

Statistical Issues in Bioavailability/Bioequivalence Studies

Shein-Chung Chow¹, Laszlo Endrenyi^{2,4*}, Eric Chi³, Lan-Yan Yang¹ and Laszlo Tothfalusi⁴

¹Duke University School of Medicine, Durham, North Carolina, USA

²University of Toronto, Toronto, Canada

³Amgen, Inc., Thousand Oaks, California, USA

⁴Semmelweis University, Budapest, Hungary

Abstract

For the approval of generic drug products, bioavailability/bioequivalence studies are often conducted to demonstrate that the drug absorption profiles in terms of the extent and rate of absorption of test products are bioequivalent to those of the innovative drug product. The bioavailability/bioequivalence studies are often conducted under a standard two-sequence, two-period (2x2) crossover design. Under the standard 2x2 crossover design, statistical methods are well established for the assessment of bioequivalence. However, it is a concern whether approved generic drug products can be used safely and interchangeably. In this article, drug interchangeability under a replicated crossover bioavailability/bioequivalence study is discussed. Several controversial statistical issues that are commonly encountered in the assessment of bioequivalence are discussed. In addition, some frequently asked questions during regulatory submissions are reviewed. Recommendations regarding possible resolutions are made whenever possible. Some concluding remarks on the feasibility of the application of current methods for bioequivalence to the assessment of biosimilarity of follow-on biologics are also presented.

Keywords: Bioequivalence; Therapeutic equivalence; One size-fits-all criterion; Drug interchangeability; Biosimilarity; Follow-on biologics

Introduction

In the pharmaceutical industry, when an innovative (brand name) drug product (for human use) is going off patent, pharmaceutical companies can file an abbreviated new drug application (ANDA) for the approval of a generic product. For this approval, most regulatory agencies, including the United States Food and Drug Administration (FDA), require that evidence of average bioequivalence (BE) (in terms of the extent and rate of drug absorption) be provided through the conduct of bioequivalence studies. These are often undertaken on healthy volunteers for characterizing drug absorption in the blood stream.

For the assessment of average bioequivalence, a standard two-sequence, two-period (2x2) crossover design is usually employed. Average BE is commonly determined by a confidence interval approach or, equivalently, by a two one-sided tests procedure [1]. FDA requires that log-transformation be performed before data analysis. The test product is then claimed to be bioequivalent to the reference product if the calculated 90% confidence interval around the ratio of geometric means of the primary study endpoint [such as the area under the blood or plasma concentration time curve (AUC) or the peak concentration (C_{max})] is totally within the bioequivalence limits (in %) of 80% to 125%.

The assessment of bioequivalence is made under the so-called Fundamental Bioequivalence Assumption. This states that if two drug products are shown to be bioequivalent then they are assumed to have the same therapeutic and adverse effects, i.e. they are therapeutically equivalent. The assumption is discussed further in Section 4.1. The assumption has been challenged by sponsors of many innovative drug products [2]. The assessment of average BE utilizes a one size-fits-all criterion on the primary pharmacokinetic parameters such as AUC and C_{max} regardless of the within-subject and between-subject variabilities of responses to the drug or the therapeutic index of the drug. This one size-fits-all criterion which lacks clinical and/or scientific justification, has been criticized and challenged by many authors. It is further

observed that analyzing the results of studies with or without log-transformation of the parameters may result in different conclusions about bioequivalence. Note that analysis with and without log-transformation of the pharmacokinetic (PK) parameters are referred to as log-transformation model and raw data model, respectively, or alternatively, as the multiplicative and additive model.

An approved generic copy can be used as a substitute of the innovative drug product. However, as more and more generic drug products of the same brand-name drug become available, it is a safety concern whether these generic drug products can be used interchangeably. Note that evidence of average bioequivalence among generic drug products is not required by the regulatory agencies. The issue of drug interchangeability has been studied extensively since early 1990s, which has led to recommendations for the assessment of population and/or individual bioequivalence under replicated crossover designs in the early 2000s [3]. In addition, many controversial issues such as that of multiplicity regarding the design and analysis of bioequivalence studies, still remain to be resolved.

In the next section, the design and analysis of bioequivalence studies will be briefly outlined. Drug interchangeability in terms of drug prescribability and drug switchability are discussed in Section 3. Section 4 presents some controversial issues that are commonly encountered when conducting bioequivalence studies for the assessment of average bioequivalence. These controversial issues include, but are not limited

***Corresponding author:** Laszlo Endrenyi, University of Toronto, Department of Pharmacology and Toxicology, 1 King's College Circle, Toronto, ON, M5S 1A8, Canada, Tel: 416-925-3779; Fax: 416-978-6395; E-mail: l.endrenyi@utoronto.ca

Received September 14, 2011; **Accepted** November 01, 2011; **Published** November 03, 2011

Citation: Chow SC, Endrenyi L, Chi E, Yang LY, Tothfalusi L (2011) Statistical Issues in Bioavailability/Bioequivalence Studies. J Bioequiv Availab S1. doi:10.4172/jbb.S1-007

Copyright: © 2011 Chow SC, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

to: (1) the challenge of the Fundamental Bioequivalence Assumption, (2) the adequacy of a one size-fits-all criterion, (3) the evaluation of BE for highly-variable drugs, and (4) the appropriateness of the log-transformation. Some frequently asked questions during the ANDA submission for generic approval are given in Section 5. Section 6 focuses on the applicability and feasibility of the application of current methods for bioequivalence to assess the biosimilarity of biological products (follow-on biologics).

Bioequivalence assessment

Study design: In order to satisfy the regulatory requirements for the declaration of bioequivalence (i.e., that the 90% confidence interval around the ratio of geometric means should be, in %, between 80% and 125%), various study designs can be considered. As indicated in the Federal Register [Vol. 42 No. 5 Sec. 320.26(b) and Sec. 320.27(b), 1977], a bioequivalence study (single-dose or multi-dose) should be crossover in design, unless a parallel or other design is more appropriate for valid scientific reasons. Thus, in practice, a standard two-sequence, two-period (or 2x2) crossover design is often applied. Denote the test product and the reference product by T and R, respectively. A 2x2 crossover design can be expressed as (TR, RT), where TR is the first sequence of treatments and RT denotes the second sequence of treatments. Under the (TR, RT) design, qualified subjects who are randomly assigned to sequence 1 (TR) will receive the test product (T) first and then are provided the reference product (R) after a sufficient length of washout period. Similarly, subjects who are randomly assigned to sequence 2 (RT) will receive the reference product (R) first and then are given the test product (T) after a sufficient length of washout period.

One of the limitations of the standard 2x2 crossover design is that it does not provide independent estimates of intra-subject variabilities since each subject will receive the same treatment only once. In the interest of assessing intra-subject variabilities, the following higher order crossover designs for comparing two drug products are often considered:

- (1) Balaam's design – (TT, RR, RT, TR);
- (2) Two-sequence, three-period (2x3) dual design – (TRT, RTR);
- (3) Two-sequence, four-period (2x4) design – (TRTR, RTRT).

In some cases, an incomplete block design or an extra-reference 2x3 or 3x3 design such as (TRR, RTR) or (TRR, RTR, RRT) may be considered depending upon the study objectives of a bioequivalence investigation.

Statistical methods

As indicated earlier, average bioequivalence is claimed if the 90% confidence interval around the ratio of the geometric means between test and reference products is (in %), for a primary PK parameter, within the bioequivalence limits of 80% and 125%. Along this line, commonly employed statistical methods are the confidence interval approach and the method of interval hypotheses testing.

For the confidence interval approach, a 90% confidence interval for the ratio of geometric means of a primary pharmacokinetic response such as AUC or C_{max} is obtained under an analysis of variance model. We claim bioequivalence if the estimated 90% confidence interval is (in %) entirely within the bioequivalence limits of 80% and 125%.

For the method of interval hypotheses testing, we would reject the null hypothesis of bioinequivalence in favor of the alternative

of bioequivalence. In practice, the interval hypotheses are often decomposed into two sets of one-sided hypotheses. The first set of hypotheses is to verify that the average bioavailability of the test product is not too low, whereas the second set of hypotheses is to verify that average bioavailability of the test product is not too high. Schuirmann's two one-sided tests procedure is commonly employed for the interval hypotheses testing for average BE [1].

Remarks

Although the assessment of average bioequivalence for generic approval has been in practice for years, it has the following limitations: (1) it focuses only on population average; (2) it ignores the distribution of the metric; (3) it does not provide independent estimates of intra-subject variabilities and ignores the subject-by-formulation interaction. Many authors criticize that the assessment of average BE does not address the question of drug interchangeability and that it may penalize drug products with lower variability.

Drug interchangeability

As indicated by the regulatory agencies, a generic drug can be used as a substitution of the brand-name drug if their bioequivalence has been demonstrated. Current regulations do not indicate that two generic copies can be used interchangeably even if both of them are bioequivalent to the same brand-name drug. Bioequivalence between generic copies of a brand-name drug is not required. Thus, one of the controversial issues is whether these approved generic drug products can be used safely and interchangeably.

Theoretically, the potential difference between two generics could be about twice as large as what is allowed between a generic and the originator's formulation. The reason is that regulatory agencies worldwide require that the reference drug to which the comparison is made should be the originator's formulation.

Drug prescribability and drug switchability

When a new drug product is administered to a patient, distinction must be made between the conditions of prescribability and switchability. Drug prescribability is defined as the physician's choice for prescribing (or a pharmacist providing) a drug product for a patient who has not taken before the drug in any of its forms. The choice is between a brand-name drug product and a number of generic drug products that have been shown to be bioequivalent to the brand-name drug product.

Drug switchability, on the other hand, involves the switch from a drug product (either a brand-name or generic drug product) to an alternative formulation (again, either a generic or the brand-name drug product) within the same subject whose concentration of the drug product has been titrated to a steady, efficacious and safe level

Population/individual bioequivalence

It was suggested that the assessment of bioequivalence should take into consideration of both prescribability and switchability [3]. Extensive discussions in the 1990s and early 2000s suggested that population bioequivalence (PBE) and individual bioequivalence (IBE) be considered for testing prescribability and switchability, respectively. However, the current position of FDA has been that declaration of average BE criterion ensures the safety and efficacy of the generic product and, therefore, its prescribability. FDA does not consider at present IBE as an applicable approach. IBE was abandoned mainly for

methodological reasons. Nevertheless, the underlying background and principles are valuable and important and will be briefly discussed here.

To address drug prescribability, FDA proposed [3] the following aggregated, scaled, moment-based, one-sided criterion for population bioequivalence (PBC):

$$PBC = \frac{(\mu_T - \mu_R)^2 + (\sigma_{TT}^2 - \sigma_{TR}^2)}{\max(\sigma_{TR}^2, \sigma_{T0}^2)} \leq \theta_p,$$

where μ_T and μ_R are the means of the test and reference drug products, respectively, σ_{TT}^2 and σ_{TR}^2 are the total variances of the test and reference drug products, respectively, σ_{T0}^2 is a regulatory constant that can be adjusted to control the probability of passing PBE, and θ_p is the bioequivalence limit for PBE. The numerator on the left-hand side of the criterion is the sum of the squared difference of the population averages and the difference between the total variances of the test and reference drug products; it measures the similarity of the marginal population distributions between the test and reference drug products. The denominator on the left-hand side of the criterion is a scaling factor that depends upon the variability of the drug class of the reference drug product. The FDA guidance suggests that θ_p be chosen as:

$$\theta_p = \frac{(\log 1.25)^2 + \varepsilon_p}{\sigma_{T0}^2},$$

where ε_p is guided by the consideration of the variability term $\sigma_{TT}^2 - \sigma_{TR}^2$ being added to the ABE criterion. As suggested by the FDA guidance, it may be appropriate that ε_p be chosen as 0.02. For the determination of σ_{T0}^2 , the guidance suggests a value of 0.04.

Similarly, to address drug switchability, FDA recommended the following aggregated, scaled, moment-based, one-sided criterion:

$$IBC = \frac{(\mu_T - \mu_R)^2 + (\sigma_D^2 + (\sigma_{WT}^2 - \sigma_{WR}^2))}{\max(\sigma_{WR}^2, \sigma_{W0}^2)} \leq \theta_I,$$

where σ_{WT}^2 and σ_{WR}^2 are the within-subject variances of the test and reference drug products, respectively, σ_D^2 is the variance component due to subject-by-formulation interaction, σ_{W0}^2 is a regulatory constant that can be adjusted to control the probability of passing IBE, and θ_I is the bioequivalence limit for IBE. The FDA guidance suggested that θ_I be chosen as:

$$\varepsilon_I = 0.05, \theta_I = \frac{(\log 1.25)^2 + \varepsilon_I}{\sigma_{W0}^2},$$

where ε_I is the variance allowance factor which can be adjusted for sample size control. The FDA guidance recommended $\varepsilon_I=0.05$. For the determination of σ_{W0}^2 , the guidance suggests a value of 0.04.

Some features of the aggregate model for individual BE were criticized. Comparison of the deviations between the means and between the within-subject variances of the two products could result both in substantial increase and decrease of the estimated IBC. Thereby not only benefits but also penalties could be incurred [4]. Moreover, the estimated variances, including the variance component σ_D^2 , are very unreliable with the typical sample sizes of BE studies.

While the model for IBE is not used at present, it gives rise to applications which are still widely used. These will be considered later.

It is important to observe that for prescribability (and population BE) total variation is important which includes both between- and within-subject variations. In contrast, for switchability (and individual BE), within-subject variation is relevant. This can be conveniently evaluated

in crossover studies. BE declared from this kind of investigation gives confidence that the two drug products can be indeed interchanged within patients and, therefore, that the Fundamental Bioequivalence Assumption, discussed below, can be readily applied.

To achieve the objective of interchangeability among bioequivalent pharmaceutical products, the criteria for assessment of bioequivalence must possess certain important properties, [5,6] outlined the desirable characteristics of bioequivalence criteria proposed by FDA (Table 1). In addition, to address the issues of intra-subject variability and subject-by-formulation interaction and to ensure drug switchability, valid statistical procedures, both estimation and hypothesis testing, should be developed from the criteria to control the consumer's risk at the pre-specified nominal level (e.g., 5%). Furthermore, the statistical methods developed from the criteria should be able to provide sample size determination; to take into consideration the nuisance design parameters, such as period or sequence effects; and to develop user-friendly computer software. The most critical characteristics for the proposed criteria will be their interpretation to scientists and clinicians and the cost of conducting bioequivalence studies to provide inference for the criteria.

Remarks

In the interest of assessing individual bioequivalence, FDA recommended that a replicated design be considered for obtaining independent estimates of intra-subject variabilities and variability due to subject-by-drug product interaction. A commonly considered replicate crossover scheme is a 2x4 crossover design, (TRTR, RTRT).

Note that FDA has abandoned the regulatory implementation of individual bioequivalence. Nevertheless, it is an important concept which highlights and advances the significance of switchability [7].

Controversial statistical issues

In this section, we shall focus on controversial statistical issues related to the Fundamental Bioequivalence Assumption, one size-fits-all criterion, the evaluation of BE for highly variable drugs, and issues related to the log-transformation of PK data prior to analysis. These controversial statistical issues are briefly described.

Fundamental bioequivalence assumption

As indicated by Chow and Liu [2], bioequivalence studies are performed under the so-called Fundamental Bioequivalence Assumption. It states that:

Comparison of both averages and variances
Assurance of switchability
Encouragement or reward of pharmaceutical companies to manufacture a better formulation
Control of type I error rate (consumer's risk) at 5%
Allowance for determination of sample size
Admission of the possibility of sequence and period effects as well as missing values
User-friendly software application for statistical methods
Provision of easy interpretation for scientists and clinicians
Minimization of increased cost for conducting bioequivalence studies

Source: [6].

Table 1: Desirable Features of Bioequivalence Criteria.

If two drug products are shown to be bioequivalent, it is assumed that they will reach the same therapeutic effect or they are therapeutically equivalent and hence can be used interchangeably.

It is implied in this assumption that the two drug products are also pharmaceutically equivalent, i.e., that the test and reference products are comparable dosage forms and contain identical amounts of the same medicinal ingredient [8,9]. To protect the exclusivity of a brand-name drug product, the sponsors of the innovator drug products will usually make every attempt to prevent generic drug products from being approved by the regulatory agencies such as the FDA. One of the strategies is to challenge the Fundamental Bioequivalence Assumption by filing a citizens' petition with scientific/clinical justification against the Fundamental Bioequivalence Assumption. Upon the receipt of a citizens' petition, the FDA has legal obligation to respond within 180 days. It, however, should be noted that the FDA will not suspend the review/approval process of a generic submission even while a citizens' petition is under review within the FDA.

One of the controversial statistical issues is that bioequivalence in drug absorption may possibly not imply therapeutic equivalence and therapeutic equivalence may not guarantee bioequivalence either. The assessment of average bioequivalence for generic approval has been criticized that it is based on legal/political deliberations rather than scientific considerations. In the past several decades, many sponsors/researchers have made an attempt to challenge this assumption with no success.

In practice, the Fundamental Bioequivalence Assumption is also applied to drug products with local action such as nasal spray products via the assessment of *in vitro* bioequivalence testing. In either *in vivo* or *in vitro* bioequivalence testing, the verification of the Fundamental Bioequivalence Assumption is often difficult, if not impossible, without the conduct of clinical trials.

The Fundamental Bioequivalence Assumption can thus be generally applied. This was affirmed in a public letter of an FDA Associate Commissioner [10,11]. Bioequivalence of drug products generally indicates indeed their therapeutic equivalence.

One size-fits-all criterion

For the assessment of average bioequivalence, FDA adopted a one size-fits-all criterion. That is, as noted earlier, a test drug product is said to be bioequivalent to a reference drug product if the estimated 90% confidence interval for the ratio of geometric means of the primary PK parameters (e.g., AUC and C_{max}) is (in %) totally within the bioequivalence limits of 80% to 125%. The one size-fits-all criterion does not take into consideration the therapeutic window and intra-subject variability of a drug which have been identified to have non-negligible impact on the safety and efficacy of generic drug products as compared to the innovative drug products.

In the past several decades, this one size-fits-all criterion has been challenged and criticized by researchers [12,13]. It was suggested that flexible criteria in terms of safety (upper bioequivalence limit) and efficacy (lower bioequivalence limit) should be developed based on the characteristics of the drug, its therapeutic window (TW) and intra-subject variability (ISV). See also Table 2.

The approach of one size-fits-all has begun to dissipate in recent years. For instance, in some jurisdictions such as Europe, Canada, and recently also in the United States, narrower BE limits have been proposed for drugs with narrow therapeutic windows [14-17].

However, FDA has maintained its usual requirement for these drugs with BE limits to be between 80% and 125% even though it has recently indicated a reconsideration of the issue [17].

On the other hand, for orally administered drugs with high within-subject variability and wide therapeutic window (Class D, highly variable drugs, see Table 2), the regulatory expectations have become, in some cases, more relaxed. They will be discussed in the next section.

The approach of one size-fits-all is useful for assessing BE involving most drugs. However, special criteria should be applied for drugs having narrow therapeutic window or high intrasubject variability.

Highly variable drugs

The approach of one-size fits all has been relaxed in recent years by various regulatory authorities for drugs which exhibit high variations, i.e. large fluctuations, within individuals. It has been very difficult to determine BE for this class of drugs unless unethically large numbers of volunteers were included in the investigations.

A drug is considered by international consensus to be highly variable if its within-subject coefficient of variation exceeds 30%. For the evaluation of the BE of highly-variable drug products, the approach of scaled average BE (SABE) was proposed [19]. It is possible to consider SABE as a special case of individual BE by assuming that the two drug products have the same within-subject variation, and that there is no subject-by-formulation interaction. In this case, the IBE criterion becomes (after taking square root):

$$(\mu_T - \mu_R)/\sigma_{WR} \leq \theta_s$$

Here θ_s is the BE limit for SABE which is a regulatory constant the value of which should be set by the authorities. An alternative form of the regulatory constant is:

$$\sigma_o = \ln(1.25)/\theta_s$$

The procedure could be used when the within-subject variation is high ($\sigma_{WR}^2 > \sigma_{W0}^2$); the value of σ_{W0}^2 is generally set to correspond to a coefficient of variation of 30%. If the intraindividual variation does not exceed 30% ($\sigma_{WR}^2 \leq \sigma_{W0}^2$) then, as usual, the two one-sided tests procedure is applied.

A working group of FDA scientists has adopted the approach of SABE for highly variable drugs. The suggested procedure was described in a publication [20]. It was suggested that the within-subject variation of the reference product be estimated, at least, from partially replicating 3-period studies (RRT/RTR/TRR). The regulatory constant was suggested to have the value of $\sigma_o = 0.25$ or $\theta_s = 0.893$. It is expected that the 90% confidence interval around the SABE criterion be within the BE limits; the confidence interval is calculated by a linearizing approximation of the SABE criterion [21]. In addition, it is proposed that a second criterion also be satisfied, namely that the point estimate of the GMR be between 0.80 and 1.25.

Class	TW	ISV	Example
A	Narrow	High	Cyclosporine
B	Narrow	Low	Theophylline
C	Wide	Low to moderate	Most drugs
D	Wide	High	Chlorpromazine, topical corticosteroids

TW, therapeutic window; ISV, intra-subject variability. Source: [5]; [18].

Table 2: Classification of Drugs.

FDA actually entertains submissions based on the criteria described in the 'informal' publication of [20].

The regulatory criteria recently published by EMA for determining the BE of highly-variable drugs [15] are related to but still different from those of FDA. EMA applies a model rearranging the SABE criterion [20]:

$$-\theta_s * \sigma_{WR} \leq (\mu_T - \mu_R) \leq \theta_s * \sigma_{WR}$$

This is like the criterion for average BE but with expanding limits (ABEL). Consequently, the 90% confidence limit can be calculated by the simple application of the two one-sided tests procedure.

The value of the regulatory constant of EMA, $\sigma_0 = 0.294$ or $\theta_s = 0.76$, differs from that of FDA. But similarly to FDA, EMA also requires the second criterion of constraining the point estimate of the ratio of geometric means (GMR) to range between 0.80 and 1.25. In addition, EMA imposes a limit on the application of ABEL. At a variation exceeding CV = 50%, again unscaled average BE must be used [15].

Properties of the two regulatory procedures were demonstrated and deviations in some of the outcomes were noted [19,23,24].

Determination of BE for highly variable drugs by scaled average BE, or by its variant of ABEL, is advantageous. Differing values were suggested by FDA and EMA for one of the regulatory constants. Their comparison indicates that the value recommended by EMA is preferable [24].

Issues related to log-transformation

In the past several decades, bioequivalence could be assessed based on either raw data or log-transformed data depending upon which model followed the assumption of normal distribution. This raised a controversial statistical issue regarding which model should be used for the assessment of bioequivalence. The sponsors could choose the model which would serve their purposes (e.g., the demonstration of bioequivalence). In some cases, the raw data model could reach a different conclusion regarding bioequivalence than the log-transformation model. This controversial statistical issue was discussed a great deal until a general understanding was reached in the early 1990s on the use of log-transformed parameters.

The 2001 FDA guidance provides a rationale for the use of logarithmic transformation of exposure measures. The guidance emphasizes that the limited sample size in a typical BE study precludes a reliable determination of the distribution of the data. For this reason, the guidance does not encourage the sponsors to test for the normality of error distribution after log-transformation, nor to use the normality of the error distribution as a reason for carrying out the statistical analysis on the original scale.

The use of the logarithmic transformation of pharmacokinetic parameters was questioned in the statistical literature [25-28]. It was stated that the log-transformed $AUC_{0-\infty}$ and C_{max} do not generally follow a normal distribution even when either the plasma concentrations or log-plasma concentrations are normally distributed [26]. It was suggested that performing a routine log-transformation of data and then applying normal, theory-based methods is not appropriate [28]. It was suggested that normal probability plots for the studentized inter-subject and intra-subject residuals be examined and that the Shapiro-Wilk method be applied to test for normality of the inter-subject and intra-subject variabilities. However, the sample size of a

typical bioequivalence study is generally too small to allow an adequate large-sample normal approximation and to enable clear discrimination between the normal and log-normal distributions of the estimated PK parameters.

In addition, the use of logarithmic pharmacokinetic (and generally kinetic) parameters has a strong basis in their multiplicative, rather than additive sense. We typically think of doubling or halving a dose or concentration and not adding or taking away some units. Similarly, in tables of kinetic parameters, whether they are rate constants (including half-lives), equilibrium constants or many other kinetic measures, we compare their orders of magnitudes and ask if one is, say, 10 times higher or lower than the other. Consequently, analyses involving kinetic parameters generally apply a multiplicative and not an additive model. An implementation of this sense is that the parameters are analyzed following their logarithmic transformation.

In any case, the protocol should clearly state any procedure which would test for the validity of either the raw data model or the log-transformed data model and then select one of them for submission. It is advisable to consult the regulatory agency before contemplating a calculational procedure which differs from that recommended in a guidance.

The choice between using raw data or logarithmic data can not be generally determined in a given BE study. Nevertheless, application of the logarithmic transformation is usually preferable and is expected by regulatory authorities.

Frequently asked questions

What if we pass AUC but fail C_{max} ?: Based on log-transformed data, FDA requires that both AUC and C_{max} meet the bioequivalence limits of 80% to 125% in order to establish average bioequivalence. In practice, however, it is not uncommon to pass AUC (the extent of absorption) but fail C_{max} (the rate of absorption). In this case, average bioequivalence cannot be claimed according to the FDA guidance.

If we pass AUC but fail C_{max} , [29] suggested considering C_{max}/AUC as an alternative bioequivalence measure for the rate of absorption. However, C_{max}/AUC is not currently selected as a required pharmacokinetic response for the approval of generic drug products by regulatory authorities. The condition of passing the regulatory requirement for AUC and not for C_{max} is less likely to arise in Canada where only the point estimate of the ratio of geometric means of C_{max} but not the 90% confidence interval must be (in %) between 80% and 125% [8,9].

It is possible that we would pass the regulatory requirement for C_{max} but not for AUC. It was suggested that we could look at partial AUC as an additional measure of bioequivalence [30,31].

What if we fail by a small margin? : In practice, it is possible that we fail BE testing for either AUC or C_{max} by only a small margin. For example, suppose that the estimated 90% confidence interval for AUC is from 79.5% to 121.3%, which is slightly outside the lower limit of the regulatory range of 80.0% to 125.0%. In this case, the FDA's position is very clear: A rule is a rule and you fail. In regulatory reviews and approvals, the FDA is very strict about this rule as described in the 2003 FDA guidance.

However, a sponsor may perform either an outlier detection analysis or a sensitivity analysis in order to resolve the issue. If a subject is found to be an outlier statistically, the data may be excluded from

the analysis but only with appropriate clinical justification. Once the identified outlier is excluded from the analysis, the recalculated 90% confidence interval could be totally within the bioequivalence limits of 80% to 125%, and the sponsor may present an argument for claiming bioequivalence.

Major regulatory agencies have recently encouraged additional design features which permit the later addition of subjects. Notably, they include group sequential extensions of the usual bioequivalence testing procedure [32,33]. The results of a study would first be evaluated by the customary procedures. However, in order to maintain the overall Type I error, the results would be assessed with adjusted significance levels which would yield confidence intervals higher than 90% [34]. If the analysis indicates that the calculated 90% confidence intervals of the PK parameters are moderately outside the regulatory BE interval of 80% to 125% then a second group of subjects could be investigated. A combined analysis of the two groups could be performed; these would apply a modified structure of the statistical computations and, again, adjusted significance levels.

Health Canada accepts also a simple add-on of at least 12 subjects [8,9]. The structure of the statistical analysis should be modified and the level of significance should be 0.025 instead of 0.05. In all cases, the intention of applying either the group sequential or add-on design as well as the details of the procedure should be specified in the protocol of the study.

Can we still assess bioequivalence if there is a significant sequence effect?: As indicated in [2], under a standard two-sequence, two-period (2x2) crossover design, a significant sequence effect is an indication of the possible (1) failure of randomization, (2) a true sequence effect, (3) a true carry-over effect, and/or (4) a true formulation-by-period effect. Under the standard 2x2 crossover design, the sequence effect is confounded with the carryover effect. Therefore, if a significant sequence effect is found, the treatment effect and its corresponding 90% confidence interval cannot be estimated unbiasedly due to the possibly unequal carryover effects. However, in the 2001 FDA guidance, the following list of conditions is provided to rule out the possibility of unequal carryover effects:

1. It is a single-dose study;
2. The drug is not an endogenous entity;
3. More than an adequate washout period has been allowed between periods of the study, and in the subsequent periods the predose biological matrix samples do not exhibit a detectable drug level in any of the subjects.
4. The study meets all scientific criteria (e.g., it is based on an acceptable study protocol and it contains a validated assay methodology).

The 2001 FDA guidance also recommends that sponsors conduct a bioequivalence study with parallel designs if unequal carryover effects become an issue.

Power and sample size calculations based on raw data model and log-transformed model are different

The calculations of the statistical power and of the sample size are different when they are based on the raw data model and on the log-transformed model. Under differing models, means, standard deviations, and coefficients of variation are also different. As

mentioned before, for the assessment of bioequivalence, all regulatory authorities including the FDA, EMA, WHO, and Japan require that log-transformation of the parameters AUC_{0-t} , $AUC_{0-\infty}$ and C_{max} be performed before the analysis and evaluation of bioequivalence. As a result, one should use differences between logarithmic means and the corresponding standard deviations or the coefficients of variation for the power analysis and sample size calculation based on the method for the log-transformed model (see, e.g., Chapter 5 of [2]).

Note that sponsors should make the decision in the protocol as to which model (the raw data model or the log-transformed data model) will be used for bioequivalence assessment. Once the model is chosen, appropriate formulas can be used to determine the sample size. Fishing around for obtaining the smallest sample size is not a good clinical practice.

Multiplicity and transitivity

The [35] guidance for general considerations requires that for AUC_{0-t} , $AUC_{0-\infty}$ and C_{max} , the following information be provided:

1. geometric means
2. arithmetic means
3. ratio of means
4. 90% confidence interval.

In addition, as already noted, the 2003 FDA guidance recommends that logarithmic transformation be provided for each measure of AUC_{0-t} , $AUC_{0-\infty}$ and C_{max} , and that, for the demonstration of average bioequivalence, each of the 90% confidence intervals for ratio of the geometric means of the two formulations must fall within the bioequivalence limits (in %) of 80% to 125%. It follows that according to the intersection-union principle [36], the type I error rate of average bioequivalence is still controlled under the nominal level of 5%. Therefore, there is no need for adjustment due to multiple pharmacokinetic measures.

Another issue involves the multiplicity of generic products of a drug. The bioequivalence of each generic formulation is determined against the reference, generally brand-name product. It is not obvious to what extent the generics could be equivalent with each other. This is particularly important when a patient is switched from one generic product to another. Anderson and Hauck [37] examined the transitivity of bioequivalence, i.e., to what extent is there a potential of drift about the declaration of bioequivalence when a number of generics are tested against an innovator's product. With two or three generic formulations, the confidence of transitive bioequivalence is fairly high. With six generic products, this confidence is low.

Assessment of biosimilarity of follow-on biologics

As indicated in the previous sections, although there are still some controversial statistical issues and frequently asked questions in regulatory submissions when assessing bioequivalence, statistical methods including criteria, study designs, and testing procedures for the assessment of average BE for drug products are well established. Thus, it is of particular interest to pharmaceutical/clinical scientists whether similar concepts can be applied directly to assessing biosimilarity of biologic drug products, i.e. of follow-on biologics. The applicability/feasibility is questionable due to some fundamental differences between small molecule drug products and biologic products.

Unlike small molecule drug products, biologic products are made by living cells which have complicated, heterogeneous structures. The mixtures of related molecules make it difficult to characterize the mechanism of actions. The biologic products are known to be variable which are usually sensitive to environmental conditions such as temperature and light. In addition, biologic products have the issue of immunogenicity. Note that the one size-fits-all criterion of (average) bioequivalence assessment does not take into consideration the heterogeneity and variability of drug products. On the other hand, the biologic products are very sensitive to variability – a small change in bias/variation could translate to a major change in clinical response. Thus, the well-established statistical methods for the assessment of bioequivalence for small molecule drug products may not be feasible and hence may not be applicable to the assessment of biosimilarity of biologic drug products (or follow-on biologics). For the approval of biosimilars in the EU community, EMA issued a guideline describing general principles for the approval of similar biological medicinal products, or biosimilars [38,39]. The guideline is accompanied by several product-specific guidances including those for human recombinant products containing erythropoietin, human growth hormone, granulocyte-colony stimulating factor, insulin, IFN-alpha and low molecular weight heparin.

On the other hand, for the approval of follow-on biologics in the United States, it depends whether the biologic product is approved under the Food, Drug, and Cosmetic Act (US FD&C) or licensed under the Public Health Service Act (US PHS). Some proteins are licensed under the PHS Act while others are approved under the FD&C Act. For products approved under an NDA (US FD&C Act), generic versions of the products can be approved under an ANDA, e.g., under Section 505(b)(2) of the FD&C Act. For products that are licensed under a BLA (US PHS Act), there exists no abbreviated BLA. For the assessment of similarity of follow-on biologics, the FDA would consider the totality of evidence including the following factors: (1) the robustness of the manufacturing process, (2) the degree to which structural similarity could be assessed, (3) the extent to which the mechanism of action is understood, (4) the existence of valid, mechanistically related pharmacodynamic assays, (5) comparative pharmacokinetics, (6) comparative immunogenicity, (7) the amount of clinical data available, and (8) the extent of experience with the original product [40,41]. In practice, there is strong industrial interest and desire for the regulatory agencies to develop review standards and for an approval process for biosimilars rather than a mere ad hoc case-by-case review of individual biosimilar applications. Under this consideration, FDA indicated that the following guidances are currently under development: (1) a guidance for industry on scientific considerations to demonstrate the safety and effectiveness of follow-on protein products, and (2) a guidance for industry on CMC issues for follow-on protein products.

As more innovative biologic products are going off-patent protection in the next few years, the FDA hosted a public hearing on Approval Pathway for Biosimilar and Interchangeable Biological Products in Silver Spring, Maryland between November 2-3, 2010. At the public hearing several scientific factors for assessing follow-on biologics were discussed. These scientific factors/issues included (1) how similar is considered similar? (2) the issue of drug interchangeability in terms of the concept of alternating and switching, and (3) quality attributes and comparability between manufacturing processes. Based on the fact that there are some fundamental differences between small molecule drug products and biologic products, research in this area is urgently needed in order to address the scientific factors/issues discussed at the FDA public hearing [42,43].

References

1. Schuirmann DJ (1987) 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.
2. Chow SC, Liu JP (2008) *Design and Analysis of Bioavailability and Bioequivalence Studies*. 3rd Edition, Chapman and Hall/CRC Press, Taylor & Francis, New York, New York.
3. FDA (2001) *Guidance on Statistical Approaches to Establishing Bioequivalence*. Center for Drug Evaluation and Research. U.S. Food and Drug Administration, Rockville, Maryland.
4. Endrenyi L, Hao Y (1998) Asymmetry of the mean-variability tradeoff raises questions about the model in investigations of individual bioequivalence. *Int J Clin Pharmacol Ther* 36: 1-8.
5. Chen ML (1995) Individual bioequivalence. Invited presentation at International Workshop: Statistical and Regulatory Issues on the Assessment of Bioequivalence. Dusseldorf, Germany 19-20.
6. Chen ML (1997) Individual bioequivalence – a regulatory update. *J Biopharm Stat* 7: 5-11.
7. Endrenyi L, Amidon GL, Midha KK, Skelly JP (1998) Individual bioequivalence: attractive in principle, difficult in practice. *Pharm Res* 15: 1321-1325.
8. Health Canada (1992) *Guidance for Industry. Conduct and Analysis of Bioavailability and Bioequivalence Studies. Part A: Oral Dosage Formulations Used for Systemic Effects*. Health Canada, Ottawa, ON.
9. Health Canada (2010) *Draft Guidance Document. Comparative Bioavailability Standards: Formulations Used for Systemic Effects*. Health Canada, Ottawa, ON.
10. Nightingale SL (FDA Associate Commissioner for Health Affairs) (1998) Therapeutic equivalence of generic drugs—letter to health practitioners.
11. <http://www.fda.gov/Drugs/DevelopmentApprovalProcess/HowDrugsareDevelopedandApproved/ApprovalApplications/AbbreviatedNewDrugApplicationANDAGenerics/ucm073182.htm>.
12. Benet L (1993) Bioavailability and bioequivalence: definitions and difficulties in acceptance criteria. In: *Bio-International 2: Bioavailability, Bioequivalence and Pharmacokinetics* (Midha KK and Blume HH, eds.), Medpharm, Stuttgart, pp. 27-35.
13. Blume HH (2009) "One size fits all" in bioavailability and bioequivalence testing. *Int J Clin Pharmacol Ther* 47: 419-420.
14. Health Canada (2006) *Guidance for Industry. Bioequivalence Requirements: Critical Dose Drugs*. Health Canada, Ottawa, ON.
15. EMA (2010) *Guideline on the Investigation of Bioequivalence*. Committee for Medicinal Products for Human Use (CHMP), London, UK.
16. FDA (2011) *FDA Briefing Information for the July 26, 2011 Meeting of the Pharmaceutical Science and Clinical Pharmacology Advisory Committee: Bioequivalence (BE) and Quality Standards for Narrow Therapeutic Index (NTI) Drug Products*.
17. <http://www.fda.gov/downloads/AdvisoryCommittees/CommitteesMeetingMaterials/Drugs/AdvisoryCommitteeForPharmaceuticalScienceandClinicalPharmacology/UCM263465.pdf>
18. Patnaik RN, Lesko LJ, Chen ML, Williams RL (1997) Individual bioequivalence – new concepts in the statistical assessment of bioequivalence metrics. *Clin Pharmacokinet* 33: 1-6.
19. Tothfalusi L, Endrenyi L, Garcia Arieta A (2009) Evaluation of bioequivalence for highly-variable drugs with scaled average bioequivalence. *Clin Pharmacokinet* 48: 725-743.
20. Haidar SH, Davit B, Chen ML, Conner D, Lee LM, et al. (2008a) Bioequivalence approaches for highly variable drugs and drug products. *Pharm Res* 25: 237-241.
21. Hyslop T, Hsuan F, Holder DJ (2000) A small sample confidence interval approach to assess individual bioequivalence. *Stat Med* 19: 2885-2897.

22. Boddy AW, Snikeris FC, Kringle RO, Wei GCG, Opperman JA, et al. (1995) An approach for widening the bioequivalence limits in the case of highly variable drugs. *Pharm Res* 12: 1865-1868.
23. Haidar SH, Makhoulouf F, Schuirman DJ, Hyslop T, Davit B, et al. (2008b) Evaluation of a scaling approach for the bioequivalence of highly variable drugs. *AAPS J* 10: 450-454.
24. Endrenyi L, Tothfalusi L (2009) Regulatory conditions for the determination of bioequivalence of highly variable drugs. *J Pharm Pharmaceut Sci* 12: 138-149.
25. Liu JP, Weng CS (1992) Estimation of direct formulation effect under log-normal distribution in bioavailability/bioequivalence studies. *Stat Med* 11: 881-896.
26. Liu JP, Weng CS (1994) Estimation of log-transformation in assessing bioequivalence. *Communications in Statistics – Theory and Methods* 23: 421-434.
27. Liu JP, Weng CS (1995) Bias of two one-sided tests procedures in assessment of bioequivalence. *Stat Med* 14: 853-861.
28. Patel HI (1994) Dose-response in pharmacokinetics. *Communications in Statistics – Theory and Methods* 23: 451-465.
29. Endrenyi L, Fritsch S, Yan W (1991) $(C_{max})/AUC$ is a clearer measure than (C_{max}) for absorption rates in investigations of bioequivalence. *Int J Clin Pharmacol Ther Toxicol* 29: 394-399.
30. Chen ML, Lesko L, Williams RL (2001) Measures of exposure versus measures of rate and extent of absorption. *Clin Pharmacokinet* 40: 565-572.
31. Chen ML, Davit B, Lionberger R, Wahba Z, Ahn HY, et al. (2011) Using partial area for evaluation of bioavailability and bioequivalence. *Pharm Res* 28: 1939-1947.
32. Gould AL (1995) Group sequential extensions of a standard bioequivalence procedure. *J Pharmacokinet Biopharm* 23: 57-86.
33. Potvin D, DiLiberti CE, Hauck WW, Parr AF, Schuirman DJ, et al. (2008) Sequential design approaches for bioequivalence studies with crossover designs. *Pharm Stat* 7: 245-262.
34. Pocock SJ (1977) Group sequential methods in the design and analysis of clinical data. *Biometrika* 64: 191-199.
35. FDA (2003) Guidance on Bioavailability and Bioequivalence Studies for Orally Administered Drug Products – General Consideration. Center for Drug Evaluation and Research. U.S. Food and Drug Administration, Rockville, Maryland.
36. Berger RL (1982) Multiparametric hypothesis testing and acceptance sampling. *Technometrics* 24: 295-300.
37. Anderson S, Hauck WW (1996) The transitivity of bioequivalence testing: potential for drift. *Int J Clin Pharmacol Ther* 9: 369-374.
38. EMA (2005) Guideline on Similar Biological Medicinal Products. Committee for Medicinal Products for Human Use (CHMP), London, UK.
39. EMA (2008) Questions and Answers on Biosimilar Medicines (Similar Biological Medicinal Products). Committee for Medicinal Products for Human Use (CHMP), London, UK.
40. Woodcock J (2005) FDA introductory comments: clinical studies design and evaluation issues. *Clin Trials* 2: 273-275.
41. Kozlowski S, Woodcock J, Midthun K, Behrman Sherman R (2011) Developing the nation's biosimilars program. *N Engl J Med* 365: 385-388.
42. Chow SC, Endrenyi L, Lachenbruch PA, Yang LY, Chi E (2011) Scientific factors for assessing biosimilarity and drug interchangeability of follow-on biologics. *Biosimilars* 1: 13-26.
43. Chow SC (1999) Individual bioequivalence – a review of the FDA draft guidance. *Drug Information Journal* 33: 435-444.