Evaluation of the Bioequivalence of Highly-Variable Drugs and Drug Products

Laszlo Tothfalusi,[1] Laszlo Endrenyi,[2,4] Kamal K. Midha,[3] Maureen J. Rawson,[3] and John W. Hubbard[3]

Received December 14, 2000; accepted March 7, 2001

Purpose. To establish procedures for the effective evaluation of bioequivalence (BE) for highly-variable drugs and drug products (HVD/P).

Methods. 2- and 4-period crossover BE studies with 24 subjects were simulated which generally assumed within-subject coefficients of variation of 40%. The relationship between the fraction of studies in which BE was accepted (the statistical power) and the ratio of geometric means (GMR) of the two formulations was evaluated for various methods of analysis. These included, primarily, scaled average BE (ABE), the corresponding approach of expanding BE limits (BEL), and, for comparison, unscaled ABE and scaled individual BE (IBE).

Results. Scaled ABE and expanding BEL showed very similar properties in both 2- and 4-period studies. They had steeper power curves than scaled IBE. Unscaled ABE had very low statistical power. The acceptance of BE by unscaled and scaled ABE and expanding BEL was almost independent of subject-by-formulation interaction and the ratio of within-subject variations of the two formulations. By contrast, the conclusions reached by scaled IBE were strongly affected by these parameters.

Conclusions. Scaled ABE and expanding BEL evaluate BE effectively for HVD/P in both 2- and 4-period investigations. However, additional, useful information can be obtained from 4-period studies.

KEY WORDS: highly-variable drugs; bioequivalence; scaled regulatory criterion; regulatory limits; 2- and 4-period crossover design; simulated clinical trials.

INTRODUCTION

The determination of the bioequivalence (BE) of two drug formulations has proved to be a difficult problem for highly-variable drugs and drug products (HVD/P) which are characterized by large within-subject variation. Customarily, two-way crossover studies have been conducted. The logarithmic means of the two kinetic responses (e.g., of AUCs, the areas under the concentration vs. time curves) have been evaluated together with the 90% confidence intervals (CI) of their difference. Bioequivalence of the formulations is declared if the CI is within preset regulatory limits. Typically, the limits have been set at ±log (1.25), with the corresponding untransformed bioequivalence limits (BEL) being between 0.80 and 1.25 for the ratio of the geometric means (GMR) of the two kinetic responses.

When the within-subject variation is large then the estimated CI is wide, and it is very difficult to remain within preset BE limits. An obvious remedy is to increase the number of subjects participating in a study and thereby to narrow the CI. However, a BE study becomes, as a result, very expensive and cumbersome.

To alleviate this difficulty, other approaches were proposed. Blume *et al.* (1) recommended that some kinetic parameters could have smaller variations if BE studies were conducted in the steady state instead of following single-dose administrations. A workshop discussing highly-variable drugs and drug products considered procedures for determining their BE (2). The approach for widening the BEL for this category drugs and drug products was particularly favored. Boddy *et al.* (3) quantitatively formalized the method. According to their procedure, the BEL expands in proportion to the estimated intrasubject coefficient of variation of the reference formulation.

Recent proposals related to the determination of individual and population BE offered a parallel approach to the regulatory criterion by scaling it with the intraindividual variance of the reference product (4–6). This variance can be evaluated in replicate-design investigations which are conducted over 3 or 4 study periods (5–7).

The same experimental design can be applied also to the determination of average bioequivalence (i.e., of the equivalence of the two mean responses) again by using reference scaling (8). It can be, however, reasonably expected that the procedure of scaling can be applied to the evaluation of average BE also from the results of 2-period crossover studies.

Therefore, the principal purpose of the present study is to compare the effectivenesses of 2- and 4-period investigations performed for the determination of average BE. The secondary goals include a contrast of the procedures of scaling and of extending the BEL, and a comparison of some characteristics of average and individual BE.

METHODS

The regulatory approaches to be discussed below will consider several regulatory models and criteria. Using either observed or simulated data, the parameters of a given model are estimated and substituted into the model expression. Confidence limits around the estimated magnitude of the model are then calculated. Therefore each of the regulatory criteria to be discussed will imply that the appropriate confidence limit(s) of the model should be within preset bioequivalence limit(s).

Unscaled Average Bioequivalence

The usual procedure for determining the bioequivalence of two formulations was evaluated for comparison with the other approaches. Accordingly, the average logarithmic kinetic responses of the test and reference products (μ_T and μ_R, respectively) are contrasted. Bioequivalence is declared if the 90% confidence limits around their logarithmically calculated difference are within preset BE limits (BEL):

$$-\text{BEL} \le \mu_T - \mu_R \le \text{BEL} \qquad (1)$$

[1] Semmelweis University of Budapest, Department of Pharmacodynamics, 1089 Budapest, Hungary.
[2] University of Toronto, Department of Pharmacology, Toronto, Ontario M5S 1A8, Canada.
[3] PharmaLytics Inc., Saskatoon Saskatchewan, S7N 5C9, Canada.
[4] To whom correspondence should be addressed. (e-mail: l.endrenyi@utoronto.ca)

0724-8741/01/0600-0728$19.50/0

which is equivalent to:

$$(\mu_T - \mu_R)^2 \leq BEL^2 \quad (2)$$

Usually, BEL = log(1.25) has been applied.

In 2-period simulated studies with N assumed subjects, the residual mean square (s^2_{Res}) with N-2 degrees of freedom, obtained in an analysis of variance (ANOVA), was used as the error term. The ANOVA contained the usual fixed components of formulations, sequences, and periods, and the random source of subjects-within-sequences. The estimated standard error for the difference between the means, SE = $s_{Res}(2/N)^{½}$, was used with the two balanced, simulated sequences to calculate the confidence limits.

In 4-period simulated studies, the ANOVA contained, in addition to sources of variation described for 2-period investigations, the subject-by-formulation interaction (within sequences) (6,9–11). The error term relevant for assessing the difference between the means was that of the interaction (s^2_{Int}) with N-2 degrees of freedom. The standard error for the difference of the means was, again for the balanced sequences, SE = $s_{Int}/(N)^{½}$.

Reference-Scaled Individual Bioequivalence

This approach was evaluated also for comparison. The regulatory model, derived by Schall and Luus (12) and adopted by FDA (6), expects that individual bioequivalence for HVD/P is declared if the upper 95% confidence limit for the sum of three scaled model components is within a preset BE limit (Θ_I):

$$[(\mu_T - \mu_R)^2 + (\sigma^2_{WT} - \sigma^2_{WR}) + \sigma^2_D]/\sigma^2_{WR} \leq \theta_I \quad (3)$$

Here σ^2_{WT} and σ^2_{WR} are the within-subject variances for the test and reference formulations, respectively, and σ^2_D is the variance component for the subject-by-formulation interaction. The denominator of σ^2_{WR} provides a scaling factor. FDA (6) suggested that the reference-scaled IBE be applied when the estimated magnitude of σ^2_{WR} exceeds a preset value of the variance (σ^2_{W0}): $\sigma^2_{WR} > \sigma^2_{W0}$. As the preset value, $\sigma^2_{W0} = 0.04$ was suggested (6). In the present study, however, properties of the reference-scaled IBE were determined regardless of the magnitude of the estimated σ^2_{WR}. This permitted the comparison of this approach with scaled average BE.

FDA (6) has taken into account that the model for individual BE (Eq. 3) contains three terms in the numerator, and not just one. Therefore, the regulatory limit is larger than that applied for average BE. It considers, in its numerator, not only the average BE limit of log(1.25) but also a variance factor (ϵ_I) accounting for the other terms in the model (Eq. 3). A value of $\epsilon_I = 0.05$ was recommended (6). Consequently, it was suggested that the regulatory limit for IBE be (6):

$$\theta_I = \{[\log(1.25)]^2 + \epsilon_I\}/\sigma^2_{W0} = 2.495 \quad (4)$$

IBE was evaluated only for 4-period, replicate-design, crossover studies.

The confidence limits were calculated by the approach of Hyslop *et al.* (13). Essentially, the regulatory model of Eq. (3) is rearranged by considering $\Theta_I.\sigma^2_{WR}$, taking it to the left side, and evaluating the distribution of each quadratic term within the square brackets. Bioequivalence is declared if the upper 95% confidence limit for the rearranged expression is negative or zero. The rearrangement of Eq. (3) is shown in the Appendix.

Scaled Average Bioequivalence

The regulatory model could be considered as:

$$(\mu_T - \mu_R)^2/\sigma^2_{SC} \leq \theta_A \quad (5)$$

σ^2_{SC} is a variance used as a scaling factor. Consequently, the squared BE limit is $BEL^2 = \Theta_A\,\sigma^2_{SC}$ (in terms of the squared difference between the means ($(\mu_T-\mu_R)^2$).

A BE criterion of $\Theta_A = 1.0$ was used, the value applied by Boddy *et al.* (3) in their modification of the regulatory model (see below). The scaling factor is the residual variance ($\sigma^2_{SC} = \sigma^2_{Res}$) in 2-period studies and the within-subject variance for the reference formulation ($\sigma^2_{SC} = \sigma^2_{WR}$) in 4-period investigations. Consequently, the regulatory model is:

$$(\mu_T - \mu_R)^2/\sigma^2_{Res} \leq \theta_A \quad (6)$$

and

$$(\mu_T - \mu_R)^2/\sigma^2_{WR} \leq \theta_A \quad (7)$$

in 2- and 4-period studies, respectively. The ANOVAs described for unscaled ABE were applied. The upper 95% confidence limit was calculated by modifying the method of Hyslop *et al.* (13); the procedure is described in the Appendix.

Expanding Bioequivalence Limits

The regulatory model, presented by Boddy *et al.* (3), rearranges the one for scaled ABE:

$$\mu_T - \mu_R \leq \theta_A{}^{1/2}\,\sigma_{SC} \quad (8)$$

Consequently, properties of the approach should correspond to those of scaled ABE. Notably, with $\Theta_A = 1.0$, the BE limit is BEL = σ_{SC}.

Simulations

A program was written in Fortran 90 (Compaq Visual Fortran, Professional Edition, version 6.1) using the appropriate statistical subroutines from the IMS library (Visual Numerics, Houston, TX). The computations were conducted on a PC with a Pentium II processor. 4-period, 2-sequence studies were simulated. Results of the first two periods recorded for both the test and reference formulation were used for assessing 2-period investigations. 24 subjects were assumed to participate in each simulated study, equally allocated to the two sequences. Consequently, a total of 96 lognormally distributed kinetic parameters (e.g., AUCs) were obtained in each investigation.

The overall average kinetic parameter for the reference formulation was arbitrarily 100 units. A true within-subject coefficient of $CV^\circ = 40\%$ was assumed for both formulations. This corresponded to a standard deviation of 0.385 in the natural logarithmic scale. In one study, however, the variation of the test product was changed in relation to the reference formulation. The between-subject variability was maintained at $CV^\circ = 40\%$. The square-rooted variance component for the subject-by-formulation interaction was either zero or had $CV^\circ_D = 30\%$. Five thousand simulations were performed under each condition.

RESULTS

Comparison of the Effectivenesses of Methods Evaluating the Bioequivalence of HVD/P

Figure 1 presents curves depicting the fraction of simulated studies in which BE was accepted ("statistical power curves") by various procedures at increasing assumed, true ratios of the two geometric means (GMR^o), i.e., at various deviations between the two logarithmic means. Average bioequivalence (ABE) as well as scaled ABE was evaluated from 2- and 4-period simulated crossover studies. The same simulated investigations permitted the determination of BE also by the approach of expanding BE limits (3). Finally, for comparison, IBE was also determined in the 4-period studies.

The results shown in Fig. 1 assumed that crossover studies with 24 subjects were performed. Both the reference and test formulations were assumed to have within-subject coefficients of variation of $CV^o_{WR} = CV^o_{WT} = 40\%$. Subject-by-formulation interaction was considered to be absent ($CV^o_D = 0.0$).

ABE determined with scaling and with expanding BEL showed very similar characteristics in both 2- and 4-period studies. The powers of investigations performed over two periods were somewhat lower than those conducted during four periods. As expected, unscaled ABE exhibited, particularly with two periods, very low statistical power, i.e., the proportion of accepted studies was small. In comparison with assessments by scaled ABE, scaled IBE had lower proportion of acceptance under the condition of true BE and showed a much more gently declining power curve. Consequently, scaled ABE and expanding BEL reached clearer decision than scaled IBE about the determination of BE for GMR.

Table I compares the proportion of accepting BE by scaled ABE when twice as many subjects were assumed to participate in 2- than in 4-period studies, i.e., when the number of measurements was the same. The proportions of accepted studies were very similar under the two conditions and were only slightly higher in the 2-period investigations.

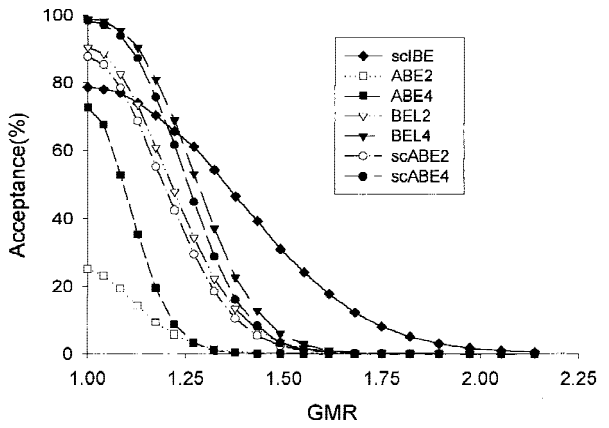

Fig. 1. Power curves for methods assessing the bioequivalence of highly-variable drugs and drug products. The dependence of the percentage of accepted BE studies is shown at various ratios of the geometric means (GMR) of the two formulations. It was assumed that the true within-subject coefficients of variation for both products were 40% in the simulations, and that the subject-formulation interaction was absent ($CV^o_D = 0$). scIBE: reference-scaled individual bioequivalence; ABE: average bioequivalence; BEL: expanding bioequivalence limits; 2 or 4 following the designation of the method refers to 2- or 4-period studies, respectively; scABE indicates scaled ABE whereas ABE denotes unscaled analysis.

Table I. Acceptance (in %) of Bioequivalence by Scaled ABE in 2- and 4-Period Studies Performed with the Same Number of Measurements[a]

GMR^o	2 periods N = 24	4 periods N = 12
1.00	88.46	86.94
1.05	84.93	83.87
1.10	76.62	73.72
1.15	64.24	62.32
1.20	50.59	48.95
1.25	36.69	35.21
1.30	24.81	24.28

[a] GMR^o: Geometric mean ratio; N: number of subjects

Effect of the Subject-by-Formulation Interaction

Figure 2 presents the proportions of accepted investigations obtained with the same methods of BE determination as shown in Fig. 1 and under the same investigational and statistical conditions except that substantial subject-by-formulation interaction was now assumed ($CV^o_D = 30\%$). The interaction had no effect on assessments by scaled ABE, and also by expanding BEL, in 2-period studies. The acceptance of scaled ABE and of expanding BEL was slightly reduced in the presence of subject-by-formulation interaction. By contrast, the acceptance of BE, and therefore the statistical power, declined very substantially when it was evaluated by unscaled ABE or scaled IBE.

Effect of the Ratio of Within-Subject Variations

Figure 3 shows the proportion of accepting BE (in %) by the various methods as the comparative variability of the two formulations (i.e., the ratios of σ_{WT}/σ_{WR} and CV^o_{WT}/CV^o_{WR}) is changed. The assumed conditions were similar to those given for Fig. 1 except that the two means were considered to be identical and the variability of the test formulation (σ^o_{WT}

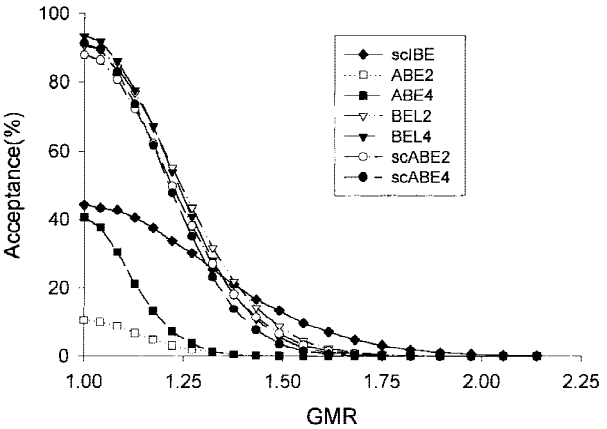

Fig. 2. Power curves for the determination of bioequivalence of highly-variable drugs and drug products. The methods and conditions are identical to those described for Fig. 1 except that large subject-by-formulation is assumed ($CV^o_D = 30\%$).

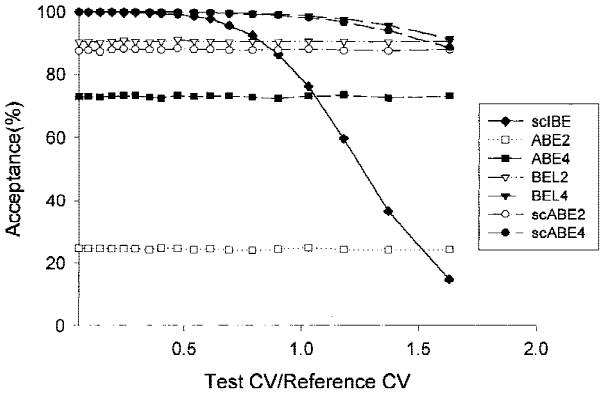

Fig. 3. Percentage of accepted bioequivalence studies at various ratios of the within-subject coefficients of variation of the two formulations. The assumed conditions are identical to those described for Fig. 1 except that the two means are identical and the ratio of variances is gradually raised. The total within-subject variance ($\sigma^{o2}_{WT} + \sigma^{o2}_{WR}$) is maintained at a constant value.

as well as CV^{o}_{WT}) was allowed to vary. However, the total within-subject variance ($\sigma^{o2}_{WT} + \sigma^{o2}_{WR}$) was maintained at a constant level which corresponded to $2CV^{o2} = 2.(40\%)^2$.

The acceptance rate of BE was independent of the ratio of variabilities when scaled ABE and the procedure of expanding BEL were applied in two-period studies, while in 4-period investigations the rate declined slightly at high ratios of CV^{o}_{WT}/CV^{o}_{WR}. The acceptance rate by unscaled ABE was also independent of the ratio of variabilities. By contrast, the acceptance rate of determining BE by scaled IBE was strongly affected by changing ratios of the variabilities.

To substantiate the results presented in Figs. 1–3, additional simulations were performed. The further conditions included [a] the assumption of 36 instead of 24 subjects for all 3 figures; [b] considering again 36 subjects for Figs. 1 and 2 and also variations within and between subjects of 50% instead of 40%; [c] assuming subject-by-formulation interactions for Fig. 2 also with $CV^{o}_{D} = 20$ and 40%; and [d] using for Fig. 3 a true ratio of geometric means $GMR^{o} = 1.1$ instead of 1.0. The results of the additional simulations closely paralleled those shown in Figs. 1–3.

DISCUSSION

Procedures evaluating the bioequivalence of highly variable drugs and drug products (HVD/P) are compared in the present study. The problem has been difficult and frustrating over many years and has often called for the use of unreasonably large numbers of subjects. The determination of average BE, i.e., the comparison of the means of the two products, is emphasized in the present investigation.

The usual approach of assessing unscaled BE, with 90% confidence limits for GMR between 0.80 and 1.25, was included for comparison. However, the present study is concerned mainly with effectiveness of scaled ABE and of the related procedure of expanding BE limits. In 2-period studies, scaling can be performed by means of the residual mean square of the relevant analysis of variance. In 4-period investigations, the within-subject variance of the reference formulation can be used for scaling.

The following conclusions can be drawn from the results:

1. The approaches of scaled ABE and expanding BEL yield very similar results. The similarity of the results yielded by the two methods is not surprising since the model for one of the procedures is readily converted to that of the other (Eqs. 5 and 8). Still, the method of the expanding BEL could be somewhat preferred because the relevant confidence limits can be calculated easily by applying the usual t-statistics. The approach of Hyslop *et al.* (13) for scaled metrics accomplishes the same goal but only indirectly. The BE limit for scaled ABE and expanding BEL is proportional to the square-rooted scaling factor.

2. The determination of ABE, with the same number of subjects, over four study periods has higher statistical power than an investigation performed during two periods. The difference between the statistical powers is reasonable since 4-period studies, with the same number of subjects, have twice as many observations as 2-period investigations. It is possible to consider also the condition when 2- and 4-period studies with the same number of measurements are performed. In this case, there are twice as many subjects in the 2- than in the 4-period investigation. Under this condition, the statistical powers are almost identical. The 2-period study has a very small edge which is due to the larger number of degrees of freedom applied in the computations. 4-Period investigations have further merits and advantages. In particular, they contain additional, useful information. They permit the comparison of the within-subject variances of the two formulations and the evaluation of the subject-by-formulation interaction. These terms are required for the determination of IBE but they provide interesting information even when ABE is evaluated. Information on these variances associated with the test and reference formulations allows assessment of the pharmaceutical quality of a new test product and its comparison with the pharmaceutical quality of the marketed brand product.

3. Scaled IBE is less sensitive than scaled ABE to deviations between the means. ABE, however, is not affected very much by differences in the test and reference variances or the subject-by-formulation interaction. The conclusion parallels that of Midha *et al.* (14). The reason is that an ABE test has a single goal, the determination of the average BE. In contrast, IBE assesses three goals simultaneously and can not devote attention to any of them with maximum effectiveness. Moreover, one of the properties of the IBE regulatory model is that it is much more sensitive to changes in estimated within-subject variations than to changes in the estimated deviation between the two means (15).

4. Large differences between the means can be accepted by scaled ABE and especially scaled IBE with substantial probabilities (Figs. 1 and 2). The observation is readily understandable. When the within-subject variability is large, the variations of the other estimated statistics are also high, including those of the two means and their difference. However, the customary framework of unscaled ABE allows only small deviations between the means, perhaps up to about 15%. This was Benet's motivation when he suggested, in discussions of IBE, that a secondary regulatory requirement be established which would constrain the GMR (16). FDA has recently adopted this approach for the evaluation of IBE (6,17). The issue appears to be relevant also to the determination of BE for HVD/P. It is explored in an investigation being undertaken. It is anticipated that an additional con-

straint will lower the probability of accepting BE. However, the effect of the constraint is expected to depend on the underlying conditions.

5. The proportion of accepted BE studies by the approaches of scaled ABE and expanding BEL in two-period studies, and by unscaled ABE, does not depend on the ratio of within-subject variations. Scaled and unscaled ABE and expanding BEL rely, in two-period investigations, on the estimation of total intraindividual variation. This is true also about unscaled ABE in 4-period studies. However, as the ratio of variances is raised, the within-subject variation for the reference formulation (σ^{o2}_{WR} and CV^{o}_{WR}) decreases. Consequently, the denominator of the regulatory model declines for reference-scaled ABE in 4-period studies [Eq. (7)], and it becomes slightly more difficult to satisfy the criterion at high ratios of the variances.

Similar considerations apply also to reference-scaled IBE. In addition, the third term in the numerator of the IBE model (Eq. 3) compares the two within-subject variances. As the ratio of these variances increases, the acceptance of IBE (as a result of the so-called mean-variability tradeoff (15,18)) moves from a favorable towards an unfavorable condition. Thus, the assessment of BE, by scaled ABE and expanding BEL, showed low sensitivity to the ratio of within-subject variances. This is a remarkable characteristic for determining the BE of HVDs which are generally safe drugs with shallow dose response curves. The replicate-design studies assumed four periods in the present investigation. It is, however, expected that the results and conclusions can be qualitatively applied also to 3-period studies.

CONCLUSIONS

Among methods evaluating the bioequivalence of highly-variable drugs and drug products, scaled average bioequivalence and the related procedure of expanding bioequivalence limits were found to be sensitive to differences between means and, consequently, highly effective for assessing the equivalence of average kinetic responses. However, these methods are insensitive to deviations between within-subject variances and to the subject-by-formulation interaction. The inefficiency of unscaled average bioequivalence, requiring large numbers of subjects in BE studies, is demonstrated. Scaled ABE can be effectively determined in both 2- and 4-period investigations. However, 4-period studies can yield additional, useful information.

ACKNOWLEDGMENTS

The research of L. Tothfalusi was supported, in part, by a grant (No. ETT-164/99) from the Hungarian Ministry of Health.

APPENDIX

Linearization for Reference-Scaled Individual Bioequivalence

FDA (6) has suggested a reference-scaled regulatory criterion for the determination individual bioequivalence [Eq. (3)]:

$$[(\mu_T - \mu_R)^2 + (\sigma^2_{WT} - \sigma^2_{WR}) + \sigma^2_D]/\sigma^2_{WR} \le \theta_I \qquad (9)$$

The criterion is useful for evaluating the IBE of highly-variable drugs and drug products. Hyslop *et al.*(13) suggested, and FDA (6) has applied, the linearization of the criterion:

$$(\mu_T - \mu_R)^2 + (\sigma^2_{WT} - \sigma^2_{WR}) + \sigma^2_D - \theta_I\sigma^2_{WR} \le 0 \qquad (10)$$

Thus, reference-scaled IBE is rejected if the upper 95% confidence limit of this expression is positive.

Confidence Limits for Scaled Average Bioequivalence

Hyslop *et al.* (13) described a procedure for evaluating the upper confidence limits in studies of individual BE. The method has been adapted for calculating the upper confidence limit in investigations of scaled average BE.

4-Period Crossover Studies

The BE criterion is (Eq. 7):

$$(\mu_T - \mu_R)^2/\sigma^2_{WR} \le \theta_A \qquad (11)$$

The expression can be rearranged to:

$$(\mu_T - \mu_R)^2 - \theta_A\,\sigma^2_{WR} \le 0 \qquad (12)$$

The two independent terms can be estimated by their respective expected values:

$$E_m = (m_T - m_R)^2 \qquad (13)$$

and

$$E_w = \theta_A\, s^2_{WR}$$

where m_T and m_R are the observed overall means of the test and reference formulations, respectively, and s^2_{WR} the estimated within-subject variance for the reference product.

The confidence limits for the two terms in the rearranged BE criterion are:

$$\begin{aligned} C_m &= [\mathrm{Abs}(m_T - m_R) + t_{1-\alpha,\,N-2}\,SE]^2 \\ C_w &= \theta_A\,(N-2)\,s^2_{WR}/\chi^2_{\alpha,\,N-2} \end{aligned} \qquad (14)$$

Here, as already noted, $SE = (s^2_{Int}/N)^{1/2}$. t and χ^2 are the respective tabulated statistics at the indicated significance levels (with $\alpha = 0.05$) and N-2 degrees of freedom.

It is worth noting that C_m and C_w, and therefore the confidence limits, depend on generally different estimated variances. The relevant error term for the comparison of the means is the estimated variance of the subject-by-formulation interaction, s^2_{Int} (which is not identical to the much-discussed variance component, σ^2_D). In contrast, the estimated within-subject variance, s^2_{WR}, is required for the calculation of the corresponding confidence limit, C_w.

The squared lengths of the confidence intervals from their respective means are:

$$\begin{aligned} L_m &= (C_m - E_m)^2 \\ L_w &= (C_w - E_w)^2 \end{aligned} \qquad (15)$$

Finally, the confidence limit (CL) for the rearranged BE criterion is:

$$CL = E_m - E_w + (L_m + L_w)^{1/2} \qquad (16)$$

Scaled average BE can be accepted if the calculated upper 95% confidence limit is negative (or, better, not positive).

2-Period Crossover Studies

The BE criterion is (Eq. 6):

$$(\mu_T - \mu_R)^2/\sigma^2_{Res} \leq \theta_A \quad (17)$$

A rearranged expression is:

$$(\mu_T - \mu_R)^2 - \theta_A \sigma^2_{Res} \leq 0 \quad (18)$$

The procedure for the calculation of the upper confidence limit is identical to that described for the 4-period studies except that the estimated residual variance, s^2_{Res}, replaces both (s^2_{Int} and s^2_{WR}) and the standard error is, as already discussed, $SE = (2s^2_{Res}/N)^{½}$.

REFERENCES

1. H. H. Blume, M. Elze, H. Potthast, and B. S. Schug. Practical strategies and design advantages in highly variable drug studies: multiple dose and replicate administration design. In H. H. Blume and K. K. Midha (eds.), *Bio-International 2: Bioavailability, Bioequivalence, and Pharmacokinetic Studies*, Medpharm, Stuttgart, 1995 pp. 117–122.
2. 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).
3. 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).
4. 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.
5. R. Schall and R. L. Williams. Towards a practical strategy for assessing individual bioequivalence. *J. Pharmacokinet. Biopharm.* **24**:133–149 (1996).
6. Food and Drug Administration. *Statistical Approaches to Establishing Bioequivalence - Guidance for Industry*, Center for Drug Evaluation and Research (CDER), Rockville, MD, 2001.
7. L. Endrenyi. A method for the evaluation of individual bioequivalence. *Int. J. Clin. Pharmacol. Ther. Toxicol.* **32**:497–508 (1994).
8. D. J. Schuirmann. Average BE criterion with replicate designs. *AAPS International Workshop on Individual Bioequivalence,* distributed handout, Montreal, Que. (1999).
9. E. K. Kimanani and D.Potvin. A parametric confidence interval for a moment-based scaled criterion for individual bioequivalence. *J. Pharmacokinet. Biopharm.* **25**:595–614 (1997).
10. J.-P. Liu. Use of the repeated cross-over designs in assessing bioequivalence. *Stat. Med.* **29**:1067–1078 (1995).
11. S.-C. Chow and J.-P. Liu. *Design and Analysis of Bioavailability and Bioequivalence Studies*, 2nd ed., Marcel Dekker, New York-Basel, 2000 pp. 277.
12. R. Schall and H. E. Luus. On population and individual bioequivalence. *Stat. Med.* **12**:1109–1124 (1993).
13. T. Hyslop, F. Hsuan, and D. J. Holder. A small sample confidence interval approach to assess individual bioequivalence. *Stat. Med*. **19**:2885–2897 (2000).
14. K. K. Midha, M. J. Rawson, and J. W. Hubbard. Individual and average bioequivalence of highly variable drugs and drug products. *J. Pharm. Sci.* **86**:1193–1197 (1997).
15. L. Endrenyi and Y. Hao. Asymmetry of the mean-variability tradeoff raises questions about the model in investigations of individual bioequivalence. *Int. J. Clin. Pharmacol. Ther.* **36**:450–457 (1998).
16. L. Benet. Individual bioequivalence: An overview. *AAPS International Workshop on Individual Bioequivalence,* distributed handout, Montreal, Que. (1999).
17. Food and Drug Administration. *Bioavailability and Bioequivalence Studies for Orally Administered Drug Products - General Considerations*, Center for Drug Evaluation and Research (CDER), Rockville, MD, 2000.
18. W. W. Hauck, M.-L. Chen, T. Hyslop, R. Patnaik, D. Schuirmann, and R. Williams. Mean difference vs. variability reduction: Trade-offs in aggregate measures for individual bioequivalence. *Int. J. Clin. Pharmacol. Biopharm.* **34**:535–541 (1996).