# Evaluating average bioequivalence using methods for high variability drugs: A case study

M.P. Sánchez[1], C. Gómez[2], J.L. Carrasco[3], J. Ocaña[4], C. von Plessing[2], C.G. Godoy[2], R. Reinbach[5] and R. Godoy[2]

*[1]Faculty of Sciences Chemicals and Pharmaceutical Sciences, University of Chile, Santiago, Chile, [2]Department of Pharmacy, Faculty of Pharmacy, University of Concepción, Concepción, Chile, [3]Department of Public Health, Faculty of Medicine, [4]Department of Statistics, Faculty of Biology, University of Barcelona, Barcelona, Spain and [5]Internal Medicine Department, University of Concepción, Concepción, Chile*

**Key words**
average bioequivalence – scaled average bioequivalence – scaled bioequivalence limits – highly variable drugs

**Abstract.** The purpose of this study was to determine if different methods for average bioequivalence in high variability drugs coincide or not in their conclusions when applied to the same dataset, and to discuss the method validity and reliability of the conclusions. Different approaches for the evaluation of average bioequivalence were applied to the results of a bioavailability trial on the diuretic drug Furosemide. These methods included widening the bioequivalence limits according to regulatory recommendations, scaling the limits and scaling the bioequivalence statistic, jointly with evaluating alternative bioavailability measures. The methods to establish the bioequivalence limits were also combined with some alternative methods to construct confidence intervals. The decision on bioequivalence depends much more on the bioavailability measures than on the statistical approach. The reliability of the final decision lies mainly in the interpretation of these measures and on the special characteristics of each drug.

Correspondence to
J.L. Carrasco
Bioestadística,
Departament de Salut Pública, Facultat de Medicina, Universitat de Barcelona, Casanova 143, 08036 Barcelona
jlcarrasco@ub.edu

## Introduction

One of the main pharmacokinetic properties of a pharmaceutical form is the capacity to release its active principle into the organism, in order to be absorbed in the best conditions and to arrive at the site of action, where it performs its expected therapeutic effect. This process is influenced by many factors, but two of them deserve special attention: the intrinsic variability of the active principle and the process of development and the technology involved in the form elaboration. These sources of variation may be translated in different characteristics of the liberation process; it may become slow or fast, complete or incomplete, etc. The above characteristics are closely related to the bioavailability (BA) and bioequivalence (BE) concepts, which thus have a special connotation in the evaluation of medicines or products.

The definition of bioavailability in FDA [2003] is: "the rate and extent to which the active ingredient or active moiety is absorbed from a drug product and becomes available at the site of action. For drug products that are not intended to be absorbed into the bloodstream, bioavailability may be assessed by measurements intended to reflect the rate and extent to which the active ingredient or active moiety becomes available at the site of action". The definition in CPMP [2001] is almost the same. This guidance adds to the definition "... is understood to be the extent and the rate at which a substance or its active moiety is delivered from a pharmaceutical form and become available in the general circulation".

For bioequivalence CPMP [2001] states: "Two medicinal products are bioequivalent if they are pharmaceutically equivalent or pharmaceutical alternatives and if their bioavailabilities after administration in the same molar dose are similar to such degree that their effects, with respect to both efficacy and safety, will be essentially the same". The 2003 FDA guidance defines it 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".

The above definitions focus on the processes by which the active ingredients or moieties are released from an oral dosage form and move to the site of action. To translate them to more operative quantitative criteria, the current FDA and EMEA regulatory standard proposes that a test formulation be considered as presenting "average bioequivalence" (ABE) with respect to a reference standard (reference formulation), and for a given BA pharmacokinetic measure, when their geometric means ratio (GMR) lies within certain pre-specified limits, currently set at 0.80/1.25 or 80/125% [CPMP 2001, FDA 2001]. The most common BA measures are the area under the curve, AUC, and the maximum reached concentration, $C_{max}$. More technical details on these measures are given below.

Obviously, the decision of declaring ABE or not must be taken on the basis of sample GMRs, computed on experimental sample data, subject to random variation. In order to find a balance between consumer or patient risk (falsely establishing ABE) and producer risk (failing to establish ABE for a truly BE product), an hypothesis test is typically formulated, with a null hypothesis of non-equivalence (say, the "true" GMR is outside the regulatory limits) and an alternative hypothesis of equivalence ("true" GMR within the regulatory limits). The usual decision criterion is based on the "interval inclusion" principle: reject non-equivalence (and so declare ABE) if an appropriate confidence interval (CI) for the GMR lies within the regulatory limits. This provides some control over consumer risk, as is discussed in more detail in the last section.

A particular problem is presented by highly variable (HV) drugs, characterized by their large intrasubject variation [Tothfalusi et al. 2001]. Drugs whose BA parameters show an intrasubject variability greater than a 30% analysis of variance - coefficient of variation (ANOVA-CV or simply CV) are generally referred to as HV drugs [Blume 1993, Midha et al. 1997, 2007, Shah et al. 1996]. Sometimes the inclusion criterion in the HV class is relaxed, as in [Boddy et al. 1995] where HV drugs are defined as those exhibiting intrasubject pharmacokinetic data variability exceeding a 25 – 30 CV%. Highly variable drug products are those exhibiting a high within-variability, possibly due to a poor pharmaceutical formulation. Their constituent drugs are not necessarily highly variable, [Midha et al. 2007].

The coefficient of variation is a very important property in defining the drug type, since it reflects the following pooled components [Midha et al. 1997] and may greatly influence the result of the ABE study: true intrasubject variability in pharmacokinetic processes, plus a component of analytical variability; within-formulation variation (e.g., tablet-to-tablet variation); a possible subject-by-formulation interaction term; and unexplained random variation.

The main problem with HV drugs lies in the low power of the standard ABE procedures [Midha et al. 1997, Boddy et al. 1995] when they are applied to these kind of data. That is, there is a low probability of BE approval even under true BE conditions, in the classic 2 × 2 cross-over design (2 periods and 2 sequences) using a moderate sample size. A large number of subjects would be required in order to achieve a suitable statistical power and to establish the condition of BE. This problem remains present in drugs with intermediate coefficients of variation, including values between 25 and 30% [Boddy et al. 1995, Shah et al. 1996], so perhaps it would be worth including these drugs in the HV category.

To cope with the problem of ABE assessment in HV drugs, some possible approaches have been proposed and discussed:

1. To reduce the confidence level of the intervals, i.e. to increase the Type I error rate or consumer risk [Boddy et al. 1995, Midha et al. 2005]. This approach is generally discarded as inadequate.
2. To conduct the BE study at steady state [Midha et al. 2005, WHO 2005].
3. To employ metabolite data. This may be adequate provided that metabolites are often less variable than the parent drug, or under special conditions, e.g. when direct drug level determination is difficult [CPMP 2001, FDA 2003, Guerra et al.

2001, Midha et al. 2004, 2005, 2007, WHO 2005].

4. To use more sophisticated experimental designs, other than the 2 × 2 crossover design, the most commonly used in bioequivalence studies [Patterson et al. 2001, Shah et al. 1996, WHO 2005].
5. To widen the ABE acceptance limits, thus redefining BE and, at the same time, trying to maintain the consumer risk at its nominal level, say 5%. Widening the acceptance limits may be achieved by broadening them to new fixed values, say 75/133% [Boddy et al. 1995, CPMP 2001, Shah et al. 1996], or by means of some sort of gradual scaling, in function of the variability and eventually other measures [Karalis et al. 2004, 2005, Kytariolos et al. 2006, Midha et al. 2005, 2007, Tothfalusi and Endrenyi 2003, Tothfalusi et al. 2001, 2003].
6. To define a new pharmacokinetic measure for the rate of absorption [Endrenyi 1993, 1997, Lacey et al. 1994, Shah et al. 1996].

Approaches 2 – 4 above imply changes in the way that the experiment is designed and conducted, and thus are not feasible in the present paper which consists in an analysis case-study using pre-existing data. Here, we illustrate the use of some possible approaches that lie in the last two categories, 5 and 6.

A drug may be HV for a given pharmacokinetic measure (usually $C_{max}$) and not for the others. This relates to the whole question of the adequacy of the BA metrics in ABE studies. Several types of area under the curve (AUC) may be good indicators of the extent of the absorption [FDA 2003, Shah et al. 1996]: $AUC_{0-t}$ (area under the curve from zero to t, where t is the last time point with a measurable concentration for each formulation) and $AUC_{0-\infty}$ (area under the curve from zero to infinite, with $AUC_{0-\infty} = AUC_{0-t} + C_t/\lambda_z$, where $C_t$ is the last measurable drug concentration and $\lambda_z$ is the terminal or elimination rate constant). The main problem lies in the measurement of the rate of absorption as the traditional indicators, $t_{max}$ (time to reach maximum plasma concentration) and $C_{max}$, are not fully reliable metrics. $t_{max}$ is highly dependent on the sampling frequency and $C_{max}$ is a measure of both extent and rate [Bois et al. 1994a,b]. Despite being a confounded measure, $C_{max}$ is often a critical metric for assessing the potential adverse effects of certain drugs. Therefore, $C_{max}$ remains a very important clinical metric and, what is more, there is evidence that $C_{max}$ and $t_{max}$ provide useful information about the absorption rate [Tothfalusi et al. 2001]. Provided the interest of $C_{max}$ but its possible drawbacks, including its associated high variabilities, a new metric, $C_{max}/AUC_{0-\infty}$, for measuring the absorption rate has been proposed [Endrenyi 1993, 1997, Lacey et al. 1994] as an adequate replacement for $C_{max}$. It may enhance specificity as is less sensitive to deviations and variations in the extent of absorption. However, in [Tozer 1997, Tozer et al. 1996] prevent against the automatic use of this measure.

In this paper, using data coming from an experiment performed in 2003 [unpublished results; available upon request to the first author], we evaluate ABE using some of the methods and criteria discussed above, i.e. those based on widening or scaling the ABE limits and the possible use of an additional BA measure. In addition, we also test a number of interesting, though in practice not widely used, confidence intervals as possible inferential alternatives to the usual "shortest" CI.

## Subjects, materials and methods

For the sake of completeness and clarity, we shortly describe the experimental data we have taken to conduct our case-study on diverse ABE methods. But it should be clear that this paper is not devoted to present these experimental results in depth, nor to discuss the correctness of the experimental process, nor to finally establish or not BE.

### *Subjects*

16 volunteers (8 male and 8 female) were enrolled in the initial average bioequivalence study performed in 2003. On the basis of their medical histories, physical examinations and laboratory tests, all the subjects were found to be in good health prior to the study. The mean

± standard deviation age of the subjects was 24 ± 4 years with a range of 21 – 36 years, the mean body weight was 62 ± 6 kg with a range of 54 – 73 kg, and the mean height was 165 ± 7 cm with a range of 155 – 178 cm.

The protocol of study was reviewed and approved by the Ethics Committee of the University of Concepción (Chile), and all subjects read and signed a written informed consent. The initial study was performed in the Teaching Pavilion of the University of Concepción, in the Hospital Las Higueras of Talcahuano (Chile) and in the Department of Pharmacy at the University of Concepción.

## *Furosemide data*

The drug under evaluation was Furosemide (4-chloro-N-furfuryl-5-sulphamoylanthranilic acid). It is a potent diuretic agent widely used in the treatment of edematous states associated with cardiac and chronic renal failure, hypertension, congestive heart failure and cirrhosis of the liver [McEvoy 1998]. The bioavailability of Furosemide from oral dosage forms is highly variable. Drug information sources generally report a bioavailability of 60 – 70% showing an incomplete but fairly rapid absorption.

The rate and extent of Furosemide absorption are complicated by high degrees of inter- and intrasubject variability [Ponto 1990] and both are reduced by food. Furosemide is extensively bound to plasma proteins, mainly to albumin, plasma concentrations ranging from 1 to 400 μg/ml are 91 – 99% bound in healthy individuals while the unbound fraction averages 2.3 – 4.1% at therapeutic concentrations [Gilman 1990, Najib et al. 2003]. Furosemide elimination takes place mainly via tubular secretion, the main part of the dose absorbed is found unchanged in the urine and a smaller fraction is excreted as the glucuronide [Hammarlund-Udenaes 1989, Vergin et al. 1998].

Furosemide was well tolerated by all the volunteers and no side effects were reported. All the subjects were, therefore, able to complete the trial and were discharged in good health.

Since its establishment in 1995, the Biopharmaceutical Classification System (BCS) [Amidon et al. 1995, Kanfer 2002, Reynolds 1996] has become an increasingly important tool for the worldwide regulation of drug products. To classify a drug according to the BCS, its solubility, dose and permeability must be known. Data regarding the solubility and permeability of Furosemide have allowed it to be classified as a class IV drug (poorly soluble, poorly permeable) in the BCS [Lindenberg et al. 2004, Lobenberg 2000, Wu 2005]. For drug products containing class IV substances, in vivo bioequivalence studies need to be conducted to evaluate the therapeutic equivalence between formulations that are equivalent pharmaceutics.

## *Study products, study design and sample collection*

The formulations employed in this trial were: Furosemide 40 mg tablets, generic (Batch No. 005084), as the test formulation, and LASIX 40 mg, Aventis® (Batch No. D 543), as the reference formulation.

The study was conducted in a randomized, single-dose, cross-over 2 × 2 design with a 1-week washout period between the two doses. During each period, the volunteers received a single test or reference tablet. Participants had been requested not to take any medication 7 days before the study or during the trial.

Blood samples (4 ml) were collected into EDTA-containing tubes before and 0.25, 0.50, 0.75, 1.00, 1.25, 1.50, 1.75, 2.00, 2.50, 3.00, 4.00, 5.00, 6.00 and 8.00 hours after the administration of each formulation. Plasma was separated by centrifugation and stored frozen at –20 ± 4 °C until quantitative analysis.

## *Analysis of plasma samples*

A high-performance liquid chromatography (HPLC) method was developed and validated for Furosemide analysis in plasma samples using an extraction procedure based on liquid-liquid extraction [Gomez et al. 2005].

## *Pharmacokinetics and statistical analysis*

The pharmacokinetics characteristics of Furosemide were determined from the plas-

ma concentration-time data. The maximum plasma concentration ($C_{max}$) and the time to reach maximum plasma concentration ($t_{max}$) were obtained directly from the plasma concentration-time data. The area under the curve plasma concentration time up to the last time (t) was determined by applying the linear trapezoidal rule. The apparent elimination rate constant ($K_e$) was calculated by the log-linear regression of the data points describing a terminal log-linear decaying phase.

The analysis of variance (ANOVA) for 2 × 2 crossover designs [Chow 2000, Senn 2002] was used to assess the effects of formulation, periods, sequences and subjects within sequence on the logarithmically transformed $AUC_{0\text{-}t}$, $AUC_{0\text{-}\infty}$, $C_{max}$ and $C_{max}/AUC_{0\text{-}\infty}$ variables. As a complementary analysis, the negligibility of the carry-over effect was tested by means of Wellek's scaled equivalence test [Wellek 2003]. This last approach tests a null hypothesis of non-negligibility vs. an alternative of negligibility of the carry-over effect and, thus, seems a more appropriate way of ruling out the presence of disturbing carry-over, as it provides direct "proof" of its negligibility.

The pharmacokinetics characteristics $AUC_{0\text{-}t}$, $AUC_{0\text{-}\infty}$, $C_{max}$, $t_{1/2}$, $K_e$, $C_{max}/AUC_{0\text{-}\infty}$ were determined based on a multiplicative model, $t_{max}$ was analyzed assuming an additive model. For the purpose of average bioequivalence analysis $AUC_{0\text{-}t}$, $AUC_{0\text{-}\infty}$, $C_{max}$, were considered as primary variables and $C_{max}/AUC_{0\text{-}\infty}$ was also computed as a measure of the rate of absorption.

The initial ABE study was performed using the standard methods: Schuirmann´s test [Schuirmann 1987] and the equivalent procedure of confidence interval inclusion using the shortest 90% CI for $AUC_{0\text{-}t}$, $AUC_{0\text{-}\infty}$, $C_{max}$, and $C_{max}/AUC_{0\text{-}\infty}$. The remaining parameters ($K_e$, $t_{max}$ and $t_{1/2}$,) were analyzed by the Wilcoxon rank sum test [Hauschke et al. 2007].

To evaluate ABE using some methods proposed in the literature for high variability drugs, a second average bioequivalence study was performed adopting the same confidence interval inclusion approach but, in addition to the 90% shortest confidence interval, three additional confidence intervals were also considered: Westlake's symmetric 95% interval [Westlake, 1976] and the two 95% intervals discussed in Hsu et al. [Hsu 1994] and designated as the symmetric interval $I_S$ and the bioequivalence specific interval $I_*$. These 95% intervals provide a bioequivalence test of size 5%, while the shortest interval must have a 90% confidence level to ensure the same 5% size. $I_S$ and $I_*$ are rarely applied to real data and have not been compared with the shortest confidence interval. The above four confidence intervals were combined with a number of methods for widening bioequivalence limits in HV drugs:

- Fixed enlarged limits according to the recommendations of the regulatory agencies [CPMP 2001, FDA 2001].
- Mixed scaled methods: unscaled BE limits up to a given CV threshold (e.g. $CV_0$ = 22.6% on the original measurement scale or, equivalently, $\sigma_0 = 0.223$ on the logarithmic scale) and scaled limits above this threshold, with a scaling factor $k$ [Midha et al. 2005, Tothfalusi and Endrenyi 2003, Tothfalusi et al. 2001]. The decision as to whether to scale or not is based on the estimated value of a scaling variability, $\hat{\sigma}_{SC}$. In 2 × 2 studies, the most appropriate choice for this variability is the residual standard deviation estimate, $\hat{\sigma}_{Res.}$
- Scaled limits by methods based on the GMR, variability considerations and leveling-off properties [Karalis et al. 2004, 2005, Kytariolos et al. 2006]. The above criteria are made explicit by means of certain functions defining the ABE limits, These functions depend on a set of preestablished parameters ($\alpha, \beta, \gamma$). In our study we selected the parameter values $\alpha = 1.25$, $\beta = 1.33$ and $\gamma = 4$, the "best" options according to the authors.
- Scaling the metric rather than the BE acceptance limits. Inference is based on two possible approaches: noncentral $t$ distribution and linearization of the regulatory criterion, described in [Midha et al. 2005, Tothfalusi and Endrenyi 2003, Tothfalusi et al. 2001].

All computations were performed on the logarithmic scale and the results exponentially transformed so that they might be discussed on the original scale.

All tests were performed at a 0.05 significance level with the statistical software SAS version 9.1 (SAS Institute Inc. Cary. NC,

USA: s.n., v. 9.1) and with R 2.4.0 (R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, http://www.R-project.org).

## Results

The main descriptive statistics of the pharmacokinetic parameters are summarized in Table 1. In general, with the exception of $t_{1/2}$, all the pharmacokinetic measures were systematically greater in the reference than in the test formulation.

Prior to bioequivalence evaluation, and as a complementary analysis to the ABE study, the means of the three main pharmacokinetic measures ($AUC_{0-\infty}$, $AUC_{0-t}$ and $C_{max}$) were compared by means of the ANOVA procedure. Neither significant drug effects nor period, nor carry-over effects were found. Additionally, the "negligibility" of the carry-over effect was tested with the equivalence method stated [Wellek 2003]. For the most problematic variable (with respect to ABE testing), $C_{max}$, carry-over was established as negligible (p-value = 0.03844). The remaining p-values were 0.1386, 0.03992 and 0.05657 for $AUC_{0-\infty}$, $AUC_{0-t}$ and $C_{max}$/AUC, respectively.

For $AUC_{0-t}$, $AUC_{0-\infty}$, and $C_{max}$, the ANOVA-CVs were 17.68%, 14.11% and 36.67% respectively, which would seem to confirm that Furosemide is a highly variable drug for $C_{max}$ , (according to Midha et al. [2005]). For the alternative measure $C_{max}$/$AUC_{0-\infty}$, the corresponding CV was 25.63%, which suggests the possible inclusion of Furosemide also in the HV category if less stringent criteria are in use. But in any case the use of this alternative measure is associated with more than a 30% of variability reduction, with respect to $C_{max}$. Table 2 shows the classic inference results for average BE.

Using the common approaches (Table 2), average bioequivalence is only concluded for $\ln AUC0_{0-\infty}$ and rejected for the remaining parameters. However, the results for $\ln AUC_{0-t}$ and $\ln C_{max}/AUC_{0-\infty}$ are very close to the lower limit of bioequivalence. In order to find out the extent to which the conclusion is sensitive to methods specially designed for HV drugs, some alternative approaches were applied to these data. The methods under consideration involved all possible combinations of approaches to bioequivalence limits redefinition with various confidence interval types.

Table 3 shows the results for $C_{max}$. None of the BE limit-widening methods allows us to confirm ABE, with the exception of the

Table 1. Statistical summary of Furosemide pharmacokinetic parameters (test and reference, n = 16).

| Parameter | Formulation | Mean | Median | Geometric mean | Standard deviation |
|---|---|---|---|---|---|
| $AUC_{0-t}$ [μg × h/ml] | Test | 2.94 | 2.86 | 2.87 | 0.67 |
| | Reference | 3.30 | 3.20 | 3.23 | 0.79 |
| $AUC_{0-\infty}$ [μg × h/ml] | Test | 3.40 | 3.28 | 3.36 | 0.58 |
| | Reference | 3.69 | 3.52 | 3.63 | 0.75 |
| $C_{max}$ [μg/ml] | Test | 1.31 | 1.14 | 1.19 | 0.63 |
| | Reference | 1.47 | 1.31 | 1.41 | 0.49 |
| $t_{max}$ [h] | Test | 1.05 | 1.00 | 0.97 | 0.43 |
| | Reference | 1.28 | 1.50 | 1.16 | 0.54 |
| $t_{1/2}$ [h] | Test | 3.49 | 2.91 | 3.17 | 1.83 |
| | Reference | 2.95 | 2.79 | 2.80 | 1.04 |
| $K_e$ [$h^{-1}$] | Test | 0.24 | 0.24 | 0.22 | 0.08 |
| | Reference | 0.26 | 0.25 | 0.25 | 0.08 |
| $C_{max}/AUC_{0-\infty}$ [$h^{-1}$] | Test | 0.37 | 0.36 | 0.36 | 0.11 |
| | Reference | 0.39 | 0.35 | 0.39 | 0.08 |

Table 2. Summary of classic methods for ABE, 5% two one-sided tests (TOST) and 90% shortest confidence intervals (SCI). All values are displayed in the original scale.

| Parameter | Geometric means | | GMRs (T to R) | Residual Variance | TOST p-value | | SCI [T/R] | ABE |
|---|---|---|---|---|---|---|---|---|
| | Reference | Test | | | Upper tail | Lower tail | | |
| $\ln AUC_{0-t}$ | 3.2278 | 2.8688 | 88.88 | 0.0308 | $< 1 \times 10^{-5}$ | 0.0559 | [79.7, 99.1] | No |
| $\ln AUC_{0-\infty}$ | 3.6306 | 3.3609 | 92.58 | 0.0197 | $1.5 \times 10^{-5}$ | 0.0054 | [84.8, 101.1] | Yes |
| $\ln C_{max}$ | 1.4059 | 1.1982 | 85.22 | 0.1262 | 0.0043 | 0.3112 | [68.3, 106.3] | No |
| $\ln[C_{max}/AUC_{0-\infty}]$ | 0.3872 | 0.3565 | 92.05 | 0.0636 | 0.0020 | 0.0689 | [78.9, 107.5] | No |

Table 3. ABE evaluation for $C_{max}$ according to different BE limits and confidence interval types.

| Bioequivalence limits | Confidence Interval | | | |
|---|---|---|---|---|
| | Shortest *I* [68.3, 106.3] | Westlake's $I_W$ [68.3, 146.5] | Hsu et al. Symmetric $I_S$ [68.3, 146.4] | Hsu et al. Specific $I_*$ [68.3, 106.3] |
| Arlington Workshop [Shah 1996] [70.0, 143.0] | Not BE | Not BE | Not BE | Not BE |
| EMEA [CPMP 2001] [75.0, 133.0] | Not BE | Not BE | Not BE | Not BE |
| Mixed Scaled limits based on residual variance [Midha et al. 2005, Tothfalusi et al. 2003] | | | | |
| $\sigma_0 = 0.20$ [67.3, 148.6] | BE | BE | BE | BE |
| $\sigma_0 = 0.223$ [70.1, 142.7] | Not BE | Not BE | Not BE | Not BE |
| $\sigma_0 = 0.294$ [76.4, 130.9] | Not BE | Not BE | Not BE | Not BE |
| Scaled limits based on GMR and variability [Karalis 2004] | | | | |
| BELscG1 [75.8, 131.9] | Not BE | Not BE | Not BE | Not BE |
| BELscG2 [77.0, 129.8] | Not BE | Not BE | Not BE | Not BE |
| Scaled limits based on GMR with leveling-off properties [Karalis 2005] | | | | |
| BELscM (Michaelis-Menten) [79.8, 125.2] | Not BE | Not BE | Not BE | Not BE |
| BELscE (Simple Exponential) [78.8, 126.9] | Not BE | Not BE | Not BE | Not BE |
| BELscW (Weibull) [78.7, 127.1] | Not BE | Not BE | Not BE | Not BE |
| Scaled limits with leveling-off properties [Kytariolos 2006] | | | | |
| $BE_{ef}W$ (Weibull limit expansion function) [75.8, 131.9] | Not BE | Not BE | Not BE | Not BE |

BE = bioequivalent; Not BE = no evidence of bioequivalence.

scaled limits with the "switching" variability threshold $\sigma_0 = 0.20$. This threshold value was suggested by the FDA [2001]. The presence of this additional $\sigma_0$ parameter has received some criticism [Midha et al. 2005], as it introduces some extra arbitrariness. Additionally, all methods would give the same result (that is, not to declare ABE) under the additional restriction imposing that the sample GMR should lie between the usual 80/125 ABE acceptance limits.

It can be seen that, for $C_{max}$, the ABE decision does not depend on the confidence interval type. In fact, all the confidence intervals share a very similar lower limit and the non acceptance of the ABE is conditioned primar-

Table 4. ABE evaluation for $C_{max}/AUC_{0-\infty}$ decision according to different BE limits and confidence interval types.

| Bioequivalence limits | Confidence Interval | | | |
|---|---|---|---|---|
| | Shortest *I* [78.7, 107.7] | Westlake's $I_W$ [78.6, 127.3] | Hsu et al. Symmetric $I_S$ [78.7, 127.1] | Hsu et al. Specific $I_*$ [78.7, 107.7] |
| Arlington Workshop [Shah 1996] [70.0, 143.0] | BE | BE | BE | BE |
| EMEA [CPMP 2001] [75.0, 133.0] | BE | BE | BE | BE |
| Mixed Scaled limits based on residual variance [Midha et al. 2005, Tothfalusi et al. 2003] | | | | |
| $\sigma_0$ = 0.20 [75.5, 132.5] | BE | BE | BE | BE |
| $\sigma_0$ = 0.223 [77.7, 128.7] | BE | BE | BE | BE |
| $\sigma_0$ = 0.294 [80, 125] (no scaling as $\hat{\sigma}_{SC}$ = 0.252 < $\sigma_0$) | Not BE | Not BE | Not BE | Not BE |
| Scaled ABE based on GMR and variability [Karalis 2004] | | | | |
| BELscG [73.8, 135.5] | BE | BE | BE | BE |
| BELscG2 [75.1, 133.2] | BE | BE | BE | BE |
| Scaled ABE based on GMR with leveling-off properties [Karalis 2005] | | | | |
| BELscM (Michaelis-Menten) [79.8, 125.3] | Not BE | Not BE | Not BE | Not BE |
| BELscE (Simple Exponential) [77.9, 128.3] | BE | BE | BE | BE |
| BELscW (Weibull) [77.9, 128.3] | BE | BE | BE | BE |
| Scaled ABE with leveling-off properties [Kytariolos 2006] | | | | |
| $BE_{ef}W$ (Weibull limit expansion function) [76.9 , 130.0] | BE | BE | BE | BE |

BE = bioequivalent; Not BE = no evidence of bioequivalence.

ily by this lower limit, which is close to, but clearly lower than, the expanded lower acceptance limit proposed in [Shah et al. 1996] – and farther from the lower acceptance limits associated with the remaining methods.

Table 4 shows the results on the metric $C_{max}/AUC_{0-\infty}$. With the exception of the scaling method BELscM, based on a Michaelis-Menten function (whose application results in limits close to the standard 0.80/1.25) and the mixed scaled limits with threshold $\sigma_0$ = 0.294 (with no scaling at all, as the observed scaling variability does not reach this threshold -in fact, if scaling were applied the limits would be narrower than 0.80/1.25), the findings for all the other approaches resulted in an ABE statement, irrespective of the confidence interval method.

As discussed earlier, what is labeled here as a "mixed scaled" method is used in Tables 3 and 4 as a mere BE limits expansion device, in conjunction with confidence intervals that are only appropriate for the difference of means under *constant* BE limits, and not limits that depend on unknown parameters. A more appropriate inferential approach may be to construct confidence limits for the scaled measure of BE [Tothfalusi and Endrenyi 2003]. A similar objection may be made of the scaled BE limits introduced in [Karalis et al. 2004, 2005, Kytariolos et al. 2006], although in these cases there is no clear inferential procedure. Table 5 shows the results when the methods based on the noncentral *t* distribution and on linearization of the regulatory criterion are applied.

Scaled ABE gives different results depending on the approach used in calculating the BE limits. In the case of $C_{max}$ all the procedures conducted to a bioinequivalence con-

Table 5. ABE decision according to inferential methods based on scaling the formulation effect measure [Midha et al. 2005]

| Noncentral *t* distribution | | Confidence interval limits | |
|---|---|---|---|
| Bioequivalence limits according to $\sigma_0$ | | $\ln C_{max}$ [–3.24,0.39] | $\ln C_{max}/AUC_{0-\infty}$ [–2.85, 0.73] |
| $\sigma_0$ = 0.20 [FDA 2001] | [–3.16, 3.16] | Not BE | BE |
| $\sigma_0$ = 0.223 [Boddy et al. 1995] | [–2.83, 2.83] | Not BE | Not BE |
| $\sigma_0$ = 0.294 [Tothfalusi et al. 2003] | [–2.15, 2.15] | Not BE | Not BE |
| Linearization of the regulatory criterion | | Upper 95% confidence limit for $(\mu_T - \mu_R)^2 - k\sigma^2_{SC}$. Bioequivalence is established if CL is not positive | |
| $\sigma_0$ = 0.200, *k* = 1.116 | | CL = 0.0839, Not BE | CL = 0.0162, Not BE |
| $\sigma_0$ = 0.223, *k* = 1.000 | | CL = 0.0856, Not BE | CL = 0.0307, Not BE |
| $\sigma_0$ = 0.294, *k* = 0.759 | | CL = 0.0912, Not BE | CL = 0.0326, Not BE |

BE = bioequivalent; Not BE = no evidence of bioequivalence; CL = upper confidence limit.

clusion (or more precisely, to not discard bioinequivalence), whereas in the case of the metric $C_{max}/AUC_{0-\infty}$ the results were not always the same. When using the noncentral *t* distribution and the threshold $\sigma_0 = 0.20$ as proposed in [FDA 2001] and $\sigma_0 = \ln(1.25) = 0.223$, in line with Boddy et al. 1995, the final outcome was declaring ABE. For $\sigma_0 = 0.294$, the final outcome was bioinequivalence. The results using the "linearization of the regulatory criterion" were in quite close agreement with the above findings, albeit this criterion would appear to be more restrictive in producing an ABE statement.

## Discussion

As has been previously pointed out, the main problem with testing ABE in HV drugs is the very low power of the standard methods. The regulatory recommendations impose sample sizes between 12 and 24, so that the sample size of 16 subjects in the data we have used to conduct this case study is within the recommendations. However, a much higher number of subjects would be necessary to evaluate ABE with a suitable power level. Specifically, for $C_{max}$ a number of 392 subjects would be required to have a 80% power for an acceptance range of 0.80/1.25 [Hauschke et al. 2007, Patterson 2006], which is much higher than the recommendation and completely unfeasible in practice. Note that in the majority of times when ABE is not declared, this decision lies mainly in the lower bound of the confidence interval, which is close but lower than the lower bound of the bioequivalence limits. It is reasonable to conjecture that a much greater sample size might result in shorter confidence intervals and perhaps lead to an ABE statement.

Provided the unfeasibility of the required sample sizes, an alternative approach would be to use statistical methods that would improve the power for a given, fixed sample size. With respect to ABE testing, the confidence interval construction method does not influence our results, so the best alternative seems to be the simple, standard "shortest" interval. In terms of formulation effect estimation, perhaps a good alternative confidence interval would be $I_*$, which is similar in length and more confident (e.g. 95% instead of 90%) than the shortest interval. There are alternative more powerful tests for ABE, for example those described in [Berger 1996, Chow 2000, Wellek 2003], but these methods have other drawbacks and are far from the approach taken in the regulations, so they are not considered here.

The final decision on ABE is slightly more dependent on the method used to widen or to scale the bioequivalence limits, but as has been pointed in the Results section, with a few marginal exceptions all methods lead to

similar results: results on $C_{max}$/AUC point to an average bioequivalence conclusion, while for $C_{max}$ the results point to a contrary conclusion. That is, for the Furosemide data under consideration, the final outcome seems to depend much more on the BA metric than on the method to determine the bioequivalence limits. So in these data the final decision on bioequivalence seems to be not clear. For absorption extent, the results on AUC strongly suggest ABE while for absorption rate the final statement will strongly depend on the metric. If $C_{max}$/AUC is considered as an acceptable measure of rate, as in [Endrenyi 1993], ABE seems clear. These authors suggest the 75/133% limits as adequate for this measure. On the other hand, if $C_{max}$ is considered, the lack of arguments favoring bioequivalence is also clear for this mixed (rate and extent) measure.

Then, what is better, to base the decision on $C_{max}$ or on $C_{max}$/AUC? In our opinion both measures are very uninformative in the specific case of Furosemide: The diuretic effect of this drug is more closely associated to the urinary excretion rate than to the plasma concentration [Hammarlund-Udenaes 1989, Vergin et al. 1998]. The excretion rate associated with maximum efficiency is about 21.5 μg $min^{-1}$, [Hammarlund-Udenaes 1985, Wakelkamp et al. 1997] and according to [Vergin et al. 1998], the plasma concentration corresponding to a maximally efficient excretion rate is approximately 200 ng/ml [Vergin et al. 1998]. If the level of Furosemide exceeds this value, there is not an enhancement of the diuretic effect as the urinary excretion remains constant.

For the reference formulation, the observed mean $C_{max}$ was (expressing it in the same unit) 1,469 ng/ml with a range of 960 to 2,900 ng/ml and for test formulation the observed mean $C_{max}$ value was 1308 ng/ml with a range of 550 to 2,990 ng/ml. So, the plasmatic concentrations largely exceeded the 200 ng/ml level, and thus $C_{max}$ and $C_{max}$/AUC seem to have very limited value (and to a lesser extent, also AUC that should be computed in a truncated manner) if therapeutic efficacy is under consideration, though they have full value for drug safety evaluation.

In our opinion, the above results suggest that all the statistical methods under consideration should help in the decision, but should not determine the decision itself. These results illustrate the complexity of making a correct decision about ABE in the case of HV drugs. Further research is needed in this field to improve methodological approaches and to have a better knowledge of their theoretical properties and their true applicability to real data.

## References


*Amidon GL, Lennernäs H, Shah VP, Crison JR.* A theoretical basis for a biopharmaceutics drug classification: the correlation of a invitro drug product dissolution and in vivo bioavailability. Pharm Res. 1995; *12:* 413-420.

*Berger RL, Hsu JC.* Bioequivalence trials, Intersection-union tests and equivalence confidence sets. Stat Sci. 1996; *11:* 283-319.

*Blume H, Midha KK.* Bio-International 92, conference on bioavailability, bioequivalence, and pharmacokinetic studies, J. Pharm. Sci. 1993; *82:* 1186-1189.

*Boddy AW, Snikeris FC, Kringle RO, Wei GCG, Oppermann JA, Midha KK.* An approach for widening the bioequivalence acceptance limits in the case of highly variable drugs. Pharm Res. 1995; *12:* 1865-1868.

*Bois F, Tozer TN, Hauck WW, Chen ML, Patnaik R, Williams RL.* Bioequivalence: performance of several measures of extent of absorption. Pharm Res. 1994a; *11:* 715-722.

*Bois F, Tozer TN, Hauck WW, Chen ML, Patnaik R, Williams RL.* Performance of Several Measures of Rate of Absorption. Pharm Res. 1994b; *11:* 966-974.

*Chow S-C, Liu J-P.* Design and analysis of bioavalability and bioequivalence studies. New York: Marcel Dekker; 2000.

*CPMP.* Note for Guidance on the Investigation of Bioavailability and Bioequivalence. CPMP/EWP/QWP/1401/98; 2001.

*Endrenyi L, Tothfalusi L.* Secondary Metrics for the Assessment of Bioequivalence. J Pharm Sci. 1997; *86:* 401-402.

*Endrenyi L, Yan W.* Variation of $C_{max}$ and $C_{max}$/AUC in investigations of Bioequivalence. Int J Clin Pharmacol Ther. 1993; *31:* 184-189.

*FDA.* Bioavailability and Bioequivalence Studies for Orally Administered Drug Products – General Considerations. Rockville, MD: CDER; 2003.

*FDA.* Statistical Approaches to Establishing Bioequivalence. Rockville, MD: CDER; 2001.

*Gilman AG, Rall TW, Nies AS et al. (eds).* Goodman and Gilman's The Pharmacological Basis of Therapeutics, 8th edition. New York: Pergamon Press; 1990.

*Gómez C, Von Plessing C, Godoy CG, Reinbach R, Godoy R.* Method validation for the determination of Furosemide in plasma by liquid-liquid Extraction and high-performance liquid Chromatography with fluorescence detection. J Chil Chem Soc. 2005; *50:* 479-482.

*Guerra P, Frias J, Ruiz B, Soto A, Carcas A, Govantes C, Montuenga C, Fernández A.* Bioequivalence of allopurinol and its metabolite oxipurinol in two tablet formulations. J Clin Pharm Ther. 2001; *26:* 113-119.

*Hammarlund-Udenaes M, Benet LZ.* Acute Tolerance to Furosemide Diuresis in Humans, Pharmacokinetics-Pharmacodynamics Modeling. J Pharmacol Exp Ther. 1985; *233:* 447-453.

*Hammarlund-Udenaes M, Benet LZ.* Furosemide pharmacokinetics and pharmacodynamics in health and disease – An update. J Pharmacokinet Biopharm. 1989; *17:* 1-46.

*Hauschke D, Steinijans V, Pigeot I.* Bioequivalence Studies in Drug Development. Methods and Applications. Chichester: John Wiley & Sons; 2007.

*Hsu J, Hwang JTG, Liu H-K, Ruberg S.* Confidence intervals associated with test for bioequivalence. Biometrika. 1994; *81:* 103-114.

*Hyslop T, Hsu F, Holder DJ.* A small sample confidence interval approach to assess individual bioequivalence. Stat Med. 2000; *19:* 2885-2897.

*Kanfer I.* Report On The International Workshop On The Biopharmaceutics Classification System (BCS): Scientific And Regulatory Aspects In Practice. J Pharm Pharmaceut Sci. 2002; *5:* 1-4.

*Karalis V, Symillides M, Macheras P.* Novel scaled average bioequivalence. Pharm Res. 2004; *21:* 1933-942.

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

*Kytariolos J, Karalis V, Macheras P, Symillides M.* Novel scaled bioequivalence limits with leveling-off properties. Pharm Res. 2006; *23:* 2657-2664.

*Lacey LF, Kenne ON, Duquesnoy C, Bye A.* Evaluation of different indirect measures of rate of drug absorption in comparative pharmacokinetic studies. J Pharm Sci. 1994; *83:* 212-215.

*Lindenberg M, Kopp S, Dressman JB.* Classification of orally administered drugs on the World Health Organization Model list of Essential Medicines according to the biopharmaceutics classification system. Eur J Pharm Biopharm. 2004; *58:* 265-278.

*Liu H-K.* Confidence intervals in bioequivalence, in proceedings of the biopharmaceutical section. Washington, DC: American Statistical Association; 1990. pp. 51-54.

*Lobenberg R, Amidon GL.* Modern bioavailability, bioequivalence and biopharmaceutics classification system. New scientific approaches to international regulatory standards. Eur J Pharm Biopharm. 2000; *50:* 3-12.

*McEvoy GK.* AHFS Drugs Information. Bethesda, MD: American Society of Health-System Pharmacists; 1998.

*Midha KK, Rawson MJ, Hubbard JW.* Individual and average bioequivalence of highly variable drugs and drug products, J Pharm Sci. 1997; *86:* 1193-1196.

*Midha KK, Rawson MJ, Hubbard JW.* The role of metabolites in bioequivalence. Pharm Res. 2004; *21:* 1331-1344.

*Midha KK, Rawson MJ, Hubbard JW.* The bioequivalence of highly variable drugs and drug products. Int J Clin Pharmacol Ther. 2005; *43:* 485-498.

*Midha KK, Shah V, Singhand GJP, Patnaik R.* Commentary, Conference Report: Bio-International 2005. J Pharm Sci. 2007; *96:* 747-754.

*Najib N, Idkaidek N, Beshtawi M, Bader M, Admour I, Alam SM, Zaman Q, Dham R.* Bioequivalence evaluation of two brands of furosemide 40 mg tablets (Salurin and Lasix) in healthy human volunteers. Biopharm Drug Dispos. 2003; *24:* 245-249.

*Patterson S, Jones B.* Bioequivalence and Statistics in Clinical Pharmacology, Boca Raton: Chapman and Hall/CRC; 2006.

*Patterson SD, Zariffa NMD, Montague TH, Howland K.* Non-traditional study design to demonstrate average bioequivalence for highly variable drug products. Eur J Clin Pharmacol. 2001; *57:* 663-679.

*Ponto LL, Schoenwald RD.* Furosemide (frusemide): A pharmacokinetic/pharmacodynamic review (Part II). Clin Pharmacokinet. 1990; *18:* 460-471.

*Reynolds JEF.* Martindale – the extra pharmacopoeia. 31st ed. London: Pharmaceutical Press; 1996.

*Schuirmann DJ.* A comparison of the two one-sided test procedures and the power approach for assessing the equivalence of average bioavaibility. J Pharmacokinet Biopharm. 1987; *15:* 657-680.

*Senn S.* Cross-over trials in clinical research. England: John Wiley & Sons; 2002.

*Shah VP, Yacobi A, Barr WH, Benet LZ, Breimer D, Dobrinska MR, Endrenyi L, Fairweather W, Gillespie W, Gonzalez MA, Hooper J, Jackson A, Lesko L, Midha KK, Noonan PK, Patnaik R, Williams RL.* Evaluation of orally administered highly variable drugs and drug formulations. Pharm Res. 1996; *13:* 1590-1594.

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

*Tothfalusi L, Endrenyi L, Midha KK, Rawson MJ, Hubbard JW.* Evaluation of the bioequivalence of highly-variable drugs and drug products. Pharm. Res. 2001; *18:* 728-733.

*Tothfalusi L, Endrenyi L, Midha KK.* Scaling or wider bioequivalence for highly variable drugs and for the special case of $C_{max}$. Int J Clin Pharmacol Ther. 2003; *41:* 217-225.

*Tozer TN, Bois F, Hauck WW, Chen ML, Williams RL.* Absorption rate vs exposure: wich is more useful for bioequivalence testing? Pharm Res. 1996; *13:* 453-456.

*Tozer TN, Hauck WW.* $C_{max}$/AUC. A Commentary. Pharm Res. 1997; *14:* 967-968.

*Vergin H, Mahr G, Metz R, Eichinger A, Nitsche V.* Investigation on the bioequivalence of 2 oral preparations containing spironolactone and furosemide. Int J Clin Pharmaco. Ther. 1998; *36:* 231-238.

*Wakelkamp M, Alván G, Paintaud G.* Natriuretic efficiency of frusemide as a consequence of drug input rate. Br J Clin Pharmacol. 1997; *43:* 481-491.

*Wellek S.* Testing statistical hypotheses of equivalence. Boca Raton: Chapman & Hall CRC; 2003.

*Westlake WJ.* Symmetrical Confidence Intervals for Bioequivalence Trials. Biometrics 1976; *32:* 741-744.

*WHO QAS/04.093/Rev4. Working Document.* Multisource (Generic) Pharmaceutical Products: Guidelines on Registration Requeriments to Establish Interchangeability. Draf Revision; 2005.

*Wu C-Y, Benet LZ.* Predicting drug disposition via application of bcs: transport/absorption/elimination interplay and development of a biopharmaceutics drug disposition classification system. Pharm Res. 2005; *22:* 11-23.