Research Article

Performance Properties of the Population Bioequivalence Approach for *In Vitro* Delivered Dose for Orally Inhaled Respiratory Products

Beth Morgan^{1,2} and Helen Strickland¹

Received 24 July 2013; accepted 24 October 2013; published online 19 November 2013

Abstract. Regulatory agencies, industry, and academia have acknowledged that *in vitro* assessments serve a role in establishing bioequivalence for second-entry drug product approvals as well as innovator post-approval drug product changes. For orally inhaled respiratory products (OIPs), the issues of correctly analyzing *in vitro* data and interpreting the results within the broader context of therapeutic equivalence have garnered significant attention. One of the recommended statistical tests for *in vitro* data is the population bioequivalence method (PBE). The current literature for assessing the PBE statistical approach for *in vitro* data assumes a log normal distribution. This paper focuses on an assessment of that assumption for *in vitro* delivered dose. Concepts in development of a statistical model are presented. The PBE criterion and hypotheses are written for the case when data follows a normal distribution, rather than log normal. Results of a simulation study are reported, characterizing the performance of the PBE approach when data are expected to be normally distributed, rather than log normal. In these cases, decisions using the PBE approach are not consistent for the same absolute mean difference that the test product is from the reference product. A conclusion of inequivalency will occur more often if the test product dose is lower than the reference product for the same deviation from target. These features suggest that more research is needed for statistical equivalency approaches for *in vitro* data.

KEY WORDS: *in vitro* delivered dose; population bioequivalence approach; simulations; statistical model.

INTRODUCTION

The determination of bioequivalence for orally inhaled products has been a topic of interest in the pharmaceutical industry in the last decade. Concepts related to bioequivalence and bioavailability are defined in the Federal Food, Drug, and Cosmetic Act in 21 USC 355(j)(8)(A), 355(j)(8)(B), and 355(j)(8)(C). From that legislation, the definition for bioequivalence was repeated in the 2003 US Food and Drug Administration (FDA) Draft Guidance for Industry: Bioavailability and Bioequivalence Studies for Nasal Aerosols and Nasal Sprays for Local Action (1), as: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.”

Bioequivalence is assessed from the comparison of bioavailability between two drug products. In this same guidance document, bioavailability was defined as:

“the rate and extent to which the active ingredient or active moiety is absorbed from a drug product and becomes available at the site of action. For drug products that are not intended to be absorbed into the bloodstream, bioavailability may be assessed by measurements intended to reflect the rate and extent to which the active ingredient or active moiety becomes available at the site of action.”

Bioequivalence and bioavailability are then used to establish therapeutic equivalence. As Lee *et al.* stated, demonstrating bioequivalence for orally inhaled products and nasal sprays is challenging because of the “incomplete understanding of the relevance of drug concentration in blood or plasma to equivalence in drug delivery to the local site(s) of action” (2). This has influenced the FDA to adopt an aggregate “weight of evidence” approach, based on demonstrating equivalent *in vitro* performance through *in vitro* studies, equivalent systemic exposure through pharmacokinetic or pharmacodynamic studies, and equivalent local action through clinical endpoint studies (3). This paper focuses on the topic of demonstrating equivalent *in vitro* performance, particularly the equivalence of delivered dose.

¹ Global Manufacturing and Supply, GlaxoSmithKline, 1011 N. Arendell Avenue, Zebulon, North Carolina 27597, USA.

² To whom correspondence should be addressed. (e-mail: beth.e.morgan@gsk.com)

Draft guidelines for equivalency of *in vitro* data were issued as part of the 1999 guidance document entitled, Guidance for Industry: Bioavailability and Bioequivalence Studies for Nasal Aerosols and Nasal Sprays for Local Action (4). Analysis approaches for different *in vitro* tests, labeled profile and non-profile, were introduced. Examples of non-profile tests included content through container life, droplet size distribution, spray pattern, and drug in small particles. Profile analyses were related to particle size distribution as measured through cascade impaction. The recommended statistical methodology for assessing equivalence in four of the *in vitro* tests was the population bioequivalence criterion (PBE). In 2003, this draft guidance document was updated. The statistical appendices, to contain details for the recommended analysis procedures, were listed in the updated document, but were not issued. More recently, at the GPhA/FDA 2012 Fall Technical Conference, the Office of Generic Drugs presented recommendations for implementing the PBE analysis for *in vitro* data as well as scientific and regulatory considerations for bioequivalence of dry powder inhalers (5). Recommendations from this conference were then reflected in recent product-specific guidance for budesonide and albuterol sulfate (6,7).

The European Medicines Agency states that limits of $\pm 15\%$ should be applied to the mean of the delivered dose when establishing drug product specifications. Comparative *in vitro* data must be provided. Exact statistical methodology is not detailed (8).

The current literature for assessing the PBE statistical approach for equivalence of *in vitro* data assumes that the log normal distribution is, fundamentally, the correct statistical model to describe the data collected. This is the case for both Chow *et al.*'s article in 2003, in which the statistical model and PBE methodology was extended to include replication per unit, and Chiu *et al.*'s article in 2010, in which a nested random effects model was proposed (9,10). This paper focuses on an assessment of that assumption for *in vitro* delivered dose and the practical impact to decision making when that assumption is not appropriate.

In the **Methods** section, concepts regarding development of a statistical model are introduced and statements regarding the processes that generated the data are reviewed under different model assumptions. This section contrasts the differences between processes that generate *in vivo* and *in vitro* data to highlight why consideration of different statistical models is warranted. Also in this section, the PBE criterion and associated hypotheses, applied to log-transformed data, are expressed as a function of parameters in the original data metric when data follow a normal distribution. The forms of those hypothesis statements, in the original metric, then establish the basis of the work presented in the **Results** section.

In the **Results** section, both a graphical review of the PBE as a function of the population parameters in the original metric and the results of a simulation study are presented and discussed. The simulation study was designed to better understand the practical impact of applying the PBE to log-transformed data when the data are expected to follow a normal distribution in the original metric.

This paper presents results of research focused on one of the primary statistical methods for equivalency assessment for *in vitro* data. A subsequent paper will follow that presents

different population requirements, representing statements of equivalency between two data distributions, and the performance properties of different statistical methodologies for those population statements.

Taken together, these papers serve to provide an initial structure in how to consider an appropriate statistical model for *in vitro* data and then, how to phrase equivalency statements as statistical hypothesis statements of the population parameters, for which the risks of different statistical methods and decision rules can then be assessed. This overall process can be extended into more complicated models for *in vitro* data or extended to other *in vitro* tests for which an assessment of statistical equivalence is required.

METHODS

In this section, a review of statistical models for delivered dose is conducted. Contrasts are made between the processes generating *in vivo* and *in vitro* data to highlight why consideration of different statistical models is warranted. An assessment of the PBE criterion as a function of parameters in the original data metric is presented.

Review of Statistical Models for *In Vitro* Delivered Dose

In vitro delivered dose is a measure of drug content available to the patient in each dose of drug product. The data are obtained from analytical testing implemented in the laboratory on units sampled from the final drug product batch and are typically reported as a percentage of the label claim dose. From each batch of drug product, a pre-defined number of units are sampled and a pre-determined number of doses are measured. Regulatory specifications are imposed on the mean delivered dose and/or on a function of individual values, depending on the intended commercial market. The registered specifications for dose, according to ICH Q6, are "critical quality standards that are proposed and justified by the manufacturer and approved by regulatory authorities as condition of approval" to be released to market (11). The detailed sampling and testing scheme that defines how the units are selected from the batch, how the analytical method is conducted, and how the product quality is demonstrated can vary. Guidance documents, regulations, and compendial standards exist to assist applicants in the development process (12–18).

A critical aspect of applying a statistical methodology for decision making is determining the statistical model and the nature of the error distribution. As with *in vivo* bioequivalence, it makes more sense to base decisions on the nature of the error distribution from first principles, rather than *p*-values from goodness-of-fit approaches.

The next sections review statements about delivered dose and the processes that generate the data. The intent is to assess the appropriateness of a statistical model from first principles. We will then show, using simulation, the impact of applying a log normal transformation to *in vitro* data when a normal distribution is more appropriate.

Normal Distribution Properties for *In Vitro* Delivered Dose

Manufacturing processes are developed and validated to deliver the label claim dose to a patient and to ensure

continuing drug product quality and performance. Analytical methods, intended to measure attributes of the units produced from these manufacturing processes, are developed to provide an accurate quantification of the chemical, physical, biological, or microbiological attributes of the units. Sources of variation, however, are acknowledged to exist in any successful manufacturing or analytical testing process (19–27).

Dose measurements often vary around the target label claim value in a symmetric manner, as reflected in the compendial specifications for delivered dose in USP Chapter 601 (17). As outlined in the process validation guidance, understanding variation is the essence of a successful validation and manufacturing program throughout the lifecycle of the product. As stated in this guidance, manufacturers should “understand the sources of variation, detect the presence and degree of variation, and understand the impact of variation on the process and ultimately on product attributes” (24).

Variation, therefore, is an inherent component of manufacturing processes and analytical testing procedures. For orally inhaled drug products, variability can be introduced due to raw material differences, manufacturing variability at different unit operation stages (e.g., micronization), blend or concentration variation, equipment performance variability, filling or mixing variation, device or component variability, and finally, analytical testing and analytical equipment variability.

In manufacturing, variation introduced from one-unit operation can often follow through other processes and result in variation in the end product. For example, uniformity issues in concentration of the active pharmaceutical ingredient could be introduced during blending, and during the next unit operation, the fill weight variation would not be impacted and the uniformity problems would, then, carry through until final drug product testing and release. Analytical variability or device variability is not expected to depend or be relative to data from a previous unit operation. In this way, the sources of variation in the final drug product quality are additive, rather than multiplicative (28). Additive sources of variation are more typically modeled by a normal distribution. This is also reflected in the nature and form of the specifications as well.

In its simplest form, the statistical model that describes data in which results vary around a target value, deviations from target are expected to occur additively and in a symmetric manner, and variation is independent of the mean is the normal distribution. In its simplest form, this model is written as:

$$y_{ij} = \mu + \varepsilon_{ij}$$

where

y_{ij} represents the j th dose measurement from the i th batch

μ = target label claim (percentage)

ε_{ij} represents the summation of all the components of variation from the unit operations in manufacturing and testing that impact the final dose measurement $\sim \text{Normal}(0, \sigma_{\text{Total}}^2)$

The model above does not directly estimate components of variation due to batch differences or variation in delivered

dose within a device or unit, sampled from a particular batch. The basic statistical model above can be extended to include those components, and therefore, the total variation would be a sum of factors that represent batch variation, unit or device variation within a batch, and remaining unexplained manufacturing and testing variation.

The normality assumption of delivered dose was further supported by an analysis of industry data from the database collected by the Inhalation Technology Focus Group (ITFG) of the American Association of Pharmaceutical Scientists (AAPS) in collaboration with the International Pharmaceutical Aerosol Consortium on Regulation and Science (IPAC-RS) (28).

This analysis showed that the overall distribution of delivered dose data was symmetrical and that, after standardization, the fit to a normal distribution was excellent. This was the case even into the tails of the distribution. The ITFG/IPAC-RS DCU database contained data for 80 products with a total of 46,816 individual determinations. In order to pool all the data from different products, all products were standardized to a mean of zero and a standard deviation of unity, in order to show the general shape of the distribution. The histogram of these data presented in the article is replicated in Fig. 1. These data show that the log transformation is not needed to achieve normality.

Beyond the overall normality assessment for all products in this database, normality was also assessed separately for individual product types (29). Data from a CFC metered dose inhaler (MDI), an HFA MDI, a pre-metered dry powder inhaler (DPI), a device-metered DPI, and a non-pressurized nasal spray were analyzed and it was concluded that a normality assumption was a better fit to the data when analyzed by product type.

Non-normality of the data was noted in this report to potentially occur, resulting in multimodal or skewed distributions (29). Multimodal distributions are often mixtures of two normal distributions that could occur, for example, due to fill weight changes during a batch. Skewed distributions could

Fig. 1. Histogram of standardized delivered dose data for all products ($n=46,816$) compared to the standardized normal distribution (Fig. 31 reproduced from a parametric tolerance interval test for improved control of delivered dose uniformity of orally inhaled and nasal drug products). ©2001 IPAC-RS. Reprinted with permission from “A Parametric Tolerance Interval Test for Improved Control of Delivered Dose Uniformity of Orally Inhaled and Nasal Drug Products”, available at http://ipacrs.com/PDFs/IPAC-RS_DDU_Proposal.PDF

occur if a multidose product with a through container life trend shows different variability for beginning and end doses. However, both these situations actually reflect a model in which the error component may be normally distributed but the mean component may need to be expressed in a more complicated form, incorporating a known bias or trend.

From a fundamental assessment that the factors which influence product variability are additive as well as from a practical assessment of raw data and the nature of defined specifications, the normality assumption and statistical model seems the most appropriate, underlying statistical model to describe delivered dose data for orally inhaled drug products.

Log Normal Distribution Properties for PK Data

This section reviews the principles by which biological data and blood serum concentration data are assumed to follow a log normal distribution and contrasts those principles compared to a model that describes *in vitro* delivered dose.

Pharmacokinetic (PK) data analysis is based on drug concentration in biological fluid over time (30). Pharmacokinetic parameters like C_{max}, C_{min}, T_{max}, and AUC are either directly observed from the original drug concentration curve or derived by mathematical calculations (31). These data are typically modeled using compartmental models that are representations of the body: one-compartment, two-compartment, and multi-compartment models (32). These compartmental models attempt to describe drug concentrations in the body at the site of action at a particular time point. The concentration is a function, however, of several processes that are related to each other: absorption, distribution, metabolism, and excretion (33).

Pharmacokinetic equations, based on these compartmental models, are multiplicative. For example, $AUC = Clearance^{-1} \cdot f \cdot dose$ where f represents the fraction absorbed (31). Clearance can be subject related and therefore, the subject effects are not additive if data are analyzed in the original metric. This compartmental model makes it clear as well that random influences that affect clearance are introduced in a multiplicative manner, not an additive manner. Log transformation of this model results in an equation: $Ln(AUC) = -Ln(Clearance) + Ln(f) + Ln(dose)$ in which subject effects and variability are now additive effects.

For PK data, to account for this multiplicative effect, the geometric mean rather than the arithmetic mean is the best descriptor (32). Additionally, because there is a lower bound below which the data cannot take on values, the data follows a skewed distribution.

An overall summary of the clinical, pharmacokinetic, and statistical rationale for why bioequivalence parameters are log-transformed was documented by Rani and Pargal (31). This is, potentially, one reason why the 2001 FDA guidance document does not recommend testing the assumption of log normality for an individual data set. The data are acknowledged to follow a log normal distribution, regardless of output from goodness of fit statistics for a particular data set based on a limited sample size (34).

The section below describes the impact on the hypothesis test and the PBE criterion when the underlying statistical model that describes the data is the normal distribution.

Assessment of Population Bioequivalence Criterion

In its simplest form, the PBE metric, θ , the scaled aggregate criterion in the FDA 2003 draft guidance for nasal aerosols and nasal sprays for local action, is defined as:

$$\theta = \frac{(\mu_T - \mu_R)^2 + (\sigma_{TT}^2 - \sigma_{TR}^2)}{\max\{\sigma_0^2, \sigma_{TR}^2\}} \quad (3.1)$$

assuming an underlying statistical model of the form:

$$y_{ij} = Ln(X_{ij}) = \mu_i + \varepsilon_{ij}$$

$$\varepsilon_{ij} \sim N(0, \sigma_{Ti}^2)$$

$i = T$ for the test product, R for the reference product

$$j = 1, 2, \dots, n_i$$

μ_i is the mean of the population ($i = T$ or R) in the Ln scale

σ_{Ti}^2 is the variance of the population ($i = T$ or R) in the Ln scale

σ_0^2 is a regulatory constant with value of 0.01, based on recent guidance (3.2)

With this model, even if measurements are taken across different batches, the batch variation is considered negligible and not directly modeled, following recommendations in the 2003 FDA draft guidance document for nasal aerosols and nasal sprays for local action and the 2012 product specific Budesonide guidance (1,6).

The PBE metric represents a combination of differences in means and variances, scaled to the reference product variance for data transformed to the natural log scale. This criterion attempts to encompass equivalency of the entire distribution in a single metric. Equivalency is met if this function of the population parameters is less than an upper acceptance limit, θ_{BE} , or if the linearized criterion, ξ , as shown below, is less than 0.

$$\xi = (\mu_T - \mu_R)^2 + (\sigma_{TT}^2 - \sigma_{TR}^2) - \theta_{BE} \max\{\sigma_0^2, \sigma_{TR}^2\} < 0 \quad (3.3)$$

For a particular data assessment, population equivalency is concluded if the upper 95% confidence bound on the linearized criterion is less than 0. Guidance exists for how to calculate that upper confidence bound (6).

Values of the upper limit for θ_{BE} are determined based on assumed values for the means and variances from the populations. The acceptance limit stated in the product-specific budesonide guidance allows a 10% shift in the mean and a doubling of the variance in the test population for specific parameter values. This is based on the means of the test and reference populations shifting from 100% to 90% and the variance, in the log-transformed scale, increasing from 0.01 to 0.02. For those given parameter values, a PBE

parameter value below 2.089 define the region of equivalency and a value above 2.089 represents non-equivalency. The value of 2.089 is derived from:

$$\begin{aligned}\theta &= \frac{[(\mu_T - \mu_R)^2 + (\sigma_{TT}^2 - \sigma_{TR}^2)]}{\max\{\sigma_0^2, \sigma_{TR}^2\}} \\ &= \frac{[Ln(100) - Ln(90)]^2 + (0.02 - 0.01)]}{0.01} \\ &= \frac{[Ln(\frac{100}{90})]^2 + (0.01)]}{0.01} = 2.089\end{aligned}\quad (3.4)$$

Using this as the definition of equivalency, the hypothesis statements are the following:

$$\begin{aligned}H_0: & \text{Population Non - equivalency} \\ \theta &= \frac{[(\mu_T - \mu_R)^2 + (\sigma_{TT}^2 - \sigma_{TR}^2)]}{\max\{\sigma_0^2, \sigma_{TR}^2\}} \geq \theta_{BE} = 2.089 \\ H_A: & \text{Population Equivalency} \\ \theta &= \frac{[(\mu_T - \mu_R)^2 + (\sigma_{TT}^2 - \sigma_{TR}^2)]}{\max\{\sigma_0^2, \sigma_{TR}^2\}} < \theta_{BE} = 2.089\end{aligned}\quad (3.5)$$

assuming log normality of the data.

If, however, the true underlying statistical model is normally distributed, then the model above is not the correct underlying model. The model based on the normal distribution is given below:

$$X_{ij} = \alpha_i + \eta_{ij} \quad (3.6)$$

where

$\eta_{ij} \sim N(0, \delta_{Ti}^2)$
 i T for the test product, R for the reference product
 j 1, 2, ..., n_i
 α_i is the mean of the population ($i = T$ or R) in the original scale
 δ_{Ti}^2 is the variance of the population ($i = T$ or R) in the original scale

Using the Delta method (which is based on a first-order Taylor series expansion), the mean and variance, approximately, of the population when log transformed can be expressed, in the original parameters, as:

$$\begin{aligned}E[Ln(X_{ij})] &\approx Ln(\alpha_i) = \mu_i \\ Var[Ln(X_{ij})] &\approx \frac{\delta_{Ti}^2}{\alpha_i^2} = \sigma_{Ti}^2\end{aligned}\quad (3.7)$$

This means that the PBE criterion value, in terms of the original parameters that generated the data, is approximately equal to this function:

$$\theta \approx \frac{[Ln(\alpha_T) - Ln(\alpha_R)]^2 + \left(\frac{\delta_{TT}^2}{\alpha_T^2} - \frac{\delta_{TR}^2}{\alpha_R^2}\right)]}{\max\left\{\frac{\delta_{TR}^2}{\alpha_R^2}, \sigma_0^2\right\}} \quad (3.8)$$

This is a complex function of population parameters for defining and visualizing equivalency between two populations.

To better understand the parameter space of the PBE criterion (θ) as a function of the original unit parameters (the mean and variance of the test and reference products, α_T , α_R , δ_{TT}^2 , δ_{TR}^2), the PBE parameter value was approximated, based on the first-order Taylor series expansion above, and calculated for a range of values for the population mean and variance, stated in terms of the original data metric, and percent label claim. These values were chosen to represent a range of possible population values for delivered dose. The reference product values encompass parameter values used in simulations from the IPAC-RS database (29). The test product values were chosen to encompass values less than, equal to, and greater than the reference product. Table I shows the parameter ranges studied.

Results of this analysis, along with output from the simulation study, are presented in the next section.

RESULTS

In this section, results from the assessment of the PBE parameter in the original parameter space as well as results from simulation work are presented and discussed.

Assessment of Population Bioequivalence Criterion in Original Parameters

Graphical output of the PBE parameter value from the combinations listed in Table I is presented in Figs. 2 and 3. The x-axis is the test product mean value. The y-axis is the PBE population criterion value, approximated, based on Eq. (3.8). The solid red vertical line represents the mean of both

Table I. Parameter Values for *In Vitro* Delivered Dose Data for Review of PBE Criterion

Description of parameter	Parameter symbol	Range studied
Mean delivered dose of the reference product	α_R	80%–120% by 10%
Mean delivered dose of the test product	α_T	$\alpha_R - 15\% \leq \alpha_T \leq \alpha_R + 15\%$ by 5%
Standard deviation of dose of the reference product	δ_{TR}	4%, 6%
Standard deviation of dose of the test product	δ_{TT}	4%, 6%, 8%

PBE population bioequivalence method

Test Product Standard Deviation: Blue=4%, Grey=6%, Red=8%

Fig. 2. Plot of PBE population criterion value: Reference Product Standard Deviation=4%

Test Product Standard Deviation: Blue=4%, Grey=6%, Red=8%

Fig. 3. Plot of PBE population criterion value: reference product standard deviation=6%

the test and reference product where no difference exists between populations; the inner, first set of dotted lines represents test product means 10% different (absolute) than the reference product mean. The outer, second set of dotted lines represents test product means 15% different (absolute). The different curves represent varying test product standard deviations, and the boundary of equivalency is shown on the vertical axis as a *horizontal* black line at the value of 2.089. These plots show how parameter values in the original metric for delivered dose for the mean and variance, approximately, map into the parameter space for the PBE criterion based on the log transformation.

These plots highlight the following:

- The PBE parameter value is asymmetric for dose parameter values in the original scale. This means that test product means that are lower than reference product means result in higher PBE values and are, therefore, defined as less equivalent than test product means that are greater than the reference mean by the same amount. As an example, for the case when the reference product mean is 100% label claim (LC) with a standard deviation of 4% and the variance is equal between products, the PBE parameter value is equal to 2.70 for a test product that delivers, on average, 15% lower dose, but is equal to 1.91 for a test product that delivers, on average, 15% more. This is partially due to the non-linear nature of the log transformation. The log transformation will decrease the range of larger values and increase the range of smaller values. This means a difference of +10% is more likely to pass equivalence compared to a difference of -10%.
- This asymmetry is more pronounced as the mean delivered dose values decreased. Test product means 10% lower than reference product means will be less likely to pass equivalence as the reference product mean decreases. Ideally, an absolute difference from target should have the same chance of being declared equivalent and should not depend on the reference product parameter value.

- As the test product variability increased, the PBE parameter value increased, ensuring that the parameter space in which the test product is more variable than the reference product is more likely to result in a conclusion of non-equivalence. In this way, the parameter is behaving as desired.
- If the variance between the test and reference product remained equal but the reference product became more variable, the PBE parameter value changed but the direction of the change depended on whether the test product delivered a higher dose or a lower dose compared to the reference product (see Table II). For situations in which the test product delivered a lower mean dose than the reference product, the PBE parameter value increased, making equivalence more difficult to declare as variation increased. However, if the test product mean was greater compared to the reference product mean dose, the PBE parameter value decreased for a more variable product, making equivalence more likely to be declared. Ideally, the equivalency metric should be consistent and should not depend on whether the test product mean is higher or lower than the reference product mean.

Simulation Study

To gain a better understanding of this practically, a simulation study was conducted, based on a sample size of $n=30$. The mean of the reference product was taken to be 90%, 100%, and 110% of label claim. The mean of the test product varied from a difference of -14%, -12%, -10%, -8%, 0%, +8%, +10%, +12%, +14% of label claim compared to the reference product mean and the variance of the test product was taken to be equal to the variance of the reference product and 1.5 times the variance of the reference product for the case in which the reference product standard deviation was 5%.

Table II. Example Values of the PBE Criterion When Standard Deviation Changes

Mean of the reference product (% LC)	Mean of the test product (% LC)	Standard deviation of the reference product (% LC)	Standard deviation of the test product (% LC)	Approximate value of the PBE criterion	Test product mean compared to reference product mean
80	70	4	4	1.8596	Less than
80	70	6	6	1.9553	
80	80	4	4	0.0000	Equal
80	80	6	6	0.0000	
80	90	4	4	1.3348	Greater than
80	90	6	6	1.2692	
90	80	4	4	1.4398	Less than
90	80	6	6	1.5053	
90	90	4	4	0.0000	Equal
90	90	6	6	0.0000	
90	100	4	4	1.0726	Greater than
90	100	6	6	1.0256	

PBE population bioequivalence method, LC label claim

Mean Reference Product=90%

Mean Reference Product=100%

Mean Reference Product=110%

Fig. 4. Plot of 95% upper confidence bound on the linearized criterion, from simulated data, as a function of the difference in the mean delivered dose between the test and the reference product

Fig. 5. Plot of proportion of simulations declaring equivalency as a function of the difference in the mean delivered dose between the test and the reference product

Data were simulated according to the following statistical model:

$$X_{ij} = \alpha_i + \eta_{ij}$$

$$\eta_{ij} \sim N(0, \delta_{Ti}^2)$$

$i = T$ for the test product, R for the reference product

$j = 1, 2, \dots, 30$

The simulations were based on 5,000 replicates. From the simulated data, the 95% upper confidence bound of the linearized criterion was calculated following the procedure stated in the draft product-specific budesonide guidance (6).

A plot of the 95% upper confidence bound of the linearized criterion as a function of the difference between the test product mean and reference product mean in the original scale appears above in Fig. 4. A quadratic function was fit to the results to visualize the average linearized upper confidence bound value as a function of the test product standard deviation.

As expected, the asymmetry noted in the PBE population parameter from Figs. 2 and 3 is reflected in the increased mean and variability of the upper 95% confidence bound for cases in which the test mean is less than the reference mean. The average upper confidence bound value increases for a more variable test product, meaning equivalence is less likely.

The statistical power of this approach, from the simulations, is presented in Fig. 5. The observed proportion of simulations in which equivalency was declared is plotted as a function of the difference between the test product and reference product mean in the original scale. The different curves represent varying combinations of reference product mean values and test product standard deviation values when the reference product standard deviation is 5%.

This plot highlights that the statistical power of the PBE approach is not symmetric for a fixed difference between the test product mean and the reference product mean. The probability of declaring equivalency is greater if the test product mean is higher than the reference product mean for the same distance away from the reference product.

Specifically, if the reference product mean is at 100% label claim, then, a test product with the same variation whose mean is 110% is estimated to be declared equivalent 99.18% of the time compared to 79.36% of the time if the test product mean is 90% so a difference in statistical power of almost 20% for the same absolute difference, given no change in variability in the test product. The difference in statistical power becomes greater as the reference product mean decreases to 90%. A test product mean at 100% for products with the same variability, is expected to pass equivalency 89.74% of the time compared to 40.36% of the time if the test product mean was 80% for a difference in statistical power of almost 50% for the same absolute difference that the test product is from the reference product. One would like an equivalency procedure that avoids this lack of symmetry when errors are normally distributed.

CONCLUSIONS

In this paper, fundamental principles for an assumed statistical model were presented for *in vitro* delivered dose. These principles should be key when implementing statistical methodology for data assessment and decision making. The discrepancy that exists between the development of appropriate statistical models and methodologies for PK data compared to *in vitro* data was highlighted, and the practical impact on equivalency decisions for *in vitro* delivered dose,

based on applying the PBE statistical approach, was reported. If data are expected to follow a normal distribution, as is the case for *in vitro* delivered dose, then, decisions using the PBE approach are not consistent for the same absolute mean difference that the test product is from the reference product. A conclusion of inequivalency will occur more often if the test product dose is lower than the reference product for the same deviation from target and this discrepancy in the conclusion grows as the reference product mean decreases. This is due to applying a non-linear log transformation to data that do not follow a log normal distribution.

These features of the PBE methodology applied to *in vitro* delivered dose data, as summarized in this paper, suggest that further research for assessing equivalence is still warranted for *in vitro* data. In particular, it may not be appropriate to take the logarithm of the data. Those approaches will be further outlined and explored in a subsequent paper.

ACKNOWLEDGMENTS

The authors would like to acknowledge the contributions of both Tammy Morris, who provided statistical programming support, and Rick Lewis, Ph.D., statistical consultant, who provided technical review and guidance, both of whom are employees of GlaxoSmithKline.

REFERENCES

1. FDA. Guidance for industry—bioavailability and bioequivalence studies for nasal aerosols and nasal sprays for local action—draft guidance. US Department of Health and Human Services. Food and Drug Administration, Center for Drug Evaluation and Research (CDER), April 2003; 2003.
2. Lee S, Adams W, Li B, Conner D, Chowdhury B, Yu L. Considerations to support bioequivalence of locally acting drugs in dry powder inhalers for lung diseases. *AAPS J*. 2009;11:414–23.
3. Scientific and regulatory considerations for bioequivalence (BE) of dry powder inhalers (DPIs). Sau Lee, GPhA/FDA Fall Technical Conference, October 4, 2011. <http://www.fda.gov/downloads/Drugs/DevelopmentApprovalProcess/HowDrugsareDevelopedandApproved/ApprovalApplications/AbbreviatedNewDrugApplicationANDAGenerics/UCM292652.pdf> (2011).
4. FDA. Statistical information from the June 1999 draft guidance and statistical information for *in vitro* bioequivalence data posted on August 18, 1999. US Department of Health and Human Services. Food and Drug Administration, Center for Drug Evaluation and Research (CDER); 1999.
5. Li B, Ren K. A step-wise procedure for population bioequivalence (PBE) analysis of orally inhaled and nasal drug product (OINDP) bioequivalence studies. 04 Oct. GPhA/FDA 2011 Fall Technical Conference; 2011.
6. OGD. Draft guidance on Budesonide. Office of Generic Drugs, FDA. Sept 2012; 2012.
7. OGD. Draft guidance on albuterol sulfate, aerosol, metered. Office of Generic Drugs, FDA. April 2013; 2013.
8. EMEA/CHMP/QWP/49313/2005. Guideline on the pharmaceutical quality of inhalation nasal products. 21 June 2006.
9. Chow S, Shao J, Wang H. Bioequivalence testing. *Stat Med*. 2003;22:55–68.
10. Chiu S, Tsai P, Liu J. Statistical evaluation of non-profile analyses for the *in vitro* bioequivalence. *J Chemom*. 2010;24:617–25.
11. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. ICH harmonised tripartite guideline: specifications: test procedures and acceptance criteria for new drug substances and new drug products: chemical substances Q6A; 06 Oct 1999.
12. 21 CFR parts 210 and 211—current good manufacturing practice in manufacturing, processing, packing or holding of drugs; general and current good manufacturing practice for finished pharmaceuticals.
13. FDA Guidance for Industry—Metered dose inhaler (MDI) and dry powder inhaler (DPI) drug products—chemistry, manufacturing, and controls documentation; 1998.
14. FDA Guidance for Industry – nasal spray and inhalation solution, suspension, and spray drug products – chemistry, manufacturing, and controls documentation; 2002.
15. EMEA—Guideline on the pharmaceutical quality of inhalation and nasal products; 2006.
16. European Pharmacopoeia – preparations for inhalation (general chapter) and 2.9.18 preparations for inhalation: aerodynamic assessment of fine particles.
17. United States Pharmacopeia (USP) chapter <601> – aerosols, nasal sprays, metered dose inhalers, and dry powder inhalers.
18. United States Pharmacopeia (USP) chapter <1151> – aerosols and inhalations.
19. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. ICH harmonised tripartite guideline: pharmaceutical development Q8 (R2); Aug 2009.
20. FDA Guidance for Industry – quality systems approach to pharmaceutical CGMP regulations, US Department of Health and Human Services. Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER), Center for Veterinary Medicine (CVM), Office of Regulatory Affairs (ORA); September 2006.
21. FDA Guidance for Industry: PAT—a framework for innovative pharmaceutical development, manufacturing, and quality assurance, U.S. Department of Health and Human Services Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Veterinary Medicine (CVM), Office of Regulatory Affairs (ORA); September 2004.
22. FDA Guidance for Industry: current good manufacturing practice for combination products -draft guidance, U.S. Department of Health and Human Services Food and Drug Administration, Office of Combination Products (OCP); September 2004.
23. FDA Guidance for Industry: powder blends and finished dosage units—stratified in-process dosage unit sampling and assessment -draft guidance, U.S. Department of Health and Human Services Food and Drug Administration, Center for Drug Evaluation and Research (CDER); October 2003.
24. FDA Guidance for Industry: process validation: general principles and practices, U.S. Department of Health and Human Services Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER), Center for Veterinary Medicine (CVM); January 2011.
25. Willig SH. Good manufacturing practices for pharmaceuticals: a plan for total quality control from manufacturer to consumer. 5th ed. New York: Marcel Dekker; 2001.
26. Warner E. Process drift and its resolution in the manufacture of drug products. 2 Dec. PQRI-FDA Workshop on Process Drift. <http://www.pqri.org/workshops/ProcDrift/imagespdfs/Warner.pdf> (2010).
27. FDA Guidance for Industry: analytical procedures and methods validation – draft guidance, U.S. Department of Health and Human Services Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER); August 2000.
28. Limpert E, Stahel WA, Abbt M. Log-normal distributions across the sciences: keys and clues. *Bioscience*. 2001;51:341–52.
29. A parametric tolerance interval test for improved control of delivered dose uniformity of orally inhaled and nasal drug products. Developed and submitted by The International Pharmaceutical Aerosol Consortium on Regulation and Science

- (IPAC-RS) in collaboration with scientists of the Inhalation Technology Focus Group (ITFG) of the American Association of Pharmaceutical Scientists (AAPS); September 2002.
30. Qayyum A. Bioequivalence studies, readings in advanced pharmacokinetics - theory, methods and applications, Dr. Ayman Noreddin (Ed.), ISBN: 978-953-51-0533-6, InTech. doi:10.5772/33826. Available from: <http://www.intechopen.com/books/readings-in-advanced-pharmacokinetics-theory-methods-and-applications/bioequivalence-studies> (2012).
 31. Rani S, Pargal A. Bioequivalence: an overview of statistical concepts. *Indian J Pharmacol.* 2004;36:209–16.
 32. Spruill WJ, Wade WE. Concepts in clinical pharmacokinetics. Bethesda: American Society of Health-System Pharmacists; 2010.
 33. Geometric statistics in PK analysis—programmer's perspective. Paper contributed to 2011 NorthEast SAS Users Group. <http://www.nesug.org/Proceedings/nesug11/ph/ph10.pdf>.
 34. FDA. Guidance for industry—statistical approaches to establishing bioequivalence. US Department of Health and Human Services. Food and Drug Administration, Center for Drug Evaluation and Research (CDER), January 2001; 2001.