Published in final edited form as:
Wiley Interdiscip Rev Comput Stat. 2014 ; 6(4): 304–312. doi:10.1002/wics.1310.

Bioavailability and Bioequivalence in Drug Development

Shein-Chung Chow, PhD
Department of Biostatistics and Bioinformatics, Duke University School of Medicine, Durham, North Carolina, USA

Abstract

Bioavailability is referred to as the extent and rate to which the active drug ingredient or active moiety from the drug product is absorbed and becomes available at the site of drug action. The relative bioavailability in terms of the rate and extent of drug absorption is considered predictive of clinical outcomes. In 1984, the United States Food and Drug Administration (FDA) was authorized to approve generic drug products under the *Drug Price Competition and Patent Term Restoration Act* based on evidence of average bioequivalence in drug absorption through the conduct of bioavailability and bioequivalence studies. This article provides an overview (from an American point of view) of definition of bioavailability and bioequivalence, *Fundamental Bioequivalence Assumption*, regulatory requirements, and process for bioequivalence assessment of generic drug products. Basic considerations including criteria, study design, power analysis for sample size determination, and the conduct of bioequivalence trial, and statistical methods are provided. Practical issues such as one size-fits-all criterion, drug interchangeability and scaled average criteria for assessment of highly variable drug products are also discussed.

Keywords

Fundamental Bioequivalence Assumption; Drug interchangeability; Highly variable drugs; Scaled average bioequivalence (SABE) criterion

1. BACKGROUND

As indicated in Chapter 21 CFR (Codes of *Federal Regulations*) Part 320.1, *bioavailability* of a drug is defined as the extent and rate to which the active drug ingredient or active moiety from the drug product is absorbed and becomes available at the site of drug action. The extent and rate of drug absorption are usually measured by the area under the blood or plasma concentration-time curve (AUC) and the maximum concentration (C_{max}), respectively. For drug products that are not intended to be absorbed into bloodstream, bioavailability may be assessed by measurements intended to reflect the rate and extent to which the active ingredient or active moiety is absorbed and becomes available at the site of action. A *comparative bioavailability* study refers to the comparison of bioavailabilities of different formulations of the same drug or different drug products. As indicated in Chow and Liu (2008), the definition of bioavailability has evolved over time with different meanings

Corresponding author: Shein-Chung Chow, PhD, Professor, Department of Biostatistics and Bioinformatics, Duke University School of Medicine, 2424 Erwin Road, Hock Suite 1102, Room 11068, Durham, NC 27710, sheinchung.chow@duke.edu.

by different individuals and organizations [1]. For example, differences are evident in the definitions by Academy of Pharmaceutical Sciences in 1972, the Office of Technology Assessment (OTA) of the Congress of the United States in 1974, and the 1984 *Drug Price Competition and Patent Restoration Act* which is amendments to the *Food, Drug, and Cosmetic Act*. For more discussion regarding the definition of bioavailability, see [2-4].

When two formulations of the same drug or two drug products are claimed *bioequivalent*, it is assumed that they will provide the same therapeutic effect or that they are therapeutically equivalent. In this case, most people interpret that they can be used interchangeably. Two drug products are considered *pharmaceutical equivalents* if they contain identical amounts of the same active ingredient. Two drugs are identified as *pharmaceutical alternatives* to each other if both contain an identical therapeutic moiety, but not necessarily in the same amount or dosage form or as the same salt or ester. Two drug products are said to be bioequivalent if they are pharmaceutical equivalents (i.e., similar dosage forms made, perhaps, by different manufacturers) or pharmaceutical alternatives (i.e., different dosage forms) and if their rates and extents of absorption do not show a significant difference to which the active ingredient or active moiety in pharmaceutical equivalents or pharmaceutical alternatives become available at the site of action when administered at the same molar dose under similar conditions in an appropriately designed study.

When an innovative (or brand-name) drug product is going off patent, pharmaceutical or generic companies may file an abbreviated new drug application (ANDA) for generic approval. Generic drug products are defined as drug products that are *identical* to an innovative (brand-name) drug which is the subject of an approved NDA with regard to active ingredient(s), route of administration, dosage form, strength, and conditions of use. Since ANDA submissions for generic applications do not require lengthy clinical evaluation of the generic drugs under investigation (see Table 1), the price of generics are usually much lower than that of the originals. On average, it is about *20*% of the price of the brand-name originals. In 1984, the United States Food and Drug Administration (FDA) was authorized to approve generic drug products under the *Drug Price Competition and Patent Term Restoration Act*. The purpose is to make less expensive, safe, and equally efficacious generics available to general public after the expiration of patent protection of expensive brand-name drugs. For approval of generic drug products, the FDA requires that evidence of average bioequivalence in drug absorption be provided through the conduct of bioavailability and bioequivalence studies. Bioequivalence assessment is considered as a surrogate for clinical evaluation of the therapeutic equivalence of drug products.

A typical process for bioequivalence assessment is to conduct a bioequivalence study with male healthy volunteers under the assumption that bioequivalence (relative bioavailability) of the drug product under investigation is predictive of clinical outcomes (i.e., safety and efficacy) of the drug product in clinical trials. A bioequivalence study is often conducted utilizing a crossover design that allows comparison within individual subjects, i.e., each subject is at his/her own control. Based on pharmacokinetic (PK) data collected, bioequivalence can then be assessed using valid statistical methods according to some pre-specified regulatory criteria for bioequivalence. As indicated by the FDA, an approved generic drug product can be used as a substitute for the brand-name drug.

In what follows, the assumption that relative bioavailability is predictive of clinical outcomes, criteria for bioequivalence, basic considerations for conduct of a bioequivalence study such as study design, sample size, and statistical methods, current issues including one-size-fits-all criterion and drug interchangeability, and recent development such as bioequivalence assessment for highly variable drugs are discussed.

2. FUNDAMENTAL BIOEQUIVALENCE ASSUMPTION

As indicated in Chow and Liu [1], bioequivalence assessment for generics approval can only be done under the so-called *Fundamental Bioequivalence Assumption*, which states that "If two drug products are shown to be bioequivalent, it is assumed that they will generally reach the same therapeutic effect or they are therapeutically equivalent." Under the Fundamental Bioequivalence Assumption, one of the controversial issues is that bioequivalence may not necessarily imply therapeutic equivalence and therapeutic equivalence does not guarantee bioequivalence either. The verification of the Fundamental Bioequivalence Assumption, however, is often difficult, if not impossible, without the conduct of clinical trials. In practice, there are four possible scenarios when assessing bioequivalence for generics approval:

1. Drug absorption profiles are similar and they are therapeutic equivalent;
2. Drug absorption profiles are not similar but they are therapeutic equivalent;
3. Drug absorption profiles are similar but they are not therapeutic equivalent;
4. Drug absorption profiles are not similar and they are not therapeutic equivalent.

Scenario (1) is the Fundamental Bioequivalence Assumption, which works if relative bioavailability in terms of the rate and extent of absorption is predictive of clinical outcomes. In this case, PK responses such as AUC and C_{max} serve as surrogate endpoints for clinical endpoints for assessment of efficacy and safety of the test product under investigation. Scenario (2) is the case where generic companies use to argue for generic approval of their drug products especially when their products fail to meet regulatory requirement for bioequivalence. In this case, it is doubtful that there is a relationship between PK responses and clinical endpoints. The innovator companies usually argue with the regulatory agency to against generic approval with scenario (3). However, more studies are necessarily conducted in order to verify scenario (3). There are no arguments with respect to scenario (4). Under the Fundamental Bioequivalence Assumption, the assessment of average bioequivalence for generic approval has been criticized that it is based on legal/political deliberations rather than scientific considerations.

3. CRITERIA FOR BIOEQUIVALENCE

Under the Fundamental Bioequivalence Assumption, the association between bioequivalence limits and clinical difference is difficult, if not impossible, to assess in practice. Bioequivalence limits or margins could be determined based on absolute change, relative change (or percent change). The specified limit could be in turn based on absolute change or relative change. Along this line, in the past several decades, the following decision rules were proposed by the FDA between 1977 and 2003 for testing the

bioequivalence in terms of average bioavailability between generic drug products and innovative drug products [5-6]. Suppose AUC and C_{max} are the primary systematic exposure measures of the extent and rate of absorption. For each parameter, the following decision rules (criteria) for assessment of average bioequivalence are usually applied.

The ±20 rule

Bioequivalence is concluded if the average bioavailability of the test formulation is within ±20% of that of the reference formulation with a certain assurance. This decision rule is based on the additive model and not on relative or percent change. Thus, it is not employed commonly for most drug products.

The 80/125 rule

Bioequivalence is concluded if the average bioavailability of the test formulation is within (80%, 125%) that of the reference formulation, with a certain assurance. From a multiplicative model for pharmacokinetic responses, the FDA 2003 guidance suggests that the logarithmic transformation on AUC(0-∞) or AUC(0-tlast) and C_{max} be considered [6]. As a result, this criterion is not symmetric about 1 on the original scale where the maximum probability of concluding average bioequivalence occurs. However, on the logarithmic scale, the criterion has a range of – 0.2231 to 0.2231, which the symmetric about 0 where the probability of concluding average bioequivalence is at maximum. Current FDA regulation adopts the 80/125 rule after log-transformation. That is, two drug products are said to be (average) bioequivalence (ABE) if the *90%* confidence interval of the *ratio of geometric means* of the primary pharmacokinetic (PK) responses (after *log-transformation*) is within the bioequivalence limits of *80%* and *125%*.

4. STUDY DESIGN

As indicated in the *Federal Register* [Vol. 42, No. 5, Sec. 320.26(b) and Sec. 320.27(b), 1977], a bioavailability 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 2×2) crossover design is often considered for a bioavailability or bioequivalence study. Denote by T and R the test product and the reference product, respectively. Thus, a 2×2 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 cross-over to receive the reference product (R) after a sufficient length of wash-out period. Similarly, subjects who are randomly assigned to sequence 2 (RT) will receive the reference product (R) first and then cross-over to receive the test product (T) after a sufficient length of wash-out period.

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

Design 1: Balaam's design – i.e., (TT, RR, RT, TR);

Design 2: Two-sequence, three-period dual design – i.e., (TRR, RTT);

Design 3: Four-period design with two sequences – i.e., (TRRT, RTTR);

Design 4: Four-period design with four sequences – i.e., (TTRR, RRTT, TRTR, RTTR).

The above study designs are also referred to as higher-order crossover designs. A higher-order crossover design is defined as a design with the number of sequences or the number of periods greater than the number of treatments to be compared.

For comparing more than two drug products, a Williams' design is often considered. For example, for comparing three drug products, a six-sequence, three-period (6×3) Williams' design is usually considered, while a 4×4 Williams' design is employed for comparing 4 drug products. Williams' design is a variance stabilizing design. More information regarding the construction and good design characteristics of Williams' designs can be found in [1].

5. SAMPLE SIZE

For sample size determination under the standard 2×2 crossover design and additive model, Phillips [7] and Liu and Chow [8] proposed performing power analysis using Schuirmann's two one-sided tests procedure [9]. For higher-order crossover designs comparing two formulations of the same drug products or two drug products, similar formulae can be derived [10]. On the other hand, sample size determination under a multiplicative model, statistical methods proposed by Hauschke et al. (1992) is commonly considered [11].

Because the power curves of Schuirmann's two one-sided tests procedure are symmetric about zero, we present only the equations for the case where $\theta = \mu_T - \mu_R \geq 0$. Let n_i, the number of subjects in each sequence i, have the same value n, and F_v denote the cumulative distribution function of the t distribution with v degrees of freedom. Then the power function, $P_k(\theta)$, of Schuirmann's tests at the α nominal level for design k is given as follows:

$$P_k(\theta)=F_{v_k}\left(\left[\frac{\Delta-\theta}{CV\sqrt{\frac{b_i}{n}}}\right]-t(\alpha,v_k)\right)-F_{v_k}\left(t(\alpha,v_k)-\left[\frac{\Delta+\theta}{CV\sqrt{\frac{b_i}{n}}}\right]\right), for\ k=1,2,3,4, \quad (1)$$

Where $v_1 = 4n - 3$, $v_2 = 4n - 4$, $v_3 = 6n - 5$, $v_4 = 12n - 5$, $b_1 = 2$, $b_2=\frac{3}{4}$, $b_3 = 11/20$, $b_4 = 1/4$.

Hence, the exact equation for determination of n required to achieve a $1-\beta$ power at the α nominal level for each design k when $\theta = 0$ is the following:

$$n \geq b_k[t(\alpha,v_k)+t(\frac{\beta}{2},v_k)]^2[CV/\Delta]^2\ for\ k=1,2,3,4, \quad (2)$$

and if $\theta > 0$ the approximate formula for n is

$$n \geq b_k[t(\alpha,v_k)+t(\beta,v_k)]^2[CV/(\Delta-\theta)]^2\ for\ k=1,2,3,4, \quad (3)$$

For the multiplicative model, we consider the (0.8, 1.25) as the bioequivalence limit for $\delta = \mu_T/\mu_R$ and μ_T and μ_R are the median bioavailabilities of the test and reference formulations, respectively. Similarly, the sample size n required to achieve a $1 - \beta$ power at the α nominal level for each corresponding design k after the logarithmic transformation is determined by the following equations [10-11]:

$$n \geq b_k[t(\alpha, v_k)+t(\frac{\beta}{2}, v_k)]^2 [CV/In(1.25)]^2 if\ \delta=1, \quad (4)$$

$$n \geq b_k[t(\alpha, v_k)+t(\beta, v_k)]^2[CV/(In(1.25) - In\delta)]^2 if\ 1<\delta<1.25, \quad (5)$$

and

$$n \geq b_k[t(\alpha, v_k)+t(\beta, v_k)]^2[CV/(\Delta - \theta)]^2 if\ 0.8<\delta<1. \quad (6)$$

where ln denotes the natural logarithm and $CV = \sqrt{exp(\sigma^2)1}$, the coefficient of variation in the multiplicative model, and σ^2 is the residual (within-subject) variance on the log-scale. However, because the degrees of freedom are usually unknown, an easy way to find the sample size is to enumerate n.

6. CONDUCT OF BIOEQUIVALENCE TRIAL

Subject selection

According to the current FDA guidance, in vivo bioequivalence studies should be conducted in individuals, 18 years or older, who are representative of the general population, taking into account for age, sex, and race. If the product is intended for use in both sexes, inclusion of similar proportions of males and females should be intended. In case where other types of populations are considered, special considerations should be taken into consideration. For example, if elderly subjects are to be included in bioequivalence studies, subjects' stress, blood loss, the status of chronic disease, and pharmacokinetic effects of altered organ function should be taken into consideration as these factors may alter the drug absorption profiles under study. Similarly, the factors of stress, blood loss, pharmacokinetic effects of disease states, concurrent medications, and special diets should be considered when the bioavailability/bioequivalence studies are intended to be conducted with patient population as these factors may inflate both the intra-subject and inter-subject variabilities and consequently result in a more heterogeneous population under study.

Washout

In bioavailability/bioequivalence studies utilizing crossover design, a sufficient length of washout period between dosing periods is necessary to wear off the possible residual effect from the previous dose that may be carried over to the next dosing period. For pivotal fasting studies, FDA requires that at least 5.5 half-lives be considered to ensure there is a sufficient length of washout for immediate release (IR) products. For controlled release (CR) products, on the other hand, FDA indicates that at least 8.5 half-lives should be considered to limit the chance of possible carry-over residual effect.

Blood sampling

In practice, it is undesirable to draw too much blood and/or too frequent from subjects under study. However, sufficient blood should be drawn at different sampling interval in order to accurately and reliably characterize the blood concentration-time curve and consequently the drug absorption profile. For this purpose, it is suggested that more blood sampling around C_{max} should be taken and sampling interval should cover at least three half-lives of the drug product.

IR product versus CR products

For IR products, FDA indicates that single dose fasting study is required, while limited food effect study may be required when needed. For CR products, the current FDA guidance recommends single-dose non-replicate fasting and food-effect studies be conducted. Multiple-dose studies are generally not expected.

7. STATISTICAL METHODS

Average bioequivalence (ABE) is claimed if the geometric means ratio (GMR) of average bioavailabilities between test and reference products is within the bioequivalence limit of (80%, 125%) with 90% assurance based on log-transformed data. 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 means of the primary pharmacokinetic response such as AUC or Cmax is obtained under an analysis of variance model. We claim bioequivalence if the obtained 90% confidence interval is totally within the bioequivalence limit of (80%, 125%).

For the method of interval hypotheses testing, the interval hypotheses that H_0 : Bioinequivalence vs. H_a : Bioequivalence

Note that the above hypotheses are usually decomposed into two sets of one-sided hypotheses. For 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. Under the two one-sided hypotheses, Schuirmann's two one-sided tests procedure is commonly employed for testing ABE [9].

In practice, other statistical methods such as Westlake's symmetric confidence interval approach, exact confidence interval based on Fieller's theorem, Chow and Shao's joint confidence region approach [12], Bayesian methods, and non-parametric methods such as Wilcoxon-Mann-Whitney two one-sided tests procedure, distribution-free confidence interval based on the Hodges-Lehmann estimator, and bootstrap confidence interval are sometimes considered (see, e.g., [1]).

8. DRUG INTERCHANGEABILITY

In practice, bioequivalence in drug absorption has been interpreted that the confidence interval for the ratio of means (of drug absorption) is within bioequivalence limits. An alternative would be to show that the tolerance intervals (or a distribution free model) overlap sufficiently. Under the above Fundamental Bioequivalence Assumption, many

practitioners interpret that generic drug products and the innovative drug product can be used interchangeably because they are therapeutically equivalent. The FDA, however, does not indicate that approved generic drug products and the innovative drug products can be used interchangeably. The FDA only indicates that an approved generic drug product can be used as a substitute to the innovative drug product.

Basically, drug interchangeability can be classified either as drug prescribability or drug switchability (see. e.g., [6, 13-14]). Drug prescribability is defined as the physician's choice for prescribing an appropriate drug product for his/her new patients 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, is related to the switch from a drug product (e.g., a brand-name drug product) to an alternative drug product (e.g., a generic copy of 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. As a result, drug switchability is considered more critical than drug prescribability in the study of drug interchangeability for patients who have been on medication for a while. Drug switchability, therefore, is exchangeability within the same subject.

Population bioequivalence for drug prescribability

As indicated in [1], in general, average bioequivalence (ABE) cannot imply either drug prescribability or drug switchability. Therefore, it is suggested that the assessment of bioequivalence should take into consideration of drug prescribability and drug switchability. To address drug interchangeability, it is recommended that population bioequivalence (PBE) and individual bioequivalence (IBE) be considered for testing drug prescribability and drug switchability (see, e.g., [12-13]), respectively. More specifically, the FDA indicates that PBE may be applied to new formulations, additional strengths, or new dosage forms in NDAs, while IBE should be considered for ANDA or AADA (abbreviated antibiotic drug application) for generic drugs.

To address drug prescribability, FDA proposed the following aggregated, scaled, moment-based, one-sided criterion:

$$PBC=\frac{(\mu_T-\mu_R)^2+(\sigma_{TT}^2-\sigma_{TR}^2)}{max(\sigma_{TR}^2,\sigma_{T0}^2)}\leq\theta_P, \quad (7)$$

where μ_T and μ_R are the mean of the test drug product and the reference drug product, respectively, σ_{TT}^2 and σ_{TR}^2 are the total variance of the test drug product and the reference drug product, respectively, σ_{T0}^2 is a 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 in total variance between the test and reference drug products which measure the similarity for the marginal population distribution 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

$$\theta_P = \frac{(log\ 1.25)^2 + \varepsilon_P}{\sigma^2_{T0}}, \quad (8)$$

where ε_P is guided by the consideration of the variability term $\sigma^2_{TT} - \sigma^2_{TR}$ added to the ABE criterion. As suggested by the FDA guidance, it may be appropriate that ε_P chosen to be 0.02.

Individual bioequivalence for drug switchability

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

$$IBC = \frac{(\mu_T - \mu_R)^2 + \sigma^2_D + (\sigma^2_{WT} - \sigma^2_{WR})}{max(\sigma^2_{WR}, \sigma^2_{W0})} \leq \theta_I, \quad (9)$$

where σ^2_{WT} and σ^2_{WR} are the within-subject variances of the test drug product and the reference drug product, respectively, σ^2_D is the variance component due to subject-by-drug interaction, σ^2_{W0} is a constant that can be adjusted to control the probability of passing IBE, and θ_I is the bioequivalence limit for IBE. The FDA guidance suggests that θ_I be chosen

$$\theta_I - \frac{(log\ 1.25)^2 + \varepsilon_I}{\sigma^2_{W0}}, \quad (10)$$

where ε_I is the variance allowance factor, which can be adjusted for sample size control. Note that the FDA guidance suggests $\varepsilon_I = 0.05$.

Remarks

The validity and feasibility of the FDA recommended criteria for assessing population bioequivalence and individual bioequivalence given in (7) and (9), respectively have been criticized my many researchers. Chow (1999) provided a comprehensive review of the merits and limitations of the criteria as described in the early version of the FDA draft guidance on the assessment of population/individual bioequivalence [15]. For statistical analysis, FDA suggests the method of small sample confidence interval approach proposed by Hyslop, Hsuan, and Holder (2000) be considered for assessment of population/individual bioequivalence [16].

9. PRACTICAL ISSUES

One-size-fits-all criterion

For the assessment of bioequivalence both *in vivo* and *in vitro*, FDA adopted a one size-fits-all criterion. That is, for *in vivo* (*in vitro*), 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 (AUC and Cmax) is totally within the bioequivalence limits of 80% to 125% (90% to 111%). 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 many researchers. It was suggested 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 Table 2).

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 expectation has become, in some cases, more relaxed. For these drugs, the approach of scaled average bioequivalence has been proposed. This method is, in fact, a special case of the procedure described earlier for individual BE when the within-subject variation is high ($\sigma^2_{WR}>\sigma^2_{W0}$). However, the current FDA guidance does not contain special provisions for this class of drugs.

Highly variable drug products

As indicated earlier, the assessment of ABE focuses on average bioavailability but ignores the variability associated with the PK responses. Thus, two drug products may fail the evaluation of ABE if the variability associated with the PK responses is large even though they have identical means. A drug with large variability is considered highly variable. FDA defines a highly variable drug (HVD) as a drug whose within-subject (or intra-subject) variation is greater than or equal to 30%. This definition based on intra-subject variation, however, rather arbitrary. One of problematic aspects of this definition is that the estimated within-subject variability depends on the metrics of pharmacokinetic responses such as AUC and Cmax. Haidar et al. [17] pointed out that HVDs show variable pharmacokinetics as a result of their inherent properties (e.g. distribution, systemic metabolism and elimination) (see also, [18-20]). A drug may have low variability if it is administered intravenously, whereas it can be highly variable after oral administration.

In practice, HVDs often fail to meet current regulatory acceptance criteria for ABE. In the past decade, the topic for evaluation of bioequivalence for HVDs has received much attention. This topic has been discussed several times at regulatory forums and international conferences, but academics, representatives of pharmaceutical industries and regulatory agencies failed to reach a consensus until recently that the approach of scaled average bioequivalence (SABE) is proposed by [17]. The approach of SABE is briefly described below.

Denoted by μ_T and μ_R respectively, are compared. The acceptance of bioequivalence is claimed it the difference between the logarithmic means is between pre-specified regulatory limits. The limits (θ_A) are generally symmetrical on the logarithmic scale and usually equal ±ln(1.25). Thus, the criterion for ABE can be expressed as follows:

$$-\theta_A \leq \mu_T - \mu_R \leq \theta_A \quad (11)$$

In a bioequivalence study, the individual kinetic responses are evaluated from the measured concentrations. The means of the logarithmic responses of the two formulations are calculated. These sample averages estimate the true population means. A variance is also estimated for each kinetic response. It is a measure of the intra-subject variance but not always identical to it. FDA suggests the above ABE could be scaled by a standard deviation as follows:

$$-\theta_s \leq \frac{(\mu_T - \mu_R)}{\sigma_W} \leq \theta_s, \quad (12)$$

where θ_S is the SABE regulatory cutoff. Here the standard deviation (σ_W) is the within-subject standard deviation. In replicate design, σ_W is generally the within-subject standard deviation of the reference formulation (denoted by σ_{WR}). Thus, the scaling factor of SABE has similar features to the scaling factor of IBE.

10. CONCLUDING REMARKS

Between early 1990 to early 2000, the FDA considered individual bioequivalence (IBE) and population bioequivalence (PBE) as a possible solution for the problem of bioequivalence for HVDs. However, the development of this approach has been abandoned. In 2004, the FDA kicked off a *Critical Path Initiative* that focused on the challenges involved in the development new drugs and generics. As part of this initiative, the FDA established a working group on the bioequivalence of HVDs for development of guidance on dealing with HVDs (see, e.g., [17], [20]). For evaluation of bioequivalence of HVDs with SABE, as indicated in [17] and [20], the bioequivalence limits for SABE can be expressed in the form of

$$\theta_s = \frac{In(1.25)}{\sigma_0}, \quad (13)$$

Where σ_0 is a so-called regulatory standardized variation, which defines the proportionality factor between the logarithmic bioequivalence limits and σ_W in the highly variable region. The value of σ_0 must be defined by the regulators. The magnitude of σ_0 defines the bioequivalence limits (θ_S). For instance, when $\sigma_0 = 0.294$, then θ_S is 0.760. Note that FDA recommended $\sigma_0 = 0.25$ as the regulatory constant (see, e.g., [17], [20]).

Although bioavailability for (*in vivo*) bioequivalence studies is usually assessed through the measures of the rate and extent to which the drug product is absorbed into the bloodstream of human subjects, for some locally acting drug products such as nasal aerosols (e.g., metered-dose inhalers) and nasal sprays (e.g., metered-dose spray pumps) 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. For those local delivery drug products, the FDA indicates that bioequivalence may be assessed, with suitable justification, by *in vitro* bioequivalence studies alone (see, e.g., Part 21 Codes of Federal Regulations Section 320.24). In practice, it is expected that *in vitro* bioequivalence testing has less variability (say <10%) due to analytical testing results, while *in vivo* bioequivalence testing typically

has larger variability (say between 20-30%). Unlike small molecule drug products, biosimilars are expected to have much larger variability (say 40-50%) [21]. The magnitude of variability has an impact on the corresponding criteria for assessment of bioequivalence or biosimilarity.

Acknowledgments

This research was partially supported by the Duke University Center for AIDS Research, an NIH-funded program (5P30 AI064518).

References

1. Chow, SC.; Liu, JP. Design and Analysis of Bioavailability and Bioequivalence Studies. 3. Chapman Hall/CRC Press, Taylor& Francis; New York, New York, USA: 2008.
2. Wagner, JG. Fundamentals of Clinical Pharmacokinetics. Drug Intelligence Publications; Hamilton, IL: 1975.
3. Balant LP. Is there a need for more precise definitions of bioavailability? European Journal of Clinical Pharmacology. 1991; 40:123–126. [PubMed: 2065692]
4. Chen ML, Lesko L, Williams RL. Measures of exposure versus measures of rate and extent of absorption. Clinical Pharmacokinetics. 2001; 40:565–572. [PubMed: 11523723]
5. Purich, E. Bioavailability/Bioequivalence Regulations: An FDA Perspective in Drug Absorption and Disposition: Statistical Considerations. Albert, KS., editor. American Pharmaceutical Association, Academy of Pharmaceutical Sciences; Washington, DC: 1980. p. 115-137.
6. FDA. Guidance on Bioavailability and Bioequivalence Studies for Orally Administrated Drug Products – General Considerations. Center for Drug Evaluation and Research, the US Food and Drug Administration; Rockville, Maryland, USA: 2003.
7. Phillips KF. Power of the two one-sided tests procedure in bioequivalence. Journal of Pharmacokinetics and Biopharmaceutics. 1990; 18:137–144. [PubMed: 2348380]
8. Liu JP, Chow SC. On power calculation of Schuirmann's two one-sided tests procedure in bioequivalence. Journal of Pharmacokinetics and Biopharmaceutics. 1992; 20:101–104. [PubMed: 1588502]
9. Schuirmann DJ. A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. Journal of Pharmacokinetics and Biopharmaceutics. 1987; 15:657–680. [PubMed: 3450848]
10. Chen KW, Li G, Chow SC. A note on sample size determination for bioequivalence studies with higher-order crossover designs. Journal of Pharmacokinetics and Biopharmaceutics. 1997; 25:753–765. [PubMed: 9697082]
11. Hauschke D, Steinijans VW, Diletti E, Durke M. Sample size determination for bioequivalence assessment using a multiplicative model. Journal of Pharmacokinetics and Biopharmaceutics. 1992; 20:557–561. [PubMed: 1287202]
12. Chow SC, Shao J. An alternative approach for the assessment of bioequivalence between two formulations of a drug. Biometrical Journal. 1990; 32:969–976.
13. Chow SC, Shao J, Wang H. Individual bioequivalence testing under 2×3 designs. Statistics in Medicine. 2002; 21:629–648. [PubMed: 11870806]
14. Chow SC, Shao J, Wang H. Statistical tests for population bioequivalence. Statistica Sinica. 2002; 13:539–554.
15. Chow SC. Individual bioequivalence – a review of the FDA draft guidance. Drug Informational Journal. 1999; 33:435–444.
16. Hyslop T, Hsuan F, Holder KJ. A small-sample confidence interval approach to assess individual bioequivalence. Statistics in Medicine. 2000; 19:2885–2897. [PubMed: 11033583]
17. Haidar SH, Makhlouf F, Schuirmann DJ, Hyslop T, Davit B, Conner D, Yu LX. Evaluation of a scaling approach for the bioequivalence of highly variable drugs. The AAPS Journal. 2008; 10:450–454. [PubMed: 18726698]

18. Tothfalusi L, Endrenyi L, Arieta AG. Evaluation of bioequivalence for highly variable drugs with scaled average bioequivalence. Clinical Pharmacokinetics. 2009; 48:725–743. [PubMed: 19817502]

19. Davit BM, Conner DP, Fabian-Fritsch B, et al. Highly variable drugs: observations from bioequivalence data submitted to the FDA for new generic drug applications. AAPS Journal. 2008; 10:148–156. [PubMed: 18446515]

20. Haidar SH, Davit BM, Chen ML, et al. Bioequivalence approaches for highly variable drugs and drug products. Pharmaceutical Research. 2008; 25:237–241. [PubMed: 17891552]

21. EMA. Concept paper on similar biological medicinal products containing recombinant follicle stimulation hormone. London, United Kingdom: 2010. A/CHMP/BMWP/94899/2010

Table 1

NDA Versus ANDA

NDA	ANDA
1. Chemistry	1. Chemistry
2. Manufacturing	2.Manufacturing
3. Controls	3.Controls
4. Testing	4.Testing
5. Labeling	5.Labeling
6. PK/bioavailability	6.PK/bioavailability
7. Animal Studies	---
8. Clinical Safety & Efficacy Trials	---

NDA = New Drug Application; ANDA = Abbreviated New Drug Application

Table 2

Classification of Drugs

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

TW = therapeutic window; ISV = intra-subject variability.