits (θ_{s1}, θ_{s2}) of the acceptance range depend on conditions given by the agency (22).

Average Bioequivalence with Expanding Limits (ABEL)

If expanding the limits is justifiable (a) based on clinical grounds, (b) the study was performed in a replicate design (i.e., at least the reference was administered twice), (c) the estimated CV_{wR} is larger than 30% and not caused by outliers, a 'switching variability' σ_{w0} based on a CV_{w0} of 30% is defined by

$$\sigma_{w0} = \sqrt{\log_e(CV_{w0}^2 + 1)} = 0.29356\ldots \qquad \text{(Eq. 6)}$$

leading to the regulatory constant *k*

$$k = \frac{\log_e(1.25)}{\sigma_{w0}} \approx 0.760 \qquad \text{(Eq. 7)}$$

The limits of the acceptance range are based on the within-subject standard deviation of the reference treatment s_{wR} and are estimated according to

$$\{\theta_{s1}, \theta_{s2}\} = \exp(\mp k \cdot s_{wR}) \qquad \text{(Eq. 8)}$$

It should be noted that there is an upper cap (*uc*) for scaling, which is at 50% (1-14) or at ~57.4% (15). If CV_{wR} exceeds the upper cap, s_{wR} in Eq. 8 must be substituted by

$$s_{wR}^{*} = \sqrt{\log_e(uc^2 + 1)} \qquad \text{(Eq. 9)}$$

effectively limiting the expansion to 69.84 – 143.19% (1-14) or to 66.7 – 150.0% (15). See also Figure 1 and the supplementary material Figure S1.

A special case is the recommendation of the Gulf Cooperation Council (GCC 16). Instead of gradually expanding the limits, wider limits are fixed at 75.00 – 133.33% for any CV_{wR} larger than 30% (without an upper cap of scaling). See also Figure 1 and the supplementary material Figure S2.

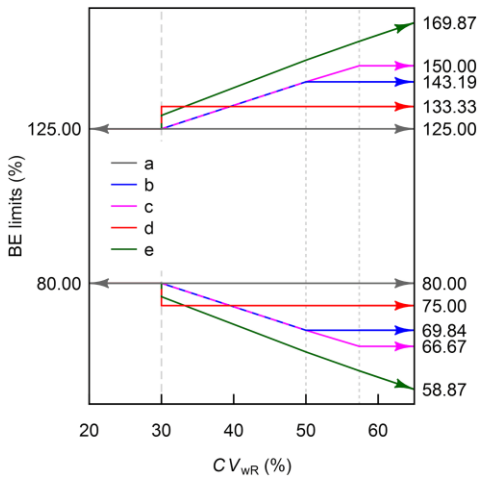

Figure 1. BE limits: a ABE (all jurisdictions), b ABEL (1–14), c ABEL (15), d ABE with Widened Limits (16), e RSABE 'implied limits' (26).

In all jurisdictions bioequivalence must be assessed by the confidence interval inclusion approach according to Eq. 5. Furthermore, the point estimate (*PE*) must lie within the conventional acceptance range of 80.00 – 125.00%.

Reference-Scaled Average Bioequivalence (RSABE)

The recommended approach (17,18) is outlined in the following. If s_{wR} <0.294, evaluate the study for ABE. Otherwise, calculate the standard error s_d of the *PE*

$$s_d = \sqrt{\frac{MSE}{seq^2}\sum\frac{1}{n_i}} \qquad \text{(Eq. 10)}$$

where *MSE* is the residual mean squares error, *seq* the number of sequences, and n_i the number of subjects in sequence *i*. The regulatory limit σ_{w0} is given by 0.25 (i.e., based on a CV_{w0} of ~25.4%), leading to the regulatory constant θ_s (23).

$$\theta_s = \frac{\log_e(1.25)}{\sigma_{w0}} = 0.89257\ldots \qquad \text{(Eq. 11)}$$

Calculate the linearised RSABE criterion

$$crit = PE^2 - s_d^2 - \theta_s^2 \cdot s_{wR}^2 \qquad \text{(Eq. 12)}$$

Calculate its approximate 95% upper confidence *bound* (24).

$$\begin{aligned} E_m &= PE^2 - s_d^2 \\ E_s &= \theta_s^2 \cdot s_{wR}^2 \\ C_m &= \left(|PE| + t_{1-\alpha,\,df} \cdot s_d\right)^2 \\ C_s &= E_s \cdot df_{RR} \,/\, \chi^2_{1-\alpha,\,df_{RR}} \\ bound &= E_m - E_s + \\ &\quad \sqrt{\left(C_m - E_m\right)^2 + \left(C_s - E_s\right)^2} \end{aligned} \qquad \text{(Eq. 13)}$$

where E_m and E_s are the estimates of the true parameters and s_d^2 acts as a bias correction (17, 25). Since the distributions of E_m and E_s are known, their upper confidence C_m and C_s can be calculated. If *bound* ≤ 0 and the *PE* lies within the conventional acceptance range of 80.00 – 125.00%, RSABE is accepted.

The so-called 'implied limits' (26) of the acceptance range are based on the within-subject standard deviation of the reference treatment s_{wR} and are estimated according to

$$\{\theta_{s1}, \theta_{s2}\} = \exp(\mp\theta_s \cdot s_{wR}) \qquad \text{(Eq. 14)}$$

See also Figure 1 and the supplementary material Figure S3.

Regulatory Landscape

Table 1 gives an overview of currently implemented methods. Following suggestions by Benet (27) 'for political reasons' all implemented methods contain a *PE* constraint of 80.00 – 125.00%.

In jurisdictions applying ABEL – except (12,13) – it must be demonstrated that outliers do not cause the high variability. It is still an open issue how that should be done. One of the authors (HS) suggested box plots of studentised model residuals – as a mere joke – at a joint symposium of the European Generic Medicines Association and the EMA, being aware of their nonparametric nature and the EMA's reluctance towards robust methods. Alas, this joke was included in the Questions & Answers document (28) without mentioning studentised residuals as Health Canada does (29). High variances are commonly associated with extreme values and

Table 1. Implemented methods.

Jurisdiction	Method	*uc*	Model	PK metric(s)	Outliers [a]
ASEAN States (8), Australia (6), Belarus (14), East African Community (7), Egypt (10), EMA (1,2), Eurasian Economic Union (9), New Zealand (11), Russia (14), WHO (4)	ABEL [b]	50%	ANOVA	C_{max}	yes
Brazil (13), Chile (12)	ABEL [b]	50%	ANOVA	C_{max}	no
EMA (3)	ABEL [b]	50%	ANOVA	C_{min} [f], C_{τ} [g], partial *AUC* [h]	yes
WHO (5)	ABEL [b]	50%	ANOVA	C_{max}, *AUC* [i]	yes
Canada (15)	ABEL [b]	57.4%	LME [j]	*AUC*	yes
Gulf Cooperation Council (16)	GCC [c]	–	ANOVA	C_{max}	yes
Kazakhstan (14), League of Arab States (19), South Africa (20)	ABE [d]	–	ANOVA	C_{max}	no
China CDE (18), U.S. FDA (17)	RSABE [e]	–	LME [j]	any	no

[a] Assessment of outliers of the reference required.
[b] Expanded limits.
[c] Widened limits depending on CV_{wR}.
[d] Fixed limits 75.00–133.33%.
[e] Scaled limits.
[f] Minimum concentration in steady state.
[g] Concentration at the end of the dosing interval in steady state.
[h] Area Under the concentration-time Curve.
[i] *AUC* if the study is performed in a full replicate design.
[j] Linear mixed effects.

property of the distribution (see supplementary material Figure S4). Hence, in jurisdictions applying RSABE assessing the data for outliers is not recommended (17,18).

Contrary to ABE, studies evaluated by any variant of SABE are not bijective, i.e., only for strict homoscedasticity ($CV_{wT} \equiv CV_{wR}$) it holds that if T is equivalent to R, R is also equivalent to T – which makes switching products with different variabilities questionable (22). Whereas the reference-scaled model is formulated in population parameters (μ_T/μ_R, σ_{wR}), in the study only their estimates (*PE*, s_{wR}) are accessible. Whereas in ABE Δ is fixed and hence, known, in a particular study assessed for SABE the *realised* Δ can be recalculated based on the lower scaled limit

$$\hat{\Delta}_r = 100(1-\theta_{s1}) \qquad \text{(Eq. 15)}$$

Assessment of the Type I Error

The Type I Error (*TIE*) is defined as the probability of falsely rejecting the true null hypothesis, i.e., erroneously claiming equivalence between Test and Reference in a PK metric of interest and the following consideration focuses on the *TIE* for a PK metric suitable for scaling. However, for a formal claim of bioequivalence, equivalence needs to be shown statistically significantly in all relevant PK metrics simultaneously (30). This translates into the consumer risk, which should not exceed the nominal α of the test (0.05). The maximal *TIE* can be assessed by exploring the power of passing BE with a true T/R-ratio (θ_0) at the limits of the acceptance range (θ_1 or θ_2). Whereas for ABE the limits are fixed and hence, an analytical solution for power exists, both ABEL and RSABE are frameworks, where the null hypothesis is generated in face of the data (the decision whether to scale or not depends on the observed variability of the reference, application of the *PE* constraint and – for ABEL – the upper cap of scaling). Therefore, a sufficiently large number of studies must be simulated, where power is obtained by the fraction of studies passing BE with a true θ_0 at the scaled limits of the acceptance range (θ_{s1} or θ_{s2}).

Monte Carlo simulations were performed based on the 'key statistics', i.e., based on the fact that s_{wR} follows a χ^2-distribution with $n - 2$ degrees of freedom and θ_0 follows a lognormal distribution (31). Since both distributions are skewed to the right, the probability of a misclassification at CV_{wR} = 30% and θ_0 = 1.25 is slightly larger than 50% (see supplementary material Figure S5). Two-sequence four-period fully replicated studies with balanced sequence groups of 24, 36, and 48 subjects were simulated under homoscedasticity (true $CV_{wR} \equiv CV_{wT}$ = 20–65%) in the R (32) package *PowerTOST* (33, functions *power.scABEL* and *power.RSABE*).

For each combination one million studies were simulated according to the respective conditions of the regulatory frameworks using the pseudo-random number generator Mersenne-Twister (34) with a fixed seed of 123456 to support reproducibility. Simulations were performed at the respective *upper* boundary of the equivalence range (although due to the symmetry in log-scale similar results are obtained

for the *lower* boundary). Due to the positive skewness of CV_{wR} and θ_0, the upper boundary represents generally the worst-case scenario. Based on the binomial test an empiric *TIE* rate above 0.05036 is considered statistically significant inflated. The standard error *SE* of the empiric *TIE* is calculated by $\sqrt{0.5 \cdot TIE_{emp} / 10^6}$. The settings of θ_0 dependent on CV_{wR} or s_{wR} are given in Table 2.

Table 2. Monte Carlo settings according to each framework. In all simulations the *PE* constraint was observed.

Method		Setting
ABEL (1–15)	$\theta_0 = 1.2500$	if $CV_{wR} \le 0.30$
	$\theta_0 = \exp(k \cdot s_{wR})$	if CV_{wR} $\{> 0.30, \ \le uc\}$
	$\theta_0 = \exp(k \cdot s^*_{wR})$	if $CV_{wR} > uc$
GCC (16)	$\theta_0 = 1.2500$	if $CV_{wR} \le 0.30$
	$\theta_0 = 1.3333\dot{}$	if $CV_{wR} > 0.30$
RSABE 'implied limits' (26)	$\theta_0 = 1.2500$	if $s_{wR} < 0.294$
	$\theta_0 = \exp(\theta_s \cdot s_{wR})$	if $s_{wR} \ge 0.294$
RSABE 'desired consumer risk model' (26)	$\theta_0 = 1.2500$	if $s_{wR} \le 0.25$
	$\theta_0 = \exp(\theta_s \cdot s_{wR})$	if $s_{wR} > 0.25$

RESULTS

Like every estimate, CV_{wR} carries some degree of uncertainty (depending on the sample size). Hence, a drug / drug product might be falsely classified as highly variable and vice versa. Even more so, the scaled limits might be wider or narrower than the ones based on the true – but unknown – CV_{wR} (Table 3).

Table 3. Parametric 95% confidence interval of CV_{wR}.

CV_{wR} [a]	n [b] = 24	n [b] = 36	n [b] = 48
20	15.4 – 28.6	16.1 – 26.4	16.6 – 25.3
25	19.2 – 35.9	20.1 – 33.1	20.7 – 31.7
30	23.0 – 43.4	24.1 – 39.9	24.8 – 38.2
35	26.8 – 51.0	28.0 – 46.8	28.8 – 44.7
40	30.5 – 58.8	31.9 – 53.9	32.9 – 51.4
45	34.1 – 66.8	35.8 – 61.0	36.9 – 58.1
50	37.8 – 75.1	39.6 – 68.3	40.8 – 65.0
55	41.4 – 83.5	43.4 – 75.8	44.7 – 71.9
60	44.9 – 92.3	47.2 – 83.4	48.6 – 79.0
65	48.4 – 101.3	50.9 – 91.2	52.5 – 86.2

[a] True CV_{wR} in percent.
[b] Sample size.

As seen in Figure 2 and Table 4, in ABE the consumer risk is strictly controlled, i.e., the *TIE* never exceeds nominal α. Since TOST is not a most powerful test (30), for high CV_{wR} together with relatively low sample sizes, it becomes conservative (Figure 2a).

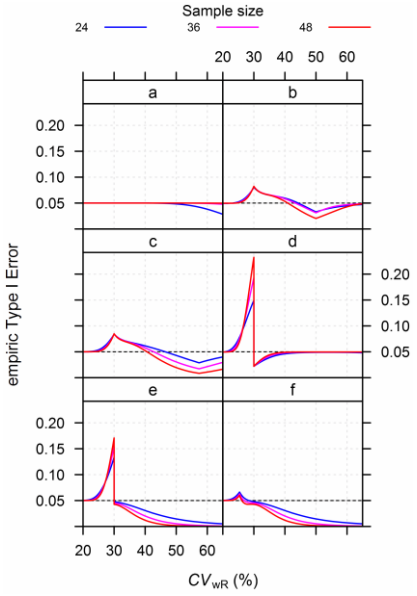

Figure 2. Empiric Type I Error. a. ABE (all jurisdictions), b. ABEL (1–14), c. ABEL (15), d. ABE with widened limits (16), e. RSABE 'implied limits' (26), f. RSABE 'desired consumer risk model' (26).

Due to potential misclassification in ABEL, i.e., if the limits are expanded based on the *observed* CV_{wR} – although based on the *true* CV_{wR} the drug / drug product is not highly variable – the *TIE* is inflated with a maximum at the switching CV_{wR} of 30%. With increasing CV_{wR} the *TIE* decreases to a minimum at the upper scaling cap and then increases again, although never exceeding nominal α even for extremely high CV_{wR} (Figure 2b and c). The maximum *TIE* increases only slightly with the sample size (Table 4).

Due to the discontinuity of the limits at CV_{wR} 30% of the GCC's approach, a misclassification has a substantially larger impact on the *TIE* than in ABEL for any CV_{wR} ≤30%. However, for any CV_{wR} >30% the method controls the consumer risk (Figure 2d). The maximum *TIE* shows a strong dependency on the sample size (Table 4).

RSABE with 'implied limits' (26) shows a large inflation of the *TIE* at CV_{wR} <30% – which can be more than twice than with ABEL. For any CV_{wR} >30% the method becomes extremely conservative thus controlling the consumer risk (Figure 2e). The maximum *TIE* shows a dependency on the sample

size, which is larger than with ABEL but smaller than the one of GCC (Table 4).

Table 4. Maximum empiric Type I Error.

Method	CV_{wR} [a]	n [b]	*TIE* [c]	*SE* [d]
ABE	any	any	0.0500	0.000158
ABEL (1–14)	30.000	24	0.0804	0.000200
		36	0.0819	0.000202
		48	0.0823	0.000203
ABEL (15)	30.000	24	0.0841	0.000205
		36	0.0846	0.000206
		48	0.0846	0.000206
GCC (16)	30.000	24	0.1493	0.000273
		36	0.1931	0.000311
		48	0.2324	0.000341
RSABE 'implied limits' (26)	30.000	24	0.1335	0.000258
		36	0.1536	0.000277
		48	0.1708	0.000292
RSABE 'desired consumer risk model' (26)	25.396	24	0.0663	0.000182
		36	0.0629	0.000177
		48	0.0600	0.000173

[a] True CV_{wR} in percent.
[b] Sample size of a 2-sequence 4-period fully replicated design.
[c] Empiric *TIE* obtained in 1 million simulations.
[d] Standard Error of empiric *TIE*.

RSABE assessed with the 'desired consumer risk model' (26) shows a moderate inflation of the *TIE* with its maximum at CV_{wR} ~25.4% (i.e., at s_{wR} = 0.25; Figure 2f). The maximum *TIE* shows a moderate dependency on the sample size, where contrasting to the other methods the *TIE* decreases with the sample size (Table 4).

Figure 3 illustrates the $\hat{\Delta}_r$ obtained in 500 simulated two-sequence four-period fully replicated studies. Based on the *true* CV_{wR}, Δ would be 22.77 and 25.38% for ABEL, and 26.17 and 29.10% for RSABE. However, based on the *observed* CV_{wR} and the expanded limits in the simulated studies the ranges of $\hat{\Delta}_r$ are for ABEL 20.00 – 29.13% (CV_{wR} = 35%) and 20.00 – 29.05% (CV_{wR} = 40%). For RSABE the ranges are 20.00 –35.56% (CV_{wR} = 35%) and 20.00 – 40.23% (CV_{wR} = 40%). By chance $\hat{\Delta}_r$ can be as large as 30.16% for ABEL and theoretically unlimited for RSABE.

To summarise, all approaches for reference-scaled average bioequivalence currently implemented in jurisdictions may not control the consumer risk – even if RSABE is assessed by the 'desired consumer risk model' (26) under a liberal condition of 0.055 (35). In the authors' opinion the 'desired consumer risk model' is no more than a mathematical prestidigitation, as products are not approved according to its conditions but the ones of the applicable guidance (17).

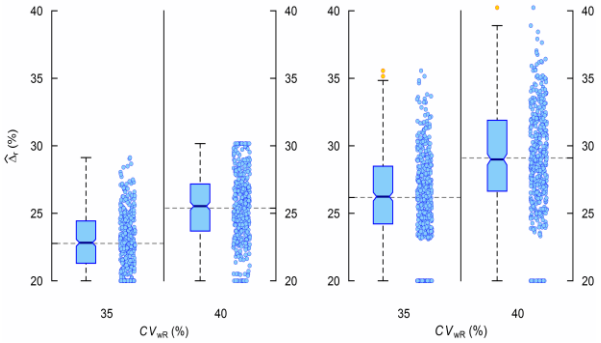

Figure 3. $\hat{\Delta}_r$ in 500 simulated studies with CV_{wR} = 35 and 40%, θ_0 = 0.90 powered to ≥80%. Left panel ABEL (n = 34 and 30), right panel RSABE (n = 28 and 24).

- Box plot: The lower edge of the box represents the 25th percentile (or first quartile), the upper edge of the box represents the 75th percentile (or third quartile) and the line within the lower edge and the upper edge of the box indicate the median. The distance from the lower edge to the upper edge of the box represents the interquartile range (IQR). A whisker is drawn above the 75th percentile to the largest data value that is less or equal to the value that is 1.5×IQR above the 75th percentile. Any data value larger than that is marked as a moderate outlier (yellow dots). A whisker is drawn below the 25th percentile to the smallest data value that is less or equal to the value that is 1.5×IQR below the 25th percentile.
- Jitter plot: The values are plotted as dots along the *y*-axis, and the dots are then shifted randomly along the *x*-axis preventing the dots to overlap.

Illustrative example

We assessed Area Under the Curve (*AUC*) data from (36) in the R (32) package *replicateBE* (37) and Phoenix WinNonlin (38) by the EMA's variant of ABEL (ANOVA) applicable for the WHO, the GCC's variant of ABEL, Health Canada's variant of ABEL (mixed-effects model), and the FDA's RSABE. In all approaches the study passed BE. For the WHO's and Health Canada's approaches the empiric *TIE* was significantly inflated, whereas for the GCC's and the FDA's approaches it was controlled (Table 5).

DISCUSSION

One should be aware that bioequivalence was never a scientific theory in the Popperian sense but an ad hoc solution to a pressing problem in the 1970s (39, 40). The commonly assumed Δ of 20% is arbitrary (as any other). Apart from case reports dealing with

Table 5. Result of the illustrative example evaluated with different approaches. CV_{wR}, limits, CI, PE, and $\hat{\Delta}_r$ in percent.

Approach	Method	CV_{wR}	Limits	CI	*bound*	*PE*	$\hat{\Delta}_r$	*TIE*
WHO (5)	ABEL	35.56	76.93–129.99	101.75–117.46	–	109.33	23.07	0.0643*
GCC (16)	ABEL	35.56	75.00–133.33	101.75–117.46	–	109.33	25.00	0.0459
Canada (15)	ABEL	35.56	76.93–129.99	101.69–117.57	–	109.34	23.07	0.0651*
U.S. FDA (17)	RSABE	35.40	–	–	-0.05552	109.78	26.41	0.0232

* Significantly inflated.

narrow therapeutic index drugs (41-43), no problems are evident switching between the originator and generics (and vice versa) in terms of lack of efficacy or compromised safety, providing decades of empiric evidence that the concept is sufficient in practice (44).

A larger Δ of 30% for C_{max} was suggested (45-48) and a wider acceptance range even for the *AUC* acceptable 'in exceptional cases' (45,48). Many products were approved in Europe according to such more liberal rules. However, the observation that the consumer risk in reference-scaled average bioequivalence is compromised is not new (22,23, 49-59). It is not surprising that the Type I Error can be inflated, since in the current approaches Δ is undefined and the conclusion whether a product passes or fails is based on the within-subject variability of the reference treatment realised in the same study. Consequently, the regulatory goalposts become random variables and each study sets its own rules, awarding ones with high variability.

Well known examples where a pre-test inflates the Type I Error is assessing variance homogenicity (60) and testing for a sequence effect in BE (61). The current situation can be regarded as selecting the final statistical model based on intermediate data-dependent decisions – where up to three decisions are possible in the respective frameworks also utilizing unblinded treatment information. In such situations multiplicity concerns could arise where the influence on the overall Type I Error is difficult to assess (62, section 5.3) but were shown to lead to an inflation of the overall Type I Error (22,23,49–59). To ensure compliance with the International Conference on Harmonisation (ICH) guideline E9 (63) regarding adjustment for multiplicity (e.g., section 5.6, which states that 'adjustment should always be considered and the details of any adjustment procedure or an explanation of why adjustment is not thought to be necessary should be set out in the analysis plan') the level of the test should be adjusted accordingly. Various methods have been proposed to address this topic (52–56,58,59). Among those approaches, two do not entirely control the Type I Error (52,56), one requires the assumption that the observed CV_{wR} is the true one (53), two require abandoning the *PE* constraint (54,55), one is computationally intensive (58), some require a modification of the regulatory approach of ABEL (52,54,55,58), and others lead to a substantial loss in power (52,59). Consequently (56,59) are considered being acceptable within the current regulatory frameworks and do not require additional assumptions as those just requiring a lower test level to be specified. There are no reasons to object employing an *apparently* more conservative α, as 0.05 is a recommendation and an applicant free to select a lower one.

In addition, the realised Δ cannot be recalculated without access to the study report and is therefore unknown to physicians, pharmacists, and patients – representing another disadvantage of scaling approaches as it limits the possibility to make an informed decision.

Furthermore, since biopharmaceutical technology improves, it is not uncommon that $CV_{wT} < CV_{wR}$. For this reason, we recommend performing pilot studies in a fully replicated design, which allows the sample size of the pivotal study to be estimated taking different variances into account, potentially requiring lower sample sizes. If the pilot study is performed in a partial replicate design one must assume homoscedasticity, which in such a case leads to a larger – ethically and economically questionable – pivotal study.

To summarise, the lack of harmonisation of statistical approaches might lead – in the hypothetical situation of submitting the same study to different agencies – to acceptance in one jurisdiction and rejection in others (57), where in addition regulatory agencies regrettably – apart from one anecdotal report (64) – pay little attention the potentially increased consumer risk which is only controlled with average bioequivalence with fixed limits. For HVD(P)s ABE with fixed wider limits was

acceptable for the EMA until 2010 (45) and is currently recommended in Kazakhstan (14), by the League of Arab States (19), and in South Africa (20).

The ICH (63) takes control of the Type I Error seriously, which is e.g., reflected in the FDA's guidance on adaptive designs (65) and consequently, this should be also the case in the field of bioequivalence. Already at the BioInternational '92 Benet (66) proposed in a keynote that 1) innovators should perform replicate studies as part of the new drug application and 2) provide information on intra- and inter-subject measures of extent and rate of bioavailability in the PK section of the package insert. A way out of the dilemma of compromising the consumer risk and lack of harmonisation would be to collect the estimated variabilities, pool them after weighting for the sample size, and obtain a better estimate of the true CV_{wR} (22, 67) which is considered to be a feasible administrative task as reference-scaling methods are applied for more than a decade, an abundance of data exists on the regulatory side, where free exchange of information between major agencies is already performed in the field of inspections and approval of biologic products / biosimilars. Likewise, fixed widened limits in product-specific guidances could be published. It would also be possible to develop overarching guidances for entire classes of HVD(P)s, e.g., proton pump inhibitors or bisphosphonates. Such an approach was already discussed at a BioInternational conference (68). Then the conventional ABE model could be employed in subsequent studies (regardless of whether in a replicate or a conventional 2×2×2 crossover design), where Δ is unambiguously defined, and the consumer risk would always be controlled at ≤5%.

As long as such an approach is not implemented in regulatory practice, we recommend that 1) regulatory agencies should ask the sponsor to include the $\hat{\Delta}_r$ in the corresponding prescribing information to ensure that stakeholders are able to make an informed decision and 2) to specify methods in the protocol to adjust α (56,59) to be in conformity with ICH E9 (63).

CONCLUSIONS

In all methods for Reference-scaled Average Bioequivalence the Type I Error can be inflated. The realised Δ varies between studies performed with the same reference product. Only Average Bioequivalence with fixed – widened – limits would both maintain the consumer risk at the 0.05 nominal level and offer an unambiguously defined Δ.

CONFLICTS OF INTEREST. HS is a lecturer at the Medical University Vienna and a consultant. Clients include private companies, associations of pharmaceutical manufacturers, and regulatory agencies. DL is an independent consultant in the field of bioequivalence and bioavailability studies. MJW is an employee of Takeda and owns Takeda stock/shares. The views and opinions expressed by MJW do not necessarily reflect those of Takeda.

AUTHOR CONTRIBUTIONS. HS: Conceptualization, methodology, software, validation, visualization, original draft preparation; DL: Conceptualization, methodology, software, validation, original draft preparation; MJW: Methodology, visualization, reviewing and editing.

SUPPLEMENTARY MATERIAL. R-scripts to reproduce Tables 3 and 4, as well as the example are available online.

ACKNOWLEDGMENT. In gratitude to László Endrényi (1933–2020), whose work on the bioequivalence of highly variable drugs inspired many scientists in the field.

REFERENCES

1. European Medicines Agency, Committee for Medicinal Products for Human Use. Guideline on the Investigation of Bioequivalence. London. 20 January 2010. https://www.ema.europa.eu/documents/scientific-guideline/guideline-investigation-bioequivalence-rev1_en.pdf (accessed 14 May 2022).
2. European Medicines Agency. Clinical pharmacology and pharmacokinetics: questions and answers. 3.1 Which statistical method for the analysis of a bioequivalence study does the Agency recommend? Annex I. London. 21 September 2016. https://www.ema.europa.eu/en/documents/other/31-annex-i-statistical-analysis-methods-compatible-ema-bioequivalence-guideline_en.pdf (accessed 14 May 2022).
3. European Medicines Agency, Committee for Medicinal Products for Human Use. Guideline on the pharmacokinetic and clinical evaluation of modified release dosage forms. London. 20

November 2014. https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-pharmacokinetic-clinical-evaluation-modified-release-dosage-forms_en.pdf (accessed 14 May 2022)

4. World Health Organization, Essential Medicines and Health Products: Multisource (generic) pharmaceutical products: guidelines on registration requirements to establish interchangeability. WHO Technical Report Series, No. 1003, Annex 6. Geneva. 28 April 2017. https://www.who.int/medicines/areas/quality_safety/quality_assurance/trs1003_annex6.pdf (accessed 14 May 2022).
5. World Health Organization, Prequalification Team: medicines. Guidance Document: Application of reference-scaled criteria for AUC in bioequivalence studies conducted for submission to PQTm. Geneva. 02 July 2021. https://extranet.who.int/pqweb/sites/default/files/documents/AUC_criteria_July2021.pdf (accessed 14 May 2022).
6. Australian Government, Department of Health, Therapeutic Goods Administration. European Union and ICH Guidelines adopted in Australia. Guideline on the Investigation of Bioequivalence with TGA Annotations. https://www.tga.gov.au/ws-sg-index?search_api_views_fulltext=bioequivalence&field_ws_sg_category1=9140 (accessed 14 May 2022).
7. East African Community, Medicines and Food Safety Unit. Compendium of Medicines Evaluation and Registration for Medicine Regulation Harmonization in the East African Community, Part III: EAC Guidelines on Therapeutic Equivalence Requirements. January 2015. http://www.rrfa.co.za/wp-content/uploads/2012/11/guidelines-on-therapeutic-equivalence-requirements-150101.pdf (accessed 14 May 2022).
8. ASEAN States Pharmaceutical Product Working Group. ASEAN Guideline for the Conduct of Bioequivalence Studies. Vientiane. March 2015. https://www.npra.gov.my/images/reg-info/BE/BE_Guideline_FinalMarch2015_endorsed_22PPWG.pdf (accessed 14 May 2022).
9. Eurasian Economic Commission. Regulations Conducting Bioequivalence Studies within the Framework of the Eurasian Economic Union. 3 November 2016, amended 4 September 2020, https://docs.cntd.ru/document/456026107?section=text (accessed 14 May 2022).
10. Ministry of Health and Population, The Specialized Scientific Committee for Evaluation of Bioavailability & Bioequivalence Studies. Egyptian Guideline For Conducting Bioequivalence Studies for Marketing Authorization of Generic Products. Cairo. February 2017. https://web.archive.org/web/20181123145351/http://www.eda.mohp.gov.eg/Files/1092_Egyptian_Guideline_Conducting_Bioequivalence_Studies.pdf (archived from the original 23 November 2018, accessed 14 May 2022).
11. New Zealand Medicines and Medical Devices Safety Authority. Guideline on the Regulation of Therapeutic Products in New Zealand. Part 6: Bioequivalence of medicines. Wellington. February 2018. https://www.medsafe.govt.nz/regulatory/Guideline/GRTPNZ/bioequivalence-of-medicines.pdf (accessed 14 May 2022).
12. Departamento Agencia Nacional de Medicamentos. Instituto de Salud Pública de Chile. Guia para La realización de estudios de biodisponibilidad comparativa en formas farmacéuticas sólidas de administración oral y acción sistémica. Santiago. December 2018. Spanish.
13. ANVISA. Critérios para a condução de estudos de biodisponibilidade relativa/bioequivalência. Consulta Pública Nº 760/2019. Brasilia, December 27, 2019. Portuguese.
14. Shohin I.E., Roshdestvenskij D.A., Medvedev V.Yu., Komarow T.N., Grebenkin D.Yu. Russia, Belarus & Kasakhstan, in Kanfer I. (ed), Bioequivalence Requirements in Various Global Jurisdictions, Springer, Cham, Switzerland, pp 199–227, 2017.
15. Health Canada. Guidance Document. Comparative Bioavailability Standards: Formulations Used for Systemic Effects. Ottawa. 08 June 2018. https://www.canada.ca/content/dam/hc-sc/migration/hc-sc/dhp-mps/alt_formats/pdf/prodpharma/applic-demande/guide-ld/bio/comparative-bioavailability-standards-formulations-used-systemic-effects.pdf (accessed 14 May 2022).
16. Executive Board of the Health Ministers' Council for GCC States. The GCC Guidelines

for Bioequivalence. May 2021. https://www.sfda.gov.sa/sites/default/files/2021-10/GCC_Guidelines_Bioequivalence.pdf (accessed 14 May 2022).
17. U.S. Food and Drug Administration, Center for Drug Evaluation and Research. Draft Guidance for Industry. Bioequivalence Studies with Pharmacokinetic Endpoints for Drugs Submitted Under an ANDA. August 2021. https://www.fda.gov/media/87219/download (accessed 14 May 2022).
18. National Medical Products Administration, Centre for Drug Evaluation. Guidance for Industry: The study of drug bioequivalence with highly variable drugs. Beijing. 29 October 2018. Chinese.
19. League of Arab States, Higher Technical Committee for Arab Pharmaceutical Industry. Harmonised Arab Guideline on Bioequivalence of Generic Pharmaceutical Products. Cairo. March 2014. https://pharmexcil.com/uploadfile/ufiles/HarmonizedArabGuidelinesonBEFinal2014_4may14.pdf (accessed 14 May 2022)
20. Medicines Control Council. Registration of Medicines: Biostudies. Pretoria. June 2015. https://www.sahpra.org.za/wp-content/uploads/2020/01/61de452d2.06_Biostudies_Jun15_v6.pdf (accessed 14 May 2022).
21. Schuirmann D.J. A comparison of the Two One-Sided Tests Procedure and the Power Approach for Assessing the Equivalence of Average Bioavailability. *J Pharmacokin Biopharm* 1987; 15: 657–680. DOI:10.1007/BF01068419.
22. Tóthfalusi L., Endrényi L., García-Arieta A. Evaluation of bioequivalence for highly variable drugs with scaled average bioequivalence. *Clin Pharmacokinet* 2009; 48: 725–743. DOI:10.2165/11318040-000000000-00000.
23. Haidar S.H., Makhlouf F., Schuirmann D.J., Hyslop T., Davit B., Conner D., Yu L.X. Evaluation of a Scaling Approach for the Bioequivalence of Highly Variable Drugs, *AAPS J* 2008; 10: 450–454. DOI:10.1208/s12248-008-9053-4.
24. Howe W.G. Approximate Confidence Limits on the Mean of *X* + *Y* Where *X* and *Y* Are Two Tabled Independent Random Variables. *J Am Stat Assoc* 1974; 69: 789–794. DOI:10.2307/2286019.
25. Schuirmann D. U.S. FDA Perspective: Statistical Aspects of OGD's Approach to Bioequivalence (BE) Assessment for Highly Variable Drugs. Presentation at the 2nd conference of The Global Harmonisation Initiative (GBHI). Rockville. September 15–16, 2016.
26. Davit B.M., Chen M.L., Conner D.P., Haidar S.H., Kim S., Lee C.H., Lionberger R.A., Makhlouf F.T., Nwakama P.E., Patel D.T., Schuirmann D.J., Yu L.X. 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: 915–924. DOI:10.1208/s12248-012-9406-x.
27. Benet L. Why Highly Variable Drugs are Safer, Presentation at the FDA Advisory Committee for Pharmaceutical Science. Rockville, MD. 06 October, 2006. https://web.archive.org/web/20180126174602/http://www.fda.gov:80/ohrms/dockets/ac/06/slides/2006-4241s2_2.ppt (archived from the original 26 January 2018, accessed 14 May 2022).
28. European Generic Medicines Association. Revised EMA Bioequivalence Guideline. 3rd EGA Symposium on Bioequivalence. London, 1 June 2010. Questions & Answers. https://www.medicinesforeurope.com/wp-content/uploads/2016/03/EGA_BEQ_QA_WEB_QA_1_32.pdf (accessed 14 May 2022).
29. Health Canada. Guidance Document. Conduct and Analysis of Comparative Bioavailability Studies. Ottawa. 08 June 2018. https://www.canada.ca/content/dam/hc-sc/documents/services/drugs-health-products/drug-products/applications-submissions/guidance-documents/bioavailability-bioequivalence/conduct-analysis-comparative.pdf (accessed 03 July 2022).
30. Berger R.L., Hsu J.C. Bioequivalence Trials, Intersection–Union Tests and Equivalence Confidence Sets. *Stat Sci* 1996; 11: 283–319. DOI:10.1214/ss/1032280304.
31. Zheng C., Wang J., Zhao L. Testing bioequivalence for multiple formulations with power and sample size calculations. *Pharm Stat* 2012; 11: 334–341. DOI:10.1002/pst.1522.

32. R Core Team, R: A language and environment for statistical computing. Vienna, Austria, 2022. https://www.r-project.org/
33. Labes D., Schütz H., Lang B. R-package PowerTOST: Power and Sample Size for (Bio)Equivalence Studies (Version 1.5-4). February 21, 2022. https://cran.r-project.org/package=PowerTOST
34. Matsumoto M., Nishimura T. Mersenne Twister: A 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Trans Model Comput Simul* 1998; 8: 3–30. DOI:10.1145/272991.272995.
35. Roebruck P., Kühn A. Comparison of tests and sample size formulae for proving therapeutic equivalence based on the difference of binomial probabilities. *Stat Med* 1995; 14: 1583–1594. DOI:10.1002/sim.4780141409.
36. Patterson S.D, Jones B. Bioequivalence and statistics in clinical pharmacology. 2nd ed., CRC Press, Boca Raton, FL, pp. 105–106, 2016.
37. Schütz H., Tomashevskiy M., Labes D. R-package replicateBE: Average Bioequivalence with Expanding Limits (ABEL) (Version 1.1.3). May 02, 2022. https://cran.r-project.org/package=replicateBE
38. Certara USA Inc., Princeton, NJ, Phoenix® WinNonlin® (Version 8.3.4). November 11, 2021.
39. Levy G., Gibaldi M. Bioavailability of Drugs, *Circulation* 1974; 49: 391–394. DOI:10.1161/01.CIR.49.3.391.
40. Skelly J.P. A History of Biopharmaceutics in the Food and Drug Administration 1968–1993. *AAPS J* 2010; 12: 44–50. DOI:10.1208/s12248-009-9154-8.
41. Berg M.J., Gross R.A., Tomaszewski K.J., Zingaro W.M., Haskins L.S. Generic substitution in the treatment of epilepsy. Case evidence of breakthrough seizures. *Neurology* 2008; 71: 525–530. DOI:10.1212/01.wnl.0000319958.37502.8e.
42. Privitera M. Generic substitution of antiepileptic drugs What's a clinician to do? *Neurol Clin Pract* 2013; 3: 161–164. DOI:10.1212/CPJ.0b013e31828d9fc9.
43. Brito J.P., Deng Y., Ross J.R., Choi N.H., Graham D.J., Qiang Y., Rantou E., Wang Z., Zhao L., Shah N.D., Lipska K.J. Rates of, and factors associated with, switching among generic levothyroxine preparations in commercially insured American adults. *Endocrine* 2022; 76: 349–358. DOI:10.1007/s12020-022-02987-z.
44. Davit B.M., Nwakama P.E., Buehler G.J., Conner D.P., Haidar S.H., Patel D.T., Yang Y., Yu L.X., Woodcock J. Comparing Generic and Innovator Drugs: A Review of 12 Years of Bioequivalence Data from the United States Food and Drug Administration. *Ann Pharmacother* 2009; 43: 1583–1597. DOI:10.1345/aph.1M141.
45. European Medicines Agency, CHMP Efficacy Working Party, Therapeutic Subgroup on Pharmacokinetics (EWP-PK). Questions & Answers on the Bioavailability and Bioequivalence Guideline. London, 27 July 2006. Doc.Ref:EMEA/CHMP/EWP/40326/2006.
46. Blume H., Mutschler E. (eds). Bioäquivalenz. Qualitätsbewertung wirkstoffgleicher Fertigarzneimittel. 6th supplemental set. Govi-Verlag, Frankfurt/Main, Germany, 1996. German.
47. Commission of the European Communities, CPMP Working Party on Efficacy of Medicinal Products. Note for Guidance: Investigation of Bioavailability and Bioequivalence. Appendix III: Technical Aspects of Bioequivalence Statistics. Brussels, December 1991.
48. Commission of the European Communities, CPMP Working Party on Efficacy of Medicinal Products. Note for Guidance: Investigation of Bioavailability and Bioequivalence. June 1992.
49. Endrényi L., Tóthfalusi L. Regulatory and Study Conditions for the Determination of Bioequivalence of Highly Variable Drugs. *J Pharm Pharmaceut Sci* 2009; 12: 138–149. DOI:10.18433/J3ZW2C.
50. 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: 497–505. DOI:10.1016/j.ejps.2011.09.008.
51. 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: 135–143. DOI:10.1007/s11095-014-1450-z.
52. 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: 1933–1943. DOI:10.1002/sim.6834.

53. 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: 2805–2814. DOI:10.1007/s11095-016-2006-1.
54. Tóthfalusi L., Endrényi L. An Exact Procedure for the Evaluation of Reference-Scaled Average Bioequivalence. *AAPS J* 2016; 18: 476–489. DOI:10.1208/s12248-016-9873-6.
55. Tóthfalusi L., Endrényi L. Algorithms for Evaluating Reference Scaled Average Bioequivalence: Power, Bias, and Consumer Risk. *Stat Med* 2017; 36: 4378–4390. DOI:10.1002/sim.7440.
56. 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: 4777–4788. DOI:10.1002/sim.7452.
57. Endrényi L., Tóthfalusi L. Bioequivalence for highly variable drugs: regulatory agreements, disagreements, and harmonization. *J Pharmacokin Pharmacodyn* 2019; 46: 117–126. DOI:10.1007/s10928-019-09623-w.
58. Deng Y., Zhou X.H. Methods to control the empirical type I error rate in average bioequivalence tests for highly variable drugs. *Stat Methods Med Res* 2019; 29: 1650–1667. DOI:10.1177/0962280219871589.
59. Ocaña J., Muñoz J. Controlling type I error in the reference-scaled bioequivalence evaluation of highly variable drugs. *Pharm Stat* 2019; 18: 583–599. DOI:10.1002/pst.1950.
60. Zimmerman D.W. A note on preliminary tests of equality of variances. *Br J Math Stat Psychol* 2004; 57(1): 173–181. DOI:10.1348/000711004849222.
61. Freeman P.R. The performance of the two-stage analysis of two-treatment, two-period crossover trials. *Stat Med* 1989; 8(12): 1421–1432. DOI:10.1002/sim.4780081202.
62. European Medicines Agency, Committee for Human Medicinal Product. Guideline on multiplicity issues in clinical trials. London. 15 December 2016. https://www.ema.europa.eu/en/documents/scientific-guideline/draft-guideline-multiplicity-issues-clinical-trials_en.pdf (accessed 04 July 2022).
63. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. Statistical Principles for Clinical Trials. E9. 5 February 1998. https://database.ich.org/sites/default/files/E9_Guideline.pdf (accessed 14 May 2022).
64. Schütz H. ABEL: Type I Error, https://forum.bebac.at/forum_entry.php?id=21947 28 September 2020. (accessed 14 May 2022).
65. U.S. Food and Drug Administration, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research. Guidance for Industry. Adaptive Designs for Clinical Trials of Drugs and Biologics. November 2019. https://www.fda.gov/media/78495/download (accessed 14 May 2022).
66. Benet L.Z. Bioavailability and Bioequivalence Definitions and Difficulties in Acceptance Criteria, in Midha, K.K., Blume, H.H. (eds), Bio-International 2. Bioavailability, Bioequivalence and Pharmacokinetics. medpharm Scientific Publishers, Stuttgart, Germany, pp 27–35, 1993.
67. Ormsby E.D. Are We Expecting too much from Replicate Designs?, in Blume, H.H., Midha, K.K. (eds), Bio-International 2. Bioavailability, Bioequivalence and Pharmacokinetic Studies. medpharm Scientific Publishers, Stuttgart, Germany, pp 129–134, 1995.
68. McGilveray I.J. An Overview of Problems and Progress at Bio-Internationals ‘89 and ‘92, in Blume, H.H., Midha, K.K. (eds), BioInternational 2. Bioavailability, Bioequivalence and Pharmacokinetic Studies. medpharm Scientific Publishers, Stuttgart, Germany, pp 109–115, 1995.

Critical Remarks on Reference-Scaled Average Bioequivalence

Helmut Schütz[1,2], Detlew Labes[3], Martin J Wolfsegger[4]

[1] BEBAC. Neubaugasse 36/11, 1070 Vienna, Austria
[2] Institute of Medical Statistics of the Medical University of Vienna, Austria, ORCID ID 0000-0002-1167-7880;
[3] Consultant. Lössauer Straße 18, 13055 Berlin, Germany, ORCID ID 0000-0003-2169-426X;
[4] Baxalta Innovations GmbH, a Takeda company. Donau-City-Straße 7, 1220 Vienna, Austria, ORCID ID 0000-0003-2419-4324.

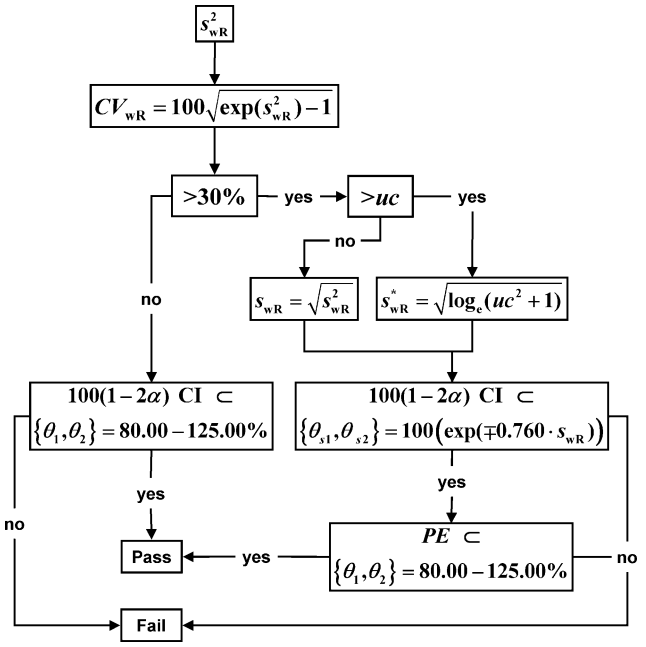

Fig. S1. Framework of Average Bioequivalence with Expanding Limits.

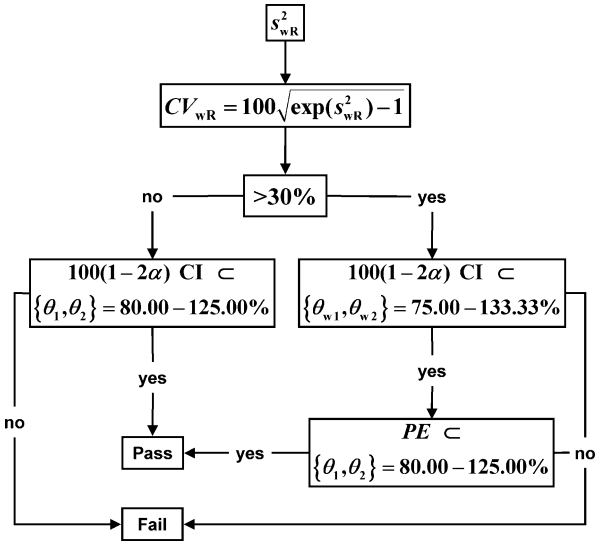

Fig. S2. Framework of the Gulf Cooperation Council.

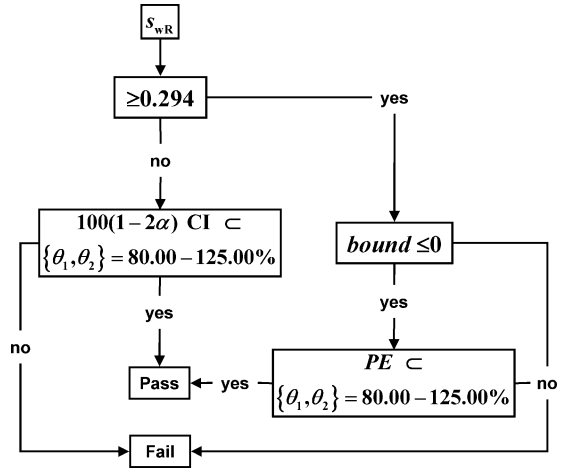

Fig. S3. Framework of Reference-Scaled Average Bioequivalence.

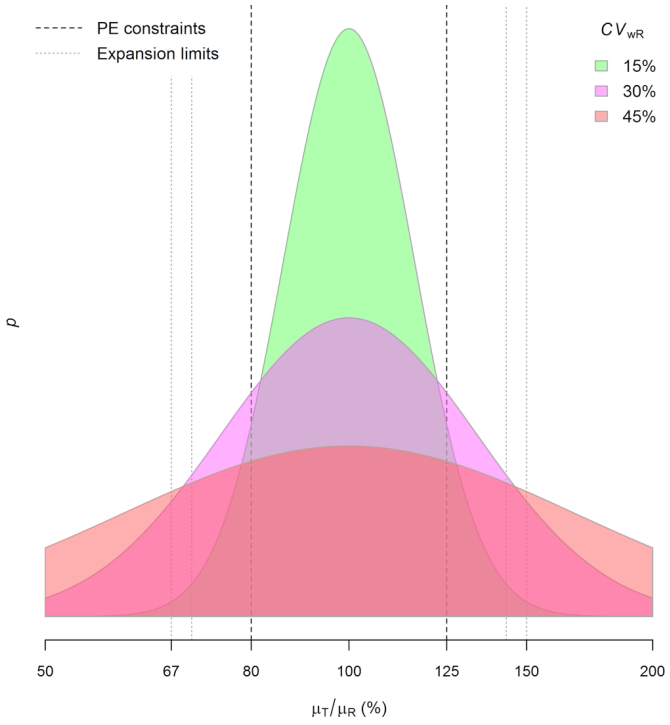

Fig. S4. Larger deviations between the geometric mean ratios are a direct consequence of higher variability. The point estimate constraint – together with the upper cap in most jurisdictions – leads to a truncated distribution whose statistical properties are essentially unknown. Hence, the test assuming the normal distribution of $\log_e$-transformed data is not correct in the strict sense.

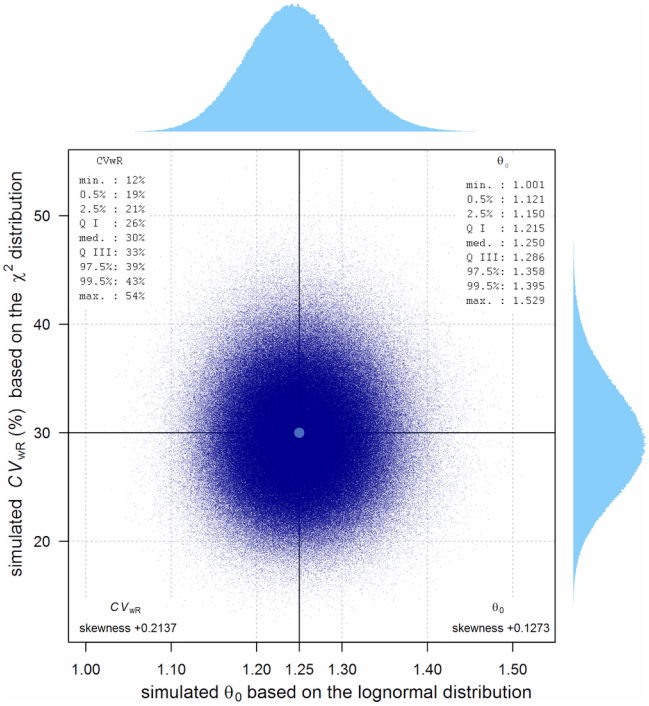

Fig. S5. Illustrative example of skewed distributions. One million simulations with true CV_{wR} = 30% and θ_0 = 1.25 (n = 24).

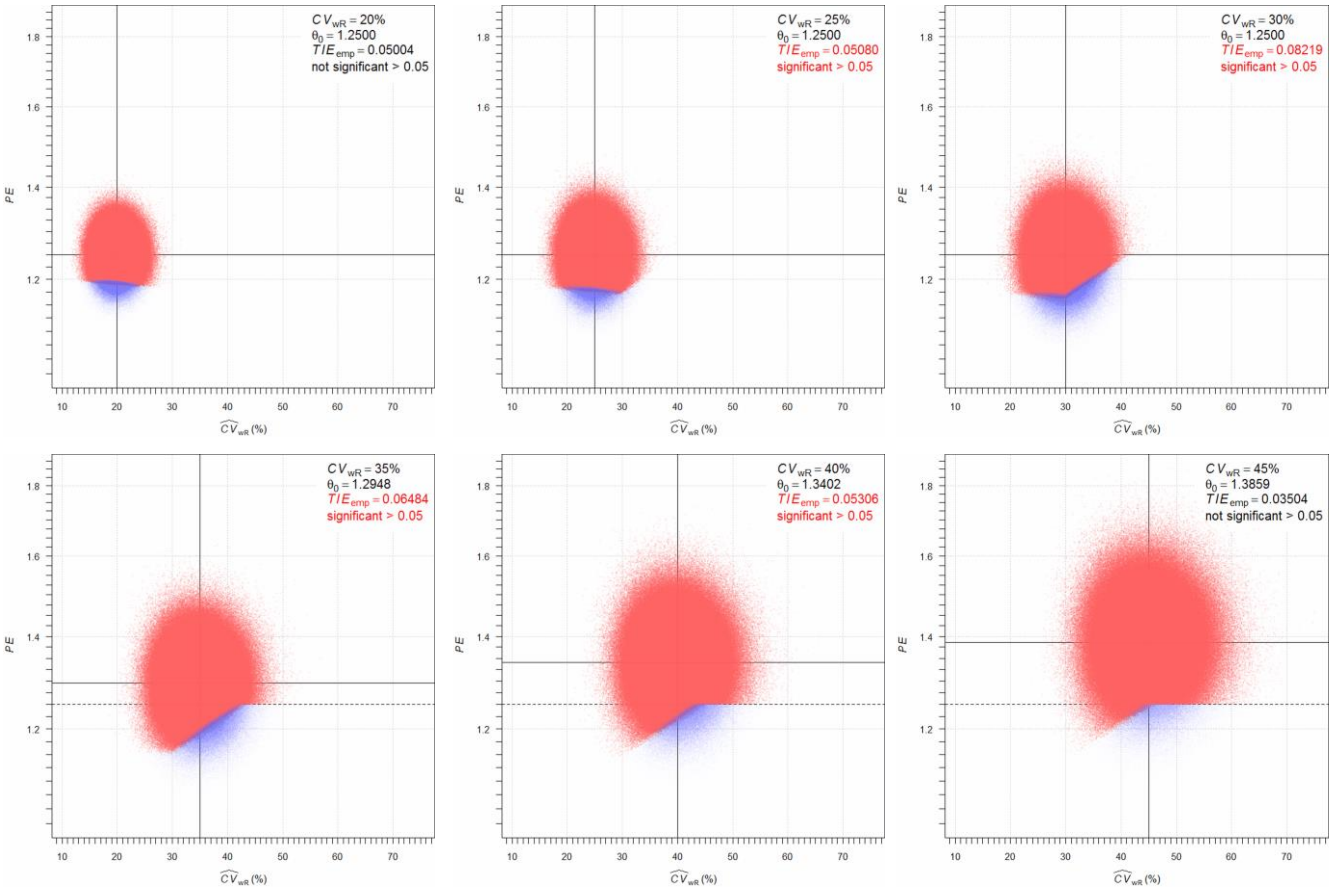

Fig. S6. One million simulations of studies in a two sequence, four period full replicate design (n = 48), assessment for ABEL (uc = 50%). True CV_{wR} = 20 – 45% → θ_0 = θ_{0s2} (solid lines). If applicable, upper PE-constraint 1.25 (dashed line). Y-axis in log-scale. Red dots represent studies which failed BE (null hypothesis correctly accepted), blue dots represent studies which passed BE (null hypothesis falsely rejected). Empiric Type I Error = number of passing studies / 10^6, significance limit of the binomial test 0. 050360.

R-script to reproduce Table 3

```
#########################################################
# Parametric 95% confidence interval of CVwR based on #
# the chi²-distribution of its associated variance.   #
#########################################################
library(PowerTOST)       # at least v1.1-0 (2013-02-08)
n    <- seq(24L, 48L, 12L)
CVwR <- seq(20L, 65L,  5L)
txt  <- "95% confidence intervals of CVwR\ndependent on the sample size.\n"
tmp  <- data.frame(n = rep(n, each = length(CVwR)), CVwR = CVwR,
                   lo = NA_real_, hi = NA_real_)
for (i in 1:nrow(tmp)) {
  tmp[i, 3:4] <- round(100*CVCL(CV = tmp$CVwR[i] / 100, df = tmp$n[i] - 2,
                                alpha = 0.05, side = "2-sided"), 1)
}
tmp <- reshape(tmp, direction = "wide", idvar = "CVwR", timevar = "n")
res <- data.frame(CVwR = CVwR, n.24 = NA_real_, n.36 = NA_real_, n.48 = NA_real_)
for (i in seq_along(CV)) {
  res$n.24[i] <- sprintf("%5.1f \u2013 %5.1f", tmp[i, 2], tmp[i, 3])
  res$n.36[i] <- sprintf("%5.1f \u2013 %5.1f", tmp[i, 4], tmp[i, 5])
  res$n.48[i] <- sprintf("%5.1f \u2013 %5.1f", tmp[i, 6], tmp[i, 7])
}
names(res)[2:4] <- paste0("    n = ", n)
cat(txt); print(res, right = FALSE, row.names = FALSE)
```

Should give:

```
   95% confidence intervals of CVwR
   dependent on the sample size.
    CVwR     n = 24        n = 36        n = 48
    20    15.4 –  28.6  16.1 –  26.4  16.6 –  25.3
    25    19.2 –  35.9  20.1 –  33.1  20.7 –  31.7
    30    23.0 –  43.4  24.1 –  39.9  24.8 –  38.2
    35    26.8 –  51.0  28.0 –  46.8  28.8 –  44.7
    40    30.5 –  58.8  31.9 –  53.9  32.9 –  51.4
    45    34.1 –  66.8  35.8 –  61.0  36.9 –  58.1
    50    37.8 –  75.1  39.6 –  68.3  40.8 –  65.0
    55    41.4 –  83.5  43.4 –  75.8  44.7 –  71.9
    60    44.9 –  92.3  47.2 –  83.4  48.6 –  79.0
    65    48.4 – 101.3  50.9 –  91.2  52.5 –  86.2
```

R-script to reproduce Table 4 (and sample sizes up to 144)

```
################################
# Empiric Type I Error at the  #
# CVwR of maximum inflation.   #
################################
library(PowerTOST)             # At least v1.5-3 (2021-01-18) for GCC
design  <- "2x2x4"             # 2-sequence 4-period full replicate
n       <- seq(24L, 144L, 12L) # Sample sizes
CV      <- c(rep(0.30, 5 * length(n)), rep(se2CV(0.25), length(n))) # True CVwR
methods <- c("ABE",            # ABE   (all jurisdictions)
             "ABEL1",          # ABEL  (EMA and others)
             "ABEL2",          # ABEL  (Health Canada)
             "ABEL3",          # ABEL  (GCC)
             "RSABE1",         # RSABE ('implied limits')
             "RSABE2")         # RSABE ('desired consumer risk model')
result  <- data.frame(method = rep(methods, each = length(n)),
                      reg = c(rep(NA, length(n)),
                              rep("EMA", length(n)),
                              rep("HC", length(n)),
                              rep("GCC", length(n)),
                              rep("FDA", 2 * length(n))),
                      CV  = CV, swR = CV2se(CV), n = n, U = 1.25,
                      TIE = NA_real_, SE = NA_real_, signif = "  no ")
limit   <- binom.test(0.05 * 1e6L, 1e6L, alternative = "less")$conf.int[[2]]
pb      <- txtProgressBar(0, 1, 0, width = NA, style = 3)
for (i in 1:nrow(result)) {
  if (result$method[i] == "ABE") {    # All jurisdictions (exact, no simulations)
    result$TIE[i] <- power.TOST(CV = result$CV[i], n = result$n[i], design = design,
                                theta0 = result$U[i])
  }
  if (result$method[i] == "ABEL1") {  # EMA and others
    result$U[i]   <- scABEL(CV = result$CV[i], regulator = result$reg[i])[["upper"]]
    result$TIE[i] <- power.scABEL(CV = result$CV[i], n = result$n[i], design = design,
                                  theta0 = result$U[i], regulator = result$reg[i], nsims = 1e6L)
  }
  if (result$method[i] == "ABEL2") {  # Health Canada
    result$U[i]   <- scABEL(CV = result$CV[i], regulator = result$reg[i])[["upper"]]
    result$TIE[i] <- power.scABEL(CV = result$CV[i], n = result$n[i], design = design,
                                  theta0 = result$U[i], regulator = result$reg[i], nsims = 1e6L)
  }
```

```
  if (result$method[i] == "ABEL3") {  # Gulf Cooperation Council
    result$U[i]   <- scABEL(CV = result$CV[i], regulator = result$reg[i])[["upper"]]
    result$TIE[i] <- power.scABEL(CV = result$CV[i], n = result$n[i], design = design,
                                  theta0 = result$U[i], regulator = result$reg[i], nsims = 1e6L)
  }
  if (result$method[i] == "RSABE1") { # FDA 'implied limits'
    result$U[i]   <- scABEL(CV = result$CV[i], regulator = result$reg[i])[["upper"]]
    result$TIE[i] <- power.RSABE(CV = result$CV[i], n = result$n[i], design = design,
                                 theta0 = result$U[i], regulator = result$reg[i], nsims = 1e6L)
    }
  if (result$method[i] == "RSABE2") { # FDA 'desired consumer risk model'
    if (result$swR[i] <= 0.25) {
      result$U[i] <- scABEL(CV = result$CV[i], regulator = result$reg[i])[["upper"]]
    } else {
      result$U[i] <- exp(result$swR[i] * log(1.25) / 0.25)
    }
    result$TIE[i] <- power.RSABE(CV = result$CV[i], n = result$n[i], design = design,
                                 theta0 = result$U[i], nsims = 1e6L)
  }
  setTxtProgressBar(pb, i / nrow(result))
}
result$signif[result$TIE > sign.limit] <- " yes "
result$SE  <- sqrt(0.5 * result$TIE / 1e6)
result$SE  <- sprintf("%.6f", result$SE)
ABE.idx    <- which(result$method == "ABE")
ABEL1.idx  <- which(result$method == "ABEL1")
ABEL2.idx  <- which(result$method == "ABEL2")
ABEL3.idx  <- which(result$method == "ABEL3")
RSABE1.idx <- which(result$method == "RSABE1")
RSABE2.idx <- which(result$method == "RSABE2")
result$CV  <- sprintf("%.3f%%", 100 * result$CV)
result$TIE <- sprintf("%.4f", result$TIE)
result$method[ABE.idx[1]]     <- "ABE (all jurisdictions)"
result$method[ABE.idx[-1]]    <- ""
result$method[ABEL1.idx[1]]   <- "ABEL (EMA and others)"
result$method[ABEL1.idx[-1]]  <- ""
result$method[ABEL2.idx[1]]   <- "ABEL (Health Canada)"
result$method[ABEL2.idx[-1]]  <- ""
result$method[ABEL3.idx[1]]   <- "ABEL (GCC)"
result$method[ABEL3.idx[-1]]  <- ""
result$method[RSABE1.idx[1]]  <- "RSABE (\u2018implied limits\u2019)"
result$method[RSABE1.idx[-1]] <- ""
result$method[RSABE2.idx[1]]  <- "RSABE (\u2018desired consumer risk model\u2019)"
result$method[RSABE2.idx[-1]] <- ""
result                        <- result[, -c(2, 4, 6)]
names(result)[2:6]            <- c("  CVwR", " n", " TIE", "   SE", "sign.")
print(result, row.names = FALSE, right = FALSE)
```

Should give:

```
   method                              CVwR   n   TIE      SE    sign.
   ABE (all jurisdictions)           30.000%  24 0.0500 0.000158   no
                                     30.000%  36 0.0500 0.000158   no
                                     30.000%  48 0.0500 0.000158   no
                                     30.000%  60 0.0500 0.000158   no
                                     30.000%  72 0.0500 0.000158   no
                                     30.000%  84 0.0500 0.000158   no
                                     30.000%  96 0.0500 0.000158   no
                                     30.000% 108 0.0500 0.000158   no
                                     30.000% 120 0.0500 0.000158   no
                                     30.000% 132 0.0500 0.000158   no
                                     30.000% 144 0.0500 0.000158   no
   ABEL (EMA and others)             30.000%  24 0.0804 0.000200  yes
                                     30.000%  36 0.0819 0.000202  yes
                                     30.000%  48 0.0823 0.000203  yes
                                     30.000%  60 0.0827 0.000203  yes
                                     30.000%  72 0.0831 0.000204  yes
                                     30.000%  84 0.0836 0.000204  yes
                                     30.000%  96 0.0834 0.000204  yes
                                     30.000% 108 0.0839 0.000205  yes
                                     30.000% 120 0.0838 0.000205  yes
                                     30.000% 132 0.0843 0.000205  yes
                                     30.000% 144 0.0840 0.000205  yes
```

```
   ABEL (Health Canada)                  30.000%  24 0.0841 0.000205  yes
                                         30.000%  36 0.0846 0.000206  yes
                                         30.000%  48 0.0846 0.000206  yes
                                         30.000%  60 0.0852 0.000206  yes
                                         30.000%  72 0.0851 0.000206  yes
                                         30.000%  84 0.0855 0.000207  yes
                                         30.000%  96 0.0855 0.000207  yes
                                         30.000% 108 0.0855 0.000207  yes
                                         30.000% 120 0.0855 0.000207  yes
                                         30.000% 132 0.0857 0.000207  yes
                                         30.000% 144 0.0857 0.000207  yes
   ABEL (GCC)                            30.000%  24 0.1493 0.000273  yes
                                         30.000%  36 0.1931 0.000311  yes
                                         30.000%  48 0.2324 0.000341  yes
                                         30.000%  60 0.2621 0.000362  yes
                                         30.000%  72 0.2664 0.000365  yes
                                         30.000%  84 0.2668 0.000365  yes
                                         30.000%  96 0.2670 0.000365  yes
                                         30.000% 108 0.2684 0.000366  yes
                                         30.000% 120 0.2683 0.000366  yes
                                         30.000% 132 0.2689 0.000367  yes
                                         30.000% 144 0.2688 0.000367  yes
   RSABE ('implied limits')              30.000%  24 0.1335 0.000258  yes
                                         30.000%  36 0.1536 0.000277  yes
                                         30.000%  48 0.1708 0.000292  yes
                                         30.000%  60 0.1871 0.000306  yes
                                         30.000%  72 0.2006 0.000317  yes
                                         30.000%  84 0.2134 0.000327  yes
                                         30.000%  96 0.2245 0.000335  yes
                                         30.000% 108 0.2336 0.000342  yes
                                         30.000% 120 0.2418 0.000348  yes
                                         30.000% 132 0.2494 0.000353  yes
                                         30.000% 144 0.2551 0.000357  yes
   RSABE ('desired consumer risk model') 25.396%  24 0.0663 0.000182  yes
                                         25.396%  36 0.0629 0.000177  yes
                                         25.396%  48 0.0600 0.000173  yes
                                         25.396%  60 0.0576 0.000170  yes
                                         25.396%  72 0.0557 0.000167  yes
                                         25.396%  84 0.0542 0.000165  yes
                                         25.396%  96 0.0532 0.000163  yes
                                         25.396% 108 0.0523 0.000162  yes
                                         25.396% 120 0.0517 0.000161  yes
                                         25.396% 132 0.0514 0.000160  yes
                                         25.396% 144 0.0508 0.000159  yes
```

R-script to reproduce the example

```
################################################################################
# AUC data from Reference (36) were evaluated in the R-package replicateBE v1.1.3 #
# (ABEL: WHO, GCC) and Phoenix WinNonlin v8.3.4 (ABEL: HC, FDA: RSABE).        #
################################################################################
library(PowerTOST) # at least v1.5-3 (2021-01-18)
data <- data.frame(subject = c(rep(1L:2L, each = 4),   rep(3L, 2),  rep(4L:13L, each = 4),
                               rep(15L:26L, each = 4), rep(27L, 2), rep(28L:40L, each = 4),
                               rep(42L:50L, each = 4), rep(52L:57L, each = 4)),
                   period = c(rep(1L:4L, 2), 1L:2L, rep(1L:4L, 10), rep(1L:4L, 12),
                              1L:2L, rep(1L:4L, 13), rep(1L:4L, 9), rep(1L:4L, 6)),
                   sequence = c(rep("RTRT", 4), rep("TRTR",  4), rep("RTRT",  2), rep("TRTR",  4),
                                rep("RTRT", 8), rep("TRTR", 12), rep("RTRT",  4), rep("TRTR",  4),
                                rep("RTRT", 4), rep("TRTR",  4), rep("RTRT",  4), rep("TRTR",  4),
                                rep("RTRT", 8), rep("TRTR",  8), rep("RTRT",  4), rep("TRTR",  8),
                                rep("TRTR", 8), rep("TRTR",  6), rep("RTRT",  8), rep("TRTR",  4),
                                rep("RTRT", 4), rep("TRTR",  8), rep("RTRT",  8), rep("TRTR", 12),
                                rep("RTRT", 8), rep("TRTR",  8), rep("RTRT",  4), rep("TRTR",  4),
                                rep("RTRT", 4), rep("TRTR",  4), rep("RTRT", 12), rep("TRTR",  4),
                                rep("RTRT", 8), rep("TRTR",  8), rep("RTRT",  4)),
                   treatment = NA_character_,
                   AUC = c(812.6, 1173.7, 889.1, 620.1, 216.3, 338.0,  502.8, 398.6, 545.1, 542.9,
                           632.6,  520.0, 716.7, 860.4, 400.0, 223.8,  173.7, 289.7, 102.1, 185.3,
                            42.0,   88.3, 596.0, 659.3, 543.8, 662.9,  402.4, 359.8, 590.8, 444.3,
                           456.7,  378.4, 477.5, 407.9, 304.5, 351.5,  520.2, 335.7, 500.7, 323.0,
                           416.3,  525.1, 176.1, 710.7, 409.5, 645.5,  160.6, 218.0, 170.1, 124.6,
                           562.4,  490.4, 504.7, 675.9, 756.0, 606.8,  477.4, 626.8, 207.5, 271.6,
                           173.7,  240.5, 571.3, 705.2, 619.0, 633.6,  511.9, 549.7, 388.2, 141.0,
                           124.0,   91.9, 113.3,  59.5, 536.1, 595.2,  445.5, 521.5, 239.7, 265.1,
                           445.9,  433.2, 609.6, 371.6, 511.3, 432.7,  449.9, 860.4, 606.8, 577.2,
                           192.5,  220.1, 233.1, 227.0, 764.4, 508.8,  757.8, 449.4, 151.9, 194.8,
                           568.1,  321.1, 338.3, 403.6, 735.6, 634.5, 1244.2, 641.9, 429.1, 391.8,
                           316.9,  335.1, 307.4, 481.8, 346.6, 369.7,  409.0, 514.6, 763.1, 406.5,
                           271.0,  221.0, 296.5, 463.7, 292.9, 431.0,  448.5, 267.8, 217.2, 332.2,
                           103.0,  127.5, 290.8, 208.6, 243.7, 489.7,  297.2, 502.0, 320.4, 334.3,
```

```
                          163.8,  232.1, 636.9, 434.9, 368.3, 292.6,  446.1, 222.3, 193.7, 202.8,
                          255.2,  244.3, 534.1, 243.1, 418.4, 441.9,  355.1, 415.2, 382.7, 334.0,
                          102.0,  282.5, 245.6, 286.2, 320.5, 233.9,  331.7, 260.5, 223.6, 645.4,
                          349.0,  507.4, 504.5, 289.9, 550.7, 244.2,  615.8, 732.1, 620.9, 665.2,
                          898.4,  924.9, 398.3, 828.3, 410.4, 329.2,  449.4, 442.1, 237.0, 505.0,
                          496.3,  580.6, 332.4, 273.6, 525.3, 293.3,  185.2, 222.9, 182.1, 194.1,
                          246.9,  620.9, 678.3, 752.2, 235.4, 190.4,  318.3, 248.4, 180.6, 174.7,
                          102.9,  117.0))
for (i in 1:nrow(data)) {      # extract treatments from sequences and periods
  treatments        <- unlist(strsplit(data$sequence[i], split = ""))
  data$treatment[i] <- treatments[data$period[i]]
}
nsims      <- 1e6L             # number of simulations for empiric Type I Error
sign.limit <- binom.test(0.05 * nsims, n = nsims, alternative = "less)$conf.int[[2]]
txt        <- paste0("Significance limit based on the binomial test for ",
                     formatC(1e6L, digits = 0, big.mark = ","), " simulations = ",
                     sprintf("%.6f", sign.limit), ".\n")
method     <- c("ABEL  (WHO)", # ANOVA (EMA 'Method A')
                "ABEL  (GCC)", # ANOVA
                "ABEL  (HC) ", # mixed-effects model
                "RSABE (FDA)") # ANDA guidance
regulator  <- c("EMA", "GCC", "HC", "FDA")
result     <- data.frame(method  = method, regulator = regulator,
                         CVwR    = c(rep(0.355648, 3), 0.353968),
                         L       = NA_real_, U = NA_real_,
                         lower   = c(rep(1.017529, 2), 1.016937, NA_real_),
                         upper   = c(rep(1.174621, 2), 1.175704, NA_real_),
                         PE      = c(rep(1.093257, 2), 1.093257, 1.097784),
                         Delta.r = NA_real_,
                         bound   = c(rep(NA_real_, 3), -0.0555212),
                         BE      = "fail", TIE = NA_real_, signif = "",
                         stringsAsFactors = FALSE) # last line required only for R <4.0
for (i in seq_along(method)) {
  if (method[i] == "ABEL  (WHO)") {
    result[i, 4:5] <- scABEL(CV = result$CVwR[i], regulator = regulator[i])
    result$Delta.r[i] <- 1 - result$L[i]
    if (result$lower[i] >= result$L[i] & result$upper[i] >= result$U[i] & # check CI inclusion
        result$PE[i] >= 0.80 & result$PE[i] <= 1.25)                      # check PE constraint
      result$BE <- "pass"
    result$TIE[i]  <- power.scABEL(CV = result$CVwR[i], n = 54, design = "2x2x4",
                                   theta0 = result$U[i], regulator = regulator[i], nsims = nsims)
    if (result$TIE[i] > limit) result$signif[i] <- "*  "                  # check significant TIE
  }
  if (method[i] == "ABEL  (GCC)") {
    result[i, 4:5] <- scABEL(CV = result$CVwR[i], regulator = regulator[i])
    result$Delta.r[i] <- 1 - result$L[i]
    if (result$lower[i] >= result$L[i] & result$upper[i] >= result$U[i] & # check CI inclusion
        result$PE[i] >= 0.80 & result$PE[i] <= 1.25)                      # check PE constraint
      result$BE <- "pass"
    result$TIE[i]  <- power.scABEL(CV = result$CVwR[i], n = 54, design = "2x2x4",
                                   theta0 = result$U[i], regulator = regulator[i], nsims = nsims)
    if (result$TIE[i] > limit) result$signif[i] <- "*  "                  # check significant TIE
  }
  if (method[i] == "ABEL  (HC) ") {
    result[i, 4:5] <- scABEL(CV = result$CVwR[i], regulator = result$regulator[i])
    result$Delta.r[i] <- 1 - result$L[i]
    if (result$lower[i] >= result$L[i] & result$upper[i] >= result$U[i] & # check CI inclusion
        result$PE[i] >= 0.80 & result$PE[i] <= 1.25)                      # check PE constraint
      result$BE <- "pass"
    result$TIE[i]   <- power.scABEL(CV = result$CVwR[i], n = 54, design = "2x2x4",
                                    theta0 = result$U[i], regulator = regulator[i], nsims = nsims)
    if (result$TIE[i] > limit) result$signif[i] <- "*  "                  # check significant TIE
  }
  if (method[i] == "RSABE (FDA)") {
    result[i, 4:5] <- scABEL(CV = result$CVwR[i], regulator = regulator[i])
    result$Delta.r[i] <- 1 - result$L[i]
    if (result$bound[i] <= 0 &                                            # check bound
        result$PE[i] >= 0.80 & result$PE[i] <= 1.25)                      # check PE constraint
      result$BE <- "pass"
    result$TIE[i]  <- power.RSABE(CV = result$CVwR[i], n = 54, design = "2x2x4",
                                  theta0 = result$U[i], regulator = regulator[i], nsims = nsims)
    if (result$TIE[i] > limit) result$signif[i] <- "*  "                  # check significant TIE
  }
}
result         <- result[, -which(names(result) == "regulator")]
result$CVwR    <- sprintf("%.2f%%", 100 * result$CVwR)
result$L       <- sprintf("%.2f%%", 100 * result$L)
result$U       <- sprintf("%.2f%%", 100 * result$U)
result$lower   <- sprintf("%.2f%%", 100 * result$lower)
result$upper   <- sprintf("%.2f%%", 100 * result$upper)
result$PE      <- sprintf("%.2f%%", 100 * result$PE)
result$Delta.r <- sprintf("%.2f%%", 100 * result$Delta.r)
result$bound   <- signif(result$bound, 4)
result$TIE     <- signif(result$TIE, 3)
result[1:3, 9] <- "\u2013     "
result[4, 5:6] <- "\u2013   "
```

```
121 cat(txt); print(result, row.names = FALSE)
```

Should give:

```
Significance limit based on the binomial test for 1,000,000 simulations = 0.050360.
      method   CVwR      L       U   lower   upper      PE Delta.r    bound   BE    TIE signif
 ABEL  (WHO) 35.56% 76.93% 129.99% 101.75% 117.46% 109.33%  23.07%   -      pass 0.0643    *
 ABEL  (GCC) 35.56% 75.00% 133.33% 101.75% 117.46% 109.33%  25.00%   -      pass 0.0459
 ABEL  (HC)  35.56% 76.93% 129.99% 101.69% 117.57% 109.33%  23.07%   -      pass 0.0651    *
 RSABE (FDA) 35.40% 73.59% 135.89%    -       -    109.78%  26.41% -0.05552 pass 0.0232
```