Bioequivalence & Bioavailability

Chow, J Bioequiv Availab 2011, S1
http://dx.doi.org/10.4172/jbb.S1-002

Research Article Open Access

Quantitative Evaluation of Bioequivalence/Biosimilarity

Shein-Chung Chow*

Biostatistics and Bioinformatics, Duke University School of Medicine Durham, North Carolina, USA

Abstract

As more biologic products are going off patent protection, the development of follow-on biologics (biosimilars) products has received much attention from both biotechnology industry and the regulatory agencies. Unlike traditional small-molecule (chemical) drug products, the development of biologic products is very different and variable with respect to the manufacturing process and environment. The complexity and heterogeneity of the molecular structure, complicated manufacturing process, different analytical methods, and possibility of severe immunogenicity reactions make quantitative evaluation of follow-on biologics a great challenge to both scientific community and regulatory agencies. In this article, an overview of current criteria, study design, and statistical methods for quantitative evaluation of bioequivalence for the traditional small molecule generic drug productsand biosimilarity for biosimilars products is provided. In addition, a general approach for development of a biosimilarity index based on the concept of reproducibility probability for quantitative evaluation of bioequivalence/biosimilarity is proposed. Some scientific factors and practical issues are also discussed.

Keywrds: Follow-on biologics; Biosimilars; Interchangeability; Switching; Alternating

Background

In the United States (US), for traditional chemical (small molecule) drug products, when an innovative (brand-name) drug product is going off patent, pharmaceutical and/or generic companies may file an abbreviated new drug application (ANDA) for approval of generic copies of the brand-name drug product. In 1984, the US Food and Drug Administration (FDA) was authorized to approve generic drug products under the *Drug Price Competition and Patent Term Restoration* Act, which is known as the Hatch and Waxman Act. For approval of small molecule generic drug products, the FDA requires that evidence in *average* of bioavailability in terms of the rate and extent of drug absorption be provided. The assessment of bioequivalence as a surrogate endpoint for quantitative evaluation of drug safety and efficacy is based on the *Fundamental Bioequivalence Assumption* that if two drug products are shown to be bioequivalent in average bioavailability, it is assumed that they will reach the same therapeutic effect or they are therapeutically equivalent and hence can be used interchangeably. Under the Fundamental Bioequivalence Assumption, regulatory requirements, study design, criteria, and statistical methods for assessment of bioequivalence have been well established (see, e.g., [1-7]).

Unlike small molecule drug products, the *generic versions* of biologic products are viewed similar biological drug products (SBDP). The SBDP are *not* generic drug products, which are drug products with *identical* active ingredient(s) as the innovative drug product. Thus, the concept for development of SBDP, which are made of living cells, is very different from that of the generic drug products for small molecule drug products. The SBDP are usually referred to as biosimilars by European Medicines Agency (EMA) of European Union (EU), follow-on biologics (FOB) by the US FDA, and subsequent entered biologics (SEB) by the Public Health Agency (PHA) of Canada. As a number of biologic products are due to expire in the next few years, the subsequent production of follow-on products has aroused interest within the pharmaceutical/biotechnology industry as biosimilar manufacturers strive to obtain part of an already large and rapidly-growing market. The potential opportunity for price reductions versus the originator biologic products remains to be determined, as the advantage of a slightly cheaper price may be outweighed by the hypothetical increased risk of side-effects from biosimilar molecules that are not exact copies of their originators.

In this article, the focus will not only be placed on the fundamental differences between small molecule drug products and biologic products, but also issues surrounding quantitative evaluation of bioequivalence (for small molecule drug products) and biosimilarity (for biosimilars or follow-on biologics). In the next section, fundamental differences between small molecule drug products and biologic drug products are briefly described. Sections 3 and 4 provide brief descriptions of current process for quantitative evaluation of bioequivalence and biosimilarity, respectively. A general approach using biosimilarity index for assessment of bioequivalence and biosimilarity, which was derived based on the concept of reproducibility probability was proposed and discussed in Section 5. Section 6 summarizes some current scientific factors and practical issues regarding the assessment of biosimilarity. Brief concluding remarks are given in the last section of this article.

Fundamental Differences

Biosimilars or follow-on biologics are fundamentally different from those of traditional chemical generic drugs. Unlike traditional chemical generic drug products which contain identical active ingredient(s), the generic versions of biologic products are made of living cells. Unlike classical generics, biosimilars are not identical to their originator products and therefore should not be brought to market using the same procedure applied to generics. This is partly a reflection of the complexities of manufacturing and safety and efficacy controls of biosimilars when compared to their small molecule generic counterparts (see, e.g., [8-11]).

***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 27705, E-mail: sheinchung.chow@duke.edu

Received August 12, 2011; **Accepted** September 06, 2011; **Published** September 08, 2011

Citation: Chow SC (2011) Quantitative Evaluation of Bioequivalence/Biosimilarity. J Bioequiv Availab S1. doi:10.4172/jbb.S1-002

Some of the fundamental differences between biosimilars and generic chemical drugs are summarized in Table 1. For example, biosimilars are known to be variable and very sensitive to the environmental conditions such as light and temperature. A small variation may translate to a drastic change in clinical outcomes (e.g., safety and efficacy). In addition to differences in the size and complexity of the active substance, important differences also include the nature of the manufacturing process. Since biologic products are often recombinant protein molecules manufactured in living cells, manufacturing processes for biologic products are highly complex and require hundreds of specific isolation and purification steps. Thus, in practice, it is impossible to produce an *identical* copy of a biologic product, as changes to the structure of the molecule can occur with changes in the manufacturing process. Since a protein can be modified during the process (e.g., a side chain may be added, the structure may have changed due to protein misfolding, and so on), different manufacturing processes may lead to structural differences in the final product, which result in differences in efficacy and safety, and may have a negative impact on the immune responses of patients. It should be noted that these issues occur also during the post-approval changes of the innovator's biological products.

Thus, SBDP are not generic products. Hence, the standard generic approach is not applicable and acceptable due to the complexity of biological/biotechnology derived products. Instead, similar biological approach depending upon the state-of-art of analytical procedures should be applied.

Quantitative Evaluation of Bioequivalence

For approval of small molecule generic drug products, the FDA requires that evidence of average bioequivalence in drug absorption in terms of some pharmacokinetic (PK) parameters such as the area under the blood and/or plasma concentration-time curve (AUC) and peak concentration (C_{max}) be provided through the conduct of bioequivalence studies. In practice, we may claim that a test drug product is bioequivalent to an innovative (reference) drug product if the 90% confidence interval for the ratio of geometric means of the primary PK parameter is completely within the bioequivalence limits of (80%, 125%). The confidence interval for the ratio of geometric means of the primary PK parameter is obtained based on log-transformed data. In what follows, study designs and statistical methods that are commonly considered in bioequivalence studies are briefly described.

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-overed 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 receive the test product (T) after a sufficient length of wash-out period.

Chemical drugs	Biologic drugs
Made by chemical synthesis	Made by living cells
Defined structure	Heterogeneous structure Mixtures of related molecules
Easy to characterize	Difficult to characterize
Relatively stable	Variable Sensitive to environmental conditions such as light and temperature
No issue of immunogenicity	Issue of immunogenicity
Usually taken orally	Usually injected
Often prescribed by a general practitioner	Usually prescribed by specialists

Table 1: Fundamental Differences.

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 will receive the same treatment only once. In the interest of assessing intra-subject variabilities, the following alternative higher-order crossover designs for comparing two drug products are often considered: (i) Balaam's design, i.e., (TT, RR, RT, TR), (ii) two-sequence, three-period dual design, e.g., (TRR,RTT), and (iii) four-sequence, four-period design, e.g., (TTRR, RRTT, TRTR, RTTR).

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 Chow and Liu [7].

In addition to the assessment of average bioequivalence (ABE), there are other types of bioequivalence assessment such as population bioequivalence (PBE) which is intended for addressing drug prescibability and individual bioequivalence (IBE) which is intended for addressing drug switchability. For assessment IBE/PBE, the FDA recommends that a *replicated* design be considered for obtaining independent estimates of intra-subject and inter-subject variabilities and variability due to subject-by-drug product interaction. A commonly considered replicate crossover design is the replicate of a 2×2 crossover design is given by (TRTR, RTRT). In some cases, an incomplete block design or an extra-reference design such as (TRR, RTR) may be considered depending upon the study objectives of the bioavailability/bioequivalence studies [12].

Statistical methods

As indicated earlier, ABE is claimed if the ratio 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 C_{max} 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

$$H_0: \text{Bioinequivalence} \quad \text{vs.} \quad H_a: \text{Bioequivalence} \quad (1)$$

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 [1].

In practice, other statistical methods such as Westlake's symmetric confidence interval approach, confidence interval based on Fieller's theorem, Chow and Shao's joint confidence region approach, 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 [7].

Remarks

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

In addition, the use of one-fits-all criterion for assessment of ABE has been criticized in the past decade. It is suggested that the one-fits-all criterion be flexible by adjusting intra-subject variability of the reference product and therapeutic window whenever possible. This has led to the proposed scaled average bioequivalence (SAB) criterion for assessment of bioequivalence for highly variable drug products (see, e.g., [13]). It, however, should be noted that the SAB criterion is a special case of the following criteria for IBE:

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

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

As indicated by the regulatory agencies, a generic drug can be used as a substitution of the brand-name drug if it has been shown to be bioequivalent to the brand-name drug. Current regulations do not indicate that two generic copies of the same brand-name drug can be used interchangeably, even though they 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.

Quantitative Evaluation of Biosimilarity

As indicated earlier, the assessment of bioequivalence is possible under the Fundamental Bioequivalence Assumption. Due to the fundamental differences between the small molecule drug products and biological products, the Fundamental Bioequivalence Assumption and the well-established standard methods may not be appropriately applied directly for assessment of biosimilarity. Based on the *Biologics Price Competition and Innovation* (BPCI) Act (as part of the *Affordable Care Act*) passed by the US Congress on March 23, 2010, quantitative evaluation of biosimilarity includes the concepts of biosimilarity and drug interchangeability, which will be briefly described below.

Biosimilarity

In the BPCI Act, a biosimilar product is defined as a product that is *highly similar* to the reference product notwithstanding minor differences in clinically inactive components and there are no clinically meaningful differences in terms of safety, purity, and potency. Based on this definition, a biological medicine is considered biosimilar to a reference biological medicine if it is highly similar to the reference in safety, purity (quality) and efficacy. However, little or no discussion regarding that' *How similar is considered highly similar*?' in the BPCI Act is given.

Basic principles: The BPCI Act seems to suggest that a biosimilar product should be highly similar to the reference drug product in all spectrums of good drug characteristics such as identity, strength, quality, purity, safety, and stability. In practice, however, it is almost impossible to demonstrate that a biosimilar product is highly similar to the reference product in all aspects of good drug characteristics in a single study. Thus, to ensure a biosimilar product is highly similar to the reference product in terms of these good drug characteristics, different biosimilar studies may be required. For example, if safety and efficacy is a concern, then a clinical trial must be conducted to demonstrate that there are no clinically meaningful differences in terms of safety and efficacy. On the other hand, to ensure highly similar in quality, assay development/validation, process control/validation, and product specification of the reference product are necessarily established. In addition, test for comparability in manufacturing process between biosimilars and the reference must be performed. In some cases, if a surrogate endpoint such as pharmacokinetic (PK), pharmacodynamics (PD), or genomic marker is predictive of the primary efficacy/safety clinical endpoint, then a PK/PD or genomic study may be used to assess biosimilarity between biosimilars and the reference product.

It should be noted that current regulatory requirements are guided based on a case-by-case basis by the following basic principles that (i) the extent of the physicochemical and biological characterization of the product, (ii) nature or possible changes in the quality and structure of the biological product due the changes in the manufacturing process (and their unexpected outcomes), (iii) clinical/regulatory experiences with the particular class of the product in question, and (iv) several factors that need to be considered for biocomparability.

Criteria, design and statistical methods for biosimilarity:

Criteria for biosimilarity: For the comparison between drug products, some criteria for the assessment of bioequivalence, similarity (e.g., the comparison of dissolution profiles), and consistency (e.g., comparisons between manufacturing processes) are available in either regulatory guidelines/guidances or the literature. These criteria, however, can be classified into either (i) absolute change versus relative change, (ii) aggregated versus disaggregated, or (iii) moment-based versus probability-based.

In practice, we may consider assessing bioequivalence or biosimilarity by comparing average and variability separately or simultaneously. This leads to the so-called disaggregated criterion and aggregated criterion. A disaggregate criterion will provide different levels of biosimilarity. For example, the study that passes criteria of both average and variability of biosimilarity provides stronger evidence of biosimilarity as compared to those studies that pass only the average biosimilarity. On the other hand, it is not clear whether an

aggregated criterion would provide a stronger evidence of biosimilarity due to potential offset (or masked) effect between the average and variability in the aggregated criterion. Further researchfor establishing the appropriate statistical testing procedures based on the aggregate criterion and comparing its performance with the disaggregate criterion may be needed.

Chow et al. [14] compared the moment-based criterion with the probability-based criterion for assessment of bioequivalence or biosimilarity under a parallel group design. The results indicate that the probability-based criterion is not only a much more stringent criterion, but also has sensitivity to any small change in variability. This justifies the use of the probability-based criterion for assessment of biosimilarity between follow-on biologics if a certain level of precision and reliability of biosimilarity is desired.

Study design: As indicated earlier, a crossover design is often employed for bioequivalence assessment. In a crossover study, each drug product is administered to each subject. Thus, estimate (approximate) within-subject variance can be sued to address switch ability and interchangeability. For a parallel-group study, each drug product is administered to a different group of subjects. Thus, we can only estimate total variance (between and within subject variances) not individual variance components. For follow-on biologics with long half-lives, crossover study would be ineffective and unethical. In this case, we need to under take study with parallel groups. However, a parallel-group study does not provide an estimate for within-subject variation (since there is no R vs. R).

Statistical methods: Similar to the assessment of average bioequivalence, Shuirmann's two one-sided tests procedure or the confidence interval are recommended for assessment of biosimilarity if similar criteria are adopted. On the other hand, if similar criteria for assessment of population/individual bioequivalence are considered, the 95% confidence upper bound can be used for assessing biosimilarity based on linearized criteria of population/individual bioequivalence.

Interchangeability

As indicated in the Subsection (b)(3) amended to the Public Health Act Subsection 351(k)(3), the term *interchangeable* or *interchangeability* in reference to a biological product that is shown to meet the standards described in subsection (k)(4), means that the biological product may be substituted for the reference product without the intervention of the health care provider who prescribed the reference product. Along this line, in what follows, definition and basic concepts of interchangeability (in terms of switching and alternating) are given.

Definition and basic concepts: As indicated in the Subsection (a)(2) amends the Public Health Act Subsection 351(k)(3), a biological product is considered to be interchangeable with the reference product if (i) the biological product is biosimilar to the reference product; and (ii) it can be expected to produce the same clinical result in *any given patient*. In addition, for a biological product that is administered more than once to an individual, the risk in terms of safety or diminished efficacy of alternating or switching between use of the biological product and the reference product is not greater than the risk of using the reference product without such alternation or switch.

Thus, there is a clear distinction between biosimilarity and interchangeability. In other words, biosimilarity does not imply interchangeability which is much more stringent. Intuitively, if a test product is judged to be interchangeable with the reference product then it may be substituted, even alternated, without a possible intervention, or even notification, of the health care provider**.** However, the Interchangeability is expected to produce the *same* clinical result in *any given patient*, which can be interpreted as that the same clinical result can be expected in *every single patient*. In reality, conceivably, lawsuits may be filed if adverse effects are recorded in a patient after switching from one product to another.

It should be noted that when FDA declares the biosimilarity of two drug products, it may not be assumed that they are interchangeable. Therefore, labels ought to state whether for a follow-on biologic which is biosimilar to a reference product, interchangeability has or has not been established. However, payers and physicians may, in some cases, switch products even if interchangeability has not been established.

Switching and alternating: Unlike drug interchangeability (in terms of prescribability and switchability [7], the US FDA has slightly perception of drug interchangeability for biosimilars. From the FDA's perspectives, interchangeability includes the concept of switching and alternating between an innovative biologic product (R) and its follow-on biologics (T). The concept of switching is referred to as not only the switch from "R to T" or "T to R" (narrow sense of switchability), but also "T to T" and "R to R" (broader sense of switchability). As a result, in order to assess switching, biosimilarity for "R to T", "T to R", "T to T", and "R to R" need to be assessed based on some biosimilarity criteria under a valid study design.

On the other hand, the concept of alternating is referred to as either the switch from T to R and then switch back to T (i.e., "T to R to T") or the switch from R to T and then switch back to R (i.e., "R to T to R". Thus, the difference between "the switch from T to R" or "the switch from R to T" and "the switch from R to T" or "the switch from T to R" needs to be assessed for addressing the concept of alternating.

Study design: For assessment of bioequivalence for chemical drug products, a standard two-sequence, two-period (2×2) crossover design is often considered, except for drug products with relatively long half-lives. Since most biosimilar products have relatively long half-lives, it is suggested that a parallel group design should be considered. However, parallel group design does not provide independent estimates of variance components such as inter- and intra-subject variabilities and variability due to subject-by-product interaction. Thus, it is a major challenge for assessing biosimilars under parallel group designs.

In order to assess biosimilarity for "R to T", "T to R", "T to T", and "R to R", the Balaam's 4×2 crossover design, i.e., (TT, RR, TR, RT) may be useful. For addressing the concept of alternating, a two-sequence, three-period dual design, i.e., (TRT, RTR) may be useful. For addressing both concepts of switching and alternating for drug interchangeability of biosimilars, a modified Balaam's crossover design, i.e., (TT, RR, TRT, RTR) is then recommended.

Remarks: With small molecule drug products, bioequivalence generally reflects therapeutic equivalence. Drug prescribability, switching, and alternating are generally considered reasonable. With biologic products, however, variations are often higher (other than pharmacokinetic factors may be sensitive to small changes in conditions). Thus, often only parallel-group design rather than crossover kinetic studies can be performed. It should be noted that very often, with follow-on biologics, biosimilarity does *not* reflect therapeutic comparability. Therefore, switching and alternating should be pursued only with substantial caution.

A General Approach for Assessment of Bioequivalence/ Biosimilarity

As indicated earlier, the concept of biosimilarity and interchangeability for follow-on biologics is very different from that of bioequivalence and drug interchangeability for small molecule drug products. It is debatable whether standard methods for assessment of bioequivalence and drug interchangeability can be applied to assessing biosimilarity and interchangeability of follow-on biologics due to the fundamental differences as described in Section 2. While appropriate criteria or standards for assessment of biosimilarity and interchangeability are still under discussion within the regulatory agencies and among the pharmaceutical industry and academia, we would like to propose the a general approach for assessing biosimilarity and interchangeability by comparing the relative difference between *"a test product vs. a reference product"* and *"the reference vs. the reference"* based on the concept of reproducibility probability of claiming biosimilairty between a test product and a reference product in a future biosimilarity study provided that the biosimilarity between the test product and the reference product has been established in the current study.

Development of biosimilarity index

Shao and Chow [15] proposed a reproducibility probability as an index for determining whether it is necessary to require a second trial when the result of the first clinical trial is strongly significant. Suppose that the null hypothesis H_0 is rejected if and only if $|T|>c$, where c is a positive known constant and T is a test statistic. Thus, the reproducibility probability of observing a significant clinical result when H_a is indeed true is given by

$$p = P\left(|T|>c \,|\, H_a\right) = P(|T|>c \,|\, \hat{\theta}), \quad (3)$$

where $\hat{\theta}$ is an estimate of θ, which is an unknown parameter or vector of parameters. Following the similar idea, a reproducibility probability can also be used to evaluate biosimilarity and interchangeability between a test product and a reference product based on any pre-specified criteria for biosimilarity and interchangeability. As an example, biosimilarity index proposed by Chow et al. [16] is illustrated based on the well-established bioequivalence criterion by the following steps:

Step 1. Assess the average biosimilarity between the test product and the reference product based on a given biosimilarity criterion. For illustration purpose, consider bioequivalence criterion as biosimilarity criterion. That is, biosimilarity is claimed if the 90% confidence interval of the ratio of means of a given study endpoint falls within the biosimilarity limit of (80%, 125%) based on log-transformed data.

Step 2. Once the product passes the test for biosimilarity in Step 1, calculate the reproducibility probability based on the observed ratio (or observed mean difference) and *variability*. We will refer to the calculated reproducibility probability as the *biosimilarity index*.

Step 3. We then claim biosimilarity if the following null hypothesis is rejected:

$$H_0: P \le p_0 \text{ vs. } H_a: P>p_0. \quad (4)$$

A confidence interval approach can be similarly applied. In other words, we claim biosimilarity if the lower 95% confidence bound of the reproducibility probability is larger than a pre-specified number p_0. In practice, p_0 can be obtained based on an estimated of reproducibility probability for a study comparing a reference product to itself (the reference product). We will refer to such a study as an R-R study.

In an R-R study, define

$$P_{TR} = P\left(\begin{array}{l}\text{concluding average biosimiliarity between the test and the}\\ \text{reference products in a future trial given that the average}\\ \text{biosimiliarity based on ABE criterion has been established}\\ \text{in first trial}\end{array}\right) \quad (5)$$

Alternatively, a reproducibility probability for evaluating the biosimilarity of the two same reference products based on ABE criterion is defined as:

$$P_{RR} = P\left(\begin{array}{l}\text{concluding average biosimiliarity of the two same reference}\\ \text{products in a future trial given that the average biosimilarity}\\ \text{based on ABE criterion have been established in first trial}\end{array}\right) \quad (6)$$

Since the idea of the biosimilarity index is to show that the reproducibility probability in a study for comparing follow-on biologic with the innovative (reference) product is higher than a reference product with the reference product. The criterion of an acceptable reproducibility probability (i.e., p_0) for assessment of biosimilarity can be obtained based on the R-R study. For example, if the R-R study suggests the reproducibility probability of 90%, i.e., P_{RR} = 90%, the criterion of the reproducibility probability for bioequivalence study could be chosen as 80% of the 90% which is p_0=80% × P_{RR}= 72%.

The above described biosimilarity index has the advantages that (i) it is robust with respect to the selected study endpoint, biosimilarity criteria, and study design, (ii) it takes variability into consideration (one of the major criticisms in the assessment of average bioequivalence), (iii) it allows the definition and assessment the degree of similarity (in other words, it provides partial answer to the question that "how similar is considered similar?" and (iv) the use of biosimilarity index will reflect the sensitivity of heterogeneity in variance.

Most importantly, the biosimilarity index proposed by Chow et al. [16] can be applied to different functional areas (domains) of biological products such as good drug characteristics such as safety (e.g., immunogenicity), purity, and potency (as described in BPCI Act), pharmacokinetics (PK) and pharmacodynamics (PD), biological activities, biomarkers (e.g., genomic markers), and manufacturing process, etc. for an assessment of *global* biosimilarity. An overall biosimilar index across domains can be obtained by the following steps:

Step 1. Obtain P_i, the probability of reproducibility for the i-th domain, i=1,.., K.

Step 2. Define the global biosimilarityindex $P = \sum_{i=1}^{K} w_i P_i$, where w_i is the weight for the i-th domain.

Step 3. Claim global biosimilarity if the lower 95% confidence bound of the reproducibility probability (P) is larger than a pre-specified number P_0, where P_0 is a pre-specified acceptable reproducibility probability.

Remarks

Hsieh et al. [17] studied the performance of the biosimilarity index under a R-R study for establishing a baseline for assessment of biosimilarity based on current criterion for average bioequivalence. The results indicate that biosimilarity index is sensitive to the variability associated with the reference product. The biosimilarity index decreases as the variability increases. As an example, Figure 1

gives reproducibility probability curves under a 2×2 crossover design with sample sizes $n_1 = n_2 = 10$, 20, 30, 40, 50, and 60 at the 0.05 level of significance and $(\theta_L, \theta_U) = (80\%, 125\%)$ when $\sigma_d = 0.2$ and 0.3, where σ_d is the standard deviation of period difference within each subject.

In practice, alternative approaches for assessment of the proposed biosimilarity index are available (see, e.g., [17,18]). The methods include maximum likelihood approach and Bayesian approach. For the Bayesian approach, let $p(\theta)$ be the power function, where θ is an unknown parameter or vector of parameters. Under this Bayesian approach, θ is random with a prior distribution assumed to be known. The reproducibility probability can be viewed as the posterior mean of the power function for the future trial

$$\int p(\theta)\pi(\theta \mid x)d\theta, \quad (7)$$

where $\pi(\theta|x)$ is the posterior density of θ, given the data set x observed for the previous trial (s). However, there may exist no explicit form for the estimation of the biosimilarity index. As a result, statistical properties of the derived biosimilarity index may not be known. In this case, the finite sample size performance of the derived biosimilarity index may only be evaluated by clinical trial simulations.

As an alternative measure for assessment of global biosimilarity across domains, we may consider $rd = \sum_{i=1}^{K} w_i rd_i$, where $rd_i = \frac{P_{TRi}}{P_{RRi}}$ which is the relative measure of biosimilarity between T and R as compared to that of between R and R. Based on rd_i, $i=1, \ldots, K$, we may conduct a *profile analysis* as described in the 2003 FDA guidance on *Bioavailability and Bioequivalence Studies for Nasal Aerosols and Nasal Sprays for Local Action* [5]. However, statistical properties of the profile analysis based on rd_i, $i=1, \ldots, K$ are not fully studied. Thus, further research is required.

Scientific Factors and Practical Issues

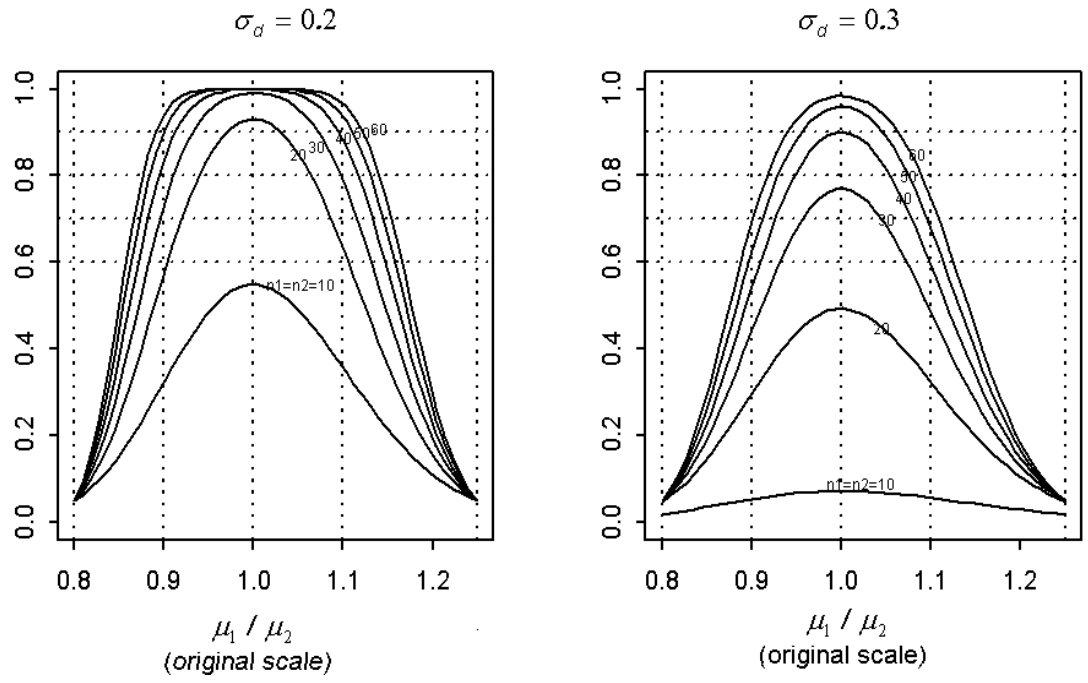
Following the passage of the BPCI Act, in order to obtain input on specific issues and challenges associated with the implementation of the BPCI Act, the US FDA conducted a two-day public hearing on *Approval Pathway for Biosimilar and Interchangeability Biological Products* held on November 2-3, 2010 at the FDA in Silver Spring, Maryland, USA. In what follows, some of the scientific factors and practical issues are briefly described.

Fundamental biosimilarity assumption

Similar to Fundamental Bioequivalence Assumption for assessment of bioequivalence, Chow et al. [14] proposed the following *Fundamental Biosimilarity Assumption* for follow-on biologics:

When a biosimilar product is claimed to be biosimilar to an innovator's product based on some well-defined product characteristics and is therapeutically equivalent provided that the well-defined product characteristics are validated and reliable predictors of safety and efficacy of the products.

For the chemical generic products, the well-defined product characteristics are the exposure measures for early, peak, and total portions of the concentration-time curve. The Fundamental Bioequivalence Assumption is assumed that the equivalence in the exposure measures implies therapeutically equivalent. However, due to the complexity of the biosimilar drug products, one has to verify that some validated product characteristics are indeed reliable predictors of the safety and efficacy. It follows that the design and analysis for evaluation of equivalence between the biosimilar drug product and innovator products are substantially different from those of the chemical generic products.



Note: The reproducibility probability decreases when μ_1/μ_2 (original scale) moves away from 1 and σ_d (log scale) is larger.

Figure 1: Impact of Variability on Reproducibility Probability.

Endpoint selection

For assessment of biosimilarity of follow-on biologics, the following questions are commonly asked. First, what endpoints should be used for assessment of biosimilarity? Second, should a clinical trial always be conducted?

To address these two questions, we may revisit the definition of biosimilarity as described in the BPCI Act. A biological product that is demonstrated to be *highly similar* to an FDA-licensed biological product may rely on certain existing scientific knowledge about safety, purity (quality), and potency (efficacy) of the reference product. Thus, if one would like to show that the safety and efficacy of a biosimilar product are highly similar to that of the reference product, then a clinical trial may be required. In some cases, clinical trials for assessment of biosimilarity may be waived if there exists substantial evidence that surrogate endpoints or biomarkers are predictive of the clinical outcomes. On the other hand, clinical trials are required for assessment of drug interchangeability in order to show that the safety and efficacy between a biosimilar product and a reference product are similar in any given patient of the patient population under study.

How similar is similar?

Current criteria for assessment of bioequivalence/biosimilarity is useful for determining whether a biosimilar product is similar to a reference product. However, it does not provide additional information regarding the *degree* of similarity. As indicated in the BPCI Act, abiosimilar product is defined as a product that is *highly similar* to the reference product. However, little or no discussion regarding the degree of similarity for highly similar was provided. Besides, it is also of concern to the sponsor that "what if a biosimilar product turns out to be superior to the reference product?". A simple answer to the concern is that superiority is not biosimilarity.

Practical issues

Since there are many critical (quality) attributes of a potential patient's response in follow-on biologics, for a given critical attribute, valid statistical methods are necessarily developed under a valid study design and a given set of criteria for similarity, as described in the previous section. Several areas can be identified for developing appropriate statistical methodologies for the assessment of biosimilarity of follow-on biologics. These areas include, but are not limited to:

Criteria for biosimilarity (in terms of average, variability, or distribution): To address the question that "how similar is similar?", we suggest establishing criteria for biosimilarity in terms of average, variability, and/or distribution.

Criteria for interchangeability: In practice, it is recognized that drug interchangeability is related to the variability due to subject-by-drug interaction. However, it is not clear whether criterion for interchangeability should be based on the variability due to subject-by-drug interaction or the variability due to subject-by-drug interaction adjusted for intra-subject variability of the reference drug.

Bridging studies for assessing biosimilarity: As most biosimilars studies are conducted using a parallel design rather than a replicated crossover design, independent estimates of variance components such as the intra-subject and the variability due to subject-by-drug interaction are not possible. In this case, bridging studies may be considered.

Other practical issues include (i) the use of a percentile method for the assessment of variability, (ii) comparability in biologic activities, (iii) assessment of immunogenicity, (iv) consistency in manufacturing processes (see, e.g., [19-21]), (v) stability testing for multiple lots and/or multiple labs (see, e.g., [19]), (vi) the potential use of sequential testing procedures and multiple testing procedures, (vii) assessing biosimilarity using a surrogate endpoint or biomarker such as genomic data (see, e.g., [22]).

Further research is needed in order to address the above mentioned scientific factors and practical issues recognized at the FDA Public Hearing.

Concluding Remarks

As indicated earlier, we claim that a test drug product is bioequivalent to a reference (innovative) drug product if the 90% confidence interval for the ratio of means of the primary PK parameter is totally within the bioequivalence limits of (80%, 125%). This one size-fits-all criterion only focuses on average bioavailability and ignores heterogeneity of variability. Thus, it is not scientifically/statistically justifiable for assessment of biosimilarity of follow-on biologics. In practice, it is then suggested that appropriate criteria, which can take the heterogeneity of variability into consideration be developed since biosimilars are known to be variable and sensitive to small variations in environmental conditions [14,23,24].

At the FDA public hearing, questions that are commonly asked are "*How similar is considered similar?*" and *"How the degree of similarity should be measured and translated to clinical outcomes (e.g., safety and efficacy)*?" These questions closely related to drug interchangeability of biosimilars or follow-on biologics which have been shown to be biosimilar to the innovative product [11,25].

For assessment of bioequivalence for chemical drug products, a crossover design is often considered, except for drug products with relatively long half-lives. Since most biosimilar products have relatively long half-lives, it is suggested that a parallel group design should be considered. However, parallel group design does not provide independent estimates of variance components such as inter- and intra-subject variabilities and variability due to subject-by-product interaction. Thus, it is a major challenge for assessing biosimilars under parallel group designs.

Although EMA of EU has published several product-specific guidances based on the concept papers (e.g., [26-34]), it has been criticized that there are no objective *standards* for assessment of biosimilars because it depends upon the nature of the products. Product-specific standards seem to suggest that a *flexible* biosimilarity criterion should be considered and the flexible criterion should be adjusted for variability and/or the therapeutic index of the innovative (or reference) product.

As described above, there are many uncertainties for assessment of biosimilarity and interchangeability of biosimilars. As a result, it is a major challenge to both clinical scientists and biostatisticians to develop valid and robust clinical/statistical methodologies for assessment of biosimilarity and interchangeability under the uncertainties. In addition, how to address the issues of quality and comparability in manufacturing process is another challenge to both the pharmaceutical scientists and biostatisticians. The proposed general approach using the biosimilarity index (derived based on the concept of reproducibility

probability) may be useful. However, further research on the statistical properties of the proposed biosimilarity index is required.

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. EMEA (2001) Note for Guidance on the Investigation of Bioavailability and Bioequivalence. The European Medicines Agency Evaluation of Medicines for Human Use. EMEA/EWP/QWP/1401/98, London, United Kingdom.
3. FDA (2001) Guidance on *Statistical Approaches to Establishing Bioequivalence*. Center for Drug Evaluation and Research, the US Food and Drug Administration, Rockville, Maryland, USA.
4. FDA (2003) 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.
5. FDA (2003).Guidance on *Bioavailability and Bioequivalence Studies for Nasal Aerosols and Nasal Sprays for Local Action*, Center for Drug Evaluation and Research, the US Food and Drug Administration, Rockville, Maryland, USA.
6. WHO (2005) World Health Organization Draft Revision on Multisource (Generic) Pharmaceutical Products: Guidelines on Registration Requirements to Establish Interchangeability, Geneva, Switzerland.
7. Chow SC, Liu JP (2008) Design and Analysis of Bioavailability and Bioequivalence Studies. 3rd edition, Chapman Hall/CRC Press, Taylor & Francis, New York, New York, USA.
8. Schellekens H (2004) How similar do 'biosimilar' need to be? Nat Biotechnol 22: 1357-1359.
9. Chirino AJ, Mire-Sluis A (2004) Characterizing biological products and assessing comparability following manufacturing changes. Nat Biotechnol 22: 1383-1391.
10. Crommelin D, Bermejo T, Bissig M, Damianns J, Kramer I, et al. (2005) Biosimilars, generic versions of the first generation of therapeutic proteins: do they exist? Contrib Nephrol 149: 287-294.
11. Roger SD, Mikhail A (2007) Biosimilars: opportunity or cause for concern? J Pharm Sci 10: 405-410.
12. Chow SC, Shao J, Wang H (2002) Individual bioequivalence testing under 2x3 crossover designs. Stat Med 21: 629-648.
13. Haidar SH, Davit B, Chen ML, Conner D, Lee L, et al. (2008) Bioequivalence approaches for highly variable drugs and drug products. Pharm Res 25: 237-241.
14. Chow SC, Hsieh TC, Chi E, Yang J (2010) A comparison of moment-based and probability-based criteria for assessment of follow-on biologics. J Biopharm Stat 20: 31-45.
15. Shao J, Chow SC (2002) Reproducibility probability in clinical trials. Stat Med 21: 1727-1742.
16. 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: 1-14.
17. Hsieh TC, Chow SC, Yang LY, Chi E (2011) Statistical test for evaluation of biosimilarity based on reproducibility probability. Submitted for publication consideration.
18. Yang LY, Chow SC, Hsieh TC, Chi E (2011) Bayesian approach for assessment of biosimilarity based on reproducibility probability. Submitted.
19. ICH (1996) Q5C Guideline on Quality of Biotechnological Products: Stability Testing of Biotechnological/Biological Products. Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research, the US Food and Drug Administration, Rockville, Maryland, USA.
20. ICH (1999) Q6B Guideline on Test Procedures and Acceptance Criteria for Biotechnological/Biological Products. Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research, the US Food and Drug Administration, Rockville, Maryland, USA.
21. ICH (2005) Q5E Guideline on Comparability of Biotechnological/Biological Products Subject to Changes in Their Manufacturing Process. Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research, the US Food and Drug Administration, Rockville, Maryland, USA.
22. Chow SC, Shao J, Li L (2004) Assessing bioequivalence using genomic data. J Biopharm Stat 14: 869-880.
23. Chow SC, Liu JP (2010) Statistical assessment of biosimilar products. J Biopharm Stat 20: 10-30.
24. Hsieh TC, Chow SC, Liu JP, Hsiao CF, Chi E (2010) Statistical test for evaluation of biosimilarity of follow-on biologics. J Biopharm Stat 20: 75-89.
25. Roger SD (2006) Biosimilars: How similar or dissimilar are they? Nephrology 11: 341-346.
26. EMEA (2003a) Note for Guidance on Comparability of Medicinal Products Containing Biotechnology-derived Proteins as Drug Substance – Non Clinical and Clinical Issues. The European Medicines Agency Evaluation of Medicines for Human Use. EMEA/CHMP/3097/02, London, United Kingdom.
27. EMEA (2003b) Rev. 1 Guideline on Comparability of Medicinal Products Containing Biotechnology-derived Proteins as Drug Substance – Quality Issues. The European Medicines Agency Evaluation of Medicines for Human Use. EMEA/CHMP/ BWP/3207/00/Rev 1, London, United Kingdom.
28. EMEA (2005a) Guideline on Similar Biological Medicinal Products. The European Medicines Agency Evaluation of Medicines for Human Use. EMEA/CHMP/437/04, London, United Kingdom.
29. EMEA (2005b) Draft Guideline on Similar Biological Medicinal Products Containing Biotechnology-derived Proteins as Drug Substance: Quality Issues. The European Medicines Agency Evaluation of Medicines for Human Use. EMEA/CHMP/49348/05, London, United Kingdom.
30. EMEA (2005c) Draft Annex Guideline on Similar Biological Medicinal Products Containing Biotechnology-derived Proteins as Drug Substance – Non Clinical and Clinical Issues – Guidance on Biosimilar Medicinal Products containing Recombinant Erythropoietins. The European Medicines Agency Evaluation of Medicines for Human Use. EMEA/CHMP/94526/05, London, United Kingdom.
31. EMEA (2005d) Draft Annex Guideline on Similar Biological Medicinal Products Containing Biotechnology-derived Proteins as Drug Substance – Non Clinical and Clinical Issues – Guidance on Biosimilar Medicinal Products containing Recombinant Granulocyte-Colony Stimulating Factor. The European Medicines Agency Evaluation of Medicines for Human Use. EMEA/CHMP /31329 /05, London, United Kingdom.
32. EMEA (2005e) Draft Annex Guideline on Similar Biological Medicinal Products Containing Biotechnology-derived Proteins as Drug Substance – Non-Clinical and Clinical Issues – Guidance on Biosimilar Medicinal Products containing Somatropin. The European Medicines Agency Evaluation of Medicines for Human Use. EMEA/CHMP/94528/05, London, United Kingdom.
33. EMEA (2005f) Draft Annex Guideline on Similar Biological Medicinal Products Containing Biotechnology-derived Proteins as Drug Substance – Non Clinical and Clinical Issues –Guidance on Biosimilar Medicinal Products containing Recombinant Human Insulin. The European Medicines Agency Evaluation of Medicines for Human Use. EMEA/CHMP/32775/05, London, United Kingdom.
34. EMEA (2005g) Guideline on the Clinical Investigating of the Pharmacokinetics of Therapeutic proteins. The European Medicines Agency Evaluation of Medicines for Human Use. EMEA/CHMP/89249/04, London, United Kingdom.