

# In Vitro Bioequivalence Testing

**Hansheng Wang**

*Peking University, Beijing, People's Republic of China*

**Ying Zhang**

*AIG American General, Neptune, New Jersey, U.S.A.*

**Jun Shao**

*University of Wisconsin–Madison, Madison, Wisconsin, U.S.A.*

**Shein-Chung Chow**

*Millennium Pharmaceuticals, Inc., Cambridge, Massachusetts, U.S.A.*

## INTRODUCTION

Bioequivalence testing is considered as a surrogate for the clinical evaluation of the therapeutic equivalence of drug products based on the *Fundamental Bioequivalence Assumption* that when two drug products (e.g., a brand name drug and its generic copy) are equivalent in bioavailability, they will reach the same therapeutic effect.<sup>[1]</sup> In other words, two drugs are called bioequivalent if they are equivalent in bioavailability. In practice, most drugs' active moiety is absorbed into the blood system before it takes action. As a result, the pharmacokinetic (PK) parameters taken from the bloodstream—such as the area under the plasma concentration curve (AUC), the maximum concentration ( $C_{max}$ ), and the time to the maximum concentration ( $T_{max}$ )—are usually considered to characterize the bioavailability of the drug. This type of bioequivalence is called *in vivo* bioequivalence. However, for some locally acting drug products, such as nasal aerosols (e.g., metered dose inhalers) and nasal sprays (e.g., metered dose spray pumps) that are not intended to be absorbed into the bloodstream, bioavailability may be assessed by measurements intended to reflect the rate and the extent to which the active ingredient or active moiety becomes available at the site of action. For these local delivery drug products, the U.S. Food and Drug Administration (FDA) indicates that bioequivalence may be assessed, with suitable justification, by *in vitro* bioequivalence studies alone (21 CFR 320.24). Although it is recognized that *in vitro* methods are less variable, easier to control, and more likely to detect differences between products if they exist, the clinical relevance of the *in vitro* tests or the magnitude of the differences in the tests is not clearly established until a guidance on bioavailability and bioequivalence studies for nasal aerosols and nasal sprays for local action has been issued by the FDA.<sup>[2]</sup>

## STUDY DESIGN AND DATA COLLECTION

For the assessment of *in vitro* bioequivalence, the FDA requires that *in vitro* testing of emitted dose uniformity, droplet size distribution, spray pattern, plume geometry, priming/repriming, and tail-off profile be done to demonstrate comparable delivery characteristics between two drug products. In this section, a brief description of study design and each of the six *in vitro* tests are given.

### Study Design

According to the FDA,<sup>[2]</sup> three lots/sublots from each product are required to be tested for *in vitro* emitted dose uniformity, droplet size distribution, spray pattern, plume geometry, priming/repriming, and tail-off profile. For each *in vitro* test, 10 samples are randomly drawn from each lot. Samples are randomized for *in vitro* tests. The analysts will not have access to the randomization codes. An automated actuation station with a fixed settings (actuation force, dose time, return time, and hold time) is usually used for the *in vitro* tests.

### Priming, Emitted Dose Uniformity, Priming/Repriming, and Tail-Off Profile

Based on the FDA comment, the priming, emitted dose uniformity, priming/repriming, and tail-off tests maybe tested in the following setting. Three individual lots of test product and reference product are evaluated. For each lot, 10 samples are then tested for pump priming, unit spray content through life, and tail-off studies. Then additional samples for each lot are evaluated for the prime hold study (reprime study).

For each sample unit, spray samples are collected for sprays 1–8 and analyzed in order to determine the minimum number of actuations required before the pump



delivers the labeled dose of drug (sprays 1–8). To characterize emitted dose uniformity at the beginning of unit life, spray 9 is collected. Sprays 10–15 are wasted by the automatic actuation station. Spray 16 is collected in the middle of unit life. Sprays 17–20 are wasted. Sprays 21–23 are collected at the end of the unit life. Additional sprays after spray 23 are collected and analyzed to determine the tail-off profile.

Ten additional samples are drawn randomly from each lot of drug product for the pump prime hold study. For each unit, the first 12 sprays (sprays 1–12) are wasted. Sprays 13 and 14 are collected as fully primed sprays. The unit is then stored undisturbed for 24 hr. Within each lot, five samples are placed in the upright position while the other five are placed in a side position. After that, sprays 15–17 are collected. The unit is then stored undisturbed in its former position for another 24 hr. After that, the doses emitted by sprays 18–20 are collected. All spray samples are weighted in order to obtain repriming characteristics.

### Spray Pattern

A spray pattern produced by a nasal spray pump evaluates in part the integrity and the performance of the orifice and pump mechanism in delivering a dose to its intended site of deposition. Measurements can be made on the diameter of the horizontal intersection of the spray plume at different distances from the actuator tip. Spray patterns are usually measured at three distances (e.g., 1, 2, and 4 cm) at both the beginning (sprays 8–10) and the end (sprays 17–19) of unit life. As a result, a total of six spray patterns is collected for each sample unit. For each spray pattern image, the diameters (the longest and shortest diameters) and the ovality (which is defined by the ratio of the longest to the shortest diameters) are measured.

### Droplet Size Distribution

For a test of droplet size distribution, methods of laser diffraction and cascade impaction are commonly used. These methods are briefly described below.

#### Laser diffraction

For a test of droplet size distribution using laser diffraction particle analyzer, each sample unit is first primed by actuating the pump eight times using an automatic actuation station. Droplet size distribution is then determined at three distances (e.g., 3, 5, and 7 cm) from the laser beam and at the beginning, the middle, and the end of unit life. At each distance, three measurements of delay times (plume formation, start of dissipation, and inter-

mediate measurements) and overall evaluation are used to characterize the droplet size. As a result, a total of 36 measurements is recorded for each sample unit.

### Cascade impaction technique

When the spray pump is actuated in the nasal cavity, a fine mist of droplets is generated. Droplets that are  $> 9 \mu\text{m}$  in diameter are considered nonrespirable and are therefore useful for nasal deposition. As recommended in the FDA 1999 Guidance, the data should be reported as follows:

- Group 1: Adaptor (expansion chamber, i.e., 5-L flask), rubber gasket, throat, and Stage 0.
- Group 2: Stage 1.
- Group 3: Stage 2 to filter.

Each sample unit is first primed by actuating the pump seven times using an automatic actuation station. Droplet size distribution is then determined at the beginning and the end of the life of the sample. Thus a total of six groups of results is reported for each spray unit.

### Plume Geometry

Plume geometry is performed on the nasal spray plume that is allowed to develop into an unconstrained space that far exceeds the volume of nasal cavity. It represents a frozen moment in spray plume development that is viewed from two axes perpendicular to the axis of plume development. The samples should be actuated vertically. Prime the pump with 10 actuations until a steady fine mist is produced from the pump. A fast-speed video camera is placed in front of the sample bottle and starts recording. Repeat the test by rotating the actuator  $90^\circ$  to the previous actuator placement so that two side views are at  $90^\circ$  to each other (two perpendicular planes) and, relative to the axis of the plume of the spray, are captured when actuated into space. Spray plumes are characterized at three stages: early upon formation, as the plume starts dissipate, and at some intermediate time. Longest vertical distance (LVD), widest horizontal distance (WHD), and plume angle (ANG) are recorded and analyzed.

## REGULATORY REQUIREMENTS AND STATISTICAL METHODS

### Noncomparative Analysis

For each *in vitro* test, the FDA requires that a non-comparative analysis be performed. Noncomparative analysis refers to the statistical summarization of the

bioavailability data by descriptive statistics. As a result, means, standard deviations, and coefficients of variation (CVs) in percentage of the six *in vitro* tests should be documented. More specifically, the overall sample means for a given formulation should be averaged over all samples (e.g., bottles/canisters), life stages (except for priming and repriming evaluations), and lots or batches. In addition to the overall means, means at each life stage for each batch averaged over all bottles/canisters and for each life stage averaged over all lots (or batches) should be presented. For profile data, means, standard deviations, and percent CVs should be reported for each stage. The between-lot (or batch), within-lot (or batch) between-sample (e.g., bottle or canister), and within-sample (e.g., bottle or canister) between-life stage variability should be evaluated through appropriate statistical models.

### Bioequivalence Limit

The following formula is given in the 1999 FDA for determining the bioequivalence (BE) limit

$$\frac{(\text{average BE limit in natural log scale})^2 + \text{variance terms offset}}{\text{scaling variance}}$$

As it can be seen, in order to obtain the BE limit, there are three quantities that need to be specified. They are 1) average BE limit, 2) variance terms offset, and 3) scaling variance, respectively. The 1999 FDA guidance indicates that the final specification of those parameters should be based on the results of the ongoing simulation study. However, the following values are recommended in the FDA's draft guidance.

As a result of the low variability of *in vitro* measurements, at the present time, the FDA recommends that the ratio of geometric means should fall within 0.90 and 1.11. As a result, a value of 0.90 is recommended as the average BE limit for *in vitro* data.

The objective of variance terms offset is to allow some difference among the total variances that may be inconsequential. As a result of the low variability of *in vitro* measurements, the FDA recommends that a value of 0 should be taken based on the guidance of population and individual bioequivalence.<sup>[3]</sup> In practice, however, a value of 0.01 may be accepted by the FDA for variance terms offset depending upon the nature of the drug products under investigation.

The purpose of scaling variance is to adjust the BE criterion depending on the reference product variance. When the reference variance is greater than the scaling variances, the limit is widened. On the other hand, the limit is narrowed when reference variance is less than

scaling variance. The FDA indicates that the choice of the scaling variance should be at least 0.1.

As a result, the specification of 0.90 for the average BE limit—0.0 for the variance offset and 0.10 for scaling standard deviation—gives the following BE limit:

$$\theta_{BE} = \frac{\log(0.9)^2 + 0}{0.1^2} = 1.11$$

### Nonprofile Analysis

The FDA<sup>[2]</sup> indicates that the *in vitro* bioequivalence of nasal aerosols and sprays can be established by six bioequivalence tests. They are classified as either the nonprofile analysis or the profile analysis. Nonprofile analysis applies to emitted dose or spray content uniformity, through container life, droplet size distribution, spray pattern, and priming/repriming. The criterion for nonprofile *in vitro* bioequivalence is described as follows.

Let  $y_T$ ,  $y_R$ , and  $y'_R$  be independent *in vitro* bioavailabilities, where  $y_T$  is from the test product and  $y_R$  and  $y'_R$  are from the reference product. The two products are said to be *in vitro* bioequivalent if  $\theta < \theta_{BE}$ , where:

$$\theta = \frac{E(y_R - y_T)^2 - E(y_R - y'_R)^2}{\max\{\sigma_0^2, E(y_R - y'_R)^2/2\}} \tag{1}$$

$\theta_{BE}$  is a prespecified BE limit and  $\sigma_0^2$  is a prespecified constant. Values of  $\sigma_0^2$  and  $\theta_{BE}$  can be found in the FDA guidance. According to the FDA guidance, *in vitro* bioequivalence can be claimed if the hypothesis that  $\theta \geq \theta_{BE}$  is rejected at the 5% level of significance, provided that the ratio of geometric means between drug products is within 0.90 and 1.11.

Let  $m_T$  and  $m_R$  be the number of canisters from the test product and the reference product, respectively, and one observation from each sample (e.g., bottle or canister) is obtained. The data can be described by the following model:

$$y_{jk} = \mu_k + \varepsilon_{jk}, \quad j = 1, \dots, m_k \tag{2}$$

where  $k = T$  for the test product,  $k = R$  for the reference product,  $\mu_T$  and  $\mu_R$  are fixed product effects,  $\varepsilon_{jk}$ 's are independent random measurement errors distributed as  $N(0, \sigma_k^2)$ ,  $k = T, R$ . Under model 2, the parameter  $\theta$  in Eq. 1 is equal to:

$$\theta = \frac{(\mu_T - \mu_R)^2 + \sigma_T^2 - \sigma_R^2}{\max\{\sigma_0^2, \sigma_R^2\}} \tag{3}$$



As a result,  $\theta < \theta_{BE}$  is equivalent to  $\zeta < 0$ , where  $\zeta$  is the linearized BE parameter given by:

$$\zeta = (\mu_T - \mu_R)^2 + \sigma_T^2 - \sigma_R^2 - \theta_{BE} \max\{\sigma_0^2, \sigma_R^2\} \quad (4)$$

The bioequivalence can be concluded by constructing an approximate 95% upper bound for  $\zeta$ . If the 95% upper bound is less than 0, *in vitro* bioequivalence is established, or otherwise rejected. In order to obtain the approximate 95% upper bound, the FDA<sup>[2]</sup> recommends the following procedure proposed by Hyslop et al.,<sup>[4]</sup> which was originally developed for the establishment of *in vivo* individual bioequivalence under a  $2 \times 4$  crossover design. Let  $\tilde{\zeta}_U$  denote the 95% upper bound for  $\zeta$  and:

$$\begin{aligned} \hat{\delta} &= \bar{y}_T - \bar{y}_R \\ s_k^2 &= \frac{1}{m_k - 1} \sum_{j=1}^{m_k} (y_{jk} - \bar{y}_k)^2 \\ \bar{y}_k &= \frac{1}{m_k} \sum_{j=1}^{m_k} y_{jk} \end{aligned}$$

Then, it follows that:

$$\tilde{\zeta}_U = \hat{\delta}^2 + s_T^2 - s_R^2 - \theta_{BE} \max\{\sigma_0^2, s_R^2\} + \sqrt{U}$$

where  $U$  is the sum of the following three quantities:

$$\left[ \left( |\hat{\delta}| + z_{0.95} \sqrt{\frac{s_T^2}{m_T} + \frac{s_R^2}{m_R}} \right)^2 - \hat{\delta}^2 \right]^2$$

$$s_T^4 \left( \frac{m_T - 1}{\chi_{0.05; m_T - 1}^2} - 1 \right)^2$$

and

$$(1 + c\theta_{BE})^2 s_R^4 \left( \frac{m_R - 1}{\chi_{0.95; m_R - 1}^2} - 1 \right)^2$$

where  $c = 1$  if  $s_R^2 \geq \sigma_0^2$ ,  $c = 0$  if  $s_R^2 < \sigma_0^2$ ,  $z_a$  is the  $a$ th quantile of the standard normal distribution, and  $\chi_{t;a}^2$  is the  $a$ th quantile of the central chi-square distribution with  $t$  degrees of freedom. As suggested by the FDA, *in vitro* bioequivalence can be claimed if  $\tilde{\zeta}_U < 0$ . This procedure is recommended by the FDA guidance.

### Profile Analysis

As indicated in the FDA,<sup>[2]</sup> profile analysis using a confidence interval approach should be applied to cascade impactor (CI) or multistage liquid impinger (MSLI) for particle size distribution. As indicated in the 1999 FDA

guidance, equivalence may be assessed based on chi-square differences. The idea is to compare the profile difference between test product and reference product samples to the profile variation between reference product samples. More specifically, let  $y_{ijk}$  denote the observation from the  $j$ th subject's  $i$ th stage in the  $k$ th treatment. Given a sample ( $j_0$ ) from test product and two samples ( $j_0, j_1$ ) from reference products and assuming that there are a total of  $S$  stages, the profile distance between test and reference is given by:

$$d_{TR} = \sum_{i=1}^S \frac{(y_{ij_0T} - 0.5(y_{ij_1R} + y_{ij_2R}))^2}{(y_{ij_0T} + 0.5(y_{ij_1R} + y_{ij_2R}))}$$

Similarly, the profile variability within reference is defined to be:

$$d_{RR} = \sum_{i=1}^S \frac{(y_{ij_1R} - y_{ij_2R})^2}{0.5(y_{ij_1R} + y_{ij_2R})}$$

For a given triplet sample of (Test, Reference 1, Reference 2), the ratio of  $d_{TR}$  and  $d_{RR}$  (i.e.,  $rd = d_{TR}/d_{RR}$ ) can then be used as a bioequivalence measure for the triplet samples between the two drug products. For a selected sample, the 95% upper confidence bound of  $E(rd) = E(d_{TR}/d_{RR})$  is then used as a bioequivalence measure for the determination of bioequivalence. In other words, if the 95% upper confidence bound is less than the bioequivalence limit, then we claim that the two products are bioequivalent.

The FDA<sup>[2]</sup> recommends a bootstrap procedure to construct the 95% upper bound for  $E(rd)$ . The procedure is described below. Assume that the samples are obtained in a two-stage sampling manner. In other words, for each treatment (test or reference), three lots are randomly sampled. Within each lot, 10 samples (e.g., bottles and canisters) are sampled. The following paragraph is quoted from the 1999 FDA guidance regarding the bootstrap procedure to establish profile bioequivalence.

For an experiment consisting of three lots each of test and reference products, and with 10 canisters per lot, the lots can be matched into six different combinations of triplets with two different reference lots in each triplet. The 10 canisters of a test lot can be paired with the 10 canisters of each of the two reference lots in  $(10 \text{ factorial})^2 = (3,628,800)^2$  combinations in each of the lot triplets. Hence a random sample of the  $N$  canister pairing of the six Test–Reference 1–Reference 2 lot triplets is needed.  $rd$  is estimated by the sample mean of the  $rd$ s calculated for the triplets in 10 selected samples of  $N$ .

In the sample guidance, the FDA recommends that a value of 500 should be taken for  $N$ .

RECENT DEVELOPMENT

Nonprofile Analysis

As indicated earlier, the FDA<sup>[2]</sup> requires at least 30 samples to be taken from each of the test and the reference drug products (i.e.,  $m_k = 30$ ). However,  $m_k = 30$  may not be enough to achieve a desired power for the bioequivalence test. In practice, there are two options to increase the power. One is to increase the sample size and the other one is to increase the replicates per sample. Increasing the sample size can certainly increase the power, but in some situations, obtaining replicates from each bottle or canister may be more practical and/or cost-effective. However, how to perform the statistical analysis based on replicates becomes a problem of interest.

Chow et al.<sup>[5]</sup> proposed the following method to assess *in vitro* bioequivalence when a replicate from each bottle or canister is available. Suppose that there are  $n_k$  replicates from each bottle or canister for product  $k$ . Let  $y_{ijk}$  be the  $i$ th replicate in the  $j$ th canister under product  $k$ . Let  $b_{jk}$  be the between-bottle or between-canister variation and let  $e_{ijk}$  be the within-bottle or within-canister measurement error. Then:

$$y_{ijk} = \mu_k + b_{jk} + e_{ijk}, \quad i = 1, \dots, n_k, \\ j = 1, \dots, m_k \tag{5}$$

where  $b_{jk}$ 's and  $e_{ijk}$ 's are independent,  $b_{jk} \sim N(0, \sigma_{Bk}^2)$ , and  $e_{ijk} \sim N(0, \sigma_{Wk}^2)$ . Under model 5, the total variances  $\sigma_T^2$  and  $\sigma_R^2$  in Eqs. 3 and 4 are equal to  $\sigma_{BT}^2 + \sigma_{WT}^2$  and  $\sigma_{BR}^2 + \sigma_{WR}^2$ , respectively (i.e., the sums of between-bottle or between-canister and within-bottle or within-canister variances). The parameter  $\theta$  in Eq. 1 is still given by Eq. 3 and  $\theta < \theta_{BE}$  if and only if  $\zeta < 0$ , where  $\zeta$  is given in Eq. 4.

Under model 5, an approximate 95% upper bound for  $\zeta$  is given by:

$$\hat{\zeta}_U = \hat{\delta}^2 + s_{BT}^2 + (1 - n_T^{-1})s_{WT}^2 - s_{BR}^2 - (1 - n_R^{-1})s_{WR}^2 \\ - \theta_{BE} \max\{\sigma_0^2, s_{BR}^2 + (1 - n_R^{-1})s_{WR}^2\} + \sqrt{U}$$

where  $U$  is the sum of the following five quantities:

$$\left[ \left( \left| \hat{\delta} \right| + z_{0.95} \sqrt{\frac{s_{BT}^2}{m_T} + \frac{s_{BR}^2}{m_R}} \right)^2 - \hat{\delta}^2 \right]^2 \\ s_{BT}^4 \left( \frac{m_T - 1}{\chi_{0.05; m_T - 1}^2} - 1 \right)^2$$

$$(1 - n_T^{-1})^2 s_{WT}^4 \left( \frac{m_T(n_T - 1)}{\chi_{0.05; m_T(n_T - 1)}^2} - 1 \right)^2$$

$$(1 + \theta_{BE})^2 s_{BR}^4 \left( \frac{m_R - 1}{\chi_{0.95; m_R - 1}^2} - 1 \right)^2$$

and

$$(1 + c\theta_{BE})^2 (1 - n_R^{-1})^2 s_{WR}^4 \left( \frac{m_R(n_R - 1)}{\chi_{0.95; m_R(n_R - 1)}^2} - 1 \right)^2$$

and  $c = 1$  if  $s_{BR}^2 + (1 - n_R^{-1})s_{WR}^2 \geq \sigma_0^2$  and  $c = 0$  if  $s_{BR}^2 + (1 - n_R^{-1})s_{WR}^2 < \sigma_0^2$ . *In vitro* bioequivalence can be claimed if  $\hat{\zeta}_U < 0$ .

If the difference between model 2 and model 5 is ignored and the confidence bound  $\hat{\zeta}_U$  with  $m_k$  replaced by  $m_k n_k$  (instead of  $\hat{\zeta}_U$ ) is used, then the asymptotic size of the test procedure is not 5%.

Profile Analysis

The bootstrap procedure described in the Section ‘‘Bioequivalence Limit’’ has received much attention and criticism since it was introduced by the FDA. The major criticisms are described below.

First, the statistical properties of this procedure are unknown. It includes two aspects. One is that the statistical model, which should be used to describe the profile data, is not clearly defined in the FDA guidance. In addition, even under an appropriate statistical model, the statistical properties of the bootstrap procedure are still unknown. More specifically, is the bootstrap sample mean a consistent estimator for  $E(rd)$ ? Is the 95% percentile of the bootstrap samples an appropriate 95% upper bound for  $E(rd)$ ? These questions are not addressed in the FDA guidance.

Second, no criteria are given regarding the passage or failure of the bioequivalence study. This is the issue that confuses most researchers/scientists in practice. After the conduction of a valid trial and an appropriate statistical analysis following the FDA guidance, the sponsor still cannot tell if its product has passed or failed the bioequivalence test. This is a direct consequence of our first point (i.e., the statistical properties of the recommended bootstrap procedure are unknown).

Third, the simulation study using different random number generation schemes may produce contradictory results. It is possible for a good product to fail the bioequivalence test simply because of ‘‘bad luck.’’ It is also possible for a bad product to pass the bioequivalence test with an ‘‘appropriate’’ choice of random number generation scheme. As a result, researchers/scientists tends to reply more on the descriptive statistics of the two drug



products (treatment and reference) in order to assess their bioequivalence instead of the bootstrap procedure. The proposed bootstrap procedure recommended by FDA is not as reliable as it should be.

As a result, further research of profile analysis becomes a problem of interest in practice. More specifically, the questions of interest include: 1) What statistical model should be used to describe the profile data? 2) Is  $E(rd)$  defined by the FDA as a good parameter for characterizing the bioequivalence between test and reference products? Can we define the test-to-reference distance and reference-to-reference variability differently? 3) What BE limit should be used? 4) What statistical procedure should we use to evaluate the *in vitro* bioequivalence between the two products based on appropriate model, parameter, and bioequivalence criterion?

### Sample Size

The FDA<sup>[2]</sup> indicates that an *in vitro* bioequivalence study should be based on testing a suitable number of bottle or canisters from each of three batches (lots) of the test

and reference drug products. The number of bottles (canisters) to be studied for each batch (lot) should be no less than 10. As a result, a minimum of 30 samples is required for each of the test and the reference drug products.

However, the FDA's requirement may not yield sufficient power for the establishment of *in vitro* bioequivalence between drug products. Chow et al.<sup>[5]</sup> propose a procedure for determining sample sizes as follows.

In practice, it is commonly preferred to choose  $m = m_T = m_R$  and  $n = n_T = n_R$  so that the power of the bioequivalence test reaches a given level  $\beta$  (say 80%) when the unknown parameter vector  $\psi = (\delta, \sigma_{BT}^2, \sigma_{BR}^2, \sigma_{WT}^2, \sigma_{WR}^2)$  is set at some initial guessing value  $\hat{\psi}$  for which the value of  $\zeta$  (denoted by  $\hat{\zeta}$ ) is negative. Let  $U$  be given in the definition of  $\hat{\zeta}_U$  and let  $U_\beta$  be the same as  $U$  but with 5% and 95% replaced by  $1 - \beta$  and  $\beta$ , respectively. Because:

$$P(\hat{\zeta}_U < \zeta + \sqrt{U} + \sqrt{U_\beta}) \approx \beta$$

the power of the bioequivalence test  $P(\hat{\zeta}_U < 0)$  is approximately larger than  $\beta$  if  $\zeta + \sqrt{U} + \sqrt{U_\beta} \leq 0$ . Let

**Table 1** Selected sample sizes  $m^*$  and  $n^*$  and the actual power  $p$  when  $\theta_{BE} = 1.125$  and  $\sigma_0 = 0.2$  (10,000 simulations)

$\sigma_{BT}$	$\sigma_{BR}$	$\sigma_{WT}$	$\sigma_{WR}$	$\delta$	Step 1	Step 2		Step 2'	
					$p$	$m^*, n^*$	$p$	$m^*, n^*$	$p$
0	0	0.25	0.25	0.0530	0.4893	55, 1	0.7658	30, 2	0.7886
				0	0.5389	47, 1	0.7546	30, 2	0.8358
		0.25	0.50	0.4108	0.6391	45, 1	0.7973	30, 2	0.8872
				0.2739	0.9138	—	—	—	—
		0.50	0.50	0.1061	0.4957	55, 1	0.7643	30, 2	0.7875
				0	0.5362	47, 1	0.7526	30, 2	0.8312
0.25	0.25	0.25	0.25	0.0750	0.4909	55, 1	0.7774	30, 3	0.7657
				0	0.5348	47, 1	0.7533	30, 2	0.7323
		0.25	0.50	0.4405	0.5434	57, 1	0.7895	30, 3	0.8489
				0.2937	0.8370	—	—	—	—
		0.50	0.50	0.1186	0.4893	55, 1	0.7683	30, 2	0.7515
				0	0.5332	47, 1	0.7535	30, 2	0.8091
0.50	0.25	0.25	0.50	0.1186	0.4903	55, 1	0.7660	30, 4	0.7586
				0	0.5337	47, 1	0.7482	30, 3	0.7778
		0.25	0.25	0.2937	0.8357	—	—	—	—
				0.50	0.5016	55, 1	0.7717	30, 4	0.7764
		0.50	0.50	0	0.5334	47, 1	0.7484	30, 3	0.7942
				0.25	0.5809	45, 1	0.7882	30, 2	0.7884
0.25	0.50	0.25	0.50	0.3873	0.9184	—	—	—	—
				0.3464	0.6766	38, 1	0.7741	30, 2	0.8661
		0.50	0.50	0.1732	0.8470	—	—	—	—
				0.3464	0.6829	38, 1	0.7842	30, 2	0.8045
		0.50	0.50	0.1732	0.8450	—	—	—	—
				0.50	0.4969	55, 1	0.7612	30, 3	0.7629
0	0.5406	47, 1	0.7534	30, 2	0.7270				

In Step 1,  $m^* = 30, n^* = 1$ .

Copyright © 2003 by Marcel Dekker, Inc. All rights reserved.

$\tilde{U}$  and  $\tilde{U}_\beta$  be  $U$  and  $U_\beta$ , respectively, with  $(\hat{\delta}, s_{\text{BT}}^2, s_{\text{BR}}^2, s_{\text{WT}}^2, s_{\text{WR}}^2)$  replaced by  $\tilde{\psi}$ . Then, the sample sizes  $m = m_T = m_R$  and  $n = n_T = n_R$  that produce a test with a power of approximately  $\beta$  should satisfy:

$$\tilde{\zeta} + \sqrt{\tilde{U}} + \sqrt{\tilde{U}_\beta} \leq 0 \quad (6)$$

As a result, having a large  $m$  and a small  $n$  is an advantage when  $mn$ , the total number of observations for one treatment, is fixed. Thus the optimal sample size and the replicates combination can be determined by the following procedure:

- Step 1. Set  $m = 30$  and  $n = 1$ . If Eq. 6 holds, stop, and the required sample sizes are  $m = 30$  and  $n = 1$ ; otherwise, go to Step 2.
- Step 2. Let  $n = 1$  and find a smallest integer  $m_*$  such that Eq. 6 holds. If  $m_* \leq m_+$  (the largest possible number of canisters in a given problem), stop, and the required sample sizes are  $m = m_*$  and  $n = 1$ ; otherwise, go to Step 3.
- Step 3. Let  $m = m_+$  and find a smallest integer  $n_*$  such that Eq. 6 holds. The required sample sizes are  $m = m_+$  and  $n = n_*$ .

If in practice it is much easier and inexpensive to obtain more replicates than to sample more canisters, then Steps 2–3 in the previous procedure can be replaced by:

- Step 2'. Let  $m = 30$  and find a smallest integer  $n_*$  such that Eq. 6 holds. The required sample sizes are  $m = 30$  and  $n = n_*$ .

Because selecting sample sizes according to Eq. 6 only produces a test with approximate power  $\beta$ , we conduct a simulation study to examine the actual power corresponding to the selected sample sizes according to Steps 1–3 (or Steps 1 and 2'). That is, for a given combination of parameter values in Table 1, we select sample sizes  $m_*$  and  $n_*$  according to Steps 1–3 or Steps 1 and 2' with  $\beta = 80\%$ , and then compute the actual power  $p$  corresponding to the selected  $m_*$  and  $n_*$  by 10,000 simulations. Note that  $m_+$  is set to be  $\infty$  in the simulation so that Step 3 is not needed.

The performance of the above sample size determination procedure is evaluated by simulation. Table 1 shows the selected sample sizes for a nominal power of 80%, as reported by Chow et al.<sup>[5]</sup> The results show that the selected sample sizes produce a test with power  $\geq 75\%$ , except for two cases where the power is about 73%.

## CONCLUSION

For the assessment of *in vitro* bioequivalence, the FDA requires that *in vitro* testing of emitted dose uniformity, droplet size distribution, spray pattern, plume geometry, priming and re-priming, and tail-off profile be done to demonstrate comparable delivery characteristics between two drug products. Those tests can be divided into two categories. They are, namely, profile and nonprofile analysis. For profile analysis, a statistical procedure similar to the one for testing individual bioequivalence<sup>[4]</sup> are adopted. However, for profile analysis, no satisfactory statistical procedure is available for establishment of *in vitro* bioequivalence. Further study is needed.

## REFERENCES

1. Chow, S.C.; Liu, J.P. *Design and Analysis of Bioavailability and Bioequivalence Studies*, 2nd Ed.; Marcel Dekker: New York, 1999.
2. FDA. *Guidance for Industry on Bioavailability and Bioequivalence Studies for Nasal Aerosols and Nasal Sprays for Local Action*; Center for Drug Evaluation and Research, Food and Drug Administration: Rockville, MD, 1999.
3. FDA. *Guidance for Industry: Bioavailability and Bioequivalence Studies for Orally Administered Drug Products—General Considerations*; Center for Drug Evaluation and Research, Food and Drug Administration: Rockville, MD, 2000.
4. Hyslop, T.; Hsuan, F.; Holder, D.J. A small sample confidence interval approach to assess individual bioequivalence. *Stat. Med.* **2000**, *19*, 2885–2897.
5. Chow, S.C.; Shao, J.; Wang, H. *In vitro* bioequivalence testing. *Stat. Med.*, *in press*.

