STATISTICS IN MEDICINE
Statist. Med. 2003; **22**:55–68 (DOI: 10.1002/sim.1345)

In vitro bioequivalence testing

Shein-Chung Chow[1], Jun Shao[2,*] and Hansheng Wang[2]

[1] *Statplus Inc., Heston Hall, Suite 206, 1790 Yardley-Langhorne Road, Yardley, PA 19067, U.S.A.*
[2] *Department of Statistics, University of Wisconsin, Madison, 1210 W. Dayton Street, WI 53706, U.S.A.*

SUMMARY

A statistical test is proposed for *in vitro* bioequivalence testing between drug products such as nasal aerosols and nasal sprays. The proposed test generalizes the one recommended in the FDA 1999 guidance to the situation where replicated observations obtained from each sampled canister or bottle of the drug product are available. The technique developed by Hyslop, Hsuan and Holder is used so that the proposed test is asymptotically accurate. The type I error probability and power of the proposed test are investigated through a simulation study. A method for determining the required sample size to achieve a desired power is also proposed. A numerical example is given for illustration. Copyright © 2003 John Wiley & Sons, Ltd.

KEY WORDS: between- and within-canister variabilities; non-profile analysis; power; type I error; sample size

1. INTRODUCTION

Bioequivalence testing is considered as a surrogate for clinical evaluation of the therapeutic equivalence of drug products based on the *fundamental bioequivalence assumption* that when two drug products (for example, a brand-name drug and its generic copy) are equivalent in bioavailability, they will reach the same therapeutic effect (Chow and Liu [1]). Bioavailability for *in vivo* bioequivalence studies is usually assessed through the measures of the rate and extent to which the drug product is absorbed into the bloodstream of human subjects. For some locally acting drug products such as nasal aerosols (for example, metered-dose inhalers) and nasal sprays (for example, 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 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

* Correspondence to: Jun Shao, Department of Statistics, University of Wisconsin, 1210 W. Dayton Street, Madison, WI 53706, U.S.A.

Received February 2001
Accepted May 2002

Copyright © 2003 John Wiley & Sons, Ltd.

tests are not clearly established until a guidance on bioavailability and bioequivalence studies for nasal aerosols and nasal sprays for local action was issued by the FDA (see reference [2]). As indicated in the FDA guidance, *in vitro* bioequivalence can be established through six *in vitro* bioequivalence tests, which are for dose or spray content uniformity through container life, droplet or particle size distribution, spray pattern, plume geometry, priming and repriming, and tail-off distribution.

The FDA classifies statistical methods for assessment of the six *in vitro* bioequivalence tests for nasal aerosols and sprays as either the non-profile analysis or the profile analysis. In this paper we focus on the non-profile analysis, which applies to tests for dose or spray content uniformity through container life, droplet size distribution, spray pattern and priming and repriming (see reference [2]). The criterion for *in vitro* bioequivalence can be 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

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

θ_{BE} is a prespecified equivalence limit, and σ_0^2 is a prespecified constant. Values of σ_0^2 and θ_{BE} can be found in the FDA guidance. According to the FDA guidance, *in vitro* bioequivalence can be claimed if the hypothesis that $\theta \geqslant \theta_{BE}$ is rejected at the 5 per cent level of significance (see reference [2]).

To ensure that the statistical test has a significant level close to the nominal level 5 per cent with a desired power, the FDA requires that at least 30 canisters or bottles (units) of each of the test and reference products be tested. When 30 canisters per product are not enough to achieve a desired power (say 80 per cent), we may increase the number of observations by either increasing the number of canisters per product or obtaining some replicates from each canister. When replicates from each canister are considered, the test procedure recommended in the FDA guidance has to be modified in order to address between- and within-canister variabilities. This is studied in Section 2.

Section 2 reviews the test procedure recommended by the FDA guidance and introduces our proposed test procedure for the case where replicates are obtained in each canister. Some simulation results are given in Section 3 to show the finite sample properties of the proposed test. In Section 4 we propose a procedure to determine the sample sizes (the number of canisters and the number of replicates within each canister) to achieve a desired power. A numerical example is given in Section 5 for illustration.

2. TESTS FOR *IN VITRO* BIOEQUIVALENCE

Suppose that m_T and m_R canisters from, respectively, the test and the reference products are randomly selected for *in vitro* bioequivalence testing and one observation from each canister is obtained. The data can be described by the following model:

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

Copyright © 2003 John Wiley & Sons, Ltd.

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

$$\theta = \frac{(\mu_\mathrm{T} - \mu_\mathrm{R})^2 + \sigma_\mathrm{T}^2 - \sigma_\mathrm{R}^2}{\max\{\sigma_0^2, \sigma_\mathrm{R}^2\}} \tag{3}$$

and $\theta < \theta_\mathrm{BE}$ if and only if $\zeta < 0$, where

$$\zeta = (\mu_\mathrm{T} - \mu_\mathrm{R})^2 + \sigma_\mathrm{T}^2 - \sigma_\mathrm{R}^2 - \theta_\mathrm{BE} \max\{\sigma_0^2, \sigma_\mathrm{R}^2\} \tag{4}$$

To test bioequivalence at level 5 per cent, it suffices to construct a 95 per cent upper confidence bound for ζ.

Under model (2), the best unbiased estimator of $\delta = \mu_\mathrm{T} - \mu_\mathrm{R}$ is

$$\hat{\delta} = \bar{y}_\mathrm{T} - \bar{y}_\mathrm{R} \sim \mathrm{N}\left(0, \ \frac{\sigma_\mathrm{T}^2}{m_\mathrm{T}} + \frac{\sigma_\mathrm{R}^2}{m_\mathrm{R}}\right)$$

where $\bar{y}_k$ is the average of y_{jk} over j for a fixed k. The best unbiased estimator of σ_k^2 is

$$s_k^2 = \frac{1}{m_k - 1} \sum_{j=1}^{m_k} (y_{jk} - \bar{y}_k)^2 \sim \frac{\sigma_k^2 \chi_{m_k - 1}^2}{m_k - 1}$$

where $k = \mathrm{T}, \mathrm{R}$ and χ_t^2 denotes the central chi-square distribution with t degrees of freedom. Using the method in reference [3] for individual bioequivalence testing, an approximate 95 per cent upper confidence bound for ζ in (4) is

$$\tilde{\zeta}_U = \hat{\delta}^2 + s_\mathrm{T}^2 - s_\mathrm{R}^2 - \theta_\mathrm{BE} \max\{\sigma_0^2, s_\mathrm{R}^2\} + \sqrt{U}$$

where U is the sum of the following three quantities:

$$\left[\left(|\hat{\delta}| + z_{0.95}\sqrt{\left\{\frac{s_\mathrm{T}^2}{m_\mathrm{T}} + \frac{s_\mathrm{R}^2}{m_\mathrm{R}}\right\}}\right)^2 - \hat{\delta}^2\right]^2$$

$$s_\mathrm{T}^4 \left(\frac{m_\mathrm{T} - 1}{\chi_{0.05; m_\mathrm{T} - 1}^2} - 1\right)^2$$

and

$$(1 + c\theta_\mathrm{BE})^2 s_\mathrm{R}^4 \left(\frac{m_\mathrm{R} - 1}{\chi_{0.95; m_\mathrm{R} - 1}^2} - 1\right)^2$$

$c = 1$ if $s_\mathrm{R}^2 \geqslant \sigma_0^2$, $c = 0$ if $s_\mathrm{R}^2 < \sigma_0^2$, z_a is the ath quantile of the standard normal distribution, and $\chi_{t;a}^2$ is the ath quantile of the central chi-square distribution with t degrees of freedom. *In vitro* bioequivalence can be claimed if $\tilde{\zeta}_U < 0$. This procedure is recommended by the FDA guidance.

Copyright © 2003 John Wiley & Sons, Ltd.

As indicated in the 1999 guidance, the FDA requires that m_k be at least 30. However, $m_k = 30$ may not be enough to achieve a desired power of the bioequivalence test in some situations (see, for example, the simulation results in Section 3). Increasing m_k can certainly increase the power, but in some situations, obtaining replicates from each canister may be more practical, and/or cost-effective. With replicates from each canister, however, the previously described test procedure is necessarily modified in order to address the between- and within-canister variabilities.

Suppose that there are n_k replicates from each canister for product k. We assume that there is no period effect so that the replicates from a canister have the same distribution. Let y_{ijk} be the ith replicate in the jth canister under product k. Let b_{jk} be the between-canister variation and e_{ijk} be the within-canister measurement error. Then

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

where b_{jk}'s and e_{ijk}'s are independent, $b_{jk} \sim \mathrm{N}(0, \sigma^2_{\mathrm{B}k})$ and $e_{ijk} \sim \mathrm{N}(0, \sigma^2_{\mathrm{W}k})$. Under model (5), the total variances σ^2_{T} and σ^2_{R} in (3) and (4) are equal to $\sigma^2_{\mathrm{BT}} + \sigma^2_{\mathrm{WT}}$ and $\sigma^2_{\mathrm{BR}} + \sigma^2_{\mathrm{WR}}$, respectively, that is, the sums of between-canister and within-canister variances. The parameter θ in (1) is still given by (3) and $\theta < \theta_{\mathrm{BE}}$ if and only if $\zeta < 0$, where ζ is given in (4).

Under model (5), the best unbiased estimator of $\delta = \mu_{\mathrm{T}} - \mu_{\mathrm{R}}$ is

$$\hat{\delta} = \bar{y}_{\mathrm{T}} - \bar{y}_{\mathrm{R}} \sim \mathrm{N}\left(0, \frac{\sigma^2_{\mathrm{BT}}}{m_{\mathrm{T}}} + \frac{\sigma^2_{\mathrm{BR}}}{m_{\mathrm{R}}} + \frac{\sigma^2_{\mathrm{WT}}}{m_{\mathrm{T}} n_{\mathrm{T}}} + \frac{\sigma^2_{\mathrm{WR}}}{m_{\mathrm{R}} n_{\mathrm{R}}}\right)$$

where $\bar{y}_k$ is the average of y_{ijk} over i and j for a fixed k.

To construct a confidence bound for ζ in (4) using the approach in reference [3], it suffices to find independent, unbiased and chi-square distributed estimators of σ^2_{T} and σ^2_{R} that are independent of $\hat{\delta}$. These estimators, however, are not available when $n_k > 1$. Note that

$$\sigma^2_k = \sigma^2_{\mathrm{B}k} + n_k^{-1}\sigma^2_{\mathrm{W}k} + (1 - n_k^{-1})\sigma^2_{\mathrm{W}k}, \quad k = \mathrm{T}, \mathrm{R}$$

$\sigma^2_{\mathrm{B}k} + n_k^{-1}\sigma^2_{\mathrm{W}k}$ can be estimated by

$$s^2_{\mathrm{B}k} = \frac{1}{m_k - 1}\sum_{j=1}^{m_k} (\bar{y}_{jk} - \bar{y}_k)^2 \sim \frac{(\sigma^2_{\mathrm{B}k} + n_k^{-1}\sigma^2_{\mathrm{W}k})\chi^2_{m_k - 1}}{m_k - 1}$$

where $\bar{y}_{jk}$ is the average of y_{ijk} over i; $\sigma^2_{\mathrm{W}k}$ can be estimated by

$$s^2_{\mathrm{W}k} = \frac{1}{m_k(n_k - 1)}\sum_{j=1}^{m_k}\sum_{i=1}^{n_k} (y_{ijk} - \bar{y}_{jk})^2 \sim \frac{\sigma^2_{\mathrm{W}k}\chi^2_{m_k(n_k-1)}}{m_k(n_k - 1)}$$

and $\hat{\delta}$, $s^2_{\mathrm{B}k}$, $s^2_{\mathrm{W}k}$, $k = \mathrm{T}, \mathrm{R}$, are independent. Thus, applying the approach in reference [3] we obtain the following approximate 95 per cent upper confidence bound for ζ in (4):

$$\begin{aligned}\hat{\zeta}_U = {} & \hat{\delta}^2 + s^2_{\mathrm{BT}} + (1 - n_{\mathrm{T}}^{-1})s^2_{\mathrm{WT}} - s^2_{\mathrm{BR}} - (1 - n_{\mathrm{R}}^{-1})s^2_{\mathrm{WR}} \\ & - \theta_{\mathrm{BE}} \max\{\sigma^2_0, s^2_{\mathrm{BR}} + (1 - n_{\mathrm{R}}^{-1})s^2_{\mathrm{WR}}\} + \sqrt{U}\end{aligned}$$

Copyright © 2003 John Wiley & Sons, Ltd.

where U is the sum of the following five quantities:

$$\left[\left(|\hat{\delta}| + z_{0.95}\sqrt{\left\{\frac{s_{\mathrm{BT}}^2}{m_{\mathrm{T}}} + \frac{s_{\mathrm{BR}}^2}{m_{\mathrm{R}}}\right\}}\right)^2 - \hat{\delta}^2\right]^2$$

$$s_{\mathrm{BT}}^4\left(\frac{m_{\mathrm{T}}-1}{\chi^2_{0.05;m_{\mathrm{T}}-1}} - 1\right)^2$$

$$(1-n_{\mathrm{T}}^{-1})^2 s_{\mathrm{WT}}^4\left(\frac{m_{\mathrm{T}}(n_{\mathrm{T}}-1)}{\chi^2_{0.05;m_{\mathrm{T}}(n_{\mathrm{T}}-1)}} - 1\right)^2$$

$$(1+\theta_{\mathrm{BE}})^2 s_{\mathrm{BR}}^4\left(\frac{m_{\mathrm{R}}-1}{\chi^2_{0.95;m_{\mathrm{R}}-1}} - 1\right)^2$$

and

$$(1+c\theta_{\mathrm{BE}})^2(1-n_{\mathrm{R}}^{-1})^2 s_{\mathrm{WR}}^4\left(\frac{m_{\mathrm{R}}(n_{\mathrm{R}}-1)}{\chi^2_{0.95;m_{\mathrm{R}}(n_{\mathrm{R}}-1)}} - 1\right)^2$$

and $c=1$ if $s_{\mathrm{BR}}^2+(1-n_{\mathrm{R}}^{-1})s_{\mathrm{WR}}^2 \geqslant \sigma_0^2$ and $c=0$ if $s_{\mathrm{BR}}^2+(1-n_{\mathrm{R}}^{-1})s_{\mathrm{WR}}^2 < \sigma_0^2$. *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 $\tilde{\zeta}_U$ 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 per cent.

3. SIMULATION RESULTS

We consider a simulation study to examine the finite sample performance (type I error probability and power) of the *in vitro* bioequivalence test described in Section 2.

For each combination of variance parameters shown in Table I, data are generated according to model (5). When the type I probability is considered, the value of $\delta=\mu_{\mathrm{T}}-\mu_{\mathrm{R}}$ is chosen so that $\zeta=0$ (that is, two products are not bioequivalent). When the power is considered, values of δ are selected so that $\zeta<0$. The values of θ_{BE} and σ_0 are chosen to be 1.125 and 0.2, respectively, according to the FDA guidance. We consider the case where $m_{\mathrm{T}}=m_{\mathrm{R}}=m$ and $n_{\mathrm{T}}=n_{\mathrm{R}}=n$ and $(m,n)=(60,1),(50,1),(40,1),(30,1),(30,2)$ and $(30,3)$. The reason to consider these combinations of sample sizes is that $(m,n)=(30,1)$ is the minimum sample size required by the FDA guidance. If $m=30$ is not enough, then we may increase either m (to $40,50$ or 60) or n (to 2 or 3).

The simulation results for the type I error probability based on 10 000 runs are given in Table I. The type I error probability ranges from 3.41 per cent to 5.17 per cent. It is under the nominal level 5 per cent in all but two cases and is generally close to 5 per cent.

Copyright © 2003 John Wiley & Sons, Ltd.

Table I. Type I error probabilities of bioequivalence test when $\theta_{BE} = 1.125$ and $\sigma_0 = 0.2$ (10 000 simulations).

σ_{BT}	σ_{BR}	σ_{WT}	σ_{WR}	Sample size (m, n)					
				(60,1)	(50,1)	(40,1)	(30,1)	(30,2)	(30,3)
0	0	0.25	0.25	0.0431	0.0427	0.0388	0.0381	0.0390	0.0471
		0.25	0.50	0.0479	0.0421	0.0433	0.0411	0.0503	0.0517
		0.50	0.50	0.0431	0.0408	0.0410	0.0382	0.0431	0.0470
0.25	0.25	0.25	0.25	0.0391	0.0430	0.0395	0.0373	0.0433	0.0423
		0.25	0.50	0.0444	0.0439	0.0449	0.0444	0.0451	0.0493
		0.50	0.50	0.0414	0.0423	0.0455	0.0424	0.0429	0.0454
0.50	0.25	0.25	0.50	0.0414	0.0412	0.0412	0.0402	0.0415	0.0436
		0.50	0.50	0.0389	0.0361	0.0408	0.0353	0.0355	0.0404
0.25	0.50	0.25	0.25	0.0434	0.0396	0.0428	0.0437	0.0401	0.0454
		0.50	0.25	0.0439	0.0442	0.0419	0.0396	0.0405	0.0391
		0.25	0.50	0.0431	0.0444	0.0446	0.0395	0.0460	0.0448
		0.50	0.50	0.0406	0.0431	0.0406	0.0379	0.0385	0.0420
0.50	0.50	0.25	0.25	0.0416	0.0430	0.0418	0.0369	0.0380	0.0392
		0.50	0.25	0.0398	0.0404	0.0360	0.0365	0.0350	0.0341
		0.25	0.50	0.0452	0.0453	0.0400	0.0391	0.0463	0.0439
		0.50	0.50	0.0468	0.0473	0.0401	0.0432	0.0425	0.0411

Figure 1 shows the power (against ζ) for different sample size combinations and variance parameters. The following are some observations from Figure 1:

1. Except for the cases where $\sigma_T^2 > \sigma_R^2$ (the total variability of the test product is larger than that of the reference product), a reasonable power (say 80 per cent) can be reached by either increasing the total number of observations or decreasing the value of ζ. For cases where $\sigma_T^2 > \sigma_R^2$ in the simulation, the chance of claiming bioequivalence is small when the two products are in fact bioequivalent.
2. The sample sizes $m = 30$ and $n = 1$ do not lead to a satisfactory power unless $\sigma_T^2 < \sigma_R^2$.
3. When the between-canister variances are non-zero, having more canisters is preferred than having more replicates within each canister (if the power of the bioequivalence test is the only concern). In all cases, the choice of $m = 60$ and $n = 1$ is better than the choice of $m = 30$ and $n = 2$ although the total number of observations are the same. The choice of $m = 60$ and $n = 1$ is even better than the choice of $m = 30$ and $n = 3$ in 11 out of 13 cases, although the latter has a larger total number of observations. In fact, the choice of $m = 30$ and $n = 3$ is even worse than the choice of $m = 50$ and $n = 1$ in some cases.
4. When $\sigma_{BT}^2 = \sigma_{BR}^2 = 0$ (that is, there is no canister-to-canister variation), the data sets for the cases of $(m, n) = (60, 1)$ and $(m, n) = (30, 2)$ are the same, but the test procedures are different, since, in the case of $(m, n) = (30, 2)$, we do not assume that one knows that $\sigma_{BT}^2 = \sigma_{BR}^2 = 0$. However, the results in Figure 1 show that the two procedures have almost the same power.

Copyright © 2003 John Wiley & Sons, Ltd.

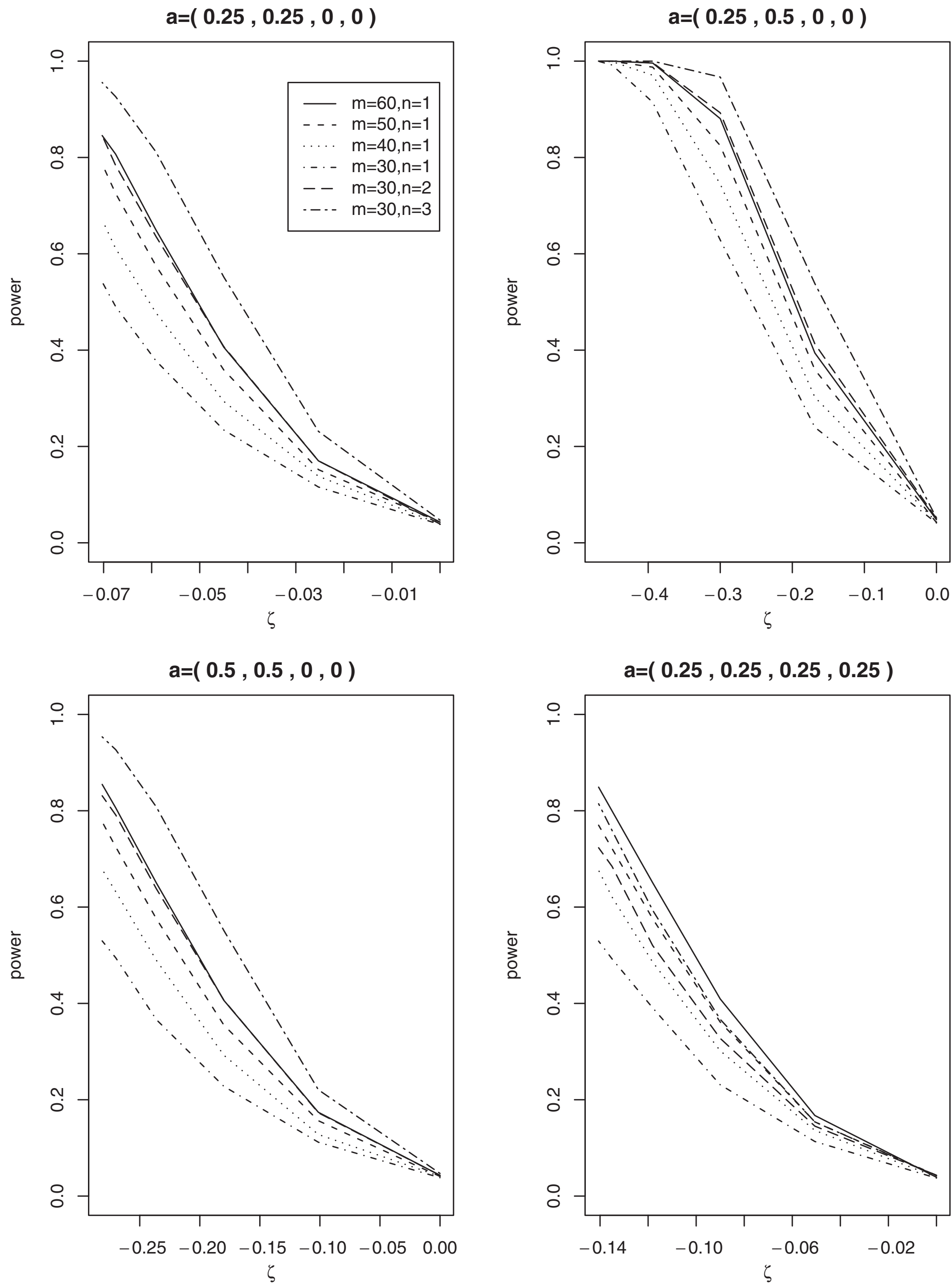

Figure 1. Power of the bioequivalence test versus ζ; $a = (\sigma_{WT}, \sigma_{WR}, \sigma_{BT}, \sigma_{BR})$.

Copyright © 2003 John Wiley & Sons, Ltd.

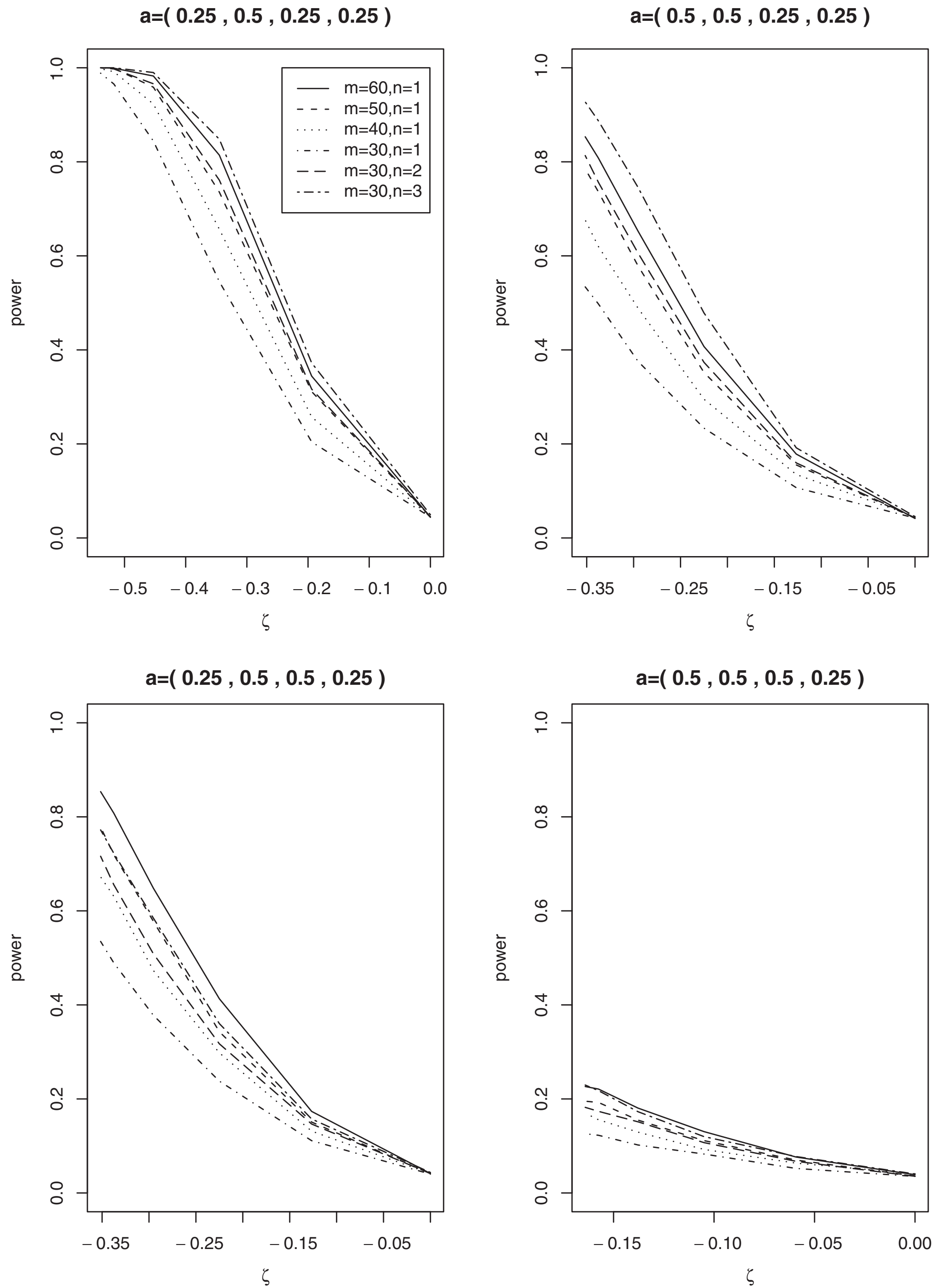

Figure 1. (*Continued*).

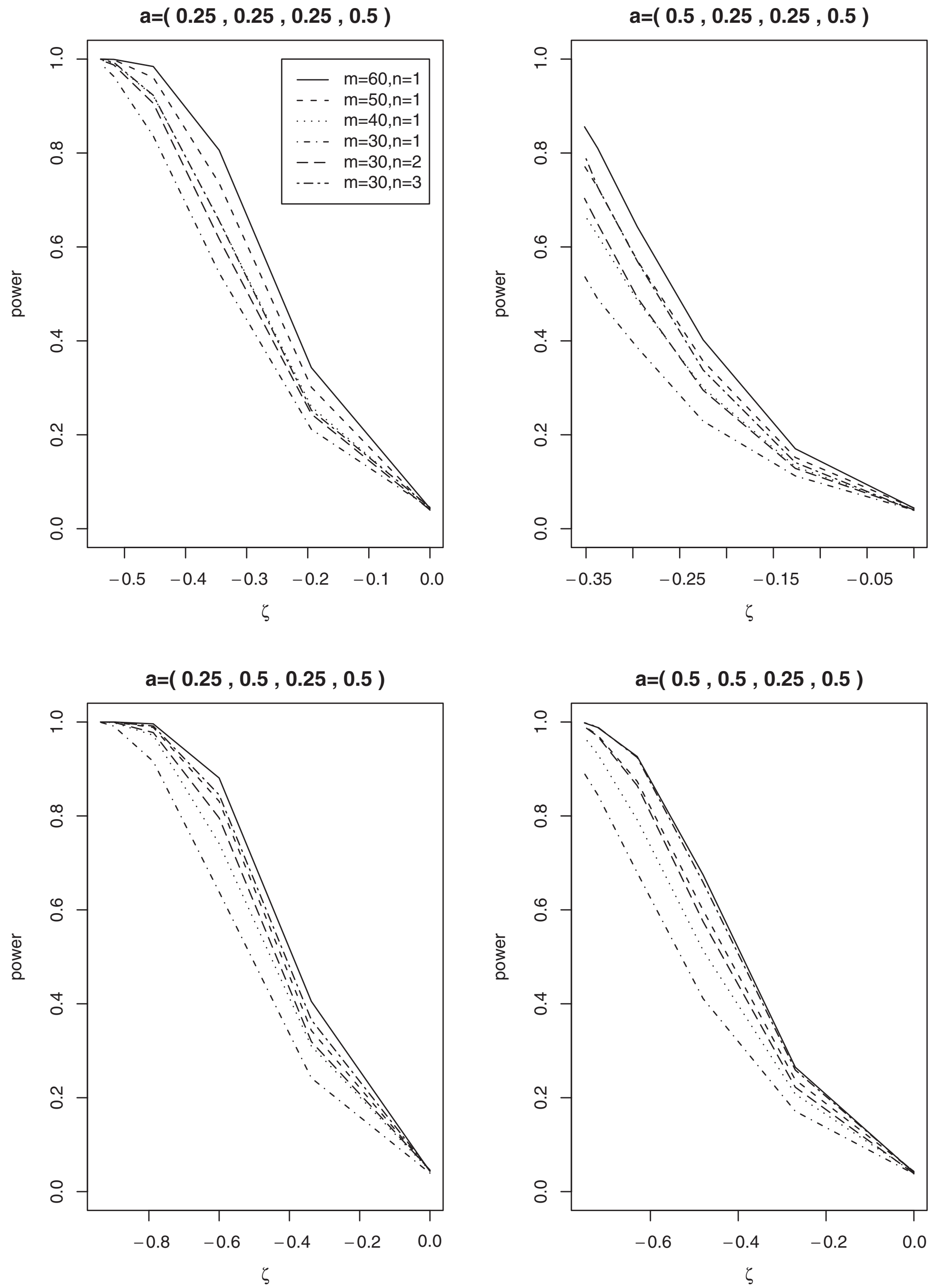

Figure 1. (*Continued*).

Copyright © 2003 John Wiley & Sons, Ltd.

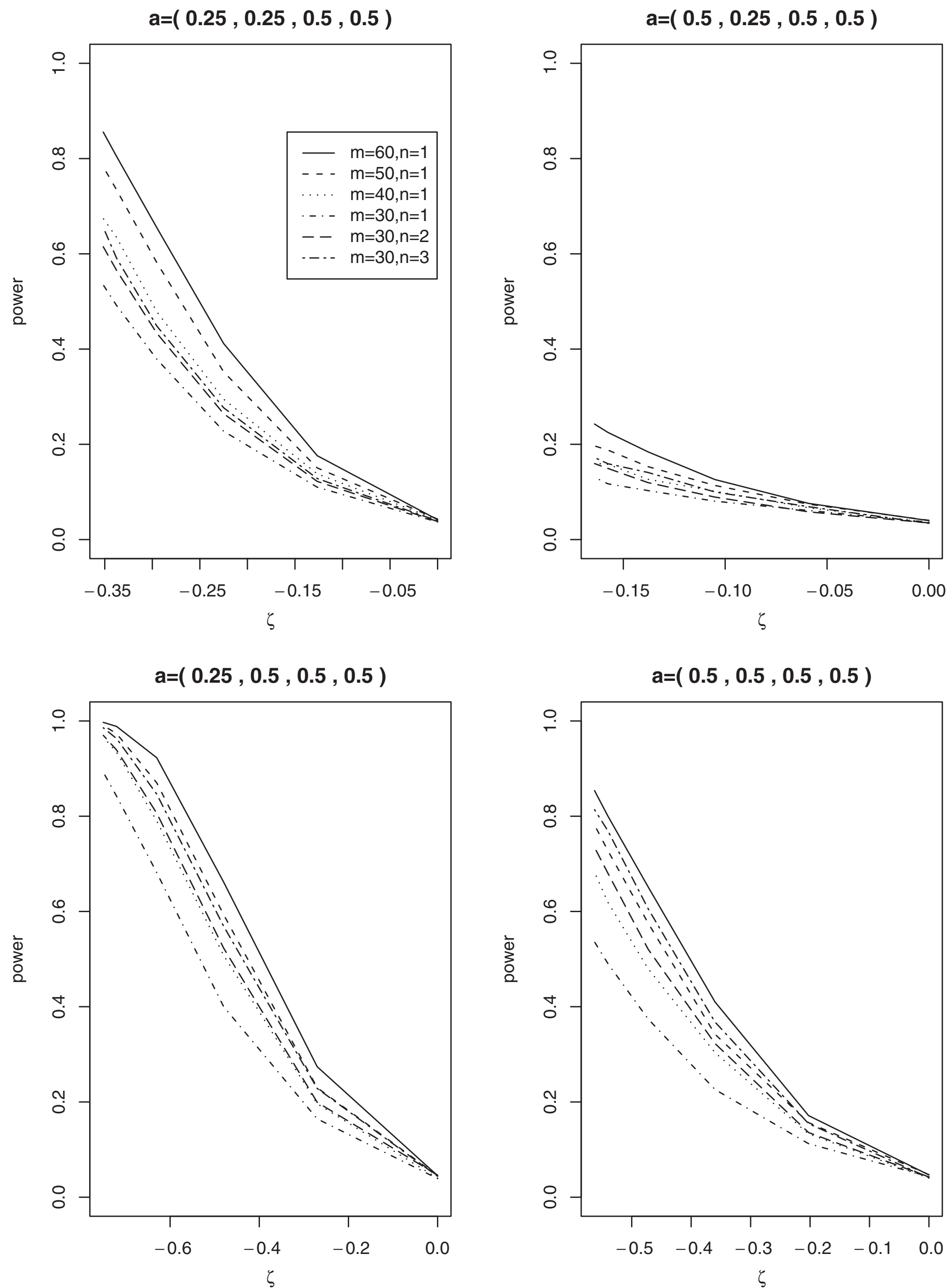

Figure 1. (*Continued*).

4. SAMPLE SIZE DETERMINATION

In view of the fact that the FDA requires $m_k \geqslant 30$ and that $m_k = 30$ and $n_k = 1$ may not produce a test with sufficient power, we propose a procedure for determining sample sizes as follows.

Typically, we would like 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 β (say 80 per cent) when the unknown parameter vector $\psi = (\delta, \sigma^2_{BT}, \sigma^2_{BR}, \sigma^2_{WT}, \sigma^2_{WR})$ is set at some initial guessing value $\tilde{\psi}$ for which the value of ζ (denoted by $\tilde{\zeta}$) is negative. Let U be given in the definition of $\hat{\zeta}_U$ and U_β be the same as U but with 5 per cent and 95 per cent replaced by $1-\beta$ and β, respectively. Since

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

the power of the bioequivalence test, $P(\hat{\zeta}_U < 0)$, is approximately larger than β if $\zeta + \sqrt{U} + \sqrt{U_\beta} \leqslant 0$. Let $\tilde{U}$ and $\tilde{U}_\beta$ be U and U_β, respectively, with $(\hat{\delta}, s^2_{BT}, s^2_{BR}, s^2_{WT}, s^2_{WR})$ replaced by $\tilde{\psi}$. Then, the sample sizes $m = m_T = m_R$ and $n = n_T = n_R$ that produce a test with power approximately β should satisfy

$$\tilde{\zeta} + \sqrt{\tilde{U}} + \sqrt{\tilde{U}_\beta} \leqslant 0 \qquad (6)$$

From the results in Section 3, having a large m and a small n is an advantage when mn, the total number of observations for one treatment, is fixed. Thus, we propose the following procedure:

Step 1. Set $m = 30$ and $n = 1$. If (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 (6) holds. If $m_* \leqslant m_+$ (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 (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 and 3 in the previous procedure can be replaced by

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

Since selecting sample sizes according to (6) only produces a test with approximate power β, 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 II, we select sample sizes m_* and n_* according to steps 1–3 or steps 1 and 2′ with $\beta = 80$ per cent 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 ∞ in the simulation so that step 3 is not needed. The results in Table II show that the selected sample sizes produce a test with power $\geqslant 75$ per cent except for two cases where the power is about 73 per cent.

Copyright © 2003 John Wiley & Sons, Ltd.

Table II. Selected sample sizes m_* and n_* and the actual power p when $\theta_{BE} = 1.125$ and $\sigma_0 = 0.2$ (10 000 simulations).

σ_{BT}	σ_{BR}	σ_{WT}	σ_{WR}	δ	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	0.50	0.25	0.25	0.2937	0.8357	—	—	—	—
		0.50	0.25	0.1186	0.5016	55,1	0.7717	30,4	0.7764
				0	0.5334	47,1	0.7484	30,3	0.7942
		0.25	0.50	0.5809	0.6416	45,1	0.7882	30,2	0.7884
				0.3873	0.9184	—	—	—	—
		0.50	0.50	0.3464	0.6766	38,1	0.7741	30,2	0.8661
				0.1732	0.8470	—	—	—	—
0.50	0.50	0.25	0.50	0.3464	0.6829	38,1	0.7842	30,2	0.8045
				0.1732	0.8450	—	—	—	—
		0.50	0.50	0.1500	0.4969	55,1	0.7612	30,3	0.7629
				0	0.5406	47,1	0.7534	30,2	0.7270

In Step 1, $m_* = 30$, $n_* = 1$.

5. A NUMERICAL EXAMPLE

We consider an *in vitro* bioequivalence study between two nasal spray products (a test and a reference). According to the FDA's guidance, the sponsor has to demonstrate similar drug delivery characteristics to the nasal cavity between the test and the reference products. This example contains data from one of the characteristics. There are 30 sampled canisters from each treatment and two replicates are obtained within each canister. A total of $2 \times 30 \times 2 = 120$ data are listed in Table III.

Using the formulae in Section 2, we obtain the following statistics and conclusion:

$$\hat{\delta} = -0.268069$$

$$s_{BT}^2 = 0.197011 \quad s_{BR}^2 = 0.185969$$

$$s_{WT}^2 = 0.182724 \quad s_{WR}^2 = 0.287895$$

Copyright © 2003 John Wiley & Sons, Ltd.

Table III. Data listing for bioequivalence testing of nasal spray products.

Test product		Reference product	
Replicate 1	Replicate 2	Replicate 1	Replicate 2
0.561666	1.125871	−0.039464	−0.432529
0.798124	1.430523	0.307254	0.267646
0.436184	−0.222413	0.253502	−0.119584
−0.024528	0.393267	−0.648004	−1.177679
−0.090290	−0.627502	0.301081	−0.494485
−0.272998	−0.153278	−0.616031	−0.191127
0.218929	−0.727645	−0.308066	0.234591
−0.540058	−0.053589	0.365324	−0.522845
−0.764153	0.368896	−0.151981	−0.391525
0.806935	1.073829	−0.313767	0.146926
0.285985	−0.079857	0.570084	−0.689592
−0.202443	−0.131901	−0.770030	−0.251785
0.443907	−0.043639	0.015629	0.458543
−0.164041	−0.688999	−0.468089	−0.501593
−1.065702	0.403384	0.167847	−0.268849
−0.320509	−0.584669	0.551757	0.733643
0.390335	−0.033335	0.069074	0.181878
−0.547364	0.351125	−0.006792	0.042096
0.113892	0.321204	−1.563092	−0.869159
−0.201228	0.978660	−0.221517	−0.407714
0.416043	−0.643256	0.624500	0.594661
−0.462978	0.780152	0.171163	1.107604
0.523478	−0.851068	0.386695	−0.810703
0.179169	0.040342	−1.273343	−0.050407
0.486552	0.564771	0.463084	0.172705
1.122766	0.070407	0.214514	0.385597
0.748173	0.236006	−0.208575	−0.301824
1.440898	0.262185	0.628295	−0.131757
−0.327088	0.314147	−0.562163	−1.000834
0.719226	0.428375	−0.794484	0.050426

$$s_{\mathrm{T}}^2 = s_{\mathrm{BT}}^2 + \tfrac{1}{2}s_{\mathrm{WT}}^2 = 0.292889 \quad s_{\mathrm{R}}^2 = s_{\mathrm{BR}}^2 + \tfrac{1}{2}s_{\mathrm{WR}}^2 = 0.329916$$

$$U = 0.062129$$

$$\hat{\zeta}_U = -0.091582 \qquad < 0$$

Conclusion: Bioequivalence

It is of interest to know what happens if we do not have replicates, that is, from each canister we only have one observation. For this purpose, we delete the second observation in each canister from Table III and apply the formulae in the FDA guidance (see Section 2). The results are given as follows:

$$\hat{\delta} = -0.252149$$

$$s_{\mathrm{T}}^2 = 0.307987 \quad s_{\mathrm{R}}^2 = 0.331149$$

Copyright © 2003 John Wiley & Sons, Ltd.

$$U = 0.120758$$

$$\tilde{\zeta}_U = 0.015377 \quad > 0$$

Conclusion: Not bioequivalence

In this example, estimates of parameters (δ and variances) based on data without replication are very similar to those based on data with replication, but the values of U, which determines the upper confidence bound for ζ, are very different in two cases. This indicates that in this example, the design of 30 canisters for each treatment without replication does not provide enough power in claiming bioequivalence when the two products are in fact bioequivalent.

ACKNOWLEDGEMENTS

The authors are grateful to two referees for their helpful comments and suggestions.

REFERENCES

1. Chow SC, Liu JP. *Design and Analysis of Bioavailability and Bioequivalence Studies*. 2nd edn. 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, Maryland, 1999.
3. Hyslop T, Hsuan F, Holder DJ. A small sample confidence interval approach to assess individual bioequivalence. *Statistics in Medicine* 2000; **19**:2885–2897.

Copyright © 2003 John Wiley & Sons, Ltd.