An Approach for Widening the Bioequivalence Acceptance Limits in the Case of Highly Variable Drugs

Alexander W. Boddy,[1,3] Fred C. Snikeris,[1] Robert O. Kringle,[1] Greg C. G. Wei,[1] James A. Oppermann,[1] and K. K. Midha[2]

Received July 5, 1995; accepted September 12, 1995

Purpose. Highly variable drugs pose a problem in bioequivalence assessment because they often fail to meet current regulatory acceptance criteria for average bioequivalence (80–125%). This paper examines alternative approaches to establishing bioequivalence.

Methods. Suggested solutions have included alternate study designs, e.g., replicate and multiple dose studies, reducing the level of the confidence interval, and widening the acceptance limits. We focus on the latter approach.

Results. A rationale is presented for defining wider acceptance limits for highly variable drugs. Two previously described methods are evaluated, and a new method having more desirable properties is proposed.

Conclusions. We challenge the "one size fits all" current definition of bioequivalence acceptance limits for highly variable drugs, proposing alternative limits or "goal posts" which vary in accordance with the intrasubject variability of the reference product.

KEY WORDS: bioequivalence; highly variable drug; intrasubject variability; acceptance limits.

INTRODUCTION

Bioequivalence (BE) studies are often conducted using a two-period crossover design. Average bioequivalence of the two formulations is concluded if the 90% confidence interval for the mean relative bioavailability falls within the prespecified limits, usually 80–125%, based upon analysis of the log-transformed AUC and Cmax data (1). The sample size for the BE study is typically based on power considerations, using available information on the AUC intrasubject variability, to ensure the sponsor a reasonable chance of demonstrating bioequivalence if it indeed exists. Highly variable drugs, those which exhibit intrasubject variability of pharmacokinetic data in excess of 25–30% CV (2,3), have a poor chance of satisfying these acceptance criteria in the typical two-period design with a moderate number of subjects. For example, two formulations of a drug with a 30% intrasubject CV would require a 52 subject study to have high chance (90%) of meeting the acceptance criteria, assuming a true relative bioavailability of 1.05.

Table I shows the required number of subjects for the conventional two-period design for intrasubject CV's of 25%–50%. Sample sizes were computed from equations given by Hauschke et al. (4), using the exact expression for the standard deviation of the log transformed data ($\sigma = [\log(CV^2 + 1)]^{1/2}$) rather than the approximation, $\sigma = CV$, which is less accurate for higher variabilities. It is clear from this table that even if the formulations are truly identical, such sample sizes would be difficult to achieve in practice. If the formulations truly differ by 10–15% (still bioequivalent by definition), the required number of subjects becomes unacceptable both practically and ethically with respect to the number of healthy volunteers who would have to be exposed to the drug. This is particularly important because highly variable drugs are often the most difficult to reformulate, and therefore differences of 10–15% in bioavailability are more likely to occur.

There have been several scientific symposia where the issue of bioequivalence of highly variable drugs has been raised. At the *Bio-international* series of symposia in 1989, 1992, and 1994, alternative study designs were proposed to overcome this difficulty using a reasonable number of volunteers. Multiple dose steady-state studies have been documented to reduce the intrasubject variability in compounds such as propafenone, nifedipine, loratadine, and verapamil (5). While multiple dose studies offer a pharmacokinetic rationale (although not completely understood) for the reduction in intrasubject variability, replicate design single dose studies (higher order crossover designs) can achieve a reduction in the number of subjects required through statistical means (replication).

While multiple-dose studies and replicate designs can be beneficial in some cases, these study designs reduce the total number of subjects exposed to the study by increasing the length of exposure of the volunteers. Thus, the overall extent of exposure is not reduced, which may still present a concern with regard to healthy volunteers. Use of patients in the target population would eliminate this ethical concern, but raises other problems, including the potential for greater variability in patients, the ability to draw a large number of blood samples, and in less common disease states the ability to recruit the required number of patients.

Two alternative methods were discussed at both *Bio-international '92* and *Bio-international '94* and offered for further study. These are

- reducing the level of the confidence interval, and
- widening the bioequivalence acceptance limits.

The first method, reducing the level of confidence, is the same as increasing the Type I error (consumer risk). Any method which substantially effects the consumer risk should be avoided. The second approach, widening the acceptance limits, is appealing in that it challenges the current *definition* of average bioequivalence for highly variable drugs. This approach of widening the "goal posts" while maintaining the current 90% confidence level was put forth as a major recommendation to FDA at the March, 1995 AAPS/FDA Workshop on Evaluation of Orally Administered Highly Variable Drugs and Drug Formulations.

In this paper a justification for widening the "goal posts" for highly variable drugs is put forth, and three methods for expanding the limits based on an estimate of the intrasubject CV are discussed together with their advantages and disadvantages. These are the "simple confidence inter-

[1] Biostatistics Dept., Sanofi Research Division, 9 Great Valley Parkway, P.O. Box 3026, Malvern, Pennsylvania 19355.

[2] College of Pharmacy and Nutrition, and College of Medicine, University of Saskatchewan, 110 Science Place, Saskatoon SK S7N 5C9, Canada.

[3] To whom correspondence should be addressed.

0724-8741/95/1200-1865$07.50/0

Table I. Sample Size to Attain a Power of 90% for Demonstrating Equivalence for the Multiplicative Model[a]

		True relative bioavailability			
CV (%)	SD (σ)	1.00	1.05	1.10	1.15
25	0.246	30	36	66	152
30	0.294	40	52	92	214
35	0.340	54	68	124	288
40	0.385	68	86	158	368
45	0.429	84	106	196	456
50	0.472	100	128	236	552

[a] The intrasubject CV and SD are for the original and log transformed data, respectively, where $\sigma = [\log(CV^2 + 1)]^{1/2}$.

val method", the "fixed sample size method", and the recommended "fixed multiple-of-CV method".

METHODS

Motivation for Expanding the Acceptance Limits

An individual receiving the reference formulation of a drug will experience varying AUC over time depending on the intrasubject CV. An interval containing 95% of the values experienced by this individual would be approximately ±2 standard deviations around the mean. For example, for a drug with a CV of 10%, intra-individual AUC's would vary from 82% to 122% of the mean. However, for a highly variable drug with a CV of 40%, intra-individual AUC's would vary from 45% to 223% of the mean. As shown in Figure 1, a 25% difference in formulation means represents a substantial shift of the distribution for the low variability drug, but for the high variability drug, the two distributions still overlap to a great extent. It can be argued, therefore, that any mean difference between two formulations which is small relative to this range of intra-individual values may not be of importance and that the acceptance limits should be scaled accordingly. The choice of a "relatively small" difference is of course subjective, but there is a logical basis to argue that the size of this difference should depend on the unique characteristics of the drug, one of the most important being the inherent intrasubject and dosage form variability.

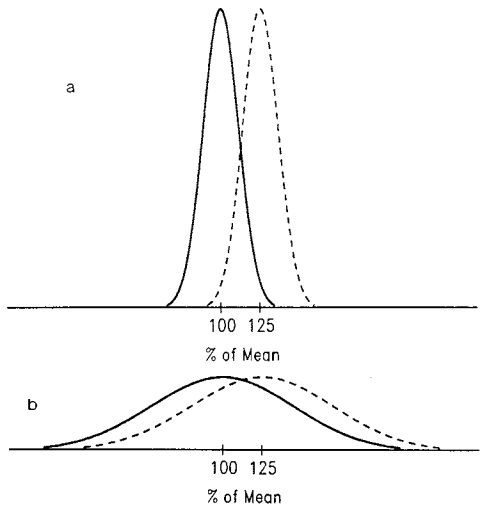

Fig. 1. A 25% shift in the mean in relation to the population of intrasubject AUCs for intrasubject variabilities of (a) 10% CV and (b) 40% CV.

Three methods for widening the acceptance limits are discussed below. These wider limits imply that for highly variable drugs, larger mean formulation differences are allowed before being considered clinically important. The first two methods (simple confidence interval and fixed sample size) define extended acceptance limits based on "a reasonable number of subjects", whereas the third method (fixed multiple of CV) directly addresses the definition of a clinically important difference in formulation means for highly variable drugs.

Simple Confidence Interval Method

This method has received attention at both *Bio-international '94* and the March, 1995 AAPS/FDA Workshop on Highly Variable Drugs. Assume the reference product is compared to itself in a two-period crossover with N subjects, producing a 90% confidence interval (L_1,U_1), that does not satisfy the 80–125% acceptance range. One possible method for extending the range is to simply use this observed confidence interval, (L_1,U_1), as the acceptance range for the subsequent test versus reference confidence interval, (L,U), based on a two period crossover with the same number of subjects, N. A variation of this method is to center the observed reference versus reference confidence interval (about zero in the log units), producing, for example, (L_2,U_2) as the new acceptance range. Neither of these methods are satisfactory from the producer viewpoint, since even when the test and reference products have identical bioavailability, there is no reason to expect that the test versus reference confidence will be narrower or wholly contained in the reference versus reference confidence interval (the second experiment is just a repeat of the first); such studies will be consistently underpowered. A simulation study was carried out which showed the power (probability of claiming equivalence) of this decision rule (using the uncentered interval) to be 17% (N = 24 for both trials and no difference in true bioavailability). Therefore, it would be unwise to use this method in practice.

Fixed Sample Size Method

The basic idea for this approach is to improve the probability of concluding bioequivalence, when it exists, while fixing the sample size at some reasonable value. An estimate of the intrasubject variability is obtained from the reference vs reference experiment(s), and based on this variability, extended acceptance limits are computed to achieve a reasonably small producer risk (Type II error) for the subsequent test versus reference study using a reasonable number of subjects, e.g., 24. A reasonably small producer risk might be 10% when there is no difference between the test and reference true bioavailabilities.

The acceptance limits [A,B] for the conventional test vs reference two period crossover can be obtained by reversing

the equation for sample size given by Hauschke et al. (4), producing

$$[A,B] = \exp[\pm(t_\alpha + t_{\beta/2})n^{-1/2}\,\hat{\sigma}] \quad (1)$$

where α and β are the consumer and producer risks, respectively, 2n is the total number of subjects desired in the subsequent test vs reference study, t is the percentile of the t-distribution with 2n-2 degrees of freedom, and $\hat{\sigma}$ is the estimated intrasubject SD from ANOVA on the log transformed reference vs reference data. Using the conventional $\alpha = 0.05$ (5% consumer risk) and $\beta = 0.10$ (10% producer risk or 90% power), and a typical sample size of $2n = 24$ subjects, this reduces to

$$[A,B] = \exp[\pm 0.997\,\hat{\sigma}] \quad (2)$$

Similar expressions can be found by varying the producer risk, the sample size, or both. Advantages of the fixed sample size approach are that the acceptance limits get wider as the intrasubject variability increases and that a "reasonable" number of subjects can be used to achieve adequate power. However, a major drawback of this approach is that the wider acceptance limits are based on controlling the sample size, rather than controlling some meaningful measure of the formulation difference. Also, since the sample size is fixed, when the true relative bioavailability differs from 100% the producer risk cannot be improved by increasing the sample size.

Fixed Multiple-of-CV Method

This proposed approach addresses both drawbacks. The central idea is to base the new acceptance limits on certain characteristics of the distribution of intra-individual bioavailabilities for the reference formulation. When intrasubject variability is high, a mean difference of 0–25% between formulations is rather small relative to the range of values an individual will experience, as shown in Figure 1b. Recall, for a highly variable drug with a CV of 40%, AUC's would vary from 45% to 223% of the mean (2 sigma limits). Accordingly, the acceptance limits should be scaled in relation to the size of the intrasubject variability. The acceptance limits [A,B] for average bioequivalence defined by this approach can be expressed as

$$[A,B] = \exp[\pm k\sigma] \quad (3)$$

where k is some multiplying factor of the intrasubject standard deviation, σ, on the log scale. Since the intrasubject standard deviation on the log-scale is related to the CV on the original scale, the limits above can also be expressed as

$$[A,B] = \exp[\pm k(\ln(CV^2 + 1))^{1/2}] \quad (4)$$

A simple choice for the multiplying factor k is one. With $k = 1$ the maximum allowable difference between the means is one standard deviation. In probability terms k represents z_p, the pth percentile of the standard normal distribution; for $k = 1$, two-thirds ($p = 67\%$) of the AUC's experienced by an individual on the reference formulation are within this range. Table II shows the acceptance limits provided by this approach for the case $k = 1$. These limits are also graphically illustrated in Figure 2.

The choice of $k = 1$ is appealing for other reasons, as well. Firstly, the acceptance limits at a CV of 25% are close to the conventional limits of 0.80 to 1.25, which are commonly applied to low to moderate variability drugs (CV < 25%). Secondly, however large the CV, the number of subjects required in the conventional two period crossover to demonstrate equivalence with 90% power is approximately 24, assuming the true relative bioavailability is one. This latter result is due to the mathematical similarity of equation (3) with equation (2) in the previously described "fixed sample size approach". Thirdly, since the limits are not related to a fixed sample size or study design, the power can be increased with higher sample sizes or with replicate designs.

Table II. Acceptance Limits, $e^{\pm k\sigma}$, as a Function of the Intrasubject Standard Deviation, σ, on the Log Scale (Approximately Equal to the CV); $k = 1$

CV (%)	SD (σ)	Lower Limit	Upper Limit
25	0.246	0.78	1.28
30	0.294	0.75	1.34
35	0.340	0.71	1.40
40	0.385	0.68	1.47
45	0.429	0.65	1.54
50	0.472	0.62	1.60

DISCUSSION

In this paper we have presented a rationale for widening the acceptance limits for average bioequivalence of highly variable drugs, and proposed a new method for doing so, with a pharmacokinetically meaningful and statistically sound basis. Given an estimate of the reference intrasubject variability (either from a previous study or a test vs reference comparative study which included replication of the

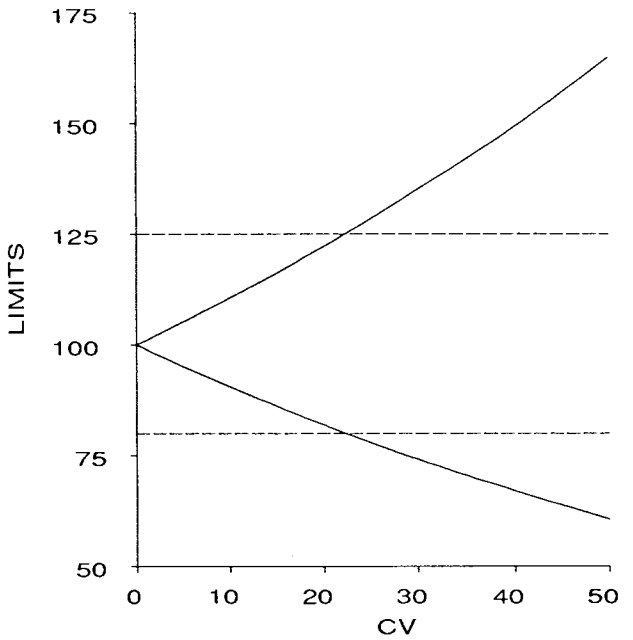

Fig. 2. Acceptance limits, $e^{\pm k\sigma}$, as a function of the intrasubject standard deviation, σ, on the log-scale (approximately equal to the CV); $k = 1$.

reference formulation), this method defines a family of acceptance limits based upon the choice of a scaling parameter, k. Although we have presented results for a value of k = 1, other choices of k could also be justified. For instance, k could be selected so that at a particular value of the CV, e.g., at 25 or 30%, the limits correspond exactly to the conventional 80% to 125% limits, thus providing a seamless transition between limits for low and high-variability drugs. In fact, it may be desirable to select different values of k for different drugs depending upon the width of the therapeutic window, a point which has been raised by regulatory agencies.

One drawback of the approaches presented in this paper is that they require an estimate of the true CV of the reference formulation. If this estimate is unreliable, particularly if it is an overestimate, then the acceptance limits will be too wide and bioequivalence may be concluded too easily. At least partial protection from this problem could be achieved by setting minimum standards for the precision of the CV estimate, and by ensuring that the CV estimated in the bioequivalence trial is consistent with the prior estimate used to set the acceptance limits. The effect on Type I error by either under or overestimation of the CV is currently under investigation.

Finally, although the focus of this paper has been on highly variable drugs, another potential use of the fixed CV approach is to derive narrower acceptance limits for drugs with low variability and narrow therapeutic windows.

REFERENCES

1. FDA Division of Bioequivalence. Guidance: Statistical Procedures for Bioequivalence Studies using a Standard Two-Treatment Crossover Design. July 1, 1992.
2. H. Blume and K. K. Midha. Report of Consensus Meeting: *Bio-international '92,* Conference on Bioavailability, Bioequivalence and Pharmacokinetic Studies, Bad Homburg, Germany, 20–22 May 1992. *Eur J Pharm Sci* 1:165–171 (1993).
3. *Bio-international '94,* 14–18 June, 1994, Munich Germany.
4. D. Hauschke, V. Steinijans, E. Diletti, and M. Burke. Sample Size Determination for Bioequivalence Assessment Using a Multiplicative model, *J Pharmacokin. Biopharm.* 20:557–561 (1992).
5. H. Blume and K. K. Midha. Practical Strategies and Design Advantages in Highly Variable Drug Studies: Multiple Dose and Replicate Administration Design, Abstract from *Bio-international '94,* 14–18 June, 1994, Munich Germany.