Journal of Pharmacokinetics and Pharmacodynamics (2019) 46:117–126
https://doi.org/10.1007/s10928-019-09623-w

REVIEW PAPER

# Bioequivalence for highly variable drugs: regulatory agreements, disagreements, and harmonization

Laszlo Endrenyi[1] · Laszlo Tothfalusi[2]

Received: 18 September 2018 / Accepted: 12 February 2019 / Published online: 23 February 2019


## Abstract

Regulatory authorities introduced procedures in the last decade for evaluating the bioequivalence (BE) for highly variable drugs. These approaches are similar in principle but differ in details. For example, the Food and Drug Administration (FDA) and the European Medicines Agency (EMA) recommend differing regulatory constants. The constant suggested by FDA results in discontinuity of the BE limits around the switching variation at 30% observed within-subject variation of the reference product. The regulatory constant of EMA does not have these problems. The Type I error reaches 6–7% around the switching variation with the EMA constant but 16–17% with the FDA constant. Various procedures were recently suggested, especially for the EMA approach, to eliminate the inflation of the Type I error. Notably, the so-called Exact algorithms try to amalgamate the positive features of both EMA and FDA procedures without their negative sides. The computational procedure for the EMA approach is simple and has a straightforward interpretation. The procedure for the FDA approach is based on an approximation, has a bias at small degrees of freedom, and requires a suitable computer program. All regulatory agencies impose a second requirement constraining the point estimate of the ratio of geometric means. In addition, EMA and Health Canada impose an upper limit for applying the recommended procedures. These expectations have psychological motivation and political rationale but no scientific foundations. Their inclusion results in incorrect and misleading interpretation of the principal criterion which involves confidence intervals. Different regulatory authorities expect to apply their approaches either to both AUC and $C_{max}$ or only to AUC or only to $C_{max}$. Rational resolution of the disharmonization is needed.

**Keywords** Bioequivalence · Highly variable drugs · Reference-scaled average bioequivalence · Regulatory constants · Type I error

## Introduction

The determination of bioequivalence (BE) for highly variable drugs was frustrating for many years. The BE metrics of these drugs have large within-subject variation, exceeding 30% by international consensus [1]. Consequently, in order to satisfy the usual regulatory criterion (the 90% confidence interval for the ratio of the geometric means of the two drug products should be between 0.80 and 1.25) very large, at times unethically large, numbers of subjects are required.

Regulatory authorities introduced new procedures in the last decade. They achieved their main goal by meaningfully lowering the needed sample sizes while preserving power. The approaches of the various agencies are similar in principle but differ in details. The differences between the regulatory requirements and procedures of the authorities have important scientific and clinical consequences.

This communication will highlight both the similarities and differences between the regulatory approaches and procedures for the determination of BE of highly variable drugs. Prospects for their resolution and harmonization will be suggested.

We focused only on the statistical and regulatory aspects without explaining the reasons of high variability. It has

✉ Laszlo Endrenyi
l.endrenyi@utoronto.ca

[1] Department of Pharmacology and Toxicology, University of Toronto, Toronto, ON M5S 1A8, Canada

[2] Semmelweis University, Budapest, Hungary

more than one source including molecular properties such as high first-pass metabolism [2, 3] and formulation factors [4–6]. In fact, instead of highly variable drugs, it is better to talk about highly variable drugs and drug products (HV/HVDP). Still, the list is not complete. The variability depends on the pharmacokinetic metrics, for example the variability of $C_{max}$ is usually higher than that of AUC. For endogenous molecules (like progesterone) the drug level depends on the endogenous concentration, and the variability of free (not protein bound) concentration can be much higher than the total plasma level.

## Average bioequivalence

The usual approach for the evaluation of BE between two drug products requires that the difference between the logarithmic means of the test ($\mu_T$) and reference ($\mu_R$) formulations should be between preset regulatory limits ($\theta_1$ and $\theta_2$) [7]:

$$\theta_1 \le \mu_T - \mu_R \le \theta_2 \quad (1)$$

The BE limits are usually symmetrical in the logarithmic scale, $-\theta_1 = \theta_2 = \theta_A$, and therefore:

$$-\theta_A \le \mu_T - \mu_R \le \theta_A \quad (2)$$

Regulatory authorities set the value of the BE limit most frequently at $\theta_A = \ln(1.25)$. Equation (2) in the original scale corresponds to the population geometric mean ratio (GMR) between the two formulations which must be between the limits:

$$0.80 \le \mathrm{GMR} = \exp(\mu_T/\mu_R) \le 1.25 \quad (3)$$

In the practical implementation of the regulatory criterion for (unscaled) average BE applies the estimated means ($m_T$ and $m_R$) of the two formulations. Symbolically:

$$-\theta_A \le m_T - m_R \le \theta_A \quad (4)$$

Actually, the last expression should be interpreted in greater detail: The 90% confidence interval of the difference between the estimated logarithmic means of the two drug products should stay between the regulatory limits, $\pm\,\theta_A$.

## Scaled average bioequivalence

The regulatory model for average BE (Eqs. 2 and 4) can be modified by applying a scaling factor [8–10]:

$$-\theta_S \le (m_T - m_R)/s_W \le \theta_S \quad (5)$$

$s_W$ is a standard deviation related to the intrasubject variation. It can be obtained, in two-period crossover studies, from the residual variation. In investigations with three or more periods, where at least one formulation is administered twice, the within-subject variation of one or both formulations can be directly estimated. The BE limit ($\theta_S$) is preset by the regulatory authorities.

The interpretation of the model for scaled average BE (SABE), in Eq. 5, is similar to that of the model for average BE in Eq. 4. The 90% confidence interval of the difference between the estimated scaled logarithmic means of the two drug products should be between the preset regulatory limits, $\pm\,\theta_S$.

SABE itself has interesting interpretations [11]. It is widely used in diverse areas of science. It is an index of standardized effect size [12] which can measure the importance and significance of clinical effects. SABE can also be derived, with simplifying assumptions, from the regulatory model for individual BE and thereby provides a measure of therapeutic switchability [13, 14].

## Implementation of scaled average bioequivalence by FDA and EMA

### Food and drug administration

The Food and Drug Administration (FDA) of the United States expects the use of SABE for determining the BE of highly variable drugs [10, 15]. The intrasubject standard deviation of the reference product should be estimated ($s_{WR}$) as a measure of within-subject variation. Consequently, the regulatory model of Eq. 5 becomes for the *reference-scaled average BE* (RSABE):

$$-\theta_S \le (m_T - m_R)/s_{WR} \le \theta_S \quad (6)$$

In order to obtain the estimated within-subject standard deviation of the reference formulation ($s_{WR}$), FDA suggests that the reference product should be measured at least twice in each subject. Therefore, replicate-design crossover investigations with either 3 or 4 periods should be undertaken.

The BE limit ($\theta_S$) is related to a regulatory constant ($\sigma_0$) according to:

$$\theta_S = \ln(1.25)/\sigma_0 \quad (7)$$

FDA recommends $\sigma_0 = 0.25$ (corresponding to a coefficient of variation of $CV_0 = 25\%$, or more precisely 25.4%) as the value of the regulatory constant, and therefore $\theta_S = 0.893$ as the BE limit in the logarithmic scale [10, 15–17].

FDA also notes that, by international agreement, a drug is considered to be highly variable if its within-subject variation is larger than 30% [1]. Therefore, FDA recommends that unscaled average BE applied if the *observed*

intraindividual variation of the reference product does not exceed $CV_{WR}$ = 30%, but that RSABE be utilized if the variation is higher than 30% [10, 15]. Consequences of this *mixed strategy* will be discussed later. These expectations should be applied to both primary pharmacokinetic metrics, AUC (the area under the curve contrasting plasma concentration with time) and $C_{max}$ (the maximum concentration).

For the calculation of confidence limits, FDA suggests a computational procedure [18]. For the computations, the RSABE model (Eq. 6) is squared and linearized:

$$(m_T - m_R)^2 - \theta_S^2 \cdot s_{WR}^2 \leq 0 \quad (8)$$

The 95% upper bound can be calculated from the distributions of the two components [19, 20]. BE is accepted if the upper bound is zero or negative and rejected if it is positive. FDA published also a SAS code [18].

FDA recommends a second regulatory requirement, additional to that based on RSABE. The point estimate of GMR should be between 0.80 and 1.25 [10, 15].

### European Medicines Agency

The approach of EMA for determining the BE of highly variable drugs [21] is similar in principle to that of FDA but differs in details. It recommends that the procedure of *Average Bioequivalence with Expanding Limits* (ABEL) be applied. Following Boddy et al. [22], the method utilizes a rearrangement of the RSABE model (Eq. 6):

$$-\theta_A = -\theta_S \times s_{WR} \leq m_T - m_R \leq \theta_S \times s_{WR} = \theta_A \quad (9)$$

Consequently, the BE limits widen in proportion with the estimated within-subject standard deviation of the reference formulation.

EMA also expects that the within-subject variation should be estimated directly from replicate-design 3- or 4-period crossover studies.

Similarly to FDA, EMA recommends the application of the mixed strategy, i.e. unscaled average BE should be applied when the observed intraindividual variation of the reference product does not exceed 30% but ABEL should be used at higher variations. However, EMA imposes an upper limit of 50% beyond which the BE limits should remain at $\pm$ ln(1.43) and unscaled average BE should be utilized again [21]. EMA, notably, suggests that, after a single oral administration, the procedure be applied only to $C_{max}$ but not to AUC.

EMA recommends $\sigma_0$ = 0.294 as the regulatory constant (corresponding to CV = 30%) and, therefore, $\theta_S$ = 0.760 as the BE limit in the logarihmic scale. The EMA published a SAS code in a Question and Answers document [23].

Similarly to FDA, EMA suggests, as a secondary regulatory criterion, that the point estimate of GMR should be between 0.80 and 1.25.

### Health Canada

The recommendations of Health Canada for the determination of BE of highly variable drugs [18] are similar to those of EMA with two exceptions. First, the requirements apply only to AUC but not to $C_{max}$. This is exactly the opposite of the expectations of EMA. Second, ABEL may be applied only to an upper limit of 57.4% for the within-subject coefficient of variation (corresponding to a maximum width of the BE limits from 66.7 to 150%).

## Comparisons of and comments on the approaches of FDA and EMA

### Regulatory constant and bioequivalence limits

As noted earlier, FDA suggests $\sigma_0$ = 0.25 as the regulatory constant for determining the BE of highly variable drugs. The corresponding BE limit is $\theta_S$ = 0.893. EMA suggests $\sigma_0$ = 0.294 with $\theta_S$ = 0.760.

The differences have important clinical and regulatory consequences. These will be discussed below.

First it is observed that, for regulatory purposes, the FDA regulatory expectation is more permissive than the EMA expectation. Under identical conditions (variations, number of subjects) it is more probable that BE can be demonstrated with the FDA than with the EMA requirement. This is a direct consequence of the differing regulatory constants. It was illustrated in simulations [9, 20, 24].

#### Continuity or discontinuity at $CV_{WR}$ = 30%

Figure 1 illustrates features of the BE limits around the *observed* within-subject variation of 30%. The EMA condition is shown on the left (Fig. 1a). Up to a variation of 30%, unscaled average BE is used and the BE limits remain constant, 0.80 and 1.25 (or − 0.223 and 0.223 in the logarithmic scale). Beyond the variation of 30%, the BE limits widen gradually. The transition between the two conditions is continuous. The reason is that the regulatory constant of $\sigma_0$ = 0.294 (corresponding to $CV_0$ = 30%) equals the variation of transition, the switching variation ($CV_S$) [20].

The FDA condition is illustrated on the right of Fig. 1 (Fig. 1b). Unscaled average BE is used again up to the *observed* variation of 30% and the BE limits remain constant. Beyond 30%, the BE limits widen but the distance

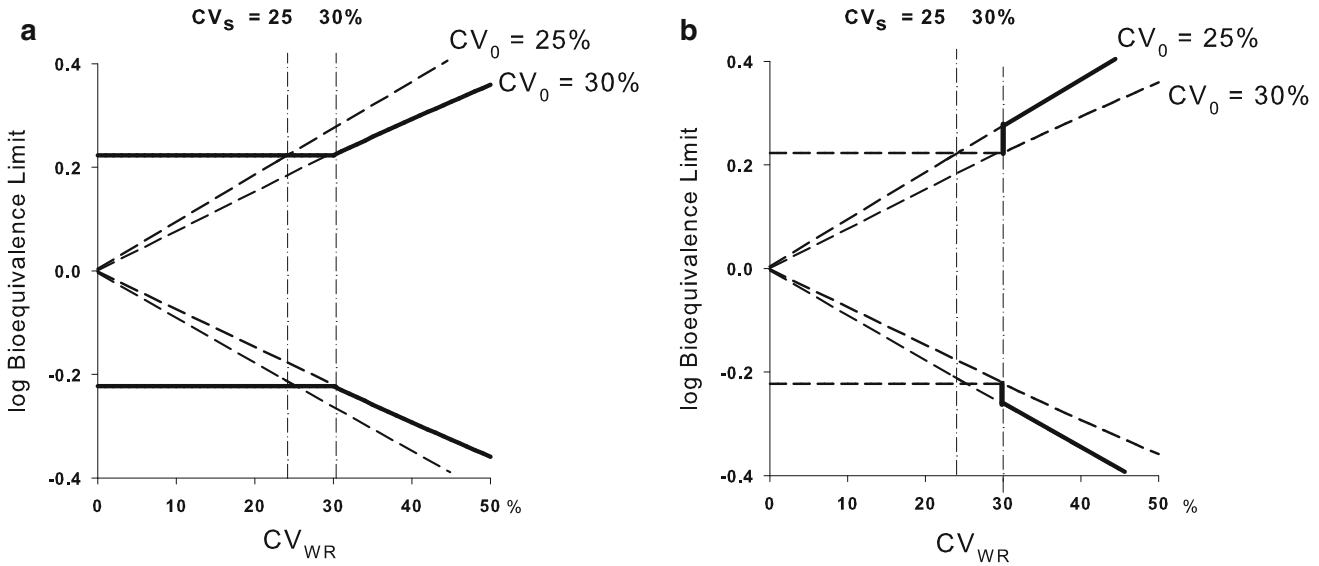


**Fig. 1** Mixed regulatory model for the determination of bioequivalence of highly-variable drugs. Thick lines show the logarithmic BE limits. They have a constant level of ± ln(1.25), and unscaled average BE is applied, when the within-subject variation of the reference product ($CV_{WR}$) does not exceed 30%. The limits widen at higher variations and either scaled average BE (SABE) or average BE with expanding limits (ABEL) is applied. The slope of the expansion (in the logarithmic scale) is determined by a regulatory constant which corresponds to a standardized variation ($CV_0$). **a** According the requirement of EMA, $CV_0$ = 30%; the BE limits are continuous. **b** Following the expectation of FDA, $CV_0$ = 25%; the BE limits are discontinuous around $CV_{WR}$ = 30%. ([25], with the permission of the publisher.)

between the two arms is larger than in the diagram on the left. Thus, the FDA condition is indeed more permissive than that of EMA. Also, the transition between the two conditions a $CV_{WR}$ = 30% is *discontinuous* now. The reason is that the regulatory constant of $\sigma_0$ = 0.25 (corresponding to $CV_0$ = 25%) differs from the switching variation [20].

Figure 2 illustrates that, when true, preset value of the within-subject variation of the reference product is assumed in simulations then the profiles for the acceptance of bioequivalence are remarkably different with the EMA and FDA regulatory constants, $\sigma_0$ = 0.294 and 0.25, respectively. Under the EMA condition, the acceptance changes continuously around $CV_{WR}$ = 30%. In contrast, there is a substantial discontinuity with the regulatory constant of FDA.

When the within-subject variation of the reference product is estimated in each simulated study then the profile of acceptance is smooth and continuous with both regulatory constants.

As discussed below, the divergence of the Type I error is a consequence of the different regulatory constants.

## Type I error

The Type I error is the probability of rejecting a null hypothesis when it is in fact correct. In the case of BE studies, the probability of accepting that the two products are bioequivalent when they are not. To protect the consumer, the error should be small, not larger than 5%. The Type I error is commonly called the consumer risk.

Therefore, it is disconcerting that larger Type I errors were found around the switching variation of 30%. As Table 1 illustrates, under the EMA condition the Type I error is about 6–7% [25]. This observation was repeatedly confirmed [26–29]. The Type I error inflation is due to multiple factors and various approaches have been proposed for the reduction and even elimination of the inflated error [26–31].

One of the reasons of the Type I error inflation around $CV_S$ = 30% is the mixed regulatory strategy. Assume, as an illustration, that the true within-subject variation of a drug is 29%, i.e. that it is not highly variable. Nevertheless, its observed variation could be by chance, say, 32% in a given investigation thereby classifying the drug as being highly variable. Figure 1 illustrates that the BE limits would then be wider and the probability of acceptance higher. Thus, the risk of false acceptance can be higher than 5%.

Table 1 illustrates also that the inflation of the Type I error around the switching variation of $CV_S$ = 30% is much larger, about 16–17%, with the FDA requirements than under the EMA conditions [25]. This observation was recently confirmed [28]. The mixed regulatory strategy accounts also for the larger inflation under the FDA than with the EMA condition. When the true within-subject

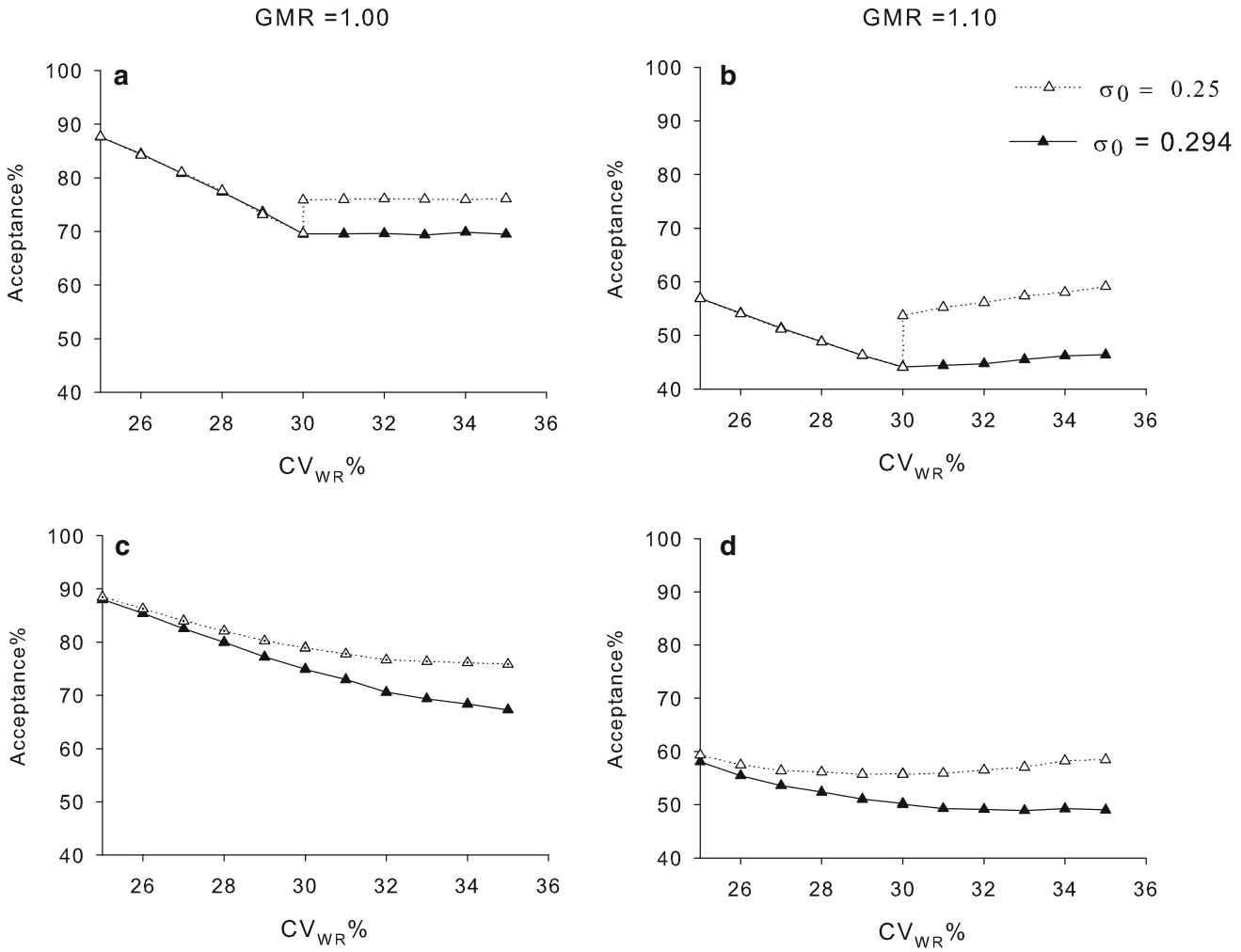


**Fig. 2** The percentage of simulated BE studies which satisfied the two one-sided test criterion at various within-subject variations of the reference product ($CV_{WR}$). The proportion of studies estimates the probability of acceptance. The three-period design TRR/RTR/RRT with 28 subjects was considered in the computations. 100,000 simulations were performed under each condition Either unscaled average BE was applied (Eq. 4) if the within-subject variation did not exceed 30%, or the recommended methods by the FDA and EMA if the variation was higher than 30%. For the value of $\sigma_0$, either 0.25 or 0.294 (i.e., either $\theta_S$ = 0.890 or 0.760) was considered. The analyses were undertaken by assuming that the ratio of the geometric mean (GMR) of their parameters was 1.00 (i.e., that the two formulations were truly equivalent) or that GMR was 1.10. **a** and **b** The true, preset within-subject variation of the reference product was used. **c** and **d** The within-subject variation estimated in each simulated study was utilized

variation is, for example, 29% than an observed variation at, say, 32% causes a much larger enhanced probability of acceptance with the FDA than under the EMA condition (comparison of Fig. 1a, b).

FDA apparently based its BE model by assuming that the regulatory constant is $\sigma_0 = 0.25$ and that the implied BE limits begin to widen at the within-subject standard deviation of 0.25 [15, 17]. It should be emphasized that the implied BE limits are plotted, in this picture, against the *true* (and not against the *observed*) intraindividual variation [15, 17]. Therefore, the inflation of the Type I error was examined and minimized at the standard deviation of 0.25 and not at 0.294.

## Comments

The differing regulatory constants, $\sigma_0 = 0.25$ by FDA and $\sigma_0 = 0.294$ by EMA, were seen to have consequences. They included the following:

- The actual BE limits change continuously around the switching variation of $CV_S = 30\%$ with the EMA assumption but discontinuously under the FDA condition.
- The EMA condition results in the inflation of the Type I error of 6–7% (and higher), the FDA assumption in 16–17% (and higher) around the switching variation. The mixed strategy inflates the Type I error because of a misclassification error.

**Table 1** Type I error for the determination of BE

| Mixed strategy | Regulatory standardized variation (%) | Type I error (%) | | |
|---|---|---|---|---|
| | | Unscaled ABE | Scaled ABE | |
| | | | Without GMR constraint | With GMR constraint |
| No | 30 | 4.95 | 5.56 | 5.56 |
| No | 25 | 4.98 | 16.50 | *16.34* |
| Yes | 30 | 5.01 | 6.98 | 6.98 |
| Yes | 25 | 4.94 | 14.78 | *14.61* |

Switching variation: $CV_S = 30\%$

Within-subject variation: $CV_W = 30\%$

Constraint on point estimate = 25% (GMR = 1.25)

Type I error was recorded in simulated investigations. For the regulatory standardized variation ($CV_0$), corresponding to the assumed regulatory constant ($\sigma_0$), either 25% (the expectation of FDA) or 30% (the requirement of EMA) was considered. The computations were performed with or without the use of the mixed regulatory strategy, and with or without the constraint of GMR = 1.25 (GMR: ratio of geometric means). ([25], with the permission of the publisher.)

Italic values indicates the results arise at the usually applied regulatory conditions

Therefore the regulatory constant of $\sigma_0 = 0.294$ suggested by EMA is favored against the constant of $\sigma_0 = 0.25$ recommended by FDA.

## Implementation of the regulatory approaches of EMA and FDA

### FDA

For calculating the confidence limits for RSABE, FDA recommends, as noted earlier, the application of the squared, linearized form (Eq. 8) of the RSABE model (Eq. 6) [18]. The procedure and its computational implementation raised various questions:

First, it was noted [30] that the approach has a bias. It had been recommended that the hypothesis of BE should be rejected if the 95% upper bound of calculated confidence interval was positive [18, 19]. However, the computed approximate upper bound is positively biased [30]. The bias depends on the number of subjects and study sequences, and leads to false rejections.

The computational implementation of the procedure by a SAS code [18] has some difficulties. It is not clear how the method should be evaluated when there are missing data. The program omits all data belonging to a subject with a missing period. The usefulness of the approach was questioned [32]. Furthermore, the evaluation of partial replicate studies with the FDA-suggested within-subject (and SAS-specific) correlation structure sometimes fails to converge. But the convergence issues can be usually easily solved with minor changes in the code, for example by altering the correlation structure.

For an average user the output of the RSABE test works as a black box providing numbers too hard to interpret. The stated decision rule is to reject the assumption of BE if the computed upper limit is positive and accept it if this number is negative. It is difficult to tell from this output that new generic product has just barely failed or if it failed very badly.

### EMA

The determination of BE by the usual approach of unscaled average BE (Eq. 4) requires the evaluation of the 90% confidence limits around the difference of the logarithmic means of the two drug products.

The same simple procedure can be applied to the determination of BE of highly variable drugs by the ABEL method (Eq. 9), the approach of EMA. However, the BE limits depend on the estimated within-subject variation of the reference product [21]. This leads to additional Type I error inflation [22, 27–31].

### Comments

To estimate the relevant parameters the EMA Bioequivalence Guideline [21] recommends using an analysis of variance (ANOVA), a least-squares type procedure, and a fixed-effect model for all terms including subjects. Volunteers participating bioequivalence trials are randomly selected to represent the population. Thus, treating them as fixed effects is conceptually questionable. Furthermore, maximum likelihood-based procedures ("mixed procedures") are often much more convenient to use. In fact, the FDA code [18] is an implementation of the second

approach for the estimation of the relevant parameters: a mixed-effect procedure with random subject effects. The two approaches can yield numerically different results for replicate-design studies with missing data. Nevertheless, the two approaches lead to the same bioequivalence conclusion in most cases. But in a borderline condition, the EMA still prefers the least-squares approach [23], thus the bioequivalence conclusion can be different from that of the FDA.

Determination of BE for highly variable drugs by ABEL is simple, straightforward. It extends the usual procedure for unscaled average BE and is easy to apply and visualize. However, Type I error inflation even if it is only 2–3%, is of concern in a very strictly regulated area such as the evaluation of bioequivalence.

Various solutions were suggested. Wonnemann et al. [27] recommended a two-stage design which, in accordance with the EMA guidance [21], analyzed results from an initial group of subjects, permitted the assessment of an additional group of volunteers, and evaluated the results from both groups of subjects in the final analysis. However, the overall, global Type I error had to be maintained at 5% or less. Labes et al. [26] kept the simple ABEL approach but recommended an additional cross-validation step to adjust the nominal alpha level to keep Type I error below 5%. Munoz et al. [28] proposed a hybrid approach which applied Hyslop's algorithm [13], i.e., a procedure recommended by FDA, but with European regulatory constants.

The proposed so-called Exact algorithms [30, 31] are computationally simple, compare to ABEL in a statistically sound way, and eliminate the Type I error inflation problem. The algorithms are based on the observation that the sampling distribution of the $(\mu_T-\mu_R)/\sigma_{WR}$ ratio can be shown to follow the noncentral t. However, if homoscedasticity cannot be assumed ($\sigma_{WR}$ is not equal to $\sigma_{WT}$) then our first version of the algorithms required to estimate the $\sigma_{WT}/\sigma_{WR}$ ratio. This can be easily done from fully replicate studies but is very cumbersome, practically impossible, from partially replicate studies. This was a serious drawback of our first proposal [30] which was overcome when, following Schall [8], we realized that this estimation step could be avoided [31]. An interesting theoretical outcome of our analysis was to show that the estimated $(m_T-m_R)/s_{WR}$ ratio is a biased estimate of the population parameters. The bias is quite severe at small degrees of freedom but is easily corrected with the formula of Hedges [33]. Simulations showed that the Exact algorithms are substantially more powerful at small degrees of freedom than Hyslop's algorithm based on the linearized form [18, 19] but keep the Type I error below the nominal level [31]. They are easily computed, and we provide the relevant R code in Supplementary Files.

## Experimental design

Full (like TRTR-RTRT) or partial replicate designs (like TRR-RTR-RRT) are the recommended trial designs for evaluating the bioequivalence of highly variable drugs.

### Comment

Simulations show [31] that regardless the applied evaluation algorithm, the Type I error can be severely inflated in partial-replicate BE trials if the within-subject variation for the Test formulation is higher than that of the Reference product. The mathematical background is not clearly understood but it is well known that one-way ANOVA is a robust procedure regarding assumptions except when there is marked heteroscedasticity and the design is unbalanced. It is reasonable to suppose that we face the same situation here. This case is quite important because a recent EMA statement [23] shows preference among 3-period design schemes for a partial replicate design over the full-replicate alternatives.

## Additional constraints

### Ratio of geometric means

All regulatory authorities expect that, in addition to a criterion based on the scaled difference between the logarithmic means (either RSABE or ABEL), a second requirement should be imposed: the point estimate for the ratio of the geometric means should be between 0.80 and 1.25.

Leslie Benet originally recommended this constraint first in the context of individual bioequivalence and, later, for highly variable drugs [34]. He noted that an additional point estimate criterion should be added for all drugs as a supplement to the bioequivalence limit criteria "in order to give patients and clinicians confidence that a generic equivalent approved by the regulatory authorities will yield the same outcome as the innovator product".

Benet stated also: "1. There is no scientific basis or rationale for the point estimate recommendations. 2. There is no belief that addition of the point estimate criteria will improve the safety of approved generic drugs. 3. The point estimate recommendations are only "political" to give greater assurance to clinicians and patients who are not familiar (don't understand) the statistics of highly variable drugs." [34]

Benet's statement is very clear and sensible. The second criterion restricting the point estimate of the GMR has had psychological motivation and political rationale.

The scientific and regulatory consequences of the point estimate are disappointing. Notably, the secondary criterion of GMR constraint distorts the meaning and interpretation of the primary criterion which calculates a confidence interval around the scaled difference between means.

First, calculation of a confidence interval involves the assumption of distribution(s). A constraint on the GMR curtails, truncates the distribution(s). Consequently, the confidence interval is computed incorrectly.

Second, in the mixture of the two regulatory criteria, the GMR constraint increasingly dominates when the within-subject variation rises [20, 24, 25]. At very high variations, the secondary criterion becomes (almost) the sole tool for regulatory decision. But the GMR constraint interferes with regulatory statements supposedly based on the primary criterion (either RSABE or ABEL) even at much smaller intraindividual variations.

Macheras and his co-workers noted the complex mathematics of two simultaneously applied tests [35–38]. They developed several interesting, combined regulatory criteria with leveling-off properties. They permit the testing of BE with both RSABE and the GMR constraint in a single step.

#### Upper limit for ABEL

As stated earlier, EMA recommends that ABEL be applied only to an upper limit of 50% of the within-subject variation. Health Canada suggests an upper limit of 57.4% for the coefficient variation. FDA does not impose an upper limit for the application of RSABE.

#### Comments

The second regulatory criterion for the constraint of GMR as well as the requirement of upper constraint appears to have psychological motivation and political rationale. The suggestions do not have scientific foundations.

Nevertheless, the second regulatory criterion appears to have wide support. At the conference of the Global Bioequivalence Harmonization Initiative (GBHI), held in Rockville, MD in September, 2016, the GMR constraint gained acceptance.

However, the preference for political instead of scientific considerations has substantial, at times inadvertent, regulatory consequences. The calculated confidence interval is, for both RSABE and ABEL, incorrect and misleading when the GMR constraint is added.

It appears that inclusion of the GMR has awkward regulatory consequences and has no scientific justification.

Similarly, the upper limit of the within-subject variation for applying ABEL is based on political and not scientific considerations. It is difficult to support the constraint on the upper limit.

### Regulation of AUC and/or $C_{max}$

FDA recommends that, after a single oral administration, RSABE should be applied to both AUC and $C_{max}$. EMA expects that ABEL be evaluated only for $C_{max}$ whereas Health Canada requires the analysis only of AUC.

#### Comment

This is an unusual illustration of a complete absence of harmonization. It is difficult to perceive that each approach is equally justified. They can be attributed, in some cases, to local traditions or committee dynamics.

The approach of FDA can be recommended for considering both metrics, AUC and $C_{max}$, in determinations of BE for highly variable drugs.

## Concluding remarks

The approach of scaled average BE has accomplished its main goal in the last decade: it greatly lowered the number of subjects which is required for the demonstration of BE for highly variable drugs.

For the implementation of the approach, the various regulatory authorities chose procedures which were similar in principle but different in some details. FDA recommends the RSABE procedure whereas EMA (and others) suggests the ABEL method.

Importantly, EMA and FDA recommend differing regulatory constants. The EMA constant corresponds to the switching variation of 30%, the FDA constant does not. The regulatory consequences are substantial.

As noted earlier, the actual BE limits change continuously around the switching variation with the EMA assumption but discontinuously with the FDA condition. As a consequence, the Type I error is 6-7% (and higher) with the EMA condition and 16-17% (and higher) with the FDA recommendation.

Therefore, the regulatory constant suggested by EMA is favored.

For the calculation of the confidence interval around the scaled difference of the two logarithmic means, ABEL uses the same simple procedure as usually applied for unscaled average BE. The FDA method utilizes the squared, linearized RSABE model. The procedure is based on an approximation which leads to notable bias (loss of power due to theoretical maximum) at small degrees of freedoms. The output of the test is yes (passed) or no (not passed) which is not informative. By contrast, ABEL

recommended by EMA is a simple procedure with informative output statistics. It has the disadvantage that the Type I error with this procedure is above the nominal level. The recently introduced new class of algorithms, the so-called Exact algorithms, try to amalgamate the positive features of both EMA and FDA procedures without their negative sides. Therefore we hope that one day they could serve as generally accepted algorithms to evaluate RSABE.

Regulatory authorities impose a second regulatory criterion, a constraint on the point estimate of GMR: it should be between 0.80 and 1.25. The expectation has no scientific foundation but only psychological motivation and political rationale. As an unintended consequence, the calculated confidence interval is, for both RSABE and ABEL, incorrect and misleading when the GMR constraint is included.

Therefore, the addition of the second regulatory criterion, the GMR constraint, is not supported.

Similarly, there is no scientific basis for the suggestion that ABEL be applied only to an upper limit of the within-subject variation.

Different regulators expect that either the RSABE or the ABEL approach should be applied either to both AUC and $C_{max}$ or only to AUC or only to $C_{max}$. The example of disharmonization is probably based on tradition and committee dynamics. The approach of FDA appears to have merit. It applies RSABE to both AUC and $C_{max}$.

So far the optimal study design to evaluate RSABE has received very little interest. We showed in the recommended partial-replicated designs the Type I error is inflated if the Test product variation is higher than that of the Reference formulation. This observation should prompt a revision of the current recommendations to design multiperiod bioequivalence trials for the evaluation of RSABE.

**Acknowledgement** We appreciate the very careful and thoughtful review of the original manuscript by two referees.

## References

1. Shah VP, Yacobi A, Barr WH, Benet LZ, Breimer D, Dobrinska MR, Endrenyi L et al (1996) Evaluation of orally administered highly variable drugs and drug formulations. Pharm Res 13:1590–1594
2. Davit BM, Conner DP, Fabian-Fritsch B, Haidar SH, Jiang X, Patel DT et al (2008) Highly variable drugs: observations from bioequivalence data submitted to the FDA for New Generic Drug Applications. AAPS J 10:148–156
3. Karalis V, Macheras P, Van Peer A, Shah VP (2008) Bioavailability and bioequivalence: focus on physiological factors and variability. Pharm Res 25:1956–1963
4. Cook CS (2011) Current issues on bioavailability and bioequivalence determination. J Bioeq Bioavail S1–003:1–5
5. Marzo A, Vuksic D, Crivelli F (2000) Bioequivalence of endogenous substances facing homeostatic equilibria: an example with potassium. Pharmacol Res 42:523–525
6. Ehmann F, Sakai-Kato K, Duncan R, de la Ossa DHP, Pita R, Vidal JM, Kohli A, Tothfalusi L et al (2013) Next-generation nanomedicines and nanosimilars: EU regulators' initiatives relating to the development and evaluation of nanomedicines. Nanomed 8:849–856
7. Schuirmann DJ (1987) A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. J Pharmacokinet Biopharm 15:657–680
8. Schall R (1995) A unified view of individual, population, and average bioequivalence. In: Blume H, Midha K (eds) Bio-International 2, bioavailability, bioequivalence and pharmacokinetics. Medpharm, Stuttgart, pp 91–106
9. Tothfalusi L, Endrenyi L, Midha K, Rawson MJ, Hubbard JW (2001) Evaluation of the bioequivalence of highly-variable drugs and drug products. Pharm Res 18:728–733
10. Haidar SH, Davit B, Chen ML, Conner D, Lee L, Li QH, Lionberger R, Makhlouf F, Patel D, Schuirmann DJ, Yu LX (2008) Bioequivalence approaches for highly variable drugs and drug products. Pharm Res 25:237–241
11. Tothfalusi L, Endrenyi L, Garcia Areta A (2009) Evaluation of bioequivalence for highly variable drugs with scaled average bioequivalence. Clin Pharmacokinet 48:725–743
12. Dragalin V, Fedorov V, Patterson S, Jones B (2003) Kullback–Leibler divergence for evaluating bioequivalence. Stat Med 22:913–930
13. Schall R, Williams RL (1996) Towards a practical strategy for assessing individual bioequivalence. J Pharmacokinet Biopharm 24:133–149
14. Patnaik RN, Lesko LJ, Chen ML, Williams RL (1997) Individual bioequivalence: new concenpts in the statististical assessment of bioequivalence metrics. Clin Pharmacokinet 33:1–6
15. Davit BM, Chen ML, Conner DP, Haidar SH, Kim S, Lee CH, Lionberger RA, Makhlouf FT, Nwakama PE, Patel DT, Schuirmann DJ, Yu LX (2012) Implementation of a reference-scaled average bioequivalence approach for highly variable generic drug products by the US Food and Drug Administration. AAPS J 14:915–924
16. Food and Drug Administration (2013) Draft guidance for industry: Bioequivalence studies with pharmacokinetic endpoints for drugs submitted under an ANDA. Center for Drug Evaluation and Research (CDER), Silver Spring, MD https://www.fda.gov/downloads/drugs/guidances/ucm377465.pdf
17. Davit BM, Conner DP (2017) The United States. In: Kanfer I (ed) Bioequivalence requirements in various global jurisdictions. Springer, Cham, pp 269–305
18. Food and Drug Administration (2011) Draft guidance for industry: Bioequivalence recommendations for progesterone oral capsules. Center for Drug Evaluation and Research (CDER), Silver Spring, MD https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM209294.pdf
19. Hyslop T, Hsuan F, Holder DJ (2000) A small sample confidence interval approach to assess individual bioequivalence. Stat Med 19:2885–2897
20. Tothfalusi L, Endrenyi L (2003) Limits for the scaled average bioequivalence of highly variable drugs and drug products. Pharm Res 20:382–389
21. European Medicines Agency (2010) Guideline on the investigation of bioequivalence. London, United Kingdom http://www.ema.europa.eu/docs/en_GB/document_library/2010/01/WC500070039.pdf

22. Boddy AW, Snikeris FC, Kringle RO, Wei GC, Oppermann JA, Midha KK (1995) An approach for widening the bioequivalence acceptance limits in the case of highly variable drugs. Pharm Res 12:1865–1868
23. European Medicines Agency (2015) Questions & Answers: positions on specific questions addressed to the Pharmacokinetics Working Party (PKWP). 19 November 2015 EMA/618604/2008 Rev. 13 http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500002963.pdf. Accessed 7 Aug 2018
24. Haidar SH, Makhlouf F, Schuirmann DJ, Hyslop T, Davit B, Conner D, Yu LX (2008) Evaluation of a scaling approach for the bioequivalence of highly variable drugs. AAPS J 10:450–454
25. Endrenyi L, Tothfalusi L (2009) Regulatory conditions for the determination of bioequivalence of highly variable drugs. J Pharm Pharmaceut Sci 12:138–149
26. Labes D (2013) 'Alpha' of scaled ABE. Bioequivalence and bioavailability forum. BEBAC Consultancy Services for Bioequivalence and Bioavailability Studies, Vienna http://forum.bebac.at/mix_entry.php?id=10202
27. Wonnemann M, Frőmke C, Koch A (2015) Inflation of the type I error: investigations on regulatory recommendations for bioequivalence of highly variable drugs. Pharm Res 32:135–143
28. Munoz J, Alcaide D, Ocana J (2016) Consumer's risk in the EMA and FDA regulatory approaches for bioequivalence in highly variable drugs. Stat Med 35:1933–1943
29. Labes D, Schütz H (2016) Inflation of type I error in the evaluation of scaled average bioequivalence, and a method for its control. Pharm Res 33:2805–2814
30. Tothfalusi L, Endrenyi L (2016) An exact procedure for the evaluation of reference scaled average bioequivalence. AAPS J 18:476–489
31. Tothfalusi L, Endrenyi L (2017) Algorithms for evaluating reference scaled average bioequivalence: power, bias, and consumer risk. Stat Med 36:4378–4390
32. Patterson SD, Jones B (2012) Viewpoint: observations on scaled average bioequivalence. Pharm Stat 11:1–7
33. Hedges LV (1981) Distribution theory for Glass's estimator of effect size and related estimator. J Educ Stat 6:107–128
34. Benet L (2006) Why highly variable drugs are safer. Meeting of FDA Committee for Pharmaceutical Science. http://www.fda.gov/ohrms/dockets/ac/06/slides/2006-4241s2_2.htm. Accessed 6 Oct 2006
35. Karalis V, Symillides M, Macheras P (2004) Novel scaled average bioequivalence limits based on GMR and variability considerations. Pharm Res 21:1933–1942
36. Karalis V, Macheras P, Symillides M (2005) Geometric mean ratio-dependent scaledbioequivalence limits with leveling-off properties. Eur J Pharm Sci 26:54–61
37. Kytariolos J, Karalis V, Macheras P, Symillides M (2006) Novel scaled bioequivalence limits with leveling-off properties. Pharm Res 23:2657–2664
38. Karalis V, Symillides M, Macheras P (2011) On the leveling-off properties of the new bioequivalence limits for highly variable drugs of the EMA guideline. Eur J Pharm Sci 44:497–505