Limits for the Scaled Average Bioequivalence of Highly Variable Drugs and Drug Products

Laszlo Tothfalusi[1] and Laszlo Endrenyi[2,3]

Received September 4, 2002; accepted January 22, 2003

Purpose. To provide a rational procedure for establishing regulatory bioequivalence (BE) limits that can be applied in determinations of scaled average BE for highly-variable (HV) drugs and drug products.
Methods. Two-period crossover BE investigations with either 24 or 36 subjects were simulated with assumptions of a coefficient of variation of 10, 20, 30, or 40%. The decline in the fraction of accepted studies was recorded as the ratio of geometric means (GMR) for the two formulations was raised from 1.00 to 1.45. Acceptance of BE was evaluated by scaled average BE, assuming various BE limits, and, for comparison, by unscaled average BE. A procedure for calculating exact confidence limits in two-period studies is presented, and an approximate method, based on the linearization of the regulatory model, is applied.
Results. A mixed model is proposed for average BE. Accordingly, at low variabilities, the BE limit is constant, $\pm BEL_o$, generally log(1.25). Beyond a logarithmic, limiting, "switching" variability (σ_o), in the region of HV drugs, the approach of scaled average BE is applied with limits of $\pm(BEL_o/\sigma_o)$. It is demonstrated that the performance of the mixed model corresponds to these expectations. The effect of σ_o and of the resulting BE limits is also demonstrated. Scaled average BE, with all reasonable limits for HV drugs, requires fewer subjects than an unscaled average BE. In two-period studies, the exact and approximate methods calculating confidence limits yield very comparable inferences.
Conclusions. Scaled average BE can be effectively applied, with the recommended limits, for determining the BE of HV drugs and drug products. The limiting, "switching" variability (σ_o) will have to be established by regulatory authorities.

KEY WORDS: highly-variable drugs; bioequivalence; scaled average bioequivalence; mixed model for average bioequivalence; regulatory limits; crossover designs.

INTRODUCTION

The problems of establishing bioequivalence (BE) for highly variable (HV) drugs and drug products are well known. It can be very difficult for these to satisfy the usual regulatory requirement that the 90% confidence interval around the estimated ratio of geometric means (GMR) of the two formulations be between 0.80 and 1.25. The approach is often referred to as the determination of average BE.

The frustrating problem of determining BE for HV drugs has been considered in recent years, but mainly in the context of individual BE. It was suggested in various studies (e.g.,1–6), culminating in a guidance published by the Food and Drug Administration (7), that a criterion recommended for individual BE could be normalized (or "scaled") by an estimated variance. It was thought that scaled individual BE would relieve the difficulties involving HV drugs. The evaluation of both unscaled and scaled individual BE requires replicate-design investigations that involve not two but three or, more typically, four study periods.

However, several investigations have found that, in practice, individual BE has unfavorable properties, and numerous objections have been raised to its implementation (e.g., 8–13). An alternative approach has also been proposed, namely, that scaling be applied not to individual but to average BE (3,14). It was suggested that scaled average BE has more favorable characteristics (i.e., with given risks, requiring fewer subjects) than scaled individual BE. Moreover, scaled average BE could be determined by both two- and four-period investigations. It was also noted (14) that the approach of scaled average BE is equivalent to expanding the BE limits as the variability increases, an approach that had been suggested by Boddy *et al.* (15).

The present communication focuses on the application of scaled average BE for HV drugs and drug products. The main purpose is to outline the procedures and alternatives for setting BE limits for the analysis of HV drugs and thereby to lay the groundwork for regulatory considerations. Attention was recently called to the need for establishing such preset limits (16). A secondary goal of the present study is to summarize the calculations that are required for the evaluation of scaled average BE.

METHODS

A summary of the notations used is provided in the Appendix.

A Mixed Model of Average BE: Unscaled and Scaled

Schall and Willams (4) recommended a mixed model for individual BE. According to this, a constant-referenced (i.e., in effect, unscaled) criterion is applied when the (within-subject reference) variability does not exceed a critical level (σ_o), and a "reference-scaled" criterion (i.e., one scaled by the within-subject reference variability) is used with higher than the critical variability, i.e., for HV drugs (4–6). The scheme has been implemented in an FDA guidance (7).

The approach can also be applied to a judicious mixture of unscaled and scaled average BE. Thus, the usual criterion of unscaled average BE prevails provided that drugs do not have unusual properties, for instance, if they do not exhibit high variability. Accordingly, under these conditions it is usually expected that the ratio of the geometric means (GMR) of the test (T) and reference (R) formulations should be between BE limits, the magnitudes of which are set by regulatory agencies. The multiplicative BE limit is often set at the value of 1.25:

$$0.80 \le \mathrm{GMR} \le 1.25 \qquad (1)$$

Here and later, such statements of regulatory expectations do not merely indicate the relationship between a regulatory model and BE limits but imply that the indicated measure is, together with its (usually 90%) confidence limits, within the BE limits.

[1] Department of Pharmacodynamics, Semmelweis University, 1089 Budapest, Hungary.
[2] Department of Pharmacology, University of Toronto, Toronto, Ontario M5S 1A8, Canada.
[3] To whom correspondence should be addressed. (e-mail: l.endrenyi@utoronto.ca)

0724-8741/03/0300-0382/0 © 2003 Plenum Publishing Corporation

In practice, BE is evaluated by calculating logarithmic quantities. Thus, means and standard deviations of the logarithmic data (μ and σ) are estimated. The usual criterion of *unscaled average BE* is applied provided that drugs do not exhibit high variability or have unusual properties otherwise. Accordingly, as long as the variability (σ) does not exceed a preset magnitude (σ_o), the difference between the two logarithmic means should be between the preset BE limits ($\pm BEL_o$):

$$-BEL_o \le \mu_T - \mu_R \le BEL_o \qquad \text{when } \sigma \le \sigma_o \qquad (2)$$

The regulatory BE limit is generally $BEL_o = \log(1.25)$.

For HV drugs, when the preset magnitude of the variability is exceeded, both the difference between the logarithmic means and the regulatory BEL_o could be normalized (scaled) by the variability. Thus, for HV drugs, evaluation of *scaled average BE* is suggested:

$$-BEL_o/\sigma_o \le (\mu_T - \mu_R)/\sigma \le BEL_o/\sigma_o \quad , \quad \text{when } \sigma > \sigma_o \qquad (3)$$

We propose these BE limits in order to be able to declare either the prevalence or the lack of BE at the limiting variability of σ_o regardless whether an unscaled or a scaled average BE [Eq. (2) or (3)] is applied.

The suggested BE limits are then $\pm(BEL_o/\sigma_o)$. Here, σ_o is to be preset by regulatory authorities. Possible alternatives and their consequences are considered later.

Features of the mixed model for average BE are illustrated in Fig. 1. Figure 1A shows the suggested BE limits with increasing variability. They remain set at BEL_o at low variabilities. Beyond the limiting, "switching" variability of σ_o, i.e., in the region of HV drugs, the BE limit for unscaled average BE would expand in proportion to the variability, and the proportionality factor is BEL_o/σ_o (Fig. 1A) (15).

Figure 1B illustrates the relationship between BE limits and variability from the point of view of scaled average BE. The scaled limits, if they were applied, would decrease in the region of small variabilities. They reach their smallest values at the "switching" variability of σ_o and, thereafter, in the region of high variabilities, are set at the constant level of BEL_o/σ_o.

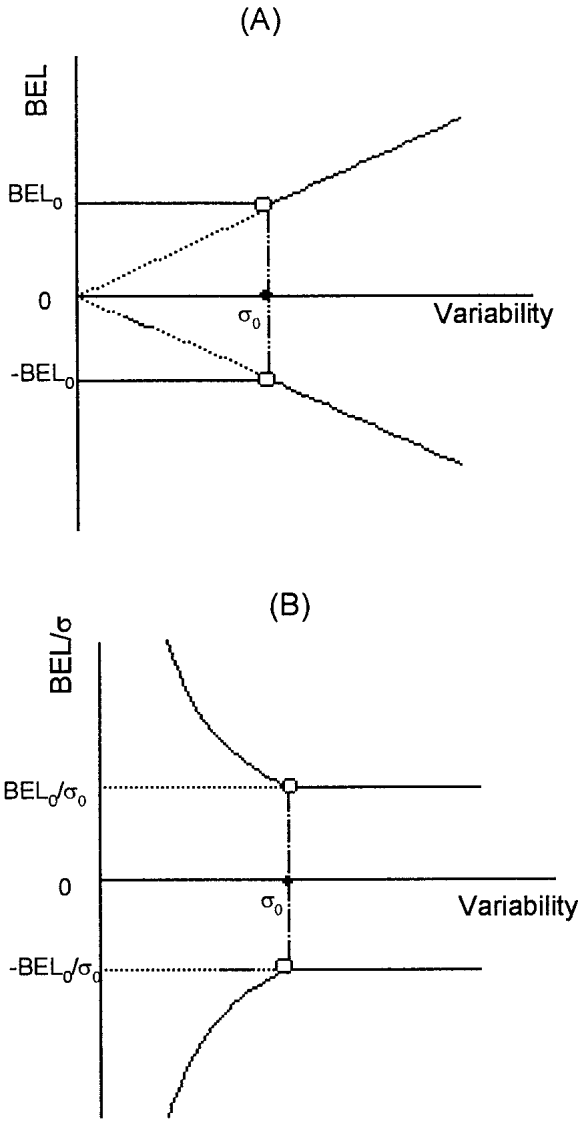

Fig. 1. Illustration of the mixed model for average BE. (A) Dependence of BE limits for unscaled average BE on the variability. At small variations, the BE limits are set by regulatory authorities at $\pm BEL_o$ [Eq. (2)]. Beyond the limiting, "switching" variability (σ_o), in the region of HV drugs, the BE limits expand in proportion to the variability [Eqs. (6) and (9)]. (B) Dependence of BE limits for scaled average BE on the variability. At small variations, the scaled limits would be large. They reach their lowest value of $\pm(BEL_o/\sigma_o)$ at the "switching" variability and maintain this level at the higher variabilities [Eq. (3)].

Rationales for Applying Scaled Average BE

High Statistical Power of Decision

As discussed in the Introduction, the determination of BE for HV drugs with the usual unscaled criterion is feasible only if an unreasonably large number of subjects is studied. In other words, if, say, 24 subjects are used in a study, the statistical power of testing for BE (i.e., the probability of its acceptance) is very low. It can be expected that, for HV drugs, the application of scaled average BE would yield a much higher power of the correct decision when the two drug products are truly bioequivalent. This feature was demonstrated in a preliminary investigation (14); its small illustration is provided later, and it will be explored in detail in the future.

A Special Case of Individual BE

The regulatory model for individual BE (5–7) examines not only the deviation between the means of the two formulations but also the difference between the corresponding (logarithmic) intraindividual variances ($\sigma_{WT}^2 - \sigma_{WR}^2$) and the importance of the subject-by-formulation (S*F) interaction (σ_D^2). Consequently, the numerator of the model contains three terms:

$$[(\mu_T - \mu_R)^2 + \sigma_D^2 + (\sigma_{WT}^2 - \sigma_{WR}^2)]/\sigma_W^2 \le \theta \qquad (4)$$

where θ is the regulatory limit for individual BE. The denominator is a "switch" between unscaled and scaled analyses or, in the phrasing of the FDA guidance, between "constant-scaled" and "reference-scaled" assessments. Accordingly, for

drugs showing moderate variability, the denominator is constant:

$$\sigma_W^2 = 0.4 \qquad \text{when } \sigma_{WR}^2 \le \sigma_o^2$$

whereas for drugs with high variability,

$$\sigma_W^2 = \sigma_{WR}^2 \qquad \text{when } \sigma_{WR}^2 > \sigma_o^2$$

The value of the "switching" constant for high variability has been defined, in this case, as $\sigma_o = 0.20$ (7). Consequently, drugs and drug products have been regarded as being HV when $\sigma_{WR}^2 > 0.04$; i.e., $\sigma_{WR} > 0.20$.

Scaled average BE can be algebraically considered as a special case of scaled individual BE. If, for HV drugs, there is no S*F interaction ($\sigma_D^2 = 0$) and the within-subject variations of the two drug products are the same ($\sigma_{WT}^2 = \sigma_{WR}^2$), then the regulatory expression becomes

$$(\mu_T - \mu_R)^2/\sigma_{WR}^2 \le \theta \tag{5}$$

This is equivalent to Eq. (3) given earlier for scaled average BE when it is observed that, with the simplifying assumptions of the special condition, $\sigma_{WR}^2 = \sigma^2$, and that $\theta = (BEL_o/\sigma_o)^2$.

For HV drugs, the simplifying assumptions are reasonable because, for these, the high variability is a property of the drug and not of the drug products. More detailed considerations are presented in the Discussion for the condition of differing variabilities between drug products.

Scaled Average BE as the First Stage of Scaled Individual BE

The regulatory condition for individual BE [Eq. (4)] is an aggregate criterion: its numerator contains three terms that correspond to the three targeted regulatory considerations.

The usefulness and effectiveness of an aggregate criterion have been repeatedly called into question (e.g., 9–13). Alternatively, a stepwise analysis could be perceived (13,17–19). The equivalence of the two means would first be determined, followed by the evaluation of the similarity of the within-subject variances, and, finally, the possible prevalence of an S*F interaction would be tested.

In this scheme, then, the determination of scaled (and also unscaled) average BE represents the first stage in the evaluation of scaled (and also unscaled) individual BE.

Expanding BE Limits

Boddy *et al.* (15) recommended that, for HV drugs, the BE limits (BEL) be widened in proportion to the standard deviation (σ). The suggestion can be generalized to yield the determination of scaled average BE with the suggested BE limits (Eq. 3).

The difference between the logarithmic mean μ and its confidence interval, for a metric (such as AUC) of the test (T) and reference (R) formulations, is expected to be within the BE limits:

$$-BEL \le \mu_T - \mu_R \le BEL \tag{6}$$

The usual approach of (unscaled) average BE expects that the BE limits are constant at a level set by the regulatory agencies (BEL_o). The condition is to be applied for drugs that do not have unusual properties, for instance, those that do not show high variability. Consequently, *unscaled average BE* should be applied up to a level of variability (σ_o), with the constant BE limits:

$$BEL = BEL_o \quad , \quad \text{when } \sigma \le \sigma_o \tag{7}$$

Therefore, with unscaled average BE, Eq. 2 follows:

$$-BEL_o \le \mu_T - \mu_R \le BEL_o \qquad \text{when } \sigma \le \sigma_o \tag{2}$$

For HV drugs, Boddy *et al.* (15) recommended that the BE limits be expanded in proportion to the variability:

$$BEL = k\sigma \tag{8}$$

where k is a constant of proportionality. As suggested above, the proportional relationship between BEL and σ should be maintained, starting from the "switching" variability (i.e., for HV drugs), and therefore:

$$BEL = (BEL_o/\sigma_o)\sigma, \qquad \text{when } \sigma > \sigma_o \tag{9}$$

Consequently, the proportionality constant is:

$$k = BEL_o/\sigma_o \tag{10}$$

This leads directly to the suggested expression for scaled average BE with its constant, preset regulatory limits:

$$-BEL_o/\sigma_o \le (\mu_T - \mu_R)/\sigma \le BEL_o/\sigma_o \quad , \quad \text{when } \sigma > \sigma_o \tag{3}$$

The BE Limits: Alternative Possibilities

Various definitions of HV drugs, with the corresponding "switching" variabilities (σ_o), have been provided in different contexts in the literature. These are presented below. In order to apply scaled average BE, regulatory agencies will have to define the limits, possibly one of those to be described.

$\sigma_o = 0.20$

As noted earlier, in applications of individual BE, a "switching" variability of $\sigma_o = 0.20$ was recommended and implemented for indicating HV drugs (7). Therefore, the corresponding BE limit for scaled average BE is $\pm \ln(1.25)/0.20 = \pm 1.116$.

$\sigma_o = ln(1.25) = 0.223$

Boddy *et al.* (15) used, by assuming type I and II errors of 5 and 10%, respectively, and a sample size of 24 in a two-period study, $k = 1.00$ for the proportionality constant k in Eq. (8). With this value, the "switching" variability is $\sigma_o = \ln(1.25) = 0.223$, and the limit for scaled average BE is $\pm \ln(1.25)/\ln(1.25) = \pm 1.00$.

$\sigma_o = 0.294$ [at $CV_o = 30\%$]

In investigations of BE, HV drugs were defined to have variabilities with a coefficient of variation exceeding $CV_o = 30\%$ (20–22). Because the coefficient of variation (CV) is related, approximately, to the standard deviation (σ) of logarithmic quantities by

$$CV = [\exp(\sigma^2) - 1]^{1/2} \tag{11}$$

the "switching" variability is $\sigma_o = 0.294$, and the limit for scaled average BE is $\pm\ln(1.25)/0.294 = \pm 0.759$.

Statistical Evaluation of Scaled Average BE

Two approaches are presented for the statistical evaluation of scaled average BE in two-period studies, i.e., for calculating the confidence interval around the scaled difference between the logarithmic means: an exact procedure based on

the noncentral *t* distribution and an approximation that linearizes the regulatory criterion.

Noncentral t *Distribution*

An exact procedure can be applied in two-period crossover investigations. This method recognizes that the estimated ratio (d/s) of a mean of normally distributed observations from *N* subjects and the standard deviation follows a noncentral t distribution with $N - 2$ degrees of freedom (df) (23). In the present case, this distribution characterizes the estimated ratio of the difference (d) between the logarithmic means and the standard deviation (s) that is obtained from the square root of the residual variance estimated in an analysis of variance.

Consequently, the estimated measure of scaled average BE, the scaled difference (d/s), follows a noncentral *t* distribution with a noncentrality parameter of $\lambda = (N/2)^{1/2}(d/s)$ and degrees of freedom of $df = N - 2$. Therefore, a symmetric 90% confidence interval around the scaled difference is:

$$T_{inv,\ 0.05}(\lambda, df) \leq d/s \leq T_{inv,\ 0.95}(\lambda, df) \qquad (12)$$

where $T_{inv,\ \alpha}(\lambda, df)$ is the inverse of the cumulative noncentral *t* distribution at the α probability level.

In order to declare BE, the 90% confidence interval should be between the limits of $(N/2)^{1/2}\log(0.80)/\sigma_o$ and $(N/2)^{1/2}\log(1.25)/\sigma_o$.

Linearization of the Regulatory Criterion

Tothfalusi *et al.* (14) described a procedure for calculating the confidence limits for scaled average BE. The method linearizes the regulatory criterion [Eq. (3)], evaluates the distribution of each resulting linear term, and determines the length of the confidence interval. The approach can be applied to both two- and four-period crossover studies. The method reformulates the linearizing procedure of Hyslop *et al.* (24), which was developed for calculating the confidence limits for unscaled and scaled individual BE.

Simulations

A program was written for the simulation of two-period crossover studies. Fortran 90 language was used (Visual Fortran, version 6.1, Compaq, Houston, TX) together with the relevant statistical subroutines from the IMS library (Visual Numerics, Sugar Land, TX).

The simulations assumed that the BE studies were performed with either 24 or 36 subjects. The simulating coefficients of variation (CVs) ranged from 10 to 40%, and lognormally distributed kinetic parameters were assumed. The true deviation between the logarithmic means was set at various values starting from zero (indicating actual BE) and gradually rising to increasing deviations from true BE; thus, the simulated GMR values were between 1.00 and 1.45. The simulations covered a wide range of conditions encountered in practice, and 15,000 simulations were performed under each condition.

RESULTS

Performance of Scaled Average BE

Figure 2 illustrates the proportion of two-period crossover studies that are accepted when evaluated by scaled average BE. The simulated ratio of the two geometric means (GMR) is 1.00 on the left-hand side, indicating true BE. GMR then gradually increases, and thereby the deviation from BE rises.

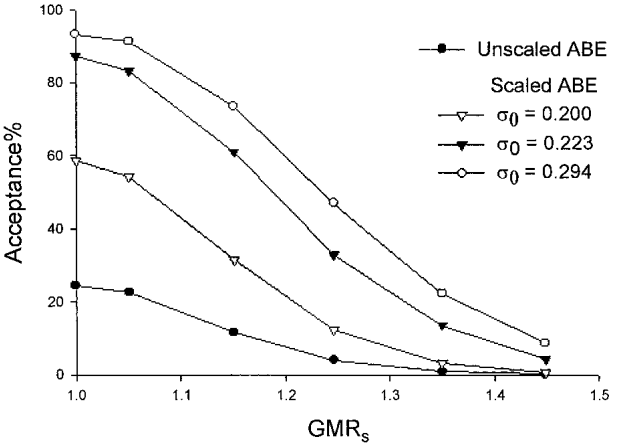

Fig. 2. Percentage of accepted BE studies at various ratios of the geometric means (GMR). Two-period crossover investigations with 24 subjects were simulated. Variability with a coefficient of variation of 40% was assumed. Characteristics of scaled average BE with various assumed "switching" variabilities [$\sigma_o = 0.20$, $\log(1.25) = 0.223$, or 0.294 (for $CV_o = 30\%$)] with the corresponding BE limits (1.116, 1.000, or 0.759, respectively) are illustrated. Acceptances by unscaled average BE are also shown for comparison.

The performances of three "switching" variabilities and of the corresponding BE limits, introduced in the section on Methods are shown. As expected, the highest proportions of accepted BE studies are observed with the smallest of the investigated "switching" variabilities ($\sigma_o = 0.20$), i.e., with the widest BE limit of ±1.116. In turn, the lowest proportions of acceptance are recorded with the largest "switching" variability ($\sigma_o = 0.294$ corresponding to $CV_o = 30\%$), i.e., with the narrowest BE limit of ±0.759. These results could be anticipated. Still, they underscore that regulatory agencies will have to select the appropriate BE limits in the future.

Figure 2 also demonstrates that unscaled average BE yields much lower proportions of accepted BE studies than scaled average BE even when the narrowest of the investigated BE limits is considered for the latter. Notably, the correct acceptances of BE by scaled average BE are, at the condition of true bioequivalence ($GMR_s = 1.00$), 94.4, 88.1, and 58.7% when switching variabilities of $\sigma_o = 0.200$, 0.223, and 0.294, respectively, are applied. By contrast, unscaled average BE accepts only 24.7% of the studies under the same condition of $GMR_s = 1.00$. Thus, under the conditions of the simulations ($n = 24$, $CV_s = 40\%$), the very unfavorable performance of unscaled average BE is stressed again for the evaluation of HV drugs.

Effect of Variation on the Performance of the Mixed Model for Average BE

Figure 3 illustrates the percentage of BE studies accepted by three methods: (1) unscaled average BE, (2) scaled average BE, and (3) the mixed approach for determining average BE. Acceptances of BE at various levels of GMR_s and pooled coefficients of variation ($CV_s = 10$, 20, 30, and 40%) are shown. A limiting, switching variability of $\sigma_o = 0.20$ was

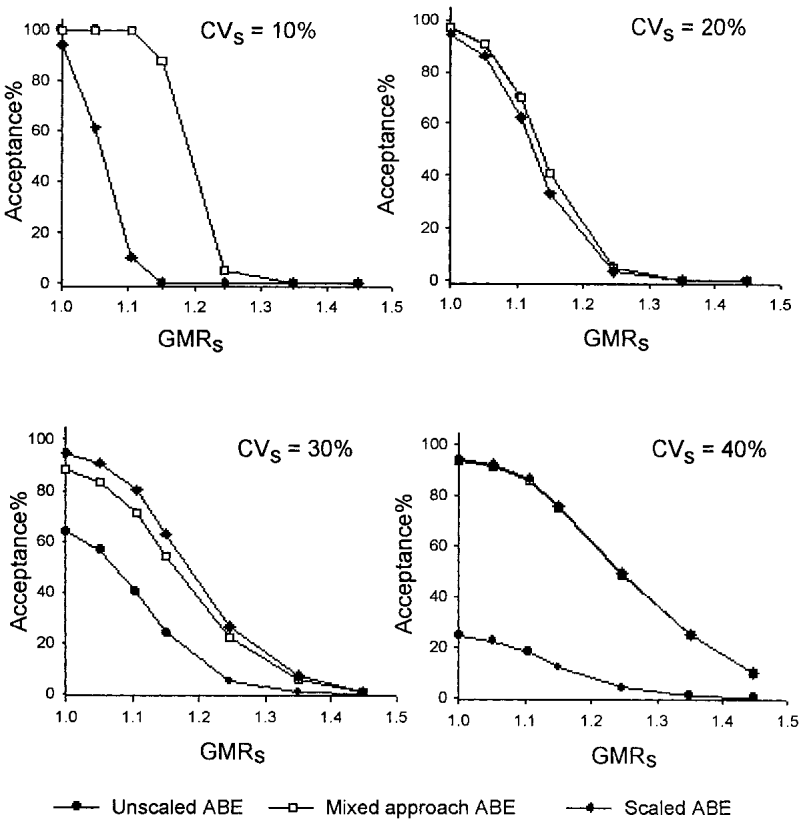

Fig. 3. Percentage of simulated two-period crossover studies in which BE was accepted by three procedures: scaled average BE *(filled squares),* unscaled average BE *(filled circles),* and the mixed approach to average BE *(open squares).* Four levels of pooled coefficients of variation were simulated: CV_s = 10, 20, 30, and 40%. BE was evaluated at several levels of GMR_s between 1.00 and 1.45. At CV_s = 10 and 20%, the results for the mixed approach almost completely overlapped with those obtained for unscaled average BE. Crossover studies with 24 subjects were simulated. A switching variability of σ_o = 0.20 was assumed for the calculation of the BE limit.

assumed in these demonstrations, and therefore, the BE limit is $\pm\ln(1.25)/0.20 = 1.116$.

At high variations (CV_s = 30% and 40%), unscaled average BE yielded substantially lower acceptance than scaled average BE. At these levels of variation, the performance of the mixed model was similar to that of scaled average BE.

In contrast, at low variation (CV_s = 10%), acceptances by unscaled average BE strongly exceeded those yielded by scaled average BE. The performance of the mixed procedure was nearly identical to that of unscaled average BE.

When the variation was similar to the switching variability (CV_s = 20% with σ_o = 0.20), the three procedures also yielded similar acceptances of BE.

Figure 4 presents additional results obtained in simulated crossover studies with 36 subjects. Other conditions were similar to those provided for Fig. 3, but the arrangements and analyses were more limited. When the simulated variation was similar to the switching variability (CV_s = 20%), the scaled and unscaled average BE yielded similar acceptances. At increasing variations, however, the deviation between the results of the two procedures spread: at a given GMR_s, acceptances obtained by scaled average BE rose, and those calculated by unscaled average BE decreased as the variability increased.

Statistical Evaluation of Scaled Average BE

Results obtained by two methods, which calculate the BE limits for scaled average BE, are compared in Table I. The proportions of accepted two-period crossover BE studies are shown at various ratios of the geometric means.

The approximate procedure, based on the linearization of the regulatory model (14,24), and the exact method, applying the noncentral *t* distribution, yield very similar results with calculated acceptances of BE deviating by less than 0.6% under the investigated conditions. Therefore, it is reasonable to use either approach in evaluations of two-period investigations. In analyses of replicate-design studies, only the linearizing procedure can be applied.

DISCUSSION

The usefulness of the approach of scaled average BE for determining the bioequivalence of HV drugs and drug products was demonstrated earlier (14). The present report defines the principles for setting the BE limits when this approach is applied. It calls attention to the importance of the "switching" variability (σ_o), the limiting condition between the uses of unscaled and scaled average BE. The application

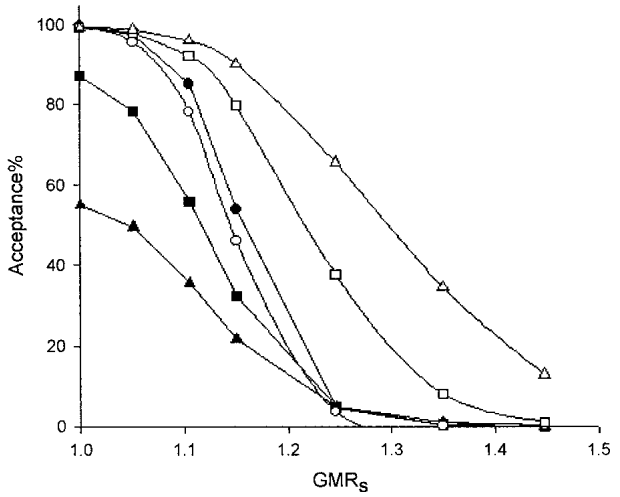

Fig. 4. Percentage of simulated two-period crossover studies with 36 subjects in which BE was accepted by the method of either unscaled or scaled average BE. Several levels of GMR_s and three levels of the pooled coefficient of variation, CV_s = 20, 30, and 40%, were simulated. A switching variability of σ_o = 0.20 was assumed.

of σ_o determines a mixed model for average BE that parallels the mixed model recommended and is implemented for individual BE (4–7).

The magnitude of σ_o defines the BE limits and will have to be established by regulatory authorities. The effects of different values are illustrated (Fig. 2). The value of σ_o strongly affects the properties of BE determination, notably the producer risk. [The producer risk is the probability of rejecting BE when the two products are in fact bioequivalent, i.e., when GMR_s = 1.00.] σ_o and the resulting BE limits affect also the permissiveness of BE evaluation, i.e., the magnitude of GMR (or the deviation between the logarithmic means) that is tolerated for a given consumer risk.

As anticipated, for HV drugs (with large pooled variations), acceptances of BE are higher when scaled rather than unscaled average BE is applied (Figs. 3 and 4). In turn, unscaled average BE yields higher acceptances when the pooled variation is small. Acceptances by the mixed approach of average BE are similar to those yielded by the dominant procedure, unscaled or scaled average BE at low or high variations, respectively (Fig. 3).

The mixed model of unscaled and scaled average BE [Eqs. (2) and (3)] is stated in terms of (logarithmic) means, standard deviations, and the BE limits. However, it is recognized that regulatory criteria imply that not only the indicated measures but also their confidence intervals should be within the BE limits. Therefore, the given rationales for applying scaled average BE should be expanded.

Table I. Acceptance of BE (in %), with Two Procedures, of Bioequivalence by Scaled Average BE in Two-Period Studies[a]

GMR_s	Noncentral T	Linearization
1.00	58.35	58.93
1.05	54.35	54.90
1.15	30.92	31.45
1.25	11.65	11.92
1.35	3.01	3.11
1.45	0.55	0.55

[a] The following conditions were assumed: N = 24, CV_s = 40%.

At the limiting, "switching" variability of σ_o, both unscaled and scaled average BE should yield the same conclusion, the same probability for accepting BE. The difference between the logarithmic means ($\mu_T - \mu_R$) may be assumed to follow a normal distribution. As discussed, the scaled difference $[(\mu_T - \mu_R)/\sigma]$ can be characterized in two-period studies by a noncentral t distribution. However, with sufficiently large number of subjects (e.g., n = 24 as demonstrated in Table I, or larger), the distribution can be approximated by a normal distribution. Consequently, the probabilities of satisfying the criterion of Eq. (2) for unscaled average BE are in accord with the probabilities for the criterion of Eq. (3) of scaled average BE at the corresponding limits of $\pm BEL_o$ and $\pm(BEL_o/\sigma_o)$. The argument of corresponding normal probabilities for unscaled and scaled average BE, at their suggested BE limits, can be applied with large samples, also to replicate-design studies.

Indeed, the preceding presentations, statements, and conclusions on the application of scaled average BE and its limits apply equally (with the exception of calculations by the noncentral t distribution) to two-period and replicate-design crossover studies. This conclusion was reached earlier (14). Thus, although demonstrations in the present report were performed, for the sake of simplicity, by assuming two-period studies, the conclusions also apply to replicate-design investigations. The main difference between the calculations for the two designs is in the estimation of the variance (σ^2). In two-period investigations, the residual variance of an analysis of variance is used for this purpose. In replicate-design studies, the scaling variance is that for within-subject variation of the reference formulation. These investigations enable the estimation of the intraindividual variances of both formulations and of the subject-by-formulation interaction. These estimates are required in calculations of the confidence limits for scaled average BE by applying the approach of Hyslop *et al.* (14,24). It is worth noting that Patterson *et al.* (25) recently recommended the application of replicate designs when unscaled average BE is used.

Thus, in two-period BE studies, the normalizing term is based on a residual variation that is estimated in an analysis of variance. This residual variance is a pooled quantity, a composite of the within-subject variations of both the test and reference products as well as the subject-by-formulation interaction. The separation of these components could be intended, for instance, when two drug products are expected to have different variations. The separation could be accomplished in replicate-design studies. However, the ratio of the two within-subject variances could also be obtained from the results of two-period crossover studies (9).

As noted in the Methods section, the regulatory model for scaled average BE [Eq. (3)] can be obtained as a special case of the model for scaled individual BE [Eq. (4)]. It is important, however, that the two models, and resulting BE criteria, are distinct in principle and operationally. In fact, the main goal of the present communication is to provide a rational basis for establishing BE limits for scaled average BE. These limits are not related to the BE limits recommended for scaled individual BE (7). Generally, the relationship between scaled average and scaled individual BE parallels the distinction between unscaled average and unscaled individual

BE. Thus, the concept of scaled average BE can be directly interpreted, outside the framework of individual BE.

Comparison of the effects of two treatments is a common challenge in the analysis of clinical studies. However, there is no universally accepted measure that would quantify the biologic impact, the substantive impressiveness, and the clinical significance of the difference. Cohen (26) proposed that the standardized effect size (the difference between the drug effects divided by a standard deviation) could usefully reflect the impact of the difference. The measure could be effectively applied in medical settings, particularly when the assumption of lognormality is reasonable (27). Consequently, the method of scaled average BE actually evaluates a standardized effect size and has thus an intuitive, medically accepted interpretation.

The approaches of expanding BE limits (15) [Eqs. (6), (8), and (9)] and scaled average BE [Eq. (3)] yield closely similar inferences (14). However, the limits of the latter method are set by the regulatory authorities and remain constant, whereas the expanding BE limits have to be estimated from the data and depend on magnitude of the evaluated variability. Therefore, the procedure of scaled average BE is, in principle, preferable.

The acceptances of two-period BE studies showed good agreement when they were calculated by the approximating linearized regulatory model [based on the method of Hyslop *et al.* (24)] and by the exact procedure involving noncentral *t* distribution of the scaled difference between the logarithmic means (Table I). This observation substantiates the usefulness and appropriateness of the approximating method in two-period investigations. The conclusion is encouraging because it suggests the effective application of the linearizing procedure in replicate-design studies (14,24) for which the alternative method is not available. Indeed, Hyslop *et al.* (24) demonstrated favorable statistical properties, such as good coverage of the confidence interval, with their approach.

In summary, it is suggested, for a variety of reasons, and demonstrated that the approach of scaled average BE with the suggested regulatory limits represents a reasonable approach for determining the BE of HV drugs and drug products. A mixed model is proposed for assessments of average BE, which would parallel the mixed model used for individual BE. Accordingly, at low variabilities, up to a limiting, "switching" value of σ_o, the currently applied unscaled average BE would be evaluated with limits of $\pm BEL_o$ [generally log(1.25)]. At variabilities exceeding σ_o, scaled average BE would be used with recommended BE limits of $\pm(BEL_o/\sigma_o)$. The value of σ_o will have to be set by regulatory authorities. Unscaled and scaled average BE can be determined from both two-period and replicate-design investigations.

ACKNOWLEDGMENT

The research of L. Tothfalusi was supported, in part, by a grant from the Hungarian Ministry of Health (No. ETT 225-2001).

APPENDIX

Symbols

d Difference between logarithmic means
df Degrees of freedom
k Proportionality constant between bioequivalence limit and logarithmic standard deviation
s Estimated standard deviation
BE Bioequivalence
BEL Logarithmic bioequivalence limit
CV Coefficient of variation
FDA Food and Drug Administration
GMR Ratio of geometric means of test and reference formulations
HV Highly variable
N Number of subjects
S*F Subject-by-formulation (interaction)
α Probability level
λ Noncentrality parameter
μ Logarithmic mean
σ Logarithmic standard deviation
θ Regulatory limit for individual bioequivalence

Subscripts

o For switching measure (BEL, CV, or σ) separating conditions for unscaled and scaled average or individual bioequivalence
s For simulated value of a measure (CV or GMR)
D For subject-by-formulation interaction
R For reference formulation
T For test formulation
W For scaling variance (σ^2) in the denominator of the model for individual bioequivalence
WR For within-subject variance (σ^2) of the reference formulation
WT For within-subject variance (σ^2) of the test formulation

REFERENCES

1. L. B. Sheiner. Bioequivalence revisited. *Stat. Med.* **11**:1777–1778 (2001).
2. R. Schall and H. G. Luus. On population and individual bioequivalence. *Stat. Med.* **12**:1109–1124 (1993).
3. R. Schall. A unified view of individual, population, and average bioequivalence. In H. H. Blume and K. K. Midha (eds.), *Bio-International 2: Bioavailability, Bioequivalence, and Pharmacokinetic Studies*, Medpharm, Stuttgart, 1995, pp. 91–106.
4. R. Schall and R. L. Williams. Towards a practical strategy for assessing individual bioequivalence. *J. Pharmacokinet. Biopharm.* **24**:133–149 (1996).
5. R. N. Patnaik, L. J. Lesko, M.-L. Chen, and R. L. Williams. Individual bioequivalence: new concepts in the statistical assessment of bioequivalence metrics. *Clin. Pharmacokin.* **33**:1–6 (1997).
6. M.-L. Chen and L. J. Lesko. Individual bioequivalence revisited. *Clin. Pharmacokin.* **40**:701–706 (2001).
7. Food and Drug Administration. *Statistical Approaches to Establishing Bioequivalence*, Center for Drug Evaluation and Research (CDER), Rockville, Maryland (2001).
8. L. Endrenyi, G. L. Amidon, K. K. Midha, and J. P. Skelly. Individual bioequivalence: attractive in principle, difficult in practice. *Pharm. Res.* **15**:1321–1325 (1998).
9. A. L. Gould. A practical approach for evaluating population and individual bioequivalence. *Stat. Med.* **19**:2721–2740 (2000).
10. J. S. Barrett, V. Batra, A. Chow, J. Cook, A. L. Gould, A. H. Heller, M. W. Lo, S. D. Patterson, B. P. Smith, J. A. Stritar, J. M. Vega, and N. Zariffa. PhRMA perspective on population and individual bioequivalence. *J. Clin. Pharmacol.* **40**:561–570 (2000).
11. L. Tothfalusi and L. Endrenyi. Evaluation of some properties of individual bioequivalence (IBE) from replicate-design studies. *Int. J. Clin. Pharmacol. Ther.* **39**:162–166 (2001).

12. S. Senn. Statistical issues in bioequivalence. *Stat. Med.* **20**:2785–2799 (2001).
13. V. W. Steinijans. Some conceptual issues in the evaluation of average, population, and individual bioequivalence. *Drug Information J.* **35**:893–899 (2001).
14. L. Tothfalusi, L. Endrenyi, K. K. Midha, M. J. Rawson, and J. W. Hubbard. Evaluation of the bioequivalence of highly variable drugs and drug products. *Pharm. Res.* **18**:728–733 (2001).
15. A. W. Boddy, F. C. Snikeris, R. O. Kringle, G. C. G. Wei, J. A. Opperman, and K. K. Midha. An approach for widening the bioequivalence limits in the case of highly variable drugs. *Pharm. Res.* **12**:1865–1868 (1995).
16. A. J. Jackson. Determination of *in vivo* bioequivalence. *Pharm. Res.* **19**:227–228 (2002).
17. J. Vuorinen and J. Turunen. A three-step procedure for assessing bioequivalence in the general mixed-model framework. *Stat. Med.* **15**:2635–2655 (1996).
18. J. Vuorinen and J. Turunen. A simple three-step procedure for parametric and nonparametric assessment of bioequivalence. *Drug Information J.* **31**:167–180 (1997).
19. S.-C. Chow. Individual bioequivalence — a review of the FDA draft guidance. *Drug Information J.* **33**:435–444 (1999).
20. H. H. Blume and K. K. Midha. Conference report. In K. K. Midha and H. H. Blume (eds.), *Bio-International 92: Bioavailability, Bioequivalence, and Pharmacokinetics*, Medpharm, Stuttgart, 1993, pp. 13–23.
21. H. H. Blume, I. J. McGilveray, and K. K. Midha. Main conference report. In H. H. Blume and K. K. Midha (eds.), *Bio-International 2: Bioavailability, Bioequivalence, and Pharmacokinetic Studies*, Medpharm, Stuttgart, 1995, pp. 15–25.
22. V. P. Shah, A. Yacobi, W. H. Barr, L. Z. Benet, D. Breimer, M. R. Dobrinska, L. Endrenyi, W. Fairweather, W. Gillespie, M. A. Gonzalez, J. Hooper, A. Jackson, L. J. Lesko, K. K. Midha, P. K. Noonan, R. Patnaik, and R. L. Williams. Evaluation of orally administered highly variable drugs and drug formulations. *Pharm. Res.* **13**:1590–1594 (1996).
23. R. Schall. Assessment of individual and population bioequivalence using the probability that bioavailabilities are similar. *Biometrics* **51**:615–626 (1995).
24. T. Hyslop, F. Hsuan, and D. J. Holder. A small sample confidence interval approach to assess individual bioequivalence. *Stat. Med.* **19**:2885–2897 (1997).
25. S. D. Patterson, N. M.-D. Zariffa, T. H. Montague, and K. Howland. Non-traditional study designs to demonstrate average bioequivalence for highly variable drug products. *Eur. J. Clin. Pharmacol.* **57**:663–670 (2001).
26. J. Cohen. *Statistical Power Analysis for the Behavioral Sciences*, Lawrence Erlbaum Associates, Hillsdale, New Jersey, 1988.
27. A. R. Feinstein. Indexes of contrast and quantitative significance for comparisons of two groups. *Stat. Med.* **18**:2557–2581 (1999).