PhRMA Perspective

Névine Zariffa, MMath
Biostatistics and Data Sciences
SmithKline Beecham Pharmaceuticals

23 September 1999

PhRMA

1

Outline

- PhRMA Expert Panel
  - » Formation and mission
  - » Membership
- PhRMA Position Paper
  - » Objectives and approval process
- PhRMA Recommendations
  - » Specific
  - » General

PhRMA

2

3547s10

Formation and mission of PhRMA Expert Panel

- FDA Blue Ribbon Panel formed in 1998 to evaluate approaches outlined in 1997 Draft FDA Guidance
- 3 separate PhRMA reps (one Biostatistics, Clinical Pharmacology and Drug Metabolism/PK) involved with Blue Ribbon Panel; disparate views
- PhRMA Expert Panel comprising representatives from Biostats, DMPK and Clin Pharm was formed January 1999 with mission of . . .
  - » Deriving PhRMA consensus on FDA guidance
  - » Investigating alternatives to proposed methods
  - » Drafting expert report on PhRMA consensus view

PhRMA

3

PhRMA Expert Panel

Jeffrey S. Barrett	DuPont Pharmaceuticals, Newark, DE
Vijay Batra	Schering Plough, Kenilworth, NJ
Andrew Chow	R.W. Johnson PRI, Raritan, NJ
Jack Cook	Parke-Davis, Ann Arbor, MI
A. Lawrence Gould	Merck & Company, West Point, PA
Allen H. Heller	Bayer Pharmaceuticals, West Haven, CT
Man-Wai Lo	Merck & Company, West Point, PA
Scott D. Patterson	SKB Pharmaceuticals, Philadelphia, PA
Brian P. Smith	Eli Lilly and Company, Indianapolis, IN
Jeffrey A. Stritar	Pfizer Inc., Groton, CT
Jose M. Vega	Merck & Company, West Point, PA
Nevine Zariffa	SKB Pharmaceuticals, Philadelphia, PA

PhRMA

4

PhRMA Position Paper
Manuscript Objectives

- Review of ABE, its properties and limitations.
- PBE and IBE criteria as proposed by the FDA along with their properties and limitations.
- Recommendation for methodology to address each of the apparent limitations of the specific PBE and IBE criteria.
- General recommendations

PhRMA

5

PhRMA Position Paper
Manuscript Approval Process

- Original and revised draft reviewed by PhRMA Expert Panel
- Final draft sent to PhRMA Section Heads for additional comments
- Revised draft sent to PhRMA for legal review and to the Journal of Clinical Pharmacology
- Manuscript cleared by PhRMA on August 26 and accepted to JCP, August 27, 1999
- PhRMA will be issuing an addendum to comment on the other points raised in the draft guidance

PhRMA

6

PhRMA Position

Properties of the proposed PBE and IBE criteria

- The clinical relevance of a subject by formulation interaction has not been demonstrated.
- To date, no association between clinical failure and subject by formulation interaction has been demonstrated.

PhRMA

7

PhRMA Position

Properties of the proposed PBE and IBE criteria

- A consequence of the aggregate criteria is that a numerical tradeoff occurs between the various terms.
- The allowable difference between test and reference means is very sensitive to differences between variances permitting large rewards/penalties, and these differences are quite likely because estimates of variances tend be quite variable

PhRMA

8

PhRMA Position

Properties of the proposed PBE and IBE criteria

- The proposed criteria do not mandate hierarchical testing (means, variances, then σ^2_D in that order).
- Successful demonstration of IBE does not imply demonstration of PBE or of ABE.

PhRMA

9

PhRMA Position

Properties of the proposed PBE and IBE criteria

- While IBE seeks to ensure switchability between test and reference products, it does nothing to ensure switchability between two test products (i.e., generic to generic switching) which is expected to occur in practice.

PhRMA

10

PhRMA Position

Properties of the proposed PBE and IBE criteria

- The lack of global harmonization on the subject of bioequivalence, at least for some transition period, will place burden on sponsors and regulators involved in worldwide submissions.

PhRMA

11

PhRMA Position

General Comments

- New criteria should be transparent to regulators, prescribing physicians, pharmacists, and patients and provide a demonstrable improvement over the current criteria either in terms of the overall performance or simply in the handling of extreme cases such as narrow therapeutic index drugs or drugs with high variability.
- The proposed criteria for assessing PBE and IBE do not represent a significant improvement, at least in any demonstrable clinical or public health sense.

PhRMA

12

PhRMA Position

General Comments

- PBE and IBE do address some of the limitations of ABE but also introduce new limitations which could, in turn, present undesirable characteristics beyond those observed with ABE.

PhRMA

13

PhRMA Position

Specific Recommendations

- The clinical relevance of σ^2_D and its use as a surrogate for switchability could be studied by a targeted clinical pharmacology trial constructed to provide the best evidence of σ^2_D.
- Simulation techniques can be used to evaluate the statistical estimation procedures and quantify associated bias.

PhRMA

14

PhRMA Position

Specific Recommendations

- Tradeoffs between parameters, scaling and the maximal allowable difference all could be addressed by the use of an ordered testing procedure.
- Generic to generic switching can be addressed through suitable simulation studies.
- FDA and PhRMA should continue to engage in dialogue with other regulatory agencies and solicit their involvement in any proposed change.

PhRMA

15

PhRMA Position

General Recommendations

- While the PBE and IBE criteria as proposed by the FDA carry a number of statistical flaws, PhRMA believes these are minor in comparison to the issues outlined above and would be resolved through focussed effort and research
- PhRMA proposes that the current standard of average bioequivalence should continue as the basis for market access until another method is scientifically demonstrated to better serve the public interest.

PhRMA

16

PhRMA Position

General Recommendations

- PhRMA believes that the trial or phase-in period should be replaced by simulation studies -- a regulatory guidance should reflect a set of current practices and not a set of proposed studies to validate the guidance itself.
- PhRMA proposes that there may be other more effective ways of addressing the Public Health concerns without the burdens of the complexity, design and analysis of the proposed criteria such as the methodology described by Gould.

PhRMA

17

PhRMA Position

General Recommendations

- PhRMA proposes that an evaluation of Gould's method and the FDA proposed criteria be undertaken and will work with FDA to identify the standards of evaluation.
- The concept of scaling is appealing and PhRMA is committed to exploring the applicability and performance of any method utilizing it.
- Examining the performance of the proposed PBE and IBE criteria and its alternatives is something that PhRMA and FDA can, and shouid, do cooperatively.

PhRMA

18

PhRMA Perspective on Population and Individual Bioequivalence

PhRMA Expert Panel: Jeffrey S. Barrett [a], Ph.D., Vijay Batra [b], Ph.D., Andrew Chow [c], Ph.D., Jack Cook [d], Ph.D., A. Lawrence Gould [e], Ph.D., Allen H. Heller [f], M.D., Man-Wai Lo [e], Ph.D., Scott D. Patterson [g], Ph.D., Brian P. Smith [h], Ph.D., Jeffrey A. Stritar [i], M.D., Jose M. Vega [e], M.D., Nevine Zariffa [g], M.S.

[a] DuPont Pharmaceuticals, Newark, DE
[b] Schering Plough Research Institute, Kenilworth, NJ
[c] R.W. Johnson Pharmaceutical Research Institute, Raritan, NJ
[d] Parke-Davis Pharmaceutical Research, Ann Arbor, MI
[e] Merck & Company, West Point, PA
[f] Bayer Institute for Preclinical Drug Development, West Haven, CT
[g] SmithKline Beecham Pharmaceuticals, Philadelphia, PA
[h] Eli Lilly and Company, Indianapolis, IN
[i] Pfizer Inc., Groton, CT

To whom correspondence can be addressed:

Jeffrey S. Barrett, Ph.D.
DuPont Pharmaceuticals
Stine-Haskell Research Center, Bld. 112/1
P.O. Box 30, 1090 Elkton Road
Newark, DE 19714
Email: jeffrey.s.barrett@dupontpharma.com

Background

PhRMA represents approximately 100 U.S. companies with a primary commitment to pharmaceutical research devoted to providing medicines for maintaining and improving human health. PhRMA has studied the question of whether the proposed change in the Food and Drug Administration Guidance on Bioequivalence is warranted.

A panel of clinical pharmacologists, pharmacokineticists and biostatisticians from PhRMA member companies has worked to compile, review and address the various key discussion points. Their report has subsequently been reviewed within the PhRMA organization as a whole and is currently being disseminated for public discussion as a White Paper outlining the PhRMA position on the matter. PhRMA considers itself a full partner in this interesting and stimulating process that has identified many issues worthy of discussion.

Abstract

The FDA is expected to issue a second draft guidance in August 1999 on the subject of *in vivo* bioequivalence which is based on the concepts of individual and population bioequivalence (IBE and PBE respectively). The intention of this guidance is to replace the 1992 guidance that requires that in vivo bioequivalence be demonstrated by average bioequivalence (ABE). Although the concepts of population and individual bioequivalence are intuitively reasonable, a detailed review of the literature has not uncovered clinical evidence to justify the additional burden to the industry (both innovator and generic companies) and the consumer that the new guidelines would impose. The criteria for bioequivalence described in the draft guidance employ aggregate statistics that combine information about differences in bioavailability between formulation means, and differences in bioavailability variation of formulations between and within subjects. The purely technical aspects of the statistical approach are reasonably sound. However, PhRMA believes that important operational issues remain that need to be resolved before any changes to current practice are implemented.

PhRMA believes that the ideals of prescribability and switchability are intuitively reasonable, but is uncertain of the extent to which the proposed guidance can achieve these goals. It is not clear whether the attainment of such goals is necessary in the evaluation of bioequivalence given the role this plays in drug development, and the lack of clinical evidence argues against a pressing need to change current practice. PhRMA is concerned that the tradeoff offered by the aggregate criteria may ultimately represent more harm than good to the public interest.

PhRMA recommends more rigorous evaluation of methods based on two-way crossover designs before moving to methods that require more complex designs. One such method is identified herein and contains procedures for estimating prescribability and switchability.

The possibility of a 'phase-in' or 'trial' period to collect replicate crossover data in order to further evaluate IBE and PBE and possibly allow market access based on these criteria *as they are being evaluated* has been proposed. PhRMA believes this is unprecedented and will offer little additional information beyond that which can be obtained by simulation or has already been collected by FDA. Simulation studies have the advantage of allowing evaluation of the sensitivity of various procedures to represent the data patterns as created within the simulation. Operating characteristics by which proposed criteria can be adequately judged have not yet been defined. The limitations of ABE for highly variable drugs and narrow therapeutic drugs are well appreciated and may be addressed by means other than a wholesale change in the current criteria.

Introduction

PhRMA companies routinely perform Bioavailability/Bioequivalence trials to evaluate new formulations, compare different routes of administration and to evaluate the effect of food. Bioequivalence trials are used primarily to infer therapeutic equivalence of different formulations based on the similarity of the pharmacokinetic characteristics of the formulations tested. This paper reflects PhRMA's commitment to continued study and evaluation of potential improvements in bioequivalence procedures and outlines the PhRMA position on recent developments in this area and specifically on the recent FDA proposed guidance on bioequivalence [1].

Evidence of differences in bioavailability from various oral formulations of the same therapeutic agents had become apparent by the early 1960s. The ensuing 30 years have produced the body of scientific opinion, debate, and policy on the subject of bioequivalence. A timeline of significant events in the evolution of bioequivalence is shown in Table 1. A more-detailed description of these events can be found in the editorial by Chow [2]. To its credit, the motivating force behind many of these events has been the continued interest of the FDA to improve the manner in which these studies are conducted, the quality of the data generated from such studies and the methods by which they are evaluated. Not captured in Table 1 or in previous reviews of this subject are all of the numerous scientific meetings and publications that have been generated by this topic.

For the most part, the regulatory authorities outside the US have followed the lead of the FDA and issued similar bioequivalence guidances. Table 1 does not reflect all of the transactions outside the US. Of note, however, is the observation that bioavailability and bioequivalence was excluded as a topic from the ICH (International Conference on Harmonization) working groups. In fact, the observation made was that bioavailability and bioequivalence guidance in Europe, the US and Canada were already 'harmonized'.

The recent CPMP guidance (draft issued Dec 98) does not however reflect the current FDA interest in new criteria for assessing bioequivalence. The non-US regulatory community appears to be less interested in approaches involving population and individual bioequivalence, preferring to widen confidence intervals for highly variable drugs (de facto scaling) and to provide a decision tree for the necessity of conducting bioequivalence studies [3,4].

While many FDA guidances refer to the bioequivalence guidance (food effect, controlled release and drug interaction to name a few), the discussion which follows will focus uniquely on the bioequivalence guidance and the notions of Average, Population and Individual Bioequivalence (ABE, PBE and IBE). We will first present ABE and its properties and limitations. Second, we present the PBE and IBE criteria as proposed by the FDA along with their properties and limitations. Third, we recommend methodology to address each of the apparent limitations of the specific PBE and IBE criteria proposed by FDA. Last, we provide general recommendations on an overall approach to evaluate the concerns that led the FDA to propose these alternative criteria for bioequivalence.

1. The Current Criteria: Average Bioequivalence

The first bioequivalence criterion was set by the FDA in 1977 [5]. When the sample size was chosen so that one had 80% power of detecting a 20% difference with type I error of 5% ($\alpha = 0.05$), bioequivalence could be claimed if the null hypothesis for AUC and C_{max} could not be rejected. The 20% difference is arbitrary, but appears to have been chosen to satisfy clinical considerations [6]. Westlake [7] and Schuirmann [8] were two early proponents of a confidence interval or what is also called a two one-sided test approach. By 1989 and then again in 1992 [9], the FDA revised their criteria to incorporate a confidence interval approach.

Measures of bioavailability for single dose studies include the area under blood, serum or plasma drug concentration-time curve (AUC) and the peak blood, serum or plasma drug concentration (Cmax). Using log-transformed data, one establishes bioequivalence by showing that the 90% confidence interval of the ratio of geometric mean response (usually AUC and Cmax) of the two formulations is contained within the limits of 0.8 to 1.25 [9]. Equivalently, one could say that bioequivalence is established if the hypothesis that the ratio of geometric means is less than or equal to 0.8 is rejected with $\alpha = 0.05$ and the hypothesis that the ratio of geometric means is greater than or equal to 1.25 is rejected with $\alpha = 0.05$. Thus, this criterion has been termed a two one-sided test procedure. Although either presentation is correct, the confidence interval appears to be the preferred option presumably because of the ease of interpretation.

The confidence interval criterion provides a reliable indicator of the likelihood that the true average responses of two formulations are within 20% of each other. It places no restrictions on the trial design. The ABE criteria can be evaluated in parallel group studies, 2-period crossover studies, replicate design studies and a host of other designs.

Properties of Average Bioequivalence

A key characteristic of the current ABE criterion is that it is straightforward to interpret for the intended audiences including regulators, prescribing physicians, pharmacists and patients. The output of the analysis presented in terms of a ratio and confidence interval is easily interpretable to the prescribing physician -- a deemed necessity to medical consumers of such data given their responsibility for choosing among alternative therapies for patients. Pharmacokinetic parameters are generally observed to be log-normally distributed so that the geometric means are the appropriate measure of central tendency. Consequently, it is sensible to express the results of a BE trial in terms of the ratio of geometric means along with the 90% confidence interval for the ratios of the 2 geometric means. The regulatory limits of +/- 20% do not add complexity to the process and are readily interpretable.

While the current methodology has proven adequate in general, the use of a single set of regulatory limits for every drug has been questioned [10-12]. Narrow therapeutic index drugs and highly variable drugs with wide broad therapeutic windows may require different regulatory limits.

At one extreme, narrow therapeutic index drugs may very well deserve tighter regulatory limits to protect public health. For example, consider phenytoin, a narrow therapeutic index drug with nonlinear pharmacokinetics. Based on phenytoin pharmacokinetic parameter values in epileptic patients, it has been predicted that, for patients maintained at the mid-range of therapeutic plasma concentrations, a 10% reduction in bioavailability would result in concentrations below the lower limit of the therapeutic range in 14% to 44% of patients. A 10% increase in bioavailability would result in steady-state phenytoin plasma concentrations above the upper limit of the therapeutic range in 61% to 90% of patients [13].

The left panel of Figure 1 presents the results of two bioequivalence trials, both of which passed the current 20% regulatory limit. Yet the mean difference in extent of absorption between Formulation 1 and the reference formulation was greater than 10% and switching between the two products may have clinically important consequences. Thus, it may be prudent to require tighter criteria for narrow therapeutic drugs.

At the other extreme, wider regulatory limits may be more appropriate for highly variable drugs with wide therapeutic windows. The right panel of Figure 1 presents individual data from 30 subjects participating in a bioequivalence trial for a highly variable drug. Visual inspection does not reveal a consistent difference between formulations. Although the mean Cmax values differed by less than 10%, the 90% confidence interval failed to meet bioequivalence criteria because of the high variability and associated lack of statistical power. For such drugs, even more subjects would be required to ensure reasonable statistical power (i.e., a reasonable chance to declare bioequivalence when it is indeed the case).

The ABE criterion concerns only the distribution means, and not the variance of the distribution or any other parameter such as the 'subject by formulation' variance component which we will define later. These additional parameters reflect the extra attributes of 'prescribability' and 'switchability' [1] also discussed later. In theory, even a subgroup by formulation interaction (which is a special case of a subject by formulation interaction) could be undetected and bioequivalence declared using the ABE criteria as currently applied. The variance component related to subgroup by formulation would be absorbed into the residual variability thus inflating it. This would result in a widening of the 90% confidence interval, which may impact the demonstration of ABE.

The current ABE criterion is simple, straightforward in its interpretation, statistically sound and applicable under a variety of study designs. For most drugs (other than highly variable drugs), *in vivo* bioequivalence under the ABE criterion can be demonstrated in a modest number of subjects. However, the ease of interpretation of the current method is

its greatest strength. Any proposed improvement to ABE should retain this feature while demonstrating clear additional benefits.

2. Proposed Bioequivalence Criteria: Population and Individual Bioequivalence

The sections that follow describe the recommendations set forth in a recently issued draft Guidance on population and individual bioequivalence to address prescribability and switchability of alternative drug formulations [1]. We provide the basic statistical model for the data, the proposed bioequivalence criteria and the statistical analysis and implementation. We also consider the properties of the proposed criteria.

Population and individual bioequivalence, considered as concepts, recognize that formulations could differ in ways other than their average bioavailabilities. Population bioequivalence expresses equality of the distributions of test and reference formulation bioavailabilities across the population of subjects, while individual bioequivalence expresses equality of the distributions of test and reference formulation bioavailabilities for individual subjects.

The bioavailability of a drug formulation could be determined for any subject after each of a number of administrations of the formulation to the subject. These values will not all be the same, but will be sampled around a 'true' bioavailability for that subject. Likewise, different subjects can be expected to have different 'true' bioavailabilities that will be distributed around a 'true' bioavailability for the population from which the subjects are drawn. Let Y_{ijk} represent the bioavailability measured after the k-th administration of formulation i (i = T, R; where T = test formulation and R = reference formulation) to subject j,

$$Y_{ijk} = \mu_i + (\mu_{ij} - \mu_i) + (Y_{ijk} - \mu_{ij}) = \mu_i + b_{ij} + e_{ijk} \qquad \text{(Eq.1)}$$

where μ_i denotes the average bioavailability of formulation i for the population from which the subjects are drawn and μ_{ij} denotes the 'true' bioavailability of formulation i for

subject j. The quantity b_{ij} denotes the deviation of the 'true' bioavailability of formulation i for subject j from the average bioavailability of formulation i for the population of subjects. The quantity e_{ijk} denotes the deviation of the measured bioavailability following the k-th administration of formulation i to subject j from that subject's 'true' bioavailability. Assume b_{ij} and e_{ijk} are independent for a subject, i.e., that the deviation of the measured bioavailability following the k-th administration of a formulation to a subject from that subject's 'true' bioavailability is independent of the deviation of the subject's 'true' bioavailability from the population mean bioavailability.

The quantities b_{ij} and e_{ijk} in (Eq.1) have zero expectation by construction. Suppose that the quantities e_{ijk} are independently distributed with variance σ^2_{Wi}, and that the quantities b_{ij} are independently distributed with variance σ^2_{Bi}. The total variance of Y_{ijk} is $\sigma_i^2 = \sigma^2_{Bi} + \sigma^2_{Wi}$. This is the mixed effects model used in the draft FDA guidance and elsewhere in the literature [14-16].

Based on this model, various bioequivalence metrics can be defined to express the difference between the formulation bioavailabilities.

The differences between the average bioavailabilities of two formulations is assessed using the metric,

$$D_A = E^2(Y_{Tj} - Y_{Rj'}) = (\mu_T - \mu_R)^2$$

which is the square of the expectation of the difference between the average bioavailabilities of the two formulations.

A number of metrics more fully reflecting the bioavailability distributions have been described in the literature [14-21]. The computations for moment-based metrics are the most tractable, and the FDA guidance [1] is based on these.

Moment-based metrics are derived from expressions about the expected mean-squared difference between the bioavailabilities of various formulations. There are two basic expressions.

$$D_P = E\{(Y_{Tj} - Y_{Rj'})^2 - (Y_{Rj} - Y_{Rj'})^2\}$$

relates the expected squared difference between the bioavailabilities of the test and reference formulations when applied to different subjects to the expected squared difference between the bioavailabilities of two administrations of the reference formulations to different subjects. The other expression,

$$D_I = E\{(Y_{Tj} - Y_{Rj})^2 - (Y_{Rj} - Y_{R'j})^2\},$$

relates the expected squared difference between the bioavailabilities of the test and reference formulations when applied to the same subject to the expected squared difference between the bioavailabilities of two administrations of the reference formulations to the same subject.

These expectations can be expressed in terms of the moments of the marginal and joint distributions of the bioavailabilities of the two formulations. It can be shown that

$$D_P = (\mu_T - \mu_R)^2 + \sigma_T^2 - \sigma_R^2 \qquad \text{(Eq. 2a)}$$

and

$$D_I = (\mu_T - \mu_R)^2 + \sigma_{WT}^2 - \sigma_{WR}^2 + Var(b_{Tj} - b_{Rj}) = (\mu_T - \mu_R)^2 + \sigma_{WT}^2 - \sigma_{WR}^2 + \sigma_D^2 \quad \text{(Eq.2b)}$$

$\sigma_D^2 = Var(b_{Tj} - b_{Rj})$, the subject by formulation interaction, expresses the lack of consistency of subjects' true bioavailabilities on the test and reference formulations. This quantity will be zero when each subject's true bioavailabilities on the test and

reference formulations are higher or lower than the corresponding true population bioavailabilities by the same amount.

The population metric D_P in (Eq. 2a) reduces to the average bioequivalence metric $D_A = (\mu_T - \mu_R)^2$ when the total variances on the test and reference formulations are equal ($\sigma_T^2 = \sigma_R^2$). The individual metric D_I in (Eq. 2b) also reduces to D_A when there is no subject by formulation interaction and the bioavailability variation within subjects is the same for the test and reference formulations ($\sigma_{WT}^2 = \sigma_{WR}^2$).

Both D_P and D_I can be scaled to express the mean squared differences relative to the variability on the reference formulation, for example,

$$D_{PS} = D_P / \sigma_R^2 \qquad \text{(Eq. 3a)}$$

and

$$D_{IS} = D_I / \sigma_{WR}^2 \qquad \text{(Eq. 3b)}$$

Metrics such as (2a,b) are unscaled, while metrics such as (3a,b) are scaled.

One concludes bioequivalence in practice when data support with high confidence the assertion that the value of a metric does not exceed some critical value. Thus, population bioequivalence, which expresses the equivalent prescribability of the different formulations, would require $D_P < c_P$ or $D_{PS} < c_{PS}$, where c is some appropriate constant. Individual bioequivalence, which expresses the switchability of the formulations within the same patient, would require $D_I < c_I$ or $D_{IS} < c_{IS}$.

The decision strategy for population bioequivalence described in the FDA Guidance uses a combination of scaled and unscaled criteria. Thus, D_P would be used if $\sigma_R^2 < \sigma_0^2$, while D_{PS} would be used if $\sigma_R^2 > \sigma_0^2$, where σ_0^2 is a known constant. The decision

strategy for individual bioequivalence also combines the unscaled and scaled criteria: D_S would be used if $\sigma^2_{WR} < \sigma^2_{W0}$, while D_{IS} would be used if $\sigma^2_{WR} > \sigma^2_{W0}$, where σ^2_{W0} is a known constant.

Population and individual bioequivalence therefore are evaluated in terms of aggregate criteria that combine differences between the test and reference formulations with respect to population average bioavailability, variability among subjects' 'true' bioavailabilities, and, for IBE only, differences between the variabilities of each subject's bioavailabilities on repeated application of the formulation. Aggregate criteria are a consequence of expressing bioequivalence in terms of the moment-based criteria. Arguments have been advanced for their utility on practical grounds [22].

The criteria for evaluating population and individual bioequivalence are expressed in terms of values of parameters of distributions. These values ordinarily will be unknown, so must be estimated from data. The need to do so has implications for the design of trials for assessing population and individual bioequivalence. The parameters used in the criteria for evaluating PBE can be estimated from the outcomes of a standard 2 x 2 crossover trial. However, this design does not provide information for estimating all of the parameters used for evaluating individual bioequivalence. In particular, estimation of the within-subject variances σ^2_{WR} and σ^2_{WT} requires the results of repeated applications of each formulation to individual subjects. This kind of information is provided, for example, by designs such as TRTR/RTRT (2 sequence groups, 4 administration periods) or TRT/RTR (2 sequence groups, 3 administration periods). More elaborate designs could be used, although the Guidance states "...use of a replicated-crossover design with no more than two sequences is important" (p. 9) to avoid ambiguities in parameter estimation."

Given appropriate data, the analysis proceeds by constructing a 90% confidence interval for the true value of the metric, with acceptance of bioequivalence if the upper bound of

this confidence interval does not exceed the corresponding critical value. As described in the December 1997 guidance [1] this confidence bound could be derived using a bootstrap procedure. An alternative approximate procedure based on the Method of Moments and the Cornish-Fisher expansion is now being considered. The current unavailability of exact calculations for confidence bound in the December 1997 draft guidance also affects sample size calculations. The Guidance includes an appendix providing recommendations based on simulation results.

Properties of the proposed PBE and IBE criteria

In contrast to ABE, the proposed aggregate criteria are an attempt to include a number of variance components into the assessment of bioequivalence. This represents one possible method to assess prescribability and switchability. In addition, the scaling features of the PBE and IBE procedures accommodate highly variable drugs without the need to study an unduly large numbers of subjects. Given the initial model and assumptions, the mathematical theory is sound.

The proposed PBE and IBE criteria, along with their implementation offer a significant departure from the current ABE framework. The complex criteria give rise to a number of diverse and complex issues which require further study.

(1.) The clinical relevance of a subject by formulation interaction as measured by σ_D^2 has not been demonstrated. To date, no association between clinical failure and this interaction has been demonstrated. The issue of "burden-of-proof" remains a difficult one given the lack of data in support (or not) of clinical failures linked to σ_D^2, even though we have observed large numerical values of σ_D^2 in historical datasets. Data sets D and O2 in Zariffa et. al.[27] are examples of this. The clinical relevance of differences in variability between formulations has not been demonstrated. Moreover, the current statistical estimation procedure for σ_D^2 carries a small but noticeable bias [23 – 25].

(2.) A consequence of the aggregate criteria is that a numerical tradeoff occurs between the various terms. For example, a substantive difference in means can be compensated by decrease in within-subject variance in the test formulation relative to the reference formulation with the proposed IBE criteria. An example of this is shown in data set I2 for Cmax in Zariffa et.al.[27] although the tradeoff in total variances is further offset by scaling. Another more direct example is that of dataset 14a published on the FDA website (http://www.fda.gov/cder/bioequivdata/index.htm). The tradeoff was initially defined as a desired property of the IBE criteria with the intention that it would reward less-variable formulations. We note, however, that the allowable difference between test and reference means is very sensitive to differences between variances permitting large rewards/penalties, and these differences are quite likely because estimates of variances tend be quite variable themselves.

(3.) The proposed criteria do not mandate hierarchical testing (means, variances, then σ_D^2 in that order). In other words, successful demonstration of IBE does not imply demonstration of PBE or of ABE. Data set A for Cmax[27] is an example of this as is the dataset 14a mentioned above. Without this hierarchical structure between the criteria, inconsistent inferences may exist, especially if a single trial is to be used to satisfy regulatory authorities worldwide. In addition, the maximum tradeoff between terms is not explicitly specified by the regulators, but determined by the procedure itself. This is certain to be problematic with respect to the ultimate inference of a given trial as there are no "built-in" checks for each parameter in the proposed procedure.

(4.) While IBE seeks to ensure switchability between test and reference products, it does nothing to ensure switchability between two test products (i.e., generic to generic switching) which is expected to occur in practice. It has been shown that the difference between 2 generic formulations shown to be bioequivalent to a same reference product under ABE would be unlikely to exceed 0.80 – 1.25 [26]. The assessment of generic to generic switches under the IBE criteria has yet to be studied in detail.

(5.) The lack of global harmonization on the subject of bioequivalence, at least for some transition period, will place burden on sponsors and regulators involved in worldwide submissions. As a consequence of lack of commutativity inherent to equations 2a and 2b, a specific drug product formulation may gain market access in one country but not another. The resolution of such discrepancies is likely to be complex and has not been addressed.

3. Recommendations

The concept of bioequivalence as a basis to ensure therapeutic equivalence (safety and efficacy) is central to regulatory approval of improved marketed formulations, changes in manufacturing processes, and the development of product line extensions (e.g., chewable tablets, oral solutions, etc.) to meet patient needs. The concept of bioequivalence is critically important to the provision of generic formulations to consumers after an innovator drug product has gone off patent.

Data from in vivo bioequivalence studies provide confidence to physicians who prescribe medications and protection to the patients who receive them. Given the need to provide the medical community and the public with clear, interpretable, and accurate information on the equivalency of two drug products, any modification to the bioequivalence criteria now in use should further this objective. If new criteria are chosen to study bioequivalence, they should be transparent to regulators, prescribing physicians, pharmacists, and patients and provide a demonstrable improvement over the current criteria either in terms of the overall performance or simply in the handling of extreme cases such as narrow therapeutic index drugs or drugs with high variability.

There has been no documented evidence of clinical failure associated with a formulation demonstrated to be equivalent to the reference product under ABE [28]. While there are

limitations to ABE as discussed earlier, the proposed criteria for assessing PBE and IBE do not represent a significant improvement, at least in any demonstrable clinical or public health sense. PBE and IBE, at least as proposed, do address some of the limitations of ABE. The proposed IBE / PBE criteria introduce new limitations which could, in turn, present undesirable characteristics beyond those observed with ABE.

PhRMA recognizes that the limitations of PBE and IBE presented earlier can be studied and their seriousness assessed in cooperation with FDA and academic centers. We propose the following method of evaluation for each item stated previously.

Specific Recommendations

(1.) The clinical relevance of σ_D^2 and its use as a surrogate for switchability could be studied by a targeted clinical pharmacology trial constructed to provide the best evidence of σ_D^2. Simulation techniques can be used to evaluate the statistical estimation procedures for σ_D^2 and quantifying the associated bias.

(2/3.) Tradeoffs between parameters, scaling and the maximal allowable difference all could be addressed by the use of an ordered testing procedure. One would first evaluate the mean differences, then the variance differences and lastly σ_D^2. Discussion of the regulatory requirements for each step would be needed but the general concept provides a framework for limiting the extent to which tradeoffs can occur. The effect of scaling cannot be studied in isolation and must be combined with the other factors at play.

(4.) Generic to generic switching can be addressed through suitable simulation studies. Generic to generic differences in bioavailability have been studied in the context of ABE [26]. While similar studies have not been carried out under IBE, it is expected that the differences allowed between various generic formulations will be greater.

(5.) In order to maintain the spirit of global harmonization, it is reasonable to expect that the FDA and PhRMA will continue to engage in dialogue with other regulatory agencies and solicit their involvement in any proposed change.

General Recommendation

PhRMA recognizes the issues stated in the 1997 draft guidance, but does not believe that the approach outlined in the guidance will effectively address these concerns. The aggregate criteria proposed in the guidance are not easily interpretable and may lead to undesirable trade-offs. The case for the additional burden of replicate designs as required for the individual bioequivalence criterion is not established. In fact, there is agreement that further information is needed as to the frequency of important subject by formulation interactions, and further information should be collected. With this in mind, the FDA's Working Group has proposed an interim period that will allow data collection to support the need for individual bioequivalence. During this period, replicate study designs would be recommended for both population and individual bioequivalence studies. It is possible that sponsors would be allowed to choose the criteria of their preference (ABE, PBE or IBE) in the protocol. Such a process is unlikely to yield meaningful results as the datasets collected may represent a narrow range of therapeutic agents and ultimately be too few in number to provide clear answers. In any event, the true distribution of the formulations' bioavailabilities would not be known. We would only have an estimated value from the sample of subjects in the particular studies. PhRMA proposes that simulation studies be undertaken to circumvent this problem. Simulations would be more valuable in evaluating the sensitivity and specificity of the proposed criteria relative to the simulation parameters. The fact that sponsors would have an option of an a priori choice of either the old or new criteria raises the concern that market access would in some cases be based on criteria under study which may in turn be found to be inappropriate and not based on scientifically established and accepted

criteria. PhRMA's view is that a Regulatory Guidance should reflect a set of current practices and not a set of proposed studies to validate the guidance itself.

PhRMA proposes that the current standard of average bioequivalence should continue as the basis for market access (especially for ANDAs and SNDAs) until another method is scientifically demonstrated to better serve the public interest.

The criteria proposed in the 1997 draft guidance present one option in terms of the implementation of what may generally be referred to as Population and Individual Bioequivalence. There may, however, very well be other more effective ways of addressing the Public Health concerns without the burdens of the complexity, design and analysis of the proposed criteria. One potentially attractive alternative is the more straightforward methodology described by Gould [28]. The method is briefly described below for the case of two-period crossover designs.

The first step in the analysis is to evaluate average bioequivalence using standard ABE methods as previously described. If ABE is demonstrated, the next step is to evaluate prescribability as follows. After log(e)-transforming all observations, first, subtract the sequence by period cell means from each observation in the appropriate sequence by period cell to derive a collection of residuals corresponding to the original observations. Then add each subject's period 1 and 2 residuals (call these values s), and calculate the test - reference difference in residuals for each subject (call these values d). Now calculate the slope of a simple regression of the s on the d values, and construct a 90% confidence interval for the slope. Prescribability could be rejected from a regulatory standpoint if the 90% confidence interval does not include a preset value (for example, 0). If prescribability is demonstrated, switchability is evaluated as follows. Calculate the sums of squares and cross products of the residuals in the 2 periods for each sequence. Pool these quantities across sequences and calculate a conventional correlation coefficient using the pooled values. Switchability could be accepted from a regulatory standpoint if the 90% confidence interval for the correlation was completely

contained above a preset value (for example, 0.5). Gould's method provides an alternative approach that can be applied with standard 2 x 2 crossover designs, and that provides evaluations of prescribability and switchability.

PhRMA proposes that an evaluation of Gould's method and the FDA proposed criteria be undertaken. PhRMA will work with FDA to identify the standards of evaluation. The methods would be compared with respect to the consistency of their decisions regarding whether average, population, or individual bioequivalence had been demonstrated under a range of simulated scenarios. Additional comparisons could be undertaken using simulation studies with known parameter values as described earlier. This may prove particularly useful in the case where the methods yield discordant results.

It has been noted that the PBE and IBE criteria as proposed by the FDA carry a number of statistical flaws. PhRMA believes these are minor in comparison to the issues outlined above and would be resolved through focussed effort and research.

If aspects of formulation distributions other than their means prove to be relevant, PhRMA recommends the use of a disaggregate procedure where each relevant parameter would be evaluated separately. We believe that it is important to require average bioequivalence as a prerequisite for population bioequivalence, and population bioequivalence as a prerequisite for individual bioequivalence. This hierarchy controls the overall probability of erroneously obtaining a successful outcome (Type 1 error) by causing the separate tests for average, population, and individual bioequivalence to form a closed set [29].

The guidance also allows for scaling to the reference product variability. This feature might be expected in principle to deal better with the problems of highly variable and narrow therapeutic index drugs (NTI). The offset of the proposed scaling rule for PBE and IBE may prove to be too liberal in the case of highly variable drugs and at the

present time has not yet been proposed for NTIs. The concept of scaling is appealing and PhRMA is committed to exploring the applicability and performance of any method utilizing it.

Conclusion

Examining the performance of the proposed PBE and IBE criteria and its alternatives is something that PhRMA and FDA can, and should, do cooperatively.

REFERENCES

1. Food and Drug Administration: Guidance for Industry: In vivo bioequivalence studies based on population and individual bioequivalence approaches 1997.
2. Chow S-C: Guest editorial note: Bioavailability and bioequivalence. *Drug Information* 1995; 29: 793-4.
3. Gleiter CH et al: When are bioavailability studies required? A German proposal. *Journal of Clinical Pharmacology* 1998; 38: 904-911.
4. European Agency for the Evaluation of Medicinal Products, Committee for Proprietary Medicinal Products: Notes for Guidance on the Investigation of Bioavailablity and Bioequivalence (Draft) 1998.
5. Federal Register 1997; 42:1648.
6. Benet LZ: In: Midha K.K, Blume H., eds Bio-International Bioavailability: Bioequivalence and Pharmacokinetics. Stuttgart: Medpharm 1993; 27-35.
7. Westlake WJ: Symmetric confidence intervals for bioequivalence trials. *Biometrics* 1976; 32:741-744.
8. 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 Biopharmaceitics* 1987; 15:657-680.
9. Food and Drug Administration: Guidance on Statistical Procedures for Bioequivalence Studies Using a Standard Two-Treatment Crossover Design. 1992.
10. McGilveray IJ et al: Consensus Report from "Bio International '89. Issues in the evaluation of bioavailability data. *Journal Pharm Sci* 1990; 10:945-946.

11. Shah VP et al: Workshop report: Evaluation of orally administered highly variable drugs and drug formulations. *Pharmaceutical Research* 1996; 13:1590-1594.

12. AAPS Workshop on scientific and regulatory issues in product quality: Narrow therapeutic index drugs and individual bioequivalence. March 16-18, 1998 Arlington, VA.

13. Ludden TM et al: Sensitivity analysis of the effect of bioavailability or dosage form content on mean steady-state phenytoin concentrations. *Therapeutic Drug Monitoring* 1991; 13:120-125.

14. Ekbohm G and Melander H: The subject-by-formulation interaction as a criterion of interchangeability of drugs. *Biometrics* 1989; 45:1249-1254.

15. Schall R and Luus HG: On population and individual bioequivalence. *Statistics in Medicine* 1993; 12:1109-1124.

16. Schall R: Assessment of individual and population bioequivalence using the probability that bioavailabilities are similar. *Biometrics* 1995; 51:615-626.

17. Liu J-P and Chow SC: On the assessment of variability in bioavailability/ bioequivalence studies. *Communications in Statistics Part A Theory and Methods* 1992; 21:2591-2607.

18. Sheiner LB: Bioequivalence revisited. *Statistics in Medicine* 1992; 11:1777-1788.

19. Guilbaud O: Exact inference about the within-subject variability in 2 x 2 crossover trials. *Journal of the American Statistical Association* 1993; 88:939-946.

20. Endrenyi L: A method for the evaluation of individual bioequivalence. *International Journal of Clinical Pharmacology and Therapeutics* 1994; 32:497-508.

21. Anderson S and Hauck WW: Consideration of individual bioequivalence. *J Pharmacokinet Biopharm* 1990; 18:259-273.

22. Hauck WW et al: Mean difference vs. variability reduction: tradeoffs in aggregate measures for individual bioequivalence. *International Journal of Clinical Pharmacology and Therapeutics* 1996; 34:535-541.

23 Endrenyi L et al: Individual bioequivalence: Attractive in principle, difficult in practice. *Pharmaceutical Research* 1998; 15:1321-1325.

24 Endrenyi L and Midha K: Individual bioequivalence – has its time come? *European Journal of Pharmaceutical Sciences* 1998; 6:271-277.

25 Endrenyi L and Tothfalusi L: Subject by formulation interaction in determinations of individual bioequivalence: Bias and prevalence. *Pharmaceutical Research* 1998; 16:186-190.

26 Anderson S and Hauck WW: The Transitivity of bioequivalence testing: potential for drift. *International Journal of Clinical Pharmacology and Therapeutics* 1996; 34:369-374.

27. Zariffa et al: Case studies, practical issues and observations on population and individual bioequivalence. *Statistics in Medicine* (In press).

28. Gould AL: A practical approach for evaluating population and individual bioequivalence. *Statistics in Medicine* (In press).

29 Marcus R et al: On closed testing procedures with special reference to ordered analysis of variance. *Biometrika* 1976; 63:655-660.

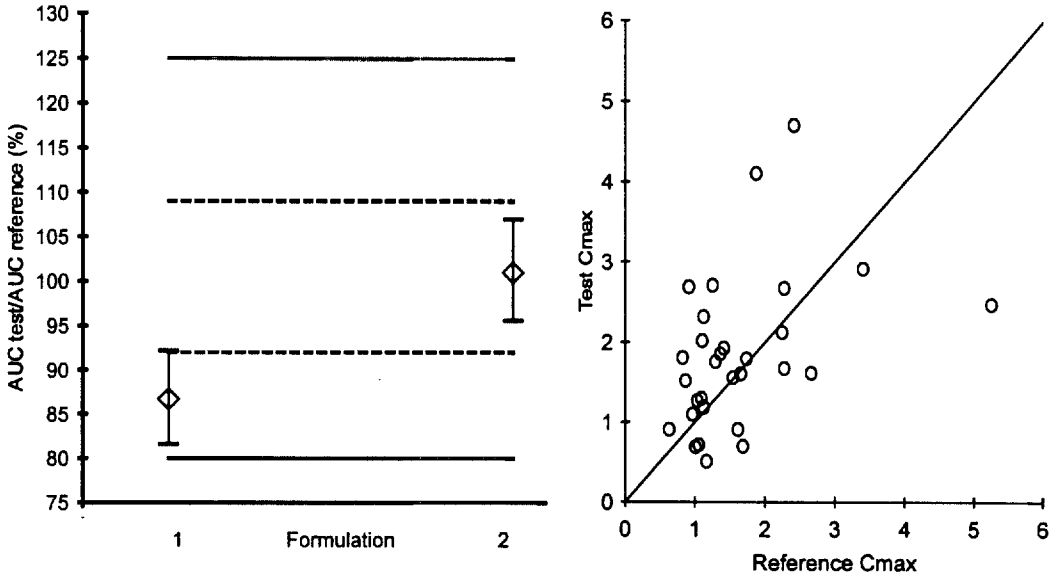
Table 1. Significant events in the historical evolution of bioequivalence

Date	Event
1970	• FDA asks for evidence of biologic availability in NDA applications.
1974	• Drug Bioequivalence Study Panel formed by the Office of Technology Assessment to examine the relationship between chemical and therapeutic equivalence of drug products. • Panel recommendations implemented by FDA and codified in 21 CFR Part 320
1977	• FDA proposed and finalized regulations for bioequivalence requirements and in vivo bioavailability procedures (42 FR 1624)
1984	• FDA authorized to approve generic drug products under the Drug Price Competition and Patent Term Restoration Act.
1986	• FDA hearing on the bioequivalence of solid dosage forms • Task Force formed to evaluate the current procedures adopted by FDA for assessment of bioequivalence between immediate release solid dosage forms.
1988	• FDA Controlled Release Workshop • Bioequivalence Task Force issues report
1992	• FDA issues guidance for the assessment of bioequivalence using a standard two-treatment crossover design • HPB issues guidance on conduct and analysis of bioavailability and bioequivalence studies
1997	• FDA issues draft guidance on in vivo bioequivalence studies based on population and individual bioequivalence approaches
1998	• AAPS/FDA Workshop on Individual Bioequivalence and NTIs • Formation of the FDA Blue Ribbon Panel to evaluate the approaches outlined in the draft guidance • CPMP (European Agency for the Evaluation of Medicinal Products) issues draft guidance on the investigation of bioavailability and bioequivalence. [3]
1999	• PhRMA forms PhRMA Expert Panel to evaluate the draft guidance and define PhRMA consensus on approaches and guidance. • FDA issues draft guidance outlining waiver of BA/BE studies for immediate release solid oral dosage forms based on a biopharmaceutical classification system • FDA issues two additional draft guidances which replace the 1997 draft guidance (expected August 1999)

Legends for Figures

Figure 1. Results of Bioequivalence Trials for a Narrow Therapeutic Index Drug (left panel) and Highly Variable Drug (right panel). Left panel: Diamonds and error bars represent the 90% confidence interval for 2 different phenytoin formulations vs Dilantin; solid lines represent FDA confidence interval criteria and dashed lines represent possibly more appropriate criteria. Right panel: Points represent individual subject data; mean Cmax ratio was 90% with a confidence interval of 78 to 104%.

Figure 1


AUC test/AUC reference (%)
130
125
120
115
110
105
100
95
90
85
80
75
1
Formulation
2
Test Cmax
Reference Cmax
0
1
2
3
4
5
6