*Pharmaceutical Research, Vol. 25, No. 1, January 2008 (© 2007)*
DOI: 10.1007/s11095-007-9434-x

*Commentary*

# Bioequivalence Approaches for Highly Variable Drugs and Drug Products

**Sam H. Haidar,[1] Barbara Davit,[1] Mei-Ling Chen,[2] Dale Conner,[1] LaiMing Lee,[1] Qian H. Li,[3] Robert Lionberger,[1] Fairouz Makhlouf,[3] Devvrat Patel,[1] Donald J. Schuirmann,[3] and Lawrence X. Yu[1,4]**

*Received June 9, 2007; accepted August 8, 2007; published online September 22, 2007*

***Abstract.*** Over the past decade, concerns have been expressed increasingly regarding the difficulty for highly variable drugs and drug products (%CV greater than 30) to meet the standard bioequivalence (BE) criteria using a reasonable number of study subjects. The topic has been discussed on numerous occasions at national and international meetings. Despite the lack of a universally accepted solution for the issue, regulatory agencies generally agree that an adjustment of the traditional BE limits for these drugs or products may be warranted to alleviate the resource burden of studying relatively large numbers of subjects in bioequivalence trials. This report summarizes a careful examination of all the statistical methods available and extensive simulations for BE assessment of highly variable drugs/products. Herein, the authors present an approach of scaling an average BE criterion to the within-subject variability of the reference product in a crossover BE study, together with a point-estimate constraint imposed on the geometric mean ratio between the test and reference products. The use of a reference-scaling approach involves the determination of variability of the reference product, which requires replication of the reference treatment in each individual. A partial replicated-treatment design with this new data analysis methodology will thus provide a more efficient design for BE studies with highly variable drugs and drug products.

**KEY WORDS:** bioequivalence; highly variable drugs; highly variable drug products; scaled average bioequivalence; statistical approach; study design.

## INTRODUCTION

In the USA, bioequivalence (BE) is defined as the absence of a significant difference in the rate and extent to which the active ingredient or active moiety in pharmaceutical equivalents or pharmaceutical alternatives becomes available at the site of drug action when administered at the same molar dose under similar conditions in an appropriately designed study (1). BE studies of systemically absorbed drug products are generally conducted by determining pharmacokinetic endpoints to compare the *in vivo* rate and extent of drug absorption of a test and a reference drug product in healthy subjects. A test product is considered bioequivalent to a reference product if the 90% confidence intervals for the geometric mean test/reference ratios of the area under the drug's plasma concentration *versus* time curve (AUC) and peak plasma concentration ($C_{max}$) both fall within the predefined BE limits of 80–125% (2).

The width of the 90% confidence interval is proportional to the estimated drug variability (in particular, within-subject variability for a crossover design) and inversely proportional to the number of subjects participating in the study. The BE limits of 80–125% are currently applied to almost all drug products regardless of the size of within-subject variability. As a result, the number of subjects required for a study of highly variable drugs or drug products can be much greater than normally needed for a typical BE study. For example, to demonstrate BE with 90% power, it was estimated that 136 subjects would be required for a drug with 60% within-subject coefficient of variation even if the test and reference products were identical (3).

It is believed that drugs with high within-subject variability generally have a wide therapeutic window; in other words, despite high variability, these products have been demonstrated to be both safe and effective (4). Thus, applying the conventional BE criteria to highly variable drugs/products may unnecessarily expose a large number of healthy subjects to a drug when this large number of subjects is not needed for assurance of BE. For these reasons, scientists and statisticians at the US Food and Drug

The opinions expressed in this report by the author's do not necessarily reflect the views or policies of Food and Drug Administration (FDA).

[1] Office of Generic Drugs, Food and Drug Administration, 7500 Standish Place, Rockville, Maryland 20855, USA.
[2] Office of Pharmaceutical Science, Food and Drug Administration, 10903 New Hampshire Avenue, Silver Spring, Maryland 20993, USA.
[3] Office of Biostatistics, Food and Drug Administration, 10903 New Hampshire Avenue, Silver Spring, Maryland 20993, USA.
[4] To whom correspondence should be addressed. (e-mail: lawrence.yu@fda.hhs.gov)

Administration (FDA) investigated various approaches available for determining BE that would reduce the sample size required for a BE study, without allowing for therapeutically inequivalent products to reach the market. In this paper, we offer the results and preliminary conclusions of this work, and present results from a research project testing a new proposal.

## BACKGROUND

### Traditional Bioequivalence Method

For systemically available drug products, FDA generally asks applicants to conduct BE studies with pharmacokinetic endpoints using a single dose, crossover design in healthy subjects. Subjects receive a single dose of test and reference products on separate occasions with random assignment to the two possible sequences of product administration. Treatments are separated by a washout period of adequate duration such that the drug of interest can no longer be detected in plasma. The FDA generally asks applicants to conduct single dose studies rather than multiple dose studies because single dose studies are generally more sensitive to detecting potential differences between products (2). For a product with multiple strengths, the highest strength is used in the BE study, unless precluded for reasons of safety. The number of subjects in the study should be sufficient to ensure adequate statistical power; most studies enroll from 24 to 36 subjects.

The bioequivalence parameters AUC and $C_{max}$, are statistically analyzed using the two one-sided tests procedure to determine whether the average values for the measures estimated after administration of the test and reference products are comparable (5). This approach involves the calculation of a 90% confidence interval for the ratio of the averages of the measures for the test and reference products (6). The choice of the current 80 to 125% acceptance limits for BE has been based on expert medical judgment and FDA experience with thousands of drug products that a difference of less than 20% in drug exposure was not clinically significant for most drugs (7). The 80% limit indicates that the test product is no less than 80% of the reference, while the 125% limit indicates that the reference product is no less than 80% of the test product (a 4:5 reference to test ratio is a 5:4 test to reference ratio).

### Highly Variable Drugs and Drug Products

Concerns have been expressed at times regarding the difficulty of meeting the standard BE criteria for highly variable drugs and/or drug products (8, 9). For our discussion of BE, we consider drugs and drug products exhibiting within-subject variability of 30% (C.V., coefficient of variation) or greater in the pharmacokinetic measures AUC and/or $C_{max}$ to be highly variable. FDA's Office of Generic Drugs (OGD) estimates that approximately 10% of the submitted BE studies from Abbreviated New Drug Applications (ANDAs) showed some evidence of high variability (10). Examples exist where a highly variable reference product failed to demonstrate BE when compared to itself in a BE study using the standard design/sample size (10).

### International Approaches for Evaluation of Bioequivalence

Although global harmonization is a general goal, the International Conference on Harmonization (ICH) to date has not accepted BE as a topic. Nonetheless, the resource and ethical concerns for highly variable drugs/products in BE are generally recognized by international regulatory agencies. It is thus useful to review the differing regulatory approaches (listed below) before an informed recommendation is made on the topic.

Health Canada applies a bioequivalence limit of 80–125% on the 90% confidence interval of the AUC ratio between the test and reference product for drugs with uncomplicated characteristics, which is similar to the FDA's practice. For $C_{max}$, however, Health Canada only requests that the mean test/reference ratio, or "point estimate," fall between 80–125% (11,12). Health Canada allows investigators to add more subjects to a BE study if random variation or a larger than expected relative difference is observed. This option may be used only when it is stated in the study protocol, the same protocol is used for the additional subjects, and consistency between the groups is demonstrated at an alpha error rate of five percent.

The European Agency for the Evaluation of Medicinal Products (EMEA) has similar BE standards to those applied by the FDA, i.e., 90% confidence limits of on AUC and $C_{max}$ must fall within 80–125% (13). However, EMEA guidelines suggest expanding the limits for $C_{max}$ (e.g., 75–133%) in certain cases, provided that there are no safety or efficacy concerns (13).

Japan's National Institute of Health, Division of Drugs also applies the bioequivalence limits of 80–125% on the 90% confidence interval for both AUC and $C_{max}$ (14). However, wider limits are allowed for "less potent" drugs. Additionally, if the study confidence intervals are outside of 80–125% limits BE may still be claimed provided that the study meets all three conditions listed below (14):

1) The total number of subjects in the initial BE study is no less than 20 ($n$=10/group), or pooled sample size of the initial and add-on studies is no less than 30;
2) The differences in average values of logarithmic AUC and $C_{max}$ between two products are between log (0.9)–log (1.11); and
3) Dissolution rates of test and reference products are determined to be equivalent under all dissolution testing conditions specified.

Japan's National Institute of Health allows the addition of subjects to increase the power of a failed BE study. However, the add-on subjects cannot be less than half the number in the original study (14).

## NEW APPROACHES EVALUATED BY THE FDA

At the April 2004 meeting of the Advisory Committee for Pharmaceutical Science (ACPS), several approaches were presented for the BE evaluation of highly variable drugs and the Committee members were asked for their opinion about each proposed method (15). The approaches discussed are listed below.

A. Direct Expansion of Bioequivalence Limits
   Sample size in BE studies is determined, in large part, by the bioavailability parameter with the high-

est variability. In most cases, $C_{max}$ has higher variability than AUC. Thus, widening of the BE limits for $C_{max}$ has been proposed to reduce the sample size needed in the evaluation of BE for highly variable drugs/products. The greater variability observed with $C_{max}$ may result from the fact that this parameter is a single point measurement, which is highly dependent on the sampling time/frequency and gastrointestinal physiology. This is the EMEA approach.

B. Expansion of Bioequivalence Limits Based on Fixed Sample Size

This method was proposed based on the notion that only a reasonable number of subjects should be required for a BE study (16). The number of subjects is fixed by a standard two-period, crossover study comparing the reference product with itself where the study fails to meet the 80–125% limit. The confidence interval obtained from the reference product in this study would become the "goalposts" for the subsequent studies comparing the test with reference product, using the same number of subjects (8,16).

C. Widening of Bioequivalence Limits Based on Reference Variability

The bioequivalence limits for these methods are not determined by the sample size. Rather, they are scaled based on the within-subject variability of the reference product. Three different methods for widening of BE limits based on reference variability are described below. For Methods 2 and 3, an additional condition to constrain the mean difference between the test and reference products has also been proposed.

Method 1: The rationale for this approach is that a mean difference of 25% is considered small relative to the range of values an individual may experience when the within-subject variability is high, e.g., 40%. Therefore, the acceptable limits may be scaled in relation to the size of within-subject variability as follows (16):

$$[\mathrm{U},\mathrm{L}] = \mathrm{Exp}[\pm k\sigma_{WR}] \tag{1}$$

where U and L are the upper and lower limits, respectively; $k$ represents the $p$-th percentile of the standard normal distribution, Zp; and $\sigma_{WR}$ is the estimated within-subject standard deviation (estimated from the ANOVA on the log scale) for the reference product. When $k$=1, ~67% of the pharmacokinetic measures (such as AUC) experienced by an individual will be within the range of [U, L]. Different $k$ values could be chosen for different drugs depending on their therapeutic windows.

Method 2: This is a scaled average BE approach (17–20). Mathematically, it may be expressed as follows:

$$(\mu_T - \mu_R)^2 \Big/ \sigma^2_{WR} \le \theta \tag{2}$$

where $\mu_T$ and $\mu_R$ are the averages of the log-transformed measure for the test and reference products, respectively; and $\theta$ is the BE limit.

Method 3: This method is an individual BE approach derived from the comparison of the distance measure between the test and reference products. The following criterion has a reference variance in the denominator, and thus is scaled to the reference variability (6):

$$\left[(\mu_T - \mu_R)^2 + \left(\sigma^2_{WT} - \sigma^2_{WR}\right) + \sigma^2_D\right] \Big/ \sigma^2_{WR} \le \theta_I \tag{3}$$

where $\sigma_{WT}$ is the estimated within-subject standard deviation for the test product; $\sigma^2_D$ is the subject-by-formulation interaction variance component; and $\theta_I$ is individual bioequivalence limit.

D. Expansion of Bioequivalence Limits Based on Sample Size and Scaling

In addition to fixing the sample size, this method takes into consideration the producer's risk (Type II error) and reference variability (16). The equation for the allowable limits is:

$$[\mathrm{U},\mathrm{L}] = \mathrm{Exp}\left[\pm\left(t_\alpha + t_{\beta/2}\right)n^{-1/2}\sigma_{WR}\right] \tag{4}$$

where $\alpha$ and $\beta$ are the consumer and producer risks, respectively; $2n$ is the number of subjects desired in the study; and t is the percentile of the $t$-distribution with $2n-2$ degrees of freedom.

## SIMULATION RESEARCH

At the conclusion of the April 2004 ACPS meeting (15), the Committee favored approach C, particularly *Method 2*, for BE evaluation of highly variable drugs/products. The FDA was encouraged to further investigate the scaled average BE approach and consider the use of a point estimate constraint to reduce the possibility of approving a drug product based on a study with a large difference in the test and reference means.

Following the ACPS recommendation, the FDA Working Group on Highly Variable Drugs (FDA Working Group) initiated simulations that compared the power of a given study design when using scaled average BE with unscaled average BE. The design used in this simulation has a three-way crossover, partial replicate design, where the reference (R) product is given twice and the test (T) product is given once (e.g., R T R). The goal of the study was to evaluate the impact of reference-scaled average BE on the statistical power, or the percent of studies passing the BE criteria, under various conditions. Several factors were tested, including different sample sizes, different values of $\sigma_{W0}$ (a constant to be defined by regulatory agencies), the presence or absence of point estimate constraints, and different levels of within-subject variability. Preliminary results of this study were presented at the October 2006 ACPS meeting (21). A more complete description of the study and the results will be provided in the future (manuscript under preparation).

## FDA PROPOSAL FOR BE EVALUATION OF HIGHLY VARIABLE DRUGS

Based on the results of the above study, the FDA proposed a method for BE evaluation of highly variable drugs at the ACPS meeting in October 2006. The Committee endorsed this proposal, with specifics regarding sample size and regulatory constants to be determined by the Agency. Additional simulations have been conducted by the FDA since then, to refine the details of the proposed approach. The results of these simulations will be reported elsewhere. Outlined below is the general approach.

### Proposed Study Design

For drugs with an expected within-subject variability of 30% or greater, a BE study with three-period, reference-replicated, crossover design with sequences of TRR, RTR, and RRT is proposed. Specifically, subjects receive a single dose of the test product once and reference product twice on separate occasions with random assignment to the three possible sequences of product administration. This partial replicate design allows for the estimation of within-subject variability for the reference product. Treatments should be separated by a washout period of adequate duration such that the drug of interest can no longer be detected in plasma. Subjects recruited for *in vivo* BE studies should be 18 years of age or older, and capable of giving informed consent unless otherwise indicated by a specific guidance. It is the sponsor's responsibility to determine the sample size needed to achieve the desired power in a study; however, the minimum number of subjects that would be acceptable is 24.

The three-period design was selected over a four-period design because of efficiency. The only advantage of the four-period design is that it allows the calculation of the variability of the test product. The test product variability is not used in the proposed statistical method. Some concern has been raised that an ANDA sponsor may produce a product that has higher variability than the reference product. However, under the recommended design, ANDA sponsors that design a product of lower variability than the reference product will need a smaller number of subjects to pass. A disadvantage of the four-period design is that the dropout rate for studies increases with the length of the study. Nevertheless, sponsors may use the four-period design to demonstrate the BE for their highly variable drug products.

### Statistical Analysis of Bioequivalence

In the analysis of a bioequivalence study, the measurements of both $C_{\max}$ and AUC are subject to the following procedure. The measurement for each subject is log-transformed and the averages, $\mu_T$ and $\mu_R$, of the test and reference products are calculated. The within subject variability of the reference product, $\sigma^2_{WR}$ , is also calculated.

There are two parts to the proposed bioequivalence criteria, a scaled average bioequivalence evaluation and a point estimate constraint. In order to demonstrate bioequivalence both parts must pass. Scaled average bioequivalence for both AUC and $C_{\max}$ is evaluated by testing the following null hypothesis

$$H_0 : \frac{(\mu_T - \mu_R)^2}{\sigma^2_{WR}} > \theta \tag{5}$$

(for given $\theta$>0) *versus* the alternative hypothesis

$$H_1 : \frac{(\mu_T - \mu_R)^2}{\sigma^2_{WR}} \leq \theta, \tag{6}$$

where $\mu_T$ and $\mu_R$ are the averages of the log-transformed measure ($C_{\max}$, AUC ) for the test and reference products, respectively; usually testing is done at level $\alpha$=0.05; and $\theta$ is the scaled average BE limit. Furthermore,

$$\theta = \frac{(\ln \Delta)^2}{\sigma^2_{W0}} \tag{7}$$

where $\Delta$ is 1.25, the usual average BE upper limit for the untransformed test/reference ratio of geometric means, and $\sigma_{W0}$=0.25. Note that rejection of the null hypothesis $H_0$ supports the conclusion of equivalence.

A 95% upper confidence bound for $\frac{(\mu_T-\mu_R)^2}{\sigma^2_{WR}}$ determined in a BE study must be $\leq\theta$, or equivalently, a 95% upper confidence bound for $(\mu_T - \mu_R)^2 - \theta\sigma^2_{WR}$ must be $\leq$0. Additionally, the point estimate (test/reference geometric mean ratio) must fall within [0.80, 1.25]. The test drug must pass both conditions before it is judged bioequivalent to the reference product.

## CONCLUSION

This report presents a proposal for the BE evaluation of highly variable drugs and drug products. This new approach addresses many of the concerns about the BE of highly variable drugs/products that have been raised for the past several years. The proposed approach adjusts the BE limits of highly variable drugs/products by scaling to the within-subject variability of the reference product in the study. The recommendation for the use of reference-scaling is based on the general concept that reference variability should be used as an index for setting the public standard expressed in the BE limit. Furthermore, for drugs and products that are highly variable, reference-scaling effectively decreases the sample size needed for demonstrating BE. The additional requirement of a point-estimate constraint will impose a limit on the difference between the test and reference means, thereby eliminating the potential that a test product would enter the market based on a bioequivalence study with a large mean difference. The use of the reference-scaling approach necessitates a study design that evaluates the reference variability, via multiple administration of the reference treatment to each subject. The recommended 3-period design is the most efficient way to obtain this information. The proposed approach will resolve a number of issues in the BE evaluation of highly variable drugs while achieving the FDA's mission of ensuring that all the drugs approved for use in U.S. are both safe and effective.

## REFERENCES


1. US Code of Federal Regulations Title 21, Section 320.1(e). US Government Printing Office, Washington, DC, 2006.
2. US Food and Drug Administration. Guidance for Industry: Bioavailability and Bioequivalence Studies for Orally Administered Drug Products—General Considerations, March 2003.
3. S. D. Patterson, N. M.-D. Zariffa, T. H. Montague, and K. Howland. Non-traditional study designs to demonstrate average bioequivalence for highly variable drug products. *Eur. J. Pharm. Sci.* **57**:663–670 (2001).
4. L. Z. Benet. Therapeutic Considerations 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_2_files/frame.htm (accessed 04/04/2007).
5. D. J. Schuirmann. A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. *J. Pharmacokinet. Biopharm.* **15**:657–680 (1987).
6. U. S. Food and Drug Administration. Guidance for Industry: Statistical Approaches to Establishing Bioequivalence, January 2001.
7. FDA Orange Book, Preface p. viii. http://www.fda.gov/cder/ob/default.htm (Accessed 03/16/2007).
8. 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).
9. Bio-international '94, Munich Germany. 14–18 June, 1994.
10. 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 3/12/2007).
11. Health Canada, Therapeutic Products Directorate (TPD). Discussion paper on 'Bioequivalence requirements—highly variable drugs and highly variable drug products: issues and options'. Expert Advisory Committee on Bioavailability and Bioequivalence (EAC-BB) Meeting, June 26–27, 2003.
12. Health Canada, Ministry of Health. Guidance for Industry: Conduct and Analysis of Bioavailability and Bioequivalence Studies—Part A: Oral Dosage Formulations Used for Systemic Effects. 1992.
13. Committee for Proprietary Medicinal Products (CPMP), the European Agency for the Evaluation of Medicinal Products (EMEA). *Note for Guidance on the Investigation of Bioavailability and Bioequivalence*. 2001.
14. Japan National Institute of Health, Division of Drugs. *Guideline for Bioequivalence Studies of Generic Drug Products*. 1997.
15. FDA Advisory Committee on Pharmaceutical Science. April 13–14, 2004. http://www.fda.gov/ohrms/dockets/ac/04/minutes/4034M1.htm (accessed April 5, 2007).
16. 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).
17. L. Tothfalusi and L. Endrenyi. Limits for the scaled average bioequivalence of highly variable drugs and drug products. *Pharm. Res.* **20**:382–389 (2003).
18. 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).
19. 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).
20. V. Karalis, M. Symillides, and P. Macheras. Novel scaled average bioequivalence limits based on GMR and variability considerations. *Pharm. Res.* **21**:1933–1942 (2004).
21. 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 3/14/2007).