Regulatory Conditions for the Determination of Bioequivalence of Highly Variable Drugs

Laszlo Endrenyi[1] and Laszlo Tothfalusi[2]

[1] University of Toronto**,** Department of Pharmacology**,** Toronto, Ontario, Canada**.** [2] Semmelweis University of Medicine**,** Department of Pharmacodynamics**,** Budapest, Hungary

ABSTRACT - Purpose. The FDA Working Group on Highly Variable (HV) Drugs recently presented procedures and conditions for determining the bioequivalence (BE) of HV drug products. They included analysis by the method of scaled average BE (SABE), a switching coefficient of variation of $CV_S = 30\%$ and a regulatory standardized variation of $CV_0 = 25\%$ for applying SABE, and the use of a secondary regulatory criterion restricting to 0.80-1.25 the point estimate for the ratio of the estimated geometric means (GMR) of the two formulations. These conditions are scrutinized in the present communication. **Methods.** 3-period BE studies were simulated with various statistical and regulatory assumptions. Power curves, obtained by gradually increasing the true GMR, compared performances of the methods of SABE, a constrained point estimate of GMR (PE/GMR), and the composite of these two approaches. The consumer risk of each procedure was evaluated. **Results.** With $CV_0 = 30\%$ and PE/GMR = 0.80-1.25, the composite criterion of BE relied on the confidence limits of SABE. In contrast, with $CV_0 = 25\%$ and/or PE/GMR = 0.87-1.15, the composite criterion approached almost completely the features of the GMR point estimate, especially at high within-subject variation. The consumer risk was near 5% with $CV_0 = 30\%$ but about 15% when $CV_0 = 25\%$. **Conclusions.** A constraint on GMR is difficult to justify scientifically. Still, if it is introduced then the condition of $CV_S = CV_0 = 30\%$ and PE/GMR = 0.80-1.25 is recommended as a composite regulatory criterion. With alternative settings of the conditions, such as the recommended $CV_0 = 25\%$ and/or PE/GMR = 0.87-1.15, the composite criterion would reflect almost entirely the GMR point estimate. Also, with $CV_0 = 25\%$, the BE limits are discontinuous at $CV_S = 30\%$ and, as consequences, the consumer risk is substantially larger than 5%, and the regulatory uncertainty for making a decision about acceptance or rejection is enhanced. These would be undesirable outcomes.

INTRODUCTION

The determination of bioequivalence (BE) for highly-variable (HV) drugs has been a difficult, frustrating problem. Most usually, two drug products are deemed to be bioequivalent if the 90% confidence interval around the estimated ratio of geometric means (GMR) of relevant metrics is between 0.80 and 1.25. [The relevant metrics are usually the area under the curve (AUC) contrasting plasma concentrations and time, and the maximum concentration (C_{max}).] In crossover studies, the length of the confidence interval is proportional to the within-subject standard deviation of the drug, and reciprocally proportional to the square-root of the number of subjects. Consequently, the regulatory BE limits of 0.80 and 1.25 are frequently penetrated when the intraindividual variation is high unless the number of subjects is also large.

Benet (1) raised concerns about exposing a large number of healthy subjects to drugs. He noted that highly-variable drugs are safe and have wide therapeutic indices, an observation recently re-emphasized by Midha et al. (2). Therefore Benet (3) questioned the use of excessive numbers of volunteers for the sake of satisfying identically preset statistical criteria ("one size fits all").

The issue of bioequivalence for HV drugs has been discussed for a long time. It was featured at several conferences over an extended period (4-8).

___

Corresponding Author: Laszlo Endrenyi, Ph.D. Department of Pharmacology, University of Toronto, Toronto, ON, M5S lA8, E-mail: l.endrenyi@utoronto.ca

Methods were recently proposed for dealing with the problem. Notably, Boddy et al. (9) suggested that, for HV drugs, the BE limits could be expanded in proportion to the within-subject variation. Tothfalusi et al. (10) recommended that the usual criterion applied for the determination of BE could be scaled by the within-subject variation.

European and American regulatory authorities have proceeded in recent years towards resolving the issue. In Europe, the topic was regularly and intensively discussed at several regulatory forums. A Concept Paper was published in June, 2006 (11) and the issue could be actively considered further.

In the United States, a Working Group was established in order to develop relevant procedures and eventual regulations. The topic was discussed at meetings of the Advisory Committee for Pharmaceutical Sciences in 1997, 1999, 2004 and 2006. At the meeting on October 6, 2006 the Working Group presented preferred regulatory and study conditions. These were recently published together with their quantitative rationale (12, 13).

The purpose of the present communication is to evaluate some of these conditions and to offer comments and suggestions about them.

BACKGROUND

Scaled Average Bioequivalence

Regulatory authorities appear to move towards adopting the approach of scaled average bioequivalence (SABE) as a tool for dealing with the problem of bioequivalence for HV drugs. Therefore, a brief background of the procedure will be summarized.

The two one-sided tests procedure (14) is generally applied for determinations of bioequivalence. In practice, BE is evaluated by calculating logarithmic quantities. Thus, means and standard deviations of the logarithmic data (μ and σ) are estimated.

Bioequivalence is declared if the difference between the logarithmic averages is between limits (BEL_A) which are preset by regulatory authorities. Therefore, *average bioequivalence* (ABE) is accepted if the following criterion is satisfied:

$$- BEL_A \leq \mu_T - \mu_R \leq BEL_A \qquad (1)$$

The most usually applied regulatory limit is:

$$BEL_A = \ln(1.25) \qquad (1A)$$

This assures the earlier stated expectation that the regulatory limits for the ratio of geometric means of metrics are 0.80 and 1.25. In practice, the 90% confidence interval around the difference between the estimated logarithmic averages should be between the regulatory limits.

Thus, regulators need to define, in the case of average BE, a single criterion for declaring bioequivalence such as that given in Eq. (1A). For highly-variable drugs, evaluated by scaled average BE, *two* quantities must be defined. They will be discussed below.

The *regulatory criterion* suggested for the application of *scaled average BE* is (10, 15):

$$-BEL_S \leq (\mu_T-\mu_R)/\sigma_W \leq BEL_S \quad (2)$$

Here a scaling standard deviation (σ_W) is related to the within-subject standard deviation of the reference formulation (σ_{WR}) or, in other views, is identical to it. This distinction will be discussed later.

Tothfalusi et al. (15) suggested that the scaled BE limits (BEL_S) should be set in the following form:

$$BEL_S = \ln(1.25)/\sigma_0 \qquad (2A)$$

Here σ_0 is the first measure which should be defined by regulators. It will be referred to as the *regulatory standardized variation*. It defines the proportionality factor between the logarithmic BE limits and σ_W in the highly-variable region (see Figure 1A). σ_0 uniquely determines BEL_s and vice versa. For example, when $\sigma_0 = 0.294$ then BEL_s is 0.759, and when $\sigma_0 = 0.246$ then BEL_s is 0.907.

Rearranging equation (2), an alternative form is obtained:

$$-BEL_S\sigma_W \leq \mu_T-\mu_R \leq BEL_S\sigma_W \qquad (2B)$$

This form represents average bioequivalence with expanding limits (ABEL). (9) Consequently, Eq. 2 and Eq. 2B, i.e. the approaches of SABE and ABEL, are (almost) identical.

Using the limits of ABEL helps to understand the properties of SABE from the perspective of ABE. In this context, the regulatory standardized variation (σ_0) defines the proportionality factor between the logarithmic ABEL limits and σ_W (Figure 1A). A representation of ABEL conveniently illustrates a *mixed regulatory strategy* that was proposed for applying the unscaled and scaled approaches to the determination of BE (Figure 1). (15)

According to the mixed regulatory strategy, a second regulatory term, the so-called *switching variation* (CV_S), separates regions of low and high variabilities. If the variation of the drug is low, i.e., when it does not exceed the switching variation ($CV_W \leq CV_S$) then, following the present practice, unscaled average BE should be evaluated. However, for HV drugs when the variability is higher than the switching variation ($CV_W > CV_S$), scaled average BE is applied.

The mixed regulatory strategy is depicted in Figure 1 where, for illustrative purposes, SABE-equivalent ABEL limits ($BEL_E*\sigma_w$) are plotted. Two different SABE-equivalent ABEL limits are shown which correspond to two different values of σ_0. How to set σ_0 is the main focus of this communication. Several aspects of Figure 1 will be discussed later.

The standard deviations (σ) can be converted, approximately, to the corresponding coefficients of variation:

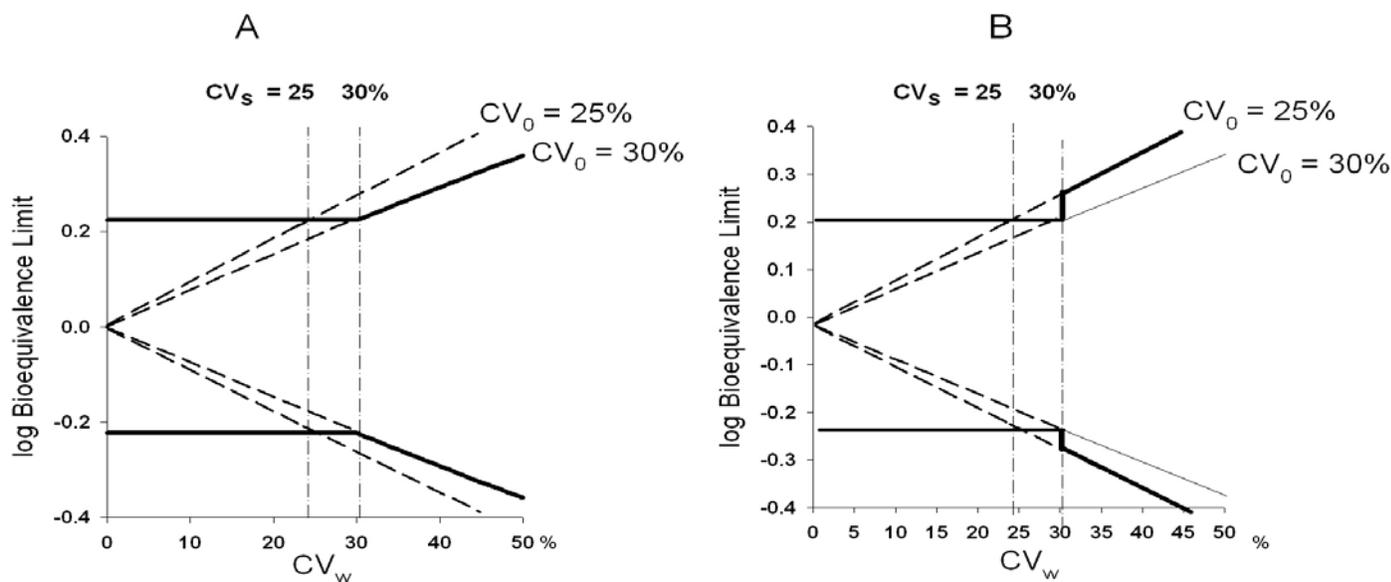$$CV = 100[\exp(\sigma^2)-1)]^{1/2} \qquad (3)$$

Figure 1. Mixed regulatory model for the determination of bioequivalence. The logarithmic BE limits, for determinations of average BE with constant and expanding limits, are shown by thick lines. If the within-subject variation (CV_W) does not exceed the switching variation (CV_S) then unscaled average BE is applied, and the BE limits have a constant level of $\pm\log(1.25)$. When the within-subject variation is higher than the switching variation then the limits widen with increasing within-subject variation, and scaled average BE can be applied. The slope (in the logarithmic scale) of the expansion is determined by the regulatory standardized variation (CV_0). The logarithmic average and the SABE-equivalent BE limits are shown by thick lines. (A) The regulatory standardized variation equals the switching variation, $CV_0 = CV_S = 30\%$. (B) The regulatory standardized variation is lower than the switching variation, $CV_0 = 25\%$ and $CV_S = 30\%$. The BE limits have a discontinuity at the switching variation.

Therefore, for unified and convenient treatment, the regulatory constants are expressed in terms of coefficients of variation. As an alternative notation, CV_0 will be used instead of σ_0 and the transformation rule between CV_0 and σ_0, given by Eq. 3, will be applied. For example, if $\sigma_0 = 0.294$ then $CV_0 = 30\%$, and when $\sigma_0 = 0.246$ then $CV_0 = 25\%$. The advantage of this unified notation is that an additional GMR restriction rule (see later) also can be expressed in relative terms. The 0.80-1.25 GMR restriction criterion becomes a regulatory constraint of 25%. Thus, in our notation, the proposed mixed approach depends on three regulatory constants, CV_s, CV_0 and CV_{GMR}, with typical values of 30%, 30% and 25%.

Considerations on the Implementation of Scaled Average Bioequivalence: the Recommendations of FDA

As noted earlier, the Advisory Committee for Pharmaceutical Sciences discussed the topic repeatedly. At its meeting, on October 6, 2006, important presentations were offered on behalf of the FDA Working Group on Highly Variable Drugs (16-18). The interim recommendations of FDA were further clarified on May 22, 2007 at an AAPS/FDA workshop (19, 20). The current proposals of FDA and their quantitative characteristics were published very recently (12, 13).

FDA has proposed to apply the approach of reference-scaled average BE for determining the BE of HV drugs (12, 17-20). This means that $\sigma_W = \sigma_{WR}$ would be adopted for scaling.

FDA suggests also that the acceptance criteria include a constraint on the point estimate for the ratio of geometric means (GMR). It recommends that GMR be limited to the range of 0.80 to 1.25. The Advisory Committee concurred with this proposal but some members actually favoured a narrower range. The rationale for this recommendation, and the resulting features, will be discussed later. FDA proposes that both AUC and C_{max} should satisfy the BE acceptance criteria (12, 18, 20).

FDA recommends that three-period BE studies be performed in which the reference product [R] is provided twice and the test product (T) is given once (12, 13, 17-20). Consequently, the possible sequences of drug administration are TRR, RRT, and RTR.

The FDA Working Group performed simulations in order to ascertain the features of the above proposals (13, 17, 19). The current FDA recommendations include a value of $\sigma_0 = 0.25$ (12, 18, 20). FDA suggests also that unscaled average BE be used if the within-subject variability is less than 30%, and that reference-scaled average BE be applied if the within-subject variability is at least 30% (12, 18, 20). These suggestions correspond to a switching coefficient of variation of $CV_S = 30\%$.

METHODS

In order to compare the performances of the procedures to be applied for the determination of BE, power curves were simulated under various assumptions and conditions. The principles and procedures of the simulations were described earlier (10, 15). Therefore, they will be summarized only briefly.

The power curves characterize the relationship between the probability of accepting the declaration of BE and the magnitude of the true GMR. Therefore, BE studies were repeatedly simulated while gradually raising the true GMR from 1.0, the condition of true BE, towards increasing deviations from this value. At each level of the true GMR, the proportion of simulated BE studies was recorded in which the declaration of BE was accepted.

A computer program written in Compaq Visual Fortran (ver. 6.1) simulated, under each condition, 10,000 BE trials. 3-period studies were considered in which the reference product was measured twice in each of 36 subjects. The within-subject coefficient of variation was assumed to be either 35 or 60%. The regulatory standardized variation was set to either $\sigma_0 = 0.246$ or 0.294; these values corresponded to coefficients of variation of $CV_0 = 25\%$ and 30%, respectively (15). The resulting BE limits for scaled average BE were $BEL_S = 0.906$ and 0.760, respectively (15). The declaration of BE was based on the linearization of Eq. 2 as suggested by Hyslop et al. (21).

The characteristics of (1) scaled average BE (SABE), (2) the point estimate of GMR, (3) the combination of SABE and the point estimate, and (4) unscaled average BE (ABE) were evaluated.

Additional simulations evaluated the consumer risk under various assumptions. The consumer risk is the probability of declaring bioequivalence when the compared drug products are actually bioinequivalent (22). Regulators wish to maintain it at a low level, usually at 5%. This means, with the usual approach of average BE, that if the true GMR is either 0.80 or 1.25 then the maximum probability of accepting a formulation with such a GMR value should not be higher than 5%. The same statement applies also when the mixed strategy is utilized since it assures the continuity of the probabilities. A million studies were simulated with each of the assumptions, and the percentage of studies declaring BE, i.e. the apparent consumer risk, was recorded.

The consumer risk was evaluated, with its quantitative definition for unscaled average BE, as described above. The calculations considered either that the mixed strategy was utilized or that it was not applied. For the sake of clarity, all variations were expressed as coefficients of variation. The switching variation was maintained at $CV_S = 30\%$ whereas the regulatory standardized variation took one of the two levels of $CV_0 = 25\%$ or 30%. The true within-subject coefficient of variation for the reference product was 30%. (As noted earlier, FDA currently recommends a value of $\sigma_0 = 0.25$ (12, 18, 20). This corresponds to $CV_0 = 25.4\%$. The deviation is negligible for the purposes and considerations of the present communication.)

The consumer risk was calculated by applying both unscaled and scaled average BE and also scaled average BE jointly with a constrained point estimate on the GMR. In the latter simulations, the constraint on the GMR was set at 25%.

RESULTS

Effects of the Switching Variation and the GMR Constraint

Figures 2 and 3 present power curves simulated by assuming that the constraint on the point estimate of GMR is either 25% (Fig. 2) or 15% (Fig. 3), i.e., that the point estimate was limited to the range of either from 0.80 to 1.25 (Fig. 2) or 0.87 to 1.15 (Fig. 3). The regulatory standardized variation and the switching variation are in some cases the same, either $CV_0 = CV_S = 30\%$ (A and B in Figs. 2 and 3) or $CV_0 = CV_S = 25\%$ (C and D in Figs. 2 and 3).

In other cases, the regulatory standard variation is lower than the switching variation, $CV_0 = 25\%$ and $CV_S = 30\%$ (E and F in Fig. 2 only). The within-subject variation is either $CV_W = 35\%$ (A, C and E) or 60% (B, D and F). $CV_W = 35\%$ rather than 30% was chosen as the lower level of intraindividual variation in order to avoid complications arising from applying the mixed model of estimation around the switching variation (15). The structure of the diagrams is to some extent similar to those presented by Haidar (13, 17, 19) and by Tothfalusi et al. (10, 15, 23).

Some general features are worth noting. First, the power curve for the point estimate of GMR shows 50% acceptance at the level of the constraint, either at a 25% (Fig.2) or 15% (Fig. 3) difference between the estimated means. Second, the proportion of acceptance for the combined criterion is always less than the corresponding acceptances of both component criteria (24). Consequently, the combined criterion is at least as demanding as the stricter of the two component criteria.

With a 25% constraint on the GMR point estimate (Figure 2), when both the regulatory standardized variation and the switching variation are $CV_0 = CV_s = 30\%$, the SABE criterion with the 90% confidence interval is stricter than the GMR constraint (Figures 2A and B). The combined criterion is somewhat more severe than the scaled approach; at moderately high within-subject variation (Fig. 2A), the two power curves almost coincide.

At a regulatory standardized variation and switching variation of $CV_0 = CV_S = 25\%$ and at moderately high variation (Fig. 2C), similar relationships prevail among the power curves: the SABE criterion is stricter than the GMR constraint, and the power curve for the former almost coincides with that of the combined criterion. In contrast, at high within-subject variation (Fig. 2D), the constrained point estimate of GMR is much more severe than SABE and it dominates the combined criterion. Some of these results are similar to those obtained by Haidar (13, 17, 19).
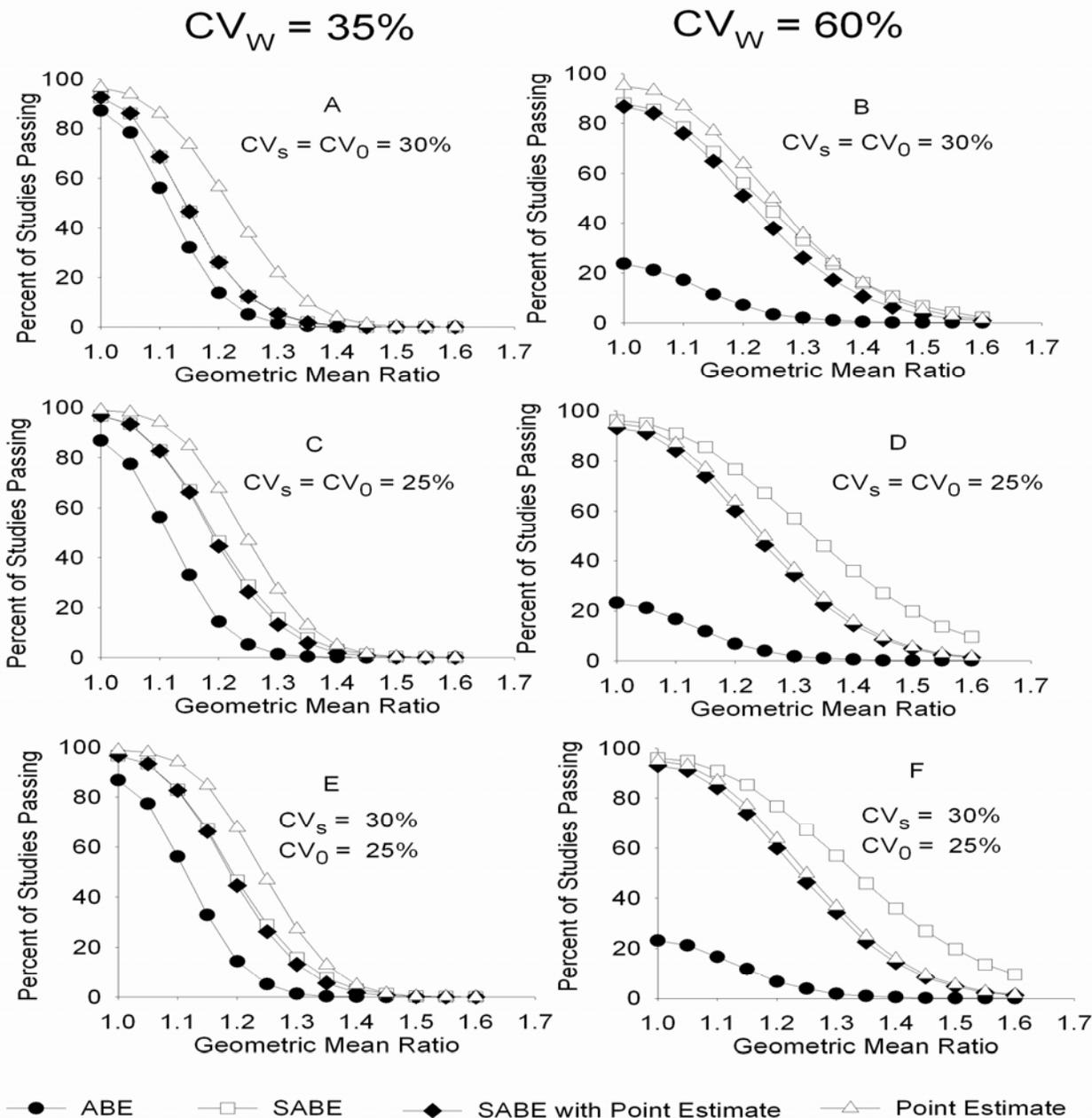
Figure 2. Power curves contrasting the proportion (in %) of simulated studies in which bioequivalence was accepted by various methods with the assumed, true ratio of geometric means (GMR) of the two drug products. Characteristics of four testing methods were compared: Unscaled average BE (ABE), scaled average BE, the point estimate of GMR, and the combination of scaled average BE with the point estimate. The point estimate of GMR was constrained to 1.25. The within-subject variation (CV_W) was either 35% (A, C and E) or 60% (B, D and F). The regulatory standardized variation equaled the switching variation, $CV_0 = CV_S$; both were either 30% (A and B) or 25% (C and D). For the last two diagrams, the regulatory standard deviation was lower than the switching variation, $CV_0 = 25\%$, $CV_S = 30\%$ (E and F).
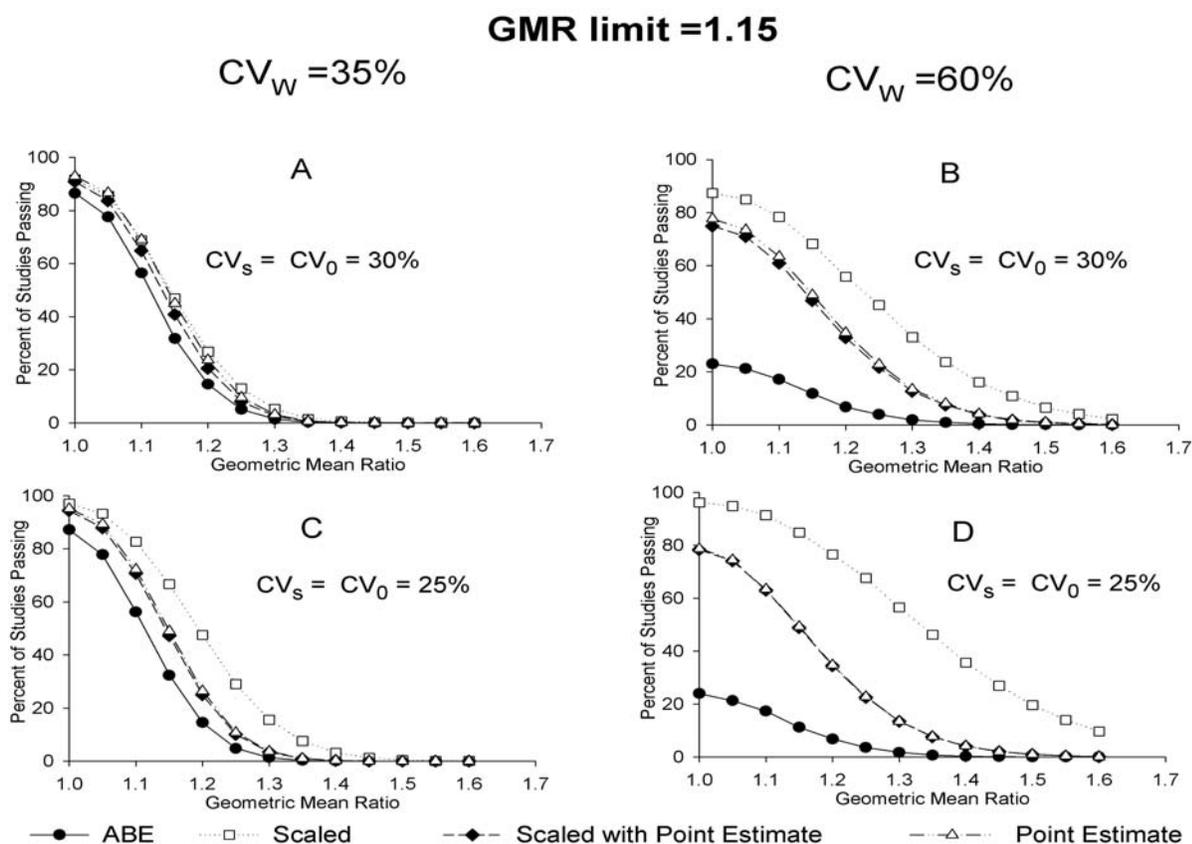
Figure 3. Power curves for the determination of bioequivalence. The conditions were the same as in Figure 2 except that the constraint on the point estimate of GMR was 1.15, and that the conditions with $CV_0 = 25\%$, $CV_S = 30\%$ (as in Figs. 2E and 2F) were not evaluated.

When the regulatory standardized variation is $CV_0 = 25\%$ and the switching variation is $CV_S = 30\%$ (Figs. 2E and 2F), the results are almost identical with those obtained with both variations being 25% (Figs. 2C and 2D). This observation indicates that deviations in features of the power curves can be attributed to changes of the regulatory standardized variation, and not of the switching variation. With a 15% constraint on the GMR point estimate (Figure 3), the GMR restriction is always more severe than the SABE approach, and it always dominates the combined criterion.

Effect of CV_0 on the Consumer Risk

Table 1 presents the consumer risks (in %) which were evaluated under various assumptions. A switching variation (CV_S) of 30% was considered.

When the mixed regulatory strategy was applied then the estimated within-subject coefficient of variation for the reference formulation (CV_{WR}) determined the decision for utilizing either unscaled or scaled average BE. Therefore, as noted earlier, unscaled ABE was used when the estimated within-subject variation did not exceed 30% ($CV_{WR} \leq 30\%$), and scaled ABE was utilized when the estimated variation was higher than 30% ($CV_{WR} > 30\%$). When the mixed regulatory strategy was not applied then the choice of the method of evaluation was determined in advance.

The most conspicuous result is that the maximum consumer risk is very high when scaled ABE is applied with a regulatory standardized variation (CV_0) of 25%. This condition is shown in Figure 1B which illustrates a discontinuity in the regulatory expectations at $CV_W = 30\%$.

Table 1. Consumer risk for the determination of BE

Mixed Strategy	Regulatory Standardized Variation (%) (CV$_0$)	Consumer Risk (%)		
		Unscaled ABE	Scaled ABE	
			Without	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	14.6

Switching variation: $CV_S = 30\%$
Within-subject variation: $CV_W = 30\%$
Constraint on point estimate of GMR: 0.80 to 1.25 (i.e., 25%)

Unscaled average BE yields a consumer risk of 5%. Scaled average BE results in a slightly higher consumer risk when $CV_0 = 30\%$. The small increase in the consumer risk is due to the approximation in calculating the linearized confidence limits of Hyslop et al. (21) and, when the mixed strategy is applied, also to the resulting uncertainties.

An additional constraint on the GMR leaves the consumer risk unchanged.

DISCUSSION

Constraint on the Point Estimate of GMR

The Concern about High and Low GMR Values, and the Possible Need for a Constraint

Benet (3) noted, first in the context of individual BE, that large deviations can occur between the means of the two drug products. Therefore he recommended (3) that "additional point estimate criteria be added for all drugs as a supplement to the bioequivalence limit criteria." He repeated his recommendation also in the case of BE for highly-variable drugs "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" (3, 24).

Benet (24), however, noted also the following: "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." We completely agree with Benet's (24) views. Some of the consequences of the scientific futility of the GMR constraint will be considered later.

As noted earlier, the FDA Working Group on HV drugs recommended the implementation of Benet's concern and suggested that the point estimates of the GMRs of AUC and C_{max} be constrained to the range of 0.80-1.25. The FDA Advisory Committee for Pharmaceutical Sciences also agreed, in 2004 and again in 2006, with the recommendation that a constraint on the point estimate be added to the confidence interval criterion when scaled average BE is applied. Some committee members preferred a narrower range of the point estimate.

Benet (24) suggested a constraint on the GMR point estimate of ±15% for AUC and ±20% for C_{max}. These recommendations were stated also in the report of the 2005 BioInternational Conference (8). Our results suggest that this constraint would be too severe since it would yield a BE criterion which would reflect only the GMR point estimate but not the 90% confidence interval around a BE regulatory model.

Effect of the Constrained GMR Point Estimate on the Evaluation of BE

Figures 2 and 3 demonstrate, under various conditions, the effect of the suggested secondary regulatory criterion, the constraint on the GMR

point estimate. It is recalled that the primary regulatory criterion expects that the 90% confidence interval of the scaled logarithmic GMR should remain within its preset limits. Thus, the diagrams illustrate the extent to which the secondary criterion alters the outcome of the primary test.

When the regulatory standardized variation is $CV_0 = 25\%$ then even a GMR constraint of 25% substantially modifies, under many conditions, the outcome of BE determination which is based only on the primary criterion (Figures 2C, 2D, 2E, and 2F). At higher variations, in fact, the result of the BE evaluation relies practically entirely on the secondary criterion and the primary, confidence interval criterion is irrelevant (Figures 2D and 2F). These conclusions are relevant regardless whether the switching variation is $CV_S =25\%$ (Figs. 2C and 2D) or 30% (Figs. 2E and 2F). These results parallel, in part, the observations of Haidar (13, 17, 19) and Tothfalusi et al. (25).

Thus, under these conditions, the determination of BE becomes a test of just assessing the GMR. This would be in analogy of the criterion expected by Health Canada which, however, is applied only to C_{max} (26). Consequently, the overall, combined regulatory criterion could amount merely to the determination of GMR. Under such conditions, it would be entirely unnecessary and even pointless to evaluate SABE. It is doubtful that such a regulatory expectation would be desirable.

By contrast, when the regulatory standardized and the switching variations are $CV_0 = CV_s = 30\%$ and the GMR constraint is 0.80-1.25 then the imposition of the constraint has generally little effect (Figures 2A and 2B). The character of the BE test remains essentially that of the primary criterion. At the same time, patients and physicians still receive assurance that even the average responses of the two drug products are not meaningfully different. Consequently, these results and considerations suggest that the regulatory standardized variation be set at the level of $CV_0 = 30\%$ and not at 25%.

Simulations were performed in order to separate the levels of the regulatory standardized ($CV_0 = 25\%$) and the switching variations ($CV_S = 30\%$) and to ascertain their distinct effects (Figures 2E and 2F). The comparative positions of the power curves were very similar to those seen when both variations were at the lower level of 25% (Figures 2C and 2D). Consequently, the effect of the GMR constraint on the BE determination is influenced by the level of the regulatory standardized variation and not by that of the switching variation.

A tighter GMR constraint, such as 0.87-1.15, accelerates the utilization of the GMR point estimate as the sole regulatory criterion (Figure 3). This feature appears to be disadvantageous.

Simple considerations support the observed relationship between the composite test and its components. SABE and the GMR point estimate can be viewed as tests which accept or do not accept bioequivalence with probabilities defined by their power curves. These power curves are shown in Figures 2 and 3. Assume that the two component tests, SABE and the GMR point estimate, accept bioequivalence, under a given condition, with probabilities of P1 and P2. If the two tests were independent then the overall success rate would be P1*P2, i.e., smaller than either component probability. Even though the assumption of independence does not hold, it is still true that the overall success rate of a composite test should be less than that of either component. If the powers of the two tests are very different then the one yielding the lower probability will be dominant. For example, if one of the tests shows complete acceptance (P1=1) then the overall success rate is 1*P2=P2.

The quantitative considerations, discussed above, reinforce the scientific futility of applying the GMR constraint. Under various conditions, such as for vey highly variable drugs and/or with $\sigma_0 = 0.25$, the GMR constraint introduces a very undesirable feature for the determination of BE since it dominates the regulatory assessment.

The mathematics of the two simultaneously applied tests is quite complex. This was studied by Karalis et al. (27, 28) and Kytariolos et al. (29). They developed several interesting, combined regulatory criteria with leveling-off properties which enable the testing of bioequivalence with SABE and a GMR constraint in a single step. However, the feasibility of these approaches still needs support.

Conclusion

A constraint on the point estimate of GMR is not justified scientifically. However, if it is applied for other reasons then a constraint of 0.80-1.25 is recommended. With the regulatory standardized and switching variations of $CV_0 = CV_s = 30\%$, the condition would continue to utilize 90% confidence intervals and essentially maintain the character of the primary regulatory criterion. Alternative conditions and restrictions could well transform the regulatory expectation to one involving merely the assessment of GMR.

The Magnitude of the Regulatory Standardized Variation (CV_0)

As noted earlier, FDA apparently suggests that the regulatory standardized variation be $CV_0 = 25.4\%$ (or $\sigma_0 = 0.25$). The rationale seems to be that this value is a compromise between the strictness of 30% and laxness of 20% (13, 17).

One of the issues is that, in this approach, the regulatory standardized variation could differ from the switching variation. As demonstrated in Figure 1B and as discussed further below, this could lead to a discontinuity in the regulatory expectations and, consequently, to the increase of consumer risk and heightened regulatory uncertainty.

Regulators could be concerned when a drug that is in fact not highly variable, is erroneously classified as highly variable. The likelihood of misclassification is highest when the true within-subject variation is near $CV_W = 30\%$ since, in this case, almost half of the assignments to be an HV drug are incorrect. The consumer risks were estimated in this worst-case condition (Table 1).

Tothfalusi and Endrenyi (15) concluded that the regulatory standardized variation and the switching variation should have the same magnitudes, otherwise the consumer risk would rise above the regulatory expectation of 5%. The results in Table 1 confirm this conclusion. When both the regulatory standard variation and the switching variation have the same value, $CV_0 = CV_S = 30\%$, then for not truly highly-variable drugs, the consumer risk is nearly 5%. In contrast, when the two variations have differing magnitudes, $CV_0 = 25\%$ and $CV_S = 30\%$, then the consumer risk triples to about 15-16%. It is noted that these results disagree with the conclusions and statements of Haidar et al. (13).

Application of ABE on one side of the switching variation and SABE on the other side should not be of concern as long as $CV_0 = 30\%$ is assumed (Figure 1A). In other words, it should not matter much whether the estimated CV_W is 29 % or 31%. The regulatory conditions are continuous around $CV_S = 30\%$, and the probabilities of acceptance and rejection are only slightly different (but "*similar*") on the two sides of CV_S (and not really more different from when, e.g., estimated CV_W's of 26% and 28% are compared or when estimated CV_W=s of 31% and 33% are contrasted). The regulators need not be concerned in this case since there is *no, or little regulatory uncertainty* for applying the mixed criterion in order to reach a decision on acceptance or rejection. Also, as observed above, the *consumer risk*, with the assumption of $CV_0 = 30\%$, *is close to 5%* (Table 1).

In contrast, if $\sigma_0 = 0.25$ is assumed, as in the proposal of FDA (12, 18, 20), then the BE limits are discontinuous at $CV_S = 30\%$ (Figure 1B). (As stated earlier, the condition recommended by FDA corresponds to $CV_0 = 25.4\%$. The difference from $CV_0 = 25.0\%$, applied in our simulation, is slight and not significant for the considerations of the present communication).

The consequences include *the enhancement of the consumer risk* for truly not-highly variable drugs (Table 1). As a consequence, the probabilities of acceptance and rejection are *different* on the two sides of CV_S. Differing probabilities prevail depending on whether the estimated CV_W is 29% or 31%, a lower probability of acceptance when the estimated CV_W is 29% than when it is 31%.

This raises a rightful uncertainty for regulators. The *heightened regulatory uncertainty should be of great concern to regulators.*

Conclusion

If, following the suggestion of FDA (12, 18, 20), a regulatory standardized variation of $CV_0 = 25.4\%$ (or $\sigma_0 = 0.25$) is assumed then the BE limits are discontinuous at $CV_S = 30\%$ and, as consequences, the consumer risk is substantially larger than 5%, and the regulatory uncertainty for making a decision about acceptance or rejection is enhanced. In contrast, if $CV_0 = 30\%$ (or $\sigma_0 = $

0.294) is assumed then the BE limits are continuous, the consumer risk is around 5% which is the conventionally accepted risk level.

REFERENCES

[1]. L. Benet. Bioavailability and bioequivalence: definitions and difficulties inacceptance criteria. In K.K. Midha and H.H. Blume (eds.), *Bio-International: Bioavailability, Bioequivalence and Pharmacokinetics,* Medpharm, Stuttgart, 1995, pp. 27-35.

[2]. K.K. Midha, M.J. Rawson, and J.W. Hubbard. The bioequivalence of highly variable drugs and drug products. *Int. J. Clin. Pharmacol. Ther.* **43:** 485-498 (2005).

[3]. L. Benet. Individual bioequivalence: an overview. *AAPS International Workshop on Individual Bioequivalence: Realities and Implementation*, Montreal, Quebec, Aug 30-Sep 1, 1999.

[4]. H.H. Blume and K.K. Midha. Bio-International 92, Conference on Bioavailability, Bioequivalence and Pharmacokinetic Studies. *J. Pharm. Sci.* **11:** 1186-1189 (1993).

[5]. H.H. Blume, I.J. McGilveray, and K.K. Midha. Bio-International 94, Conference on Bioavailability, Bioequivalence and Pharmacokinetic Studies. *Eur. J. Pharm. Sci.* **3:** 113-124 (1995).

[6]. V.P. Shah, A. Yacobi, W.H. Barr, 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 drugs and drug formulations. *Pharm. Res.* **13:** 1590-1594 (1996).

[7]. K.K. Midha, T. Nagai, H.H. Blume, J.W. Hubbard, I.J. McGilveray, and R.L. Williams. Conference Report: Bio-International '96, Conference on Bioavailability, Bioequivalence and Pharmacokinetic Studies, Tokyo, Japan, April 22-24, 1996. In K.K. Midha and T. Nagai (eds.), *Bioavailability, Bioequivalence and Pharmacokinetic Studies*, Academic Societies Japan, Tokyo, 1996, pp. 1-8.

[8]. K.K. Midha, V.P. Shah, G.J.P. Singh, and R. Patnaik. Conference report: Bio-International 2005. *J. Pharm. Sci.* **96:** 747-754 (2007).

[9]. 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 acceptance limits in the case of highly variable drugs. *Pharm. Res.* **12:** 1865-1868 (1995).

[10]. 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).

[11]. BEBAC reviews. Highly variable drugs, EU-EMEA. http://bebac.at/news/2006-04-27.htm (accessed 11/17/2008).

[12]. S.H. Haidar, B. Davit, M.-L. Chen, D. Conner, L.M. Lee, Q.H. Li, R. Lionberger, F. Makhlouf, D. Patel, D.J. Schuirmann, and L.X. Yu. Bioequivalence approaches for highly variable drugs and drug products. *Pharm. Res.* **15:** 237-241 (2008).

[13]. S.H. Haidar, F. Makhlouf, D.J. Schuirmann, T. Hyslop, B. Davit, D. Conner, and L.X. Yu. Evaluation of a scaling approach for the bioequivalence of highly variable drugs. *AAPS J.* **10:** 450-454 (2008).

[14]. D. Schuirmann. A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. *J. Pharmacokin. Biopharm.* **15:** 657-680 (1987).

[15]. L. Tothfalusi, and L. Endrenyi. Limits for the scaled average bioequivalence of highly variable drugs and drug products. *Pharm. Res.* **20:** 382-389 (2003).

[16]. L.X. Yu. Bioequivalence of highly variable drugs. Meeting of FDA Committee for Pharmaceutical Science, October 6, 2006. http://www.fda.gov/ohrms/dockets/ac/06/slides/2006-4241s2_1.htm (accessed 11/17/2008).

[17]. S.H. Haidar. Evaluation of the scaling approach for highly variable drugs. Meeting of FDA Committee for Pharmaceutical Science, October 6, 2006. http://www.fda.gov/ohrms/dockets/ac/06/slides/2006-4241s2_4.htm (accessed 11/17/2008).

[18]. B.M. Davit. Highly variable drugs – bioequivalence issues: FDA proposal under consideration. Meeting of FDA Committee for Pharmaceutical Science, October 6, 2006. http://www.fda.gov/ohrms/dockets/ac/06/slides/2006-4241s2_5.htm (accessed 11/17/2008).

[19]. S.H. Haidar. BE for highly variable drugs – FDA perspective. *AAPS/FDA Workshop on BE, BCS, and Beyond*, North Bethesda, MD, May 22, 2007. http://www.aapspharmaceutica.com/meetings/files/90/22Haidar.pdf (accessed 11/17/2008).

[20]. B.M. Davit. Highly variable drugs – bioequivalence issues: FDA proposal under consideration. *AAPS/FDA Workshop on BE, BCS, and Beyond*, North Bethesda, MD, May 22, 2007.

[21]. T. Hyslop, F. Hsuan, and D.J. Holder. A small sample confidence interval approach to assess individual bioequivalence. *Stat. Med.* **19:** 2885-2897 (2000).

[22]. D. Hauschke, V. Steinijans, and I. Pigeot. *Bioequivalence Studies in Drug Development.* Wiley, Chichester, 2007, pp. 44-46.

[23]. L. Endrenyi and L. Tothfalusi. Evaluation of bioequivalence of highly variable drugs. In I. Kanfer and L. Shargel (eds.), *Generic Drug Product Development – Bioequivalence Issues,* Informa, New York, London, 2008, pp. 97-121.

[24]. L. Benet. Why highly variable drugs are safer. Meeting of FDA Committee for Pharmaceutical Science, October 6, 2006. http://www.fda.gov/ohrms/dockets/ac/06/slides/ 2006-4241s2_2.htm (accessed 11/17/2008).

[25]. L. Tothfalusi, L. Endrenyi, and K.K. Midha. Scaling or wider bioequivalence limits for highly variable drugs and for the special case of

C_{max}. *Int. J. Clin. Pharmacol. Ther.* **41:** 217-225 (2003).

[26]. Health Canada. Guidance for Industry. Conduct and Analysis of Bioavailability and Bioequivalence Studies. Part A: Oral Dosage Formulatios Used for Systemic Effects. Health Canada, Ottawa, 1992.

[27]. V. Karalis, M. Symillides, and P. Macheras. Novel scaled average bioequivalence limits based on GMR and variability considerations. *Pharm. Res.* **21:** 1933-1942 (2004).

[28]. V. Karalis, P. Macheras, and M. Symillides. Geometric mean ratio-dependent scaled bioequivalence limits with leveling-off properties. *Eur. J. Pharm. Sci.* **26:** 54-61 (2005).

[29]. J. Kytariolos, V. Karalis, P. Macheras, and M. Symillides. Novel scaled bioequivalence limits with leveling-off properties. *Pharm. Res.* **23:** 2657-2664 (2006).