

Background Information for Advisory Committee for Pharmaceutical Science

October 20, 2004

Concept and Criteria of BioINequivalence

Introduction

Bioequivalence is defined as “the absence of a significant difference in the rate and extent to which the active ingredient or active moiety in pharmaceutical equivalents or pharmaceutical alternatives becomes available at the site of drug action when administered at the same molar dose under similar conditions in an appropriately designed study...”. To evaluate bioequivalence, the U.S. Food and Drug Administration (FDA) has employed a testing procedure termed *the two one-sided tests procedure* (i) to determine whether the average values for the pharmacokinetic measures from the test and reference products are comparable. This procedure involves the calculation of a confidence interval for the ratio between the average values of the test and reference product. FDA considers a test product to be bioequivalent to a reference product if the 90% confidence interval of the geometric mean ratio of AUC and C_{\max} between the test and reference fall within 80-125% (ii).

Recently, the FDA has received several studies intended to show bio-inequivalence between two drug products, for example an innovator company might conduct a study to challenge FDA’s approval of generic versions of its drug product. Although there has not been a formal definition of the concept of bio-inequivalence in the regulation, intuitively, the concept of bio-inequivalence is not hard to perceive, given the well-defined concept of bioequivalence. However, there are no clear criteria to guide sponsors in conducting bio-inequivalence studies and FDA reviewers in assessing the validity of such bio-inequivalence studies. Because of a lack of a clear definition of bio-inequivalence, there has been some confusion and misunderstanding by the public.

Many questions arise when evaluating a bio-inequivalence claim. A typical question is if it is appropriate to claim bio-inequivalence when the two-sided 90% confidence intervals for the ratios of the PK endpoints do not fall inside the bioequivalence interval? There are numerous literature reports that claim bio-inequivalence based on a failed bioequivalence study without identification of the causes of the study failure. There are many ways that a bioequivalence study can fail, including an insufficient number of subjects. Many products that were claimed to be bio-inequivalent in the literature might well be bioequivalent if the studies were conducted appropriately. Therefore, it is imperative to develop and establish a bio-inequivalence criterion to clarify confusion and misunderstanding in the public.

In these presentations, we first introduce the concept of bio-inequivalence and present a statistical explanation for the proposed criterion to assess bio-inequivalence. We then

discuss several statistical strategies to assess bio-inequivalence studies with three pharmacokinetic endpoints (C_{max} , AUC_t and AUC_{∞}). The goal is to propose a set of criteria that are scientifically sound, statistically valid, and easy to use and to provide sufficient information to stimulate discussion on the evaluation of bio-inequivalence.

The concept of bio-inequivalence and test criteria

FDA's bioequivalence criteria require the 90% confidence interval of the ratio of the geometric means of the test and reference drug products to be within the bioequivalence interval [80%, 125%]. The definition of the bio-inequivalence region then is simply the region that lies outside the bioequivalence interval, i.e., (0, 80%) or (125%, ∞). Now the question is why a study failing to show bioequivalence cannot be used to claim bio-inequivalence. Once this question is answered, it will be a little easier to understand the statistical criteria proposed for bio-inequivalence claims.

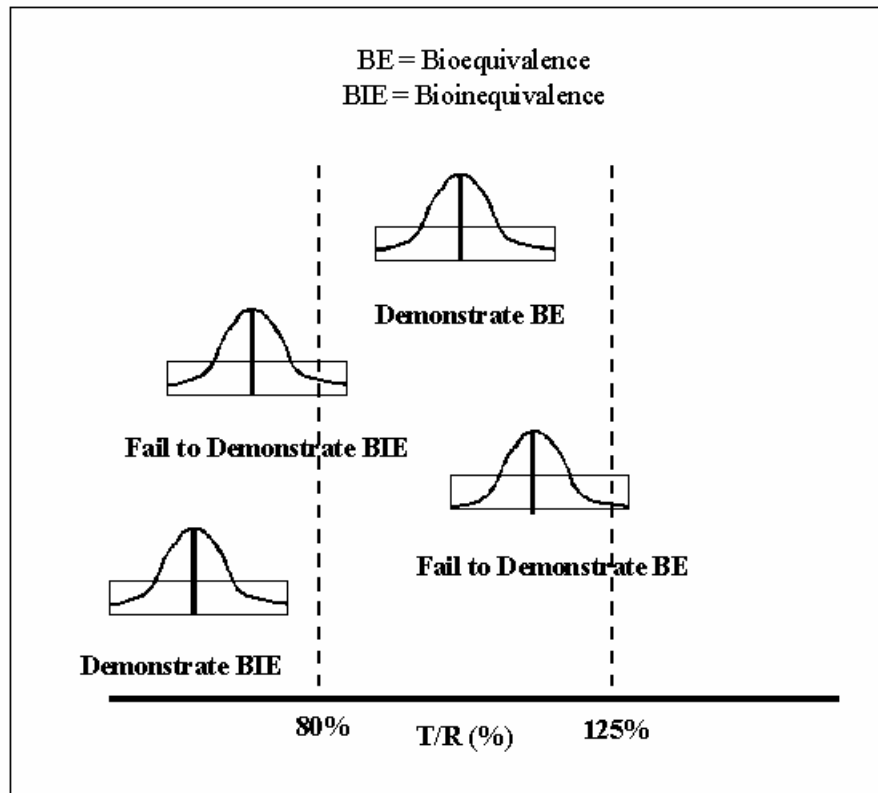
To answer this question, we need to understand statistically how the criteria for bioequivalence are formed. To test bioequivalence, the null hypothesis is set to be the bio-inequivalence region and the alternative hypothesis to be the bioequivalence interval. The goal is to see if bio-inequivalence can be rejected so that we may conclude that bioequivalence is true. For this purpose, it is important for the probability of an error that wrongfully rejects bio-inequivalence, and therefore falsely concludes bioequivalence, to be small. This error is usually controlled at the level of 0.05, which is the so-called significance level or the type I error rate. To reject the bio-inequivalence region, we need to perform two one-sided tests, each controlling the type I error rate at the level of 0.05. The maximum error rate in the two tests are actually controlled at the level of 0.05. The statistical criteria for rejecting bio-inequivalence and claiming bioequivalence are to have two-sided 90% confidence intervals (for the geometric mean ratio for each of the three PK endpoints) that are each within the bioequivalence interval. This procedure based on 90% confidence intervals is identical to carrying out the two one-sided tests described above.

To address whether failing to show bioequivalence demonstrates bio-inequivalence, we need to understand that in a bioequivalence test we usually do not control the error of wrongfully failing to conclude bioequivalence. If this error were controlled at a very low level, this would be equivalent to having very high power in a bioequivalence test. In order for both the significance level and power to be controlled at high level, a large sample size will generally be required, which will increase the cost of the study. For example, if we set the power to be 85%, and assuming the variance is 0.04, the sample size required is about 22, given the ratio of the two geometric means deviates from 1 by no more than 5%. In this case, the test could have about a 15% chance to fail to show bioequivalence even when the two drugs are truly equivalent. If the variance is larger than 0.04 and the ratio of the two geometric means deviates from 1 by more than 5% but still within the bioequivalence interval, the power could be much lower than 85% for the given sample size of 22. That is, the chance of failing to show bioequivalence would be much higher than 15% even when the two drugs are equivalent. Therefore, because there

is less control over the probability of failing to show bioequivalence, it is inappropriate to use a study that fails to show bioequivalence to claim bio-inequivalence.

Then why should the bio-inequivalence criterion be that the upper (lower) limit of the two-sided 90% CI should be less (greater) than 80% (125%)? As mentioned before, usually it is not realistic to control both types of errors, i.e., wrongfully rejecting bio-inequivalence and bioequivalence. A reasonable study only tightly controls one type of error. Therefore when testing for bio-inequivalence, we would like to control the error of wrongfully rejecting bioequivalence to be small. To be consistent with the bioequivalence testing, the error rate is also chosen at the level of 0.05. To reject bioequivalence, we also need to perform two one-sided tests, however, the level of each test may need to be 0.05. For one of the two tests to be significant at the 0.05 level, either the upper limit of the two-sided 90% CI has to be less than 80% or the lower limit to be above 125%.

Theoretically, it is possible for the type I error to reach 0.10 when a two-sided 90% CI is used to assess bio-inequivalence. However, this is true only when the variance of the estimated treatment difference (the ratio of geometric means) is very large. For typical crossover bio-inequivalence trials, such a large variance may not be a realistic possibility. Therefore, the type I error rate should be maintained at the level of 0.05 when two-sided 90% CI is used.



The above figure illustrates the different possible outcomes. A study with the two-sided 90% confidence interval completely between 80-125% demonstrates bioequivalence and allows market access. A study with the two-sided 90% confidence interval completely outside 80-125% demonstrates bio-inequivalence and may be grounds for market exclusion. A study with the point estimate within 80-125% but the two-sided 90% confidence interval outside of 80-125% fails to demonstrate bioequivalence. A study with the point estimate outside 80-125% but the two-sided 90% confidence interval overlapping 80-125% fails to demonstrate bio-inequivalence. Both of the failing cases would require studies with larger sample sizes to draw a definitive regulatory conclusion.

Evaluating the three PK endpoints collectively:

As mentioned earlier, based on the interpretation of regulation, FDA usually requires three pharmacokinetic endpoints (C_{max} , AUC_t , and AUC_{∞}) to show bioequivalence. All the two-sided 90% confidence intervals for the ratios of the geometric means for the three pharmacokinetic endpoints must be within the bioequivalence interval to demonstrate bioequivalence. If the 90% confidence interval for just one of the three pharmacokinetic endpoints does not fall completely within the bioequivalence interval, the study has not demonstrated that the two drugs are bioequivalent. However, the statistical criteria for testing bio-inequivalence using all the three pharmacokinetic endpoints will not be as simple. Here we discuss several strategies that potentially can be used for assessing bio-inequivalence using three pharmacokinetic endpoints. The evaluation of the strategies is based on both the error rate of wrongfully rejecting bioequivalence and power for detecting bio-inequivalence under various correlation structures.

One strategy that seems intuitive is to have at least one of the three pharmacokinetic endpoints satisfy the statistical criteria for bio-inequivalence, i.e., the upper (lower) limit of the two-sided 90% CI to be less (greater) than 80% (125%). However, this strategy could potentially inflate the error rate of wrongfully rejecting bioequivalence above the level of 0.05 if the three pharmacokinetic endpoints are not highly correlated.

The second strategy that is just the opposite of the first one discussed above is to require all the three pharmacokinetic endpoints to satisfy the statistical criteria for bio-inequivalence. This strategy can certainly control the error rate of wrongfully rejecting bioequivalence under all correlation structures. However, it may not always provide adequate power under alternatives that are of interest.

The third strategy that could protect the error rate of wrongfully rejecting the bioequivalence is to pre-specify one pharmacokinetic endpoint for bio-inequivalence testing. For example, one could pre-specify AUC_t and completely ignore the results of the other two pharmacokinetic endpoints. However, this strategy only has good power when AUC_t is the endpoint most likely to demonstrate bio-inequivalence. If only C_{max}

of the two drugs were bioinequivalent, then pre-specifying AUCt would give the test zero power to detect bio-inequivalence.

It is possible to develop a compromise approach. Instead of requiring all the three pharmacokinetic endpoints to satisfy the statistical criteria for bio-inequivalence with two-sided 90% confidence intervals as the measurement, we could have flexible width of the one-sided confidence intervals, while controlling the error rate at the level of 0.05 under all correlation structures. For example, it is possible to have one pharmacokinetic endpoint use a two-sided 92% confidence interval (slightly wider than 90% confidence interval) to show bio-inequivalence, while the second pharmacokinetic endpoint uses a two-sided 86% confidence interval (narrower than 90% confidence interval) and the third pharmacokinetic endpoint uses two-sided 80% confidence interval (much narrower than 90% confidence interval). For this strategy, it does not matter which pharmacokinetic endpoints uses which confidence interval. The advantage of this strategy is to use narrower confidence intervals to increase power to show bio-inequivalence, although at the cost of slightly widening one pharmacokinetic endpoint's confidence interval. Notice this strategy is developed using the assumption of a normal distribution. If the normal assumption is inadequate, it is possible to derive slightly different widths of confidence intervals under other distributions.

Summary of Bio-inequivalence discussions on April 14, 2004 ACPS Meeting

At the April 14, 2004 meeting, the FDA presented the concept of bio-inequivalence as well as statistical arguments for the support of using a two-sided 90% CI completely outside the bioequivalence interval as the criterion to establish bio-inequivalence claim. Our main focus for the discussion of bio-inequivalence criteria was on statistical issues related to power and error. The criterion to use two-sided 90% CI was established to use one PK endpoint to assess bio-inequivalence, given that the variance used to test bio-inequivalence is not large. The issue of using three PK endpoints to assess bio-inequivalence was also touched upon on April's ACPS meeting. We presented the pros and cons of several strategies to collectively evaluate the three PK endpoints for bio-inequivalence assessment. At the April 14, 2004 meeting, the following two questions were discussed:

1. Does the ACPS agree with the distinction between demonstrating bioINequivalence and failure to demonstrate bioequivalence?

Committee's comments: The Committee felt that there was a need to establish criteria for bioINequivalence evaluation and the criteria should not be just as failure of the bioequivalence test. The members argued it was important to focus on the clinical relevance with the therapeutic index. The Committee discussed both Area under the Curve (AUC) and Cmax as metrics important for bioequivalence and bioINequivalence.

2. Does the ACPS recommend a preferred method for evaluating the three pharmacokinetic endpoints for bioINequivalence?

- If bioINequivalence is demonstrated for any one pharmacokinetic endpoint, then bioINequivalence is demonstrated for the products.
- BioINequivalence must be demonstrated for all three pharmacokinetic endpoints for bioINequivalence to be demonstrated for the products.
- There should be one pre-selected pharmacokinetic endpoint used for bioINequivalence testing. If so, which one?
- The three pharmacokinetic endpoints should be evaluated for bioINequivalence with statistical corrections to the level of significance for each endpoint in order to maintain an overall significance level of 0.05.

Committee's comments: The Committee agreed on a general understanding of bioINequivalence to move forward by recognizing it is not a simple matter. In addition, the members felt this is an important concept, especially how it applies to the entire regulatory scenario. There was no consensus at this point as to a final criteria pertaining to the three pharmacokinetic endpoints.

Outline of Bioinequivalence discussion for Oct. 20 ACPS Meeting

The discussion on bioinequivalence will be focused on using three PK endpoints, AUC_t, AUC_∞, and C_{max}, to establish bioinequivalence between drug products. We will discuss strategies and recommendations under different scenarios for assessing bio-inequivalence using three PK endpoints.

If one of the three PK endpoints fail to satisfy the equivalence interval [80%, 125%], by definition, the bioequivalence is false and bioinequivalence is true. To test bioinequivalence using three PK endpoints, we would also like to control the error rate of wrongfully rejecting bioequivalence at the level of 5%.

As mentioned before, it is tempting to claim bioinequivalence between two drug products if one of the three endpoints demonstrates inequivalence by showing that a two-sided 90% CI falls in the bioinequivalence regions. This strategy could potentially inflate the overall type I error of wrongfully rejecting bioequivalence. For example, the error rate can be as high as 14.7% if no correlation exists ($\rho=0$), about 8.2% when pairwise correlation is 0.90, and 5.9% when the three endpoints are highly correlated ($\rho=0.99$), given the variance of test statistics is not too large. Therefore, this strategy is not acceptable for assessing bioinequivalence using three endpoints.

However, if one is certain that one PK endpoint could have the highest power to demonstrate inequivalence among the three, it is possible to pre-specify the PK endpoint to test for bioinequivalence using one PK endpoint at the level of 5%. This strategy will protect the error rate of wrongfully rejecting bioequivalence at the level of 5% and perform well if we know, at the design stage of study protocol, which PK endpoint is most likely to demonstrate bioinequivalence. Pre-specifying one PK endpoint is important in this strategy in order to protect the error rate. The disadvantage of this

strategy is that it could end up having low power if the pre-specified PK endpoint in fact does not have high power to show inequivalence.

If one can not pre-specify a PK endpoint, three PK endpoints should be used to demonstrate bioequivalence. For this scenario, the strategy that requires all three two-sided 90% CIs to fall outside the equivalence interval should not be recommended. Though this strategy protects the error rate tightly, it is overly stringent to show bioequivalence in most situations. An enhanced approach for this scenario is to use unequal width of CIs, such as 92%, 86%, 80% for the three CIs. This approach controls the error rate at 5% level regardless the correlation structure of the three endpoints, and can have enhanced power for the situations that all three PK endpoints may have some power to demonstrate bioequivalence, but it is uncertain which one would have the best power.

In this presentation, we will explain that so far there is not a universally powerful approach that can be used to evaluate three PK endpoints for bioequivalence. The sponsor of the bioequivalence studies should be able to make their own choice in selecting a strategy in evaluating bioequivalence. However, the choice must be pre-specified in bioequivalence study protocol before study is conducted.

Discussion Topic for ACPS Meeting

October 20, 2004

What is your preferred method for evaluating the three pharmacokinetic parameters for bioequivalence?

- If bioequivalence is demonstrated for any one pharmacokinetic parameter that is prespecified, then bioequivalence is demonstrated for the products.
- Bioequivalence must be demonstrated for all three pharmacokinetic parameters for bioequivalence to be demonstrated for the products, where the error rate is controlled at 5%.
- FDA should allow sponsors of bioequivalence studies to make their own choice on picking up strategies; sponsors should prespecify the choice in study protocols before the study is conducted.

References

- i D.J. Schuirmann. A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. *J. Pharmacokinet. Biopharm.* 15: 657-680 (1987).
- ii U.S. Food and Drug Administration, Center for Drug Evaluation and Research. Guidance for Industry: Bioavailability and Bioequivalence Studies for Orally Administered Drug Products--General Considerations, Office of Training and Communications, Division of Communications Management, Drug Information Branch, HFD-210, Rockville MD 20857, March 2003.