



On Testing of Individual Bioequivalence

Weizhen Wang

Journal of the American Statistical Association, Vol. 94, No. 447. (Sep., 1999), pp. 880-887.

Stable URL:

<http://links.jstor.org/sici?sici=0162-1459%28199909%2994%3A447%3C880%3AOTOIB%3E2.0.CO%3B2-W>

Journal of the American Statistical Association is currently published by American Statistical Association.

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/astata.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is an independent not-for-profit organization dedicated to creating and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact support@jstor.org.

On Testing of Individual Bioequivalence

Weizhen WANG

Individual bioequivalence has received much attention in the recent literature. One wants to determine whether it is suitable for a given individual to switch formulations from a reference drug to a test drug. Some researchers have argued that the formulation means, subject-by-formulation interaction, and within-subject variances should be included in a single measurement of inequivalence, and due to the complexity of the hypotheses there is a need to use a 2×3 or even higher-order crossover design for studies of this type. In this article exact level- α tests are first provided under normality assumptions, and a 2×3 crossover design is shown to be sufficient for assessing individual bioequivalence. Some simulations for the proposed test and two examples are presented.

KEY WORDS: Crossover design; Noncentral t distribution; Power; Reparametrization.

1. INTRODUCTION

Basically, there are two general concepts for bioequivalence: population bioequivalence and individual bioequivalence. As the term implies, population bioequivalence emphasizes whether two drugs have similar effects on the entire population of patients. There are two aspects of population bioequivalence: average bioequivalence, which is recommended by the U.S. Food and Drug Administration (FDA), and variance bioequivalence. Anderson and Hauck (1983), Brown, Hwang, and Munk (1997), and Schuirmann (1987) proposed several testing procedures to address the first problem. For the variance problem, two testing procedures have been provided by Liu and Chow (1992) and Wang (1997a). Wang (1997a) also pointed out that bioequivalences in mean and in intrasubject variability are indeed one problem.

However, physicians and patients may have more interest in whether the two drugs have similar effects on each individual. More precisely, for a given individual, is it suitable to switch to the generic drug from the brand name drug? Anderson and Hauck (1990), in their very influential work, first identified this problem and proposed the concept of individual bioequivalence. They assumed the following simplified 2×2 crossover design without the period effects:

$$Y_{iT} = \mu_T + b_{iT} + \varepsilon_{iT}$$

and

$$Y_{iR} = \mu_R + b_{iR} + \varepsilon_{iR}, \quad (1)$$

where Y_{ij} is the response of the i th subject for the j th formulation, where $j = R$ (reference formulation) or T (test formulation), $i = 1, 2, \dots, n$; μ_j is the population average of the j th formulation; b_{ij} is the mean deviation from the population average of a given individual; and ε_{ij} is the (within-subject) random error in observing Y_{ij} . It is assumed that b_{ij} and ε_{ij} are mutually independent. Define

$$p_{TR} = \Pr(|\mu_T + b_{iT} - \mu_R - b_{iR}| \leq r), \quad (2)$$

where r is a predetermined constant. Consider the hypotheses

$$H_0: p_{TR} \leq p_0 \quad \text{vs} \quad H_A: p_{TR} > p_0, \quad (3)$$

where p_0 is a constant larger than $1/2$. If the null hypothesis is rejected on more than $100p_0\%$ patients, then the patient can switch to the generic drug, and individual bioequivalence is established. Anderson and Hauck (1990) also provided a nonparametric testing procedure for (3) using a binomial distribution. This procedure is not always valid, and the size of the test may be much larger than the test level. A mild condition that guarantees the validity of their test was given by Hwang and Wang (1997). Under the normal assumptions on Y_{ij} , much more powerful tests than Anderson and Hauck's for (3) have been given by Wang (1995) and Wang and Hwang (1997).

One shortcoming of individual bioequivalence hypotheses (3) is that the within-subject variability has not been taken into account. Noticing this, Schall and Luus (1993) and Schall (1995) suggested using a 2×3 crossover design in which the test formulation and the reference formulation are administered on each subject once and twice,

$$Y_{iT} = \mu_T + b_{iT} + \varepsilon_{iT},$$

$$Y_{iR} = \mu_R + b_{iR} + \varepsilon_{iR},$$

and

$$Y'_{iR} = \mu_R + b_{iR} + \varepsilon'_{iR}, \quad (4)$$

so that one can compare the difference between two formulations with the variation of the reference formulation on each subject. Let $\sigma_T^2 = \text{var}(\varepsilon_{iT})$, $\sigma_R^2 = \text{var}(\varepsilon_{iR}) = \text{var}(\varepsilon'_{iR})$, and $\sigma_D^2 = \text{var}(b_{iT} - b_{iR})$. For mathematical convenience, Schall and Luus (1993) proposed a moment-based approach to assess individual bioequivalence comparing the second moments of $Y_{iT} - Y_{iR}$ and of $Y'_{iR} - Y_{iR}$ that is equal to $2\sigma_R^2$. More precisely, consider the hypotheses

$$H_0: E|Y_{iT} - Y_{iR}|^2 \geq 2\gamma_m \sigma_R^2 \quad \text{versus}$$

$$H_A: E|Y_{iT} - Y_{iR}|^2 < 2\gamma_m \sigma_R^2, \quad (5)$$

Weizhen Wang is Assistant Professor, Department of Mathematics and Statistics, Wright State University, Dayton, OH 45435 (E-mail: wwang@math.wright.edu). The author gratefully acknowledges two referees, the associate editor, and editor for their helpful comments and suggestions that substantially improved the article.

© 1999 American Statistical Association
Journal of the American Statistical Association
September 1999, Vol. 94, No. 447, Theory and Methods

where γ_m is a prespecified constant. Individual bioequivalence is established if we successfully reject the null hypothesis. More generally, the FDA (U.S. Food and Drug Administration 1997) draft guidance on bioequivalence suggests considering

$$H_0: \frac{\theta^2 + c_1\sigma_D^2 + c_2(\sigma_T^2 - \sigma_R^2)}{\max\{\sigma_R^2, a\}} \geq \gamma_1 \quad \text{versus} \\ H_A: \frac{\theta^2 + c_1\sigma_D^2 + c_2(\sigma_T^2 - \sigma_R^2)}{\max\{\sigma_R^2, a\}} < \gamma_1, \quad (6)$$

where $\theta = \mu_T - \mu_R$ and c_1, c_2, γ_1 , and a are nonnegative constants (see also Anderson and Hauck 1996). Schall and Luus (1993) extended the idea of Anderson and Hauck (1990) and proposed another type of individual bioequivalence using probability instead of moment; see (20). It is clear that (6) reduces to (5) if we choose $a = 0$ and choose c_i and γ_1 appropriately. The new types of individual bioequivalence hypotheses in both (5) and (6) have better interpretation than (3), because variation among the responses of the reference formulation has been taken into account. These follow a so-called aggregate approach (see Anderson and Hauck 1996; Chen 1997), because univariate criteria are used. Liu and Chow (1996) used a disaggregate approach in which three sets of hypotheses with regard to three parameters $\sigma_T^2/\sigma_R^2, \sigma_D^2$, and θ are considered simultaneously (see also Liu 1995). However, it would be difficult to declare individual bioequivalence if using the disaggregate approach. Thus in this article I focus on the aggregate approaches.

Example 1. Some real datasets from bioequivalence studies are available on the FDA web site: <http://www.fda.gov/cder/bioequivdata/index.htm>.

Monoamine oxidase (MAO) inhibitor (drug 14), a generic drug to treat depression, and is compared with the reference drug using the following 2×4 crossover design:

	Period			
	I	II	III	IV
Sequence 1	T	R	R	T
Sequence 2	R	T	T	R

which is similar to that recommended by the FDA (U.S. Food and Drug Administration 1997). The pharmacokinetic parameters, AUC_{0-t} , $AUC_{0-\infty}$, and C_{\max} (see U.S. Food and Drug Administration 1997 for definitions), are recorded from 38 subjects on 4 components A, B, C, and D of the MAO inhibitor. The test formulation and the reference formulation are both administered twice to each subject so that the subject-by-formulation interaction and the within-subject variation of the reference formulation can be detected. However, as I show later, such a 2×4 design is not necessary to achieve this. The goal here is to establish the individual bioequivalence for the MAO inhibitor by performing data analysis on one or several pharmacokinetic parameters.

Now, two questions arise:

1. Is the 2×3 crossover design (4) sufficient to assess individual bioequivalence?
2. How does one construct tests for (5)?

In the case of no subject-by-formulation interaction (i.e., $\sigma_D^2 = 0$), Wang (1997b) showed that the 2×2 crossover design (1) is sufficient for assessing bioequivalence (5). He considered a general class of testing problems that includes (5) for $\sigma_D^2 = 0$ and provided exact α -level tests. However, the subject-by-formulation interaction cannot be ignored in general. Here I will show that the 2×3 crossover design (4) is sufficient to assess bioequivalence (5) under the following normality assumptions:

$$b_{iT} - b_{iR} \stackrel{\text{iid}}{\sim} N(0, \sigma_D^2), \quad \varepsilon_{iT} \stackrel{\text{iid}}{\sim} N(0, \sigma_T^2), \quad \varepsilon'_{iR} \stackrel{\text{iid}}{\sim} N(0, \sigma_R^2),$$

with the b 's independent of the ε 's. So far, there is no solid statistical testing procedure for (5) other than bootstrap results (see Schall 1995) when $\sigma_D^2 \neq 0$. In a bioequivalence study the sample size is quite limited, typically from 20 to 30. Exact testing procedures based on a finite sample would certainly be desirable. In this article I derive exact level- α tests for (5) using an approach similar to that specified in an earlier work (Wang 1997b).

Section 2 contains all preliminaries. Section 3 constructs two tests for (5) and presents the numerical calculations. Section 4 discusses two real examples, including Example 1, and gives a detailed step-by-step description of how to implement the proposed test. Section 5 applies the method developed in Section 3 to the other types of individual bioequivalence hypotheses, including (6) and (20). The Appendix provides all proofs.

2. PRELIMINARIES

I first introduce a reparametrization:

$$\theta = \mu_T - \mu_R, \quad \sigma^2 = \sigma_D^2 + \sigma_T^2 + \sigma_R^2/2, \quad \beta = \sigma_R^2/\sigma^2.$$

Then by straightforward calculation, (5) can be written as

$$H_0: \left| \frac{\theta}{\sigma} \right| \geq H_m(\beta) \quad \text{versus} \quad H_A: \left| \frac{\theta}{\sigma} \right| < H_m(\beta), \quad (7)$$

where $H_m(\beta) = \sqrt{(2\gamma_m - .5)\beta - 1}$ and $\beta \in (1/(2\gamma_m - .5), 2]$, so that H_A is not empty. To obtain test statistics, the following data transformations are needed:

$$V_i = Y_{iT} - \frac{Y'_{iR} + Y_{iR}}{2}, \quad U_i = Y'_{iR} - Y_{iR}, \quad (8)$$

for $i = 1, \dots, n$. It is clear that V_i and U_i are independent and that

$$V_i \stackrel{\text{iid}}{\sim} N(\theta, \sigma^2), \quad U_i \stackrel{\text{iid}}{\sim} N(0, 2\beta\sigma^2). \quad (9)$$

To make inferences about θ/σ , let $\hat{\theta}$ be the sample mean of the sample $\{V_1, \dots, V_n\}$ and let $\hat{\sigma}^2$ be its sample variance.

Because the V_i 's are iid $N(\theta, \sigma^2)$, we obtain

$$\frac{\hat{\theta}}{\hat{\sigma}/\sqrt{n}} \sim t_{n-1} \left(\frac{\sqrt{n}\theta}{\sigma} \right),$$

where $t_{n-1}(\sqrt{n}\theta/\sigma)$ denotes a random variable that has a noncentral t distribution with $n - 1$ df and a noncentrality parameter, $\sqrt{n}\theta/\sigma$. To make inferences about β , I introduce

$$\hat{\beta} = \frac{(\sum_{i=1}^n U_i^2)/(2n)}{\hat{\sigma}^2}, \quad (10)$$

then $\hat{\beta}/\beta$ follows an F distribution with n and $n - 1$ df. These two statistics, $\hat{\theta}/\hat{\sigma}$ and $\hat{\beta}$, serve as the test statistics. It is clear that any test of (7) based on these two statistics is invariant under the group of scale change. The parameter space now is $\Omega_m = \{(\theta/\sigma, \beta): \beta \in (1/(2\gamma_m - .5), 2]\}$, which is a horizontal strip in \mathcal{R}^2 . Figure 1 gives the parameter space Ω_m , H_0 and H_A of (7), when $\gamma_m = 1.5$.

Which kind of rejection region should the desired test have? The form of hypotheses (7) suggests the following.

Lemma 1. For any fixed value of β , the power function, as a function of θ/σ , of a test with the rejection region

$$\left| \frac{\hat{\theta}}{\hat{\sigma}/\sqrt{n}} \right| < T(\hat{\beta}), \quad (11)$$

where T is any nonnegative function, is unimodal and symmetric about 0.

Therefore, for any given nonnegative function T , one can check whether it defines a level- α test by evaluating its type I error at the boundary of H_0 , $\partial H_0 = \{(\sqrt{(2\gamma_m - .5)\beta - 1}, \beta): \beta \in (1/(2\gamma_m - .5), 2]\}$, which is part of parabolic curve. If the type I errors are all no larger

than α , then the function T defines a level- α test; if the supremum of the type I error is equal to α , then T defines a size- α test. With the help of a modern computer, the numerical evaluation can easily be done. Finally, we provide guidance on how to choose the function T .

Lemma 2. Consider testing the hypotheses (7) for any fixed value of β . There is a uniformly most powerful invariant level- α test based on the V_i 's for (7) with the rejection region

$$\left| \frac{\hat{\theta}}{\hat{\sigma}/\sqrt{n}} \right| < T_0(\beta), \quad (12)$$

where $T_0(\beta)$ is determined by the equation

$$\Pr(|t_{n-1}(\sqrt{n}H_m(\beta))| < T_0(\beta)) = \alpha. \quad (13)$$

This lemma is not realistic, because β is an unknown parameter. However, the function T should be somewhat related to T_0 .

3. TESTS OF INDIVIDUAL BIOEQUIVALENCE

A naive choice of T is $T(\hat{\beta}) = T_0(\hat{\beta})$. The numerical calculations show that the type I error if T is chosen in this way is much larger than the test level, which indicates that $T_0(\hat{\beta})$ overestimates $T_0(\beta)$ systematically. That T_0 is an increasing function is trivial because H_m is increasing. Therefore, $T_0(\beta)$ can be underestimated by underestimating

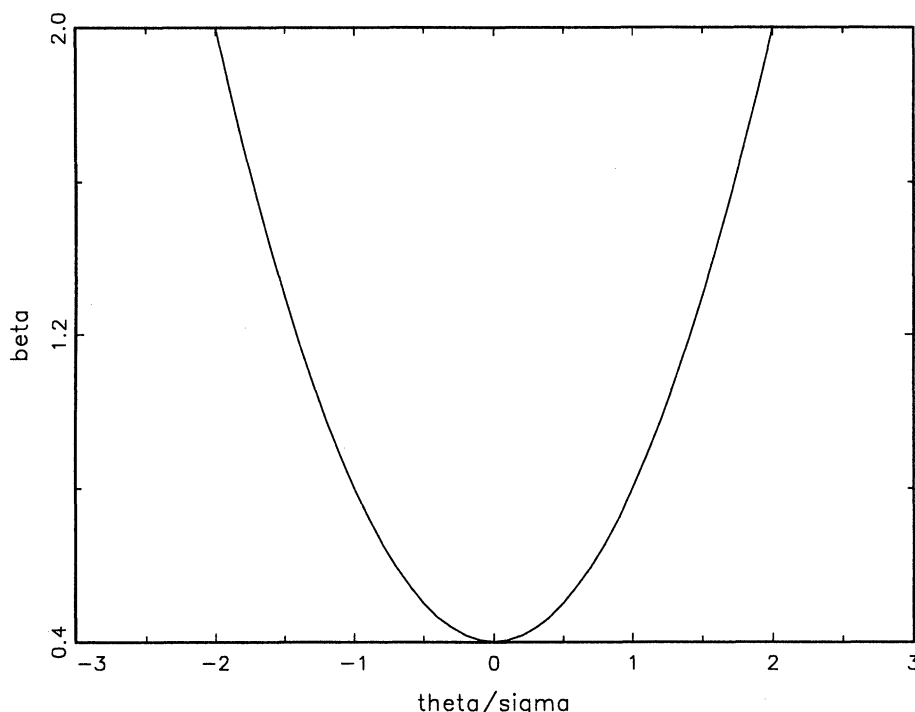


Figure 1. The Hypothesis Spaces of (7) With $\gamma_m = 1.5$. Ω_m is the horizontal strip $.4 < \beta \leq 2$, the solid parabolic curve is ∂H_0 , and H_A is inside the curve.

β . I define a rejection region

$$\left| \frac{\hat{\theta}}{\hat{\sigma}/\sqrt{n}} \right| < T(\hat{\beta}) = \begin{cases} 0 & \text{if } k\hat{\beta} \leq 1/(2\gamma_m - .5) \\ T_0(2) & \text{if } k\hat{\beta} > 2 \\ T_0(k\hat{\beta}) & \text{otherwise,} \end{cases} \quad (14)$$

where k is a positive constant to be determined. The rationale behind the definition is that β is underestimated by $k\hat{\beta}$. Because H_A is empty when $\beta \leq 1/(2\gamma_m - .5)$, H_0 cannot be rejected when $k\hat{\beta}$ is small. Similarly, when β is close to 2, one is basically testing $|\theta/\sigma| \geq H_m(2)$. It is clear that a larger k yields a more powerful test. Therefore, the largest k is chosen so that the test has level α . To do so, partition the interval $[1/(2\gamma_m - .5), 2]$ as $\beta_0 = 1/(2\gamma_m - .5), \dots, \beta_\nu = 2$ with equal increment. Simulate the type I error of the test (14) at each of $\omega_i = (\sqrt{(2\gamma_m - .5)\beta_i - 1}, \beta_i)$ at ∂H_0 for $i = 0, \dots, \nu$ for an initial value of k . Throughout this article, each power or type I error is based on 100,000 simulations coded in Gauss software. If the maximum type I error is less than the test level α , then choose a larger k for the next round of simulations; otherwise, choose a smaller value for k . For example, if $\gamma_m = 1.5, n = 24$ and $\alpha = .05$, choose $\nu = 50$, which makes the points β_i dense enough in the interval $[.4, 2]$. I simulate 51 type I errors for each value of k and keep track of the maximum type I error (i.e., the size of the test). I choose $k = .618$ at last, which gives the maximum type I error .0497 attained at point $(0, .4)$ on the boundary of H_0 . Figure 2 shows the rejection region of this test, the area between two solid curves. To simulate the power or the type I error of (14), I use

$$f(\theta/\sigma, \beta) \stackrel{\text{def}}{=} P\left(\left| \frac{\hat{\theta}}{\hat{\sigma}/\sqrt{n}} \right| < T(\hat{\beta})\right) \quad (15)$$

$$= P\left(\left| \frac{Z + \sqrt{n}\theta/\sigma}{\chi_{n-1}/\sqrt{n-1}} \right| < T\left(\beta \frac{\chi_n^2/n}{\chi_{n-1}^2/(n-1)}\right)\right), \quad (16)$$

where Z is $N(0, 1)$, χ_u^2 follows a chi-squared distribution with u df for $u = n, n-1$, and they are all independent.

It is important that γ_m be chosen neither too small nor too large. If γ_m is no larger than 1 (i.e., the second moment of $Y_{iT} - Y_{iR}$ is controlled to be at most that of $Y'_{iR} - Y_{iR}$), then it is very difficult to have a reasonable power for establishing individual bioequivalence. On the other hand, if γ_m is too large, then some non-individually bioequivalent drug may have the chance to establish bioequivalence. I recommend that γ_m be chosen from the interval $[1.5, 2]$. Table 1 provides the values of the constant k , the size and the maximum power of (14) when $\gamma_m = 1.5$ and the sample size n is 18–38, typical numbers of subjects in bioequivalence studies. Numerical calculations show that the size and the maximum power of (14) are always achieved at points $(\theta/\sigma, \beta) = (0, .4)$ and $(\theta/\sigma, \beta) = (0, 2)$. In terms of the original parameters, $(\theta/\sigma, \beta) = (0, .4)$ if $\mu_T = \mu_R$ and $\sigma_D^2 + \sigma_T^2 = 2\sigma_R^2$, and $(\theta/\sigma, \beta) = (0, 2)$ if $\mu_T = \mu_R$ and $\sigma_D^2 + \sigma_T^2 = 0$.

I now provide a detailed simulation study on the power function and the type I error of (14) when $n = 24$. Here the test level α is chosen to be .05. If $\gamma_m = 1.5$, then $\partial H_0 = \{(\theta/\sigma, \beta): \beta \in (.4, 2], \theta/\sigma = \sqrt{2.5\beta - 1}\}$. Table 2 presents the type I error at ∂H_0 and the power at $\theta/\sigma = 0, .4$ and 1 when β changes from .4 to 2 with 10 equal increments. It can be seen that all type I errors are less than the test level .05. By Lemma 1, the test is known to be a .05-level test if its type I error at ∂H_0 is at most .05. Notice that the type I error decreases when β increases and the type I error is only .0004 when $\beta = 2$, so it is possible to uniformly improve the power of (14) by enlarging its rejection region,

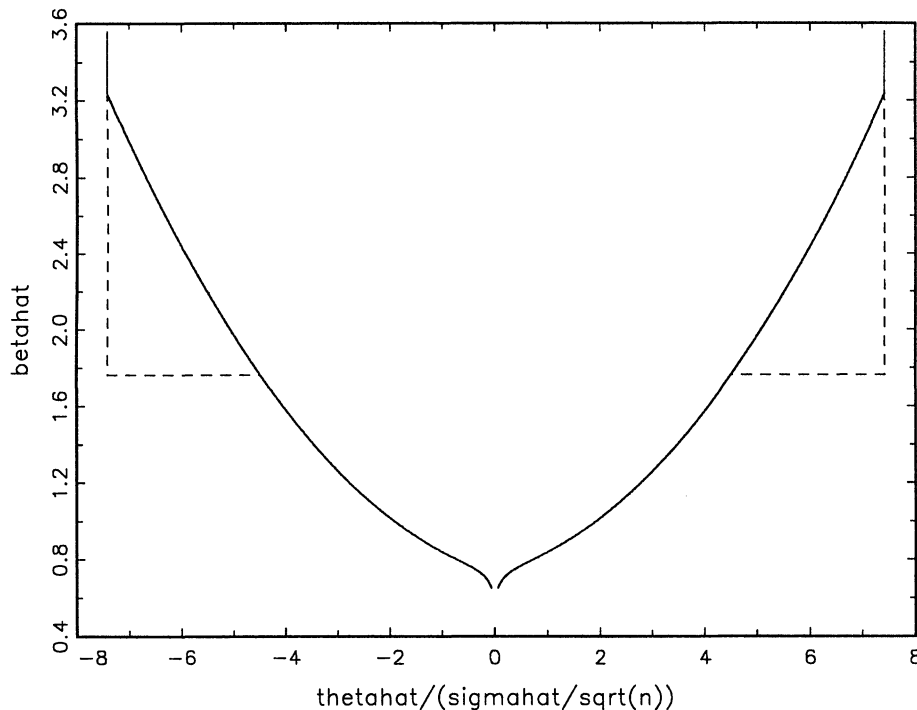


Figure 2. The Rejection Regions of Tests (14) and (17) When $\gamma_m = 1.5, n = 24$, and $\alpha = .05$. The region inside the solid curve is for (14). The union of this region and two dashed regions is for (17).

Table 1. The Constant k , Size, and Maximum Power of the Test (14) When $\gamma_m = 1.5$, the Level $\alpha = .05$, and the Number of Subjects n Varies

n	k	Size	Maximum power
18	.586	.0500	.9487
20	.597	.0499	.9636
22	.607	.0497	.9754
24	.618	.0497	.9840
26	.626	.0497	.9890
28	.633	.0496	.9929
30	.641	.0496	.9952
32	.648	.0499	.9970
34	.655	.0500	.9980
36	.660	.0499	.9986
38	.666	.0500	.9992

Table 2. The Type I Error at the Boundary of H_0 and the Power at $\theta/\sigma = 1, .4$, and 1 for the Test (14) When $\alpha = .05$, $n = 24$, $\gamma_m = 1.5$, and $\beta = .4 + .16i$ for $i = 0, \dots, 10$

i	$(\sigma_D^2 + \sigma_T^2)/\sigma_R^2$	Type I error	$\theta/\sigma = 1$	$\theta/\sigma = .04$	$\theta/\sigma = 0$
0	2	.0499	4e-05	.0151	.0500
1	1.29	.0213	.0008	.0785	.1955
2	.89	.0133	.0052	.1997	.3967
3	.64	.0090	.0196	.3531	.5859
4	.46	.0063	.0494	.5047	.7305
5	.33	.0043	.0987	.6360	.8298
6	.24	.0029	.1647	.7394	.8938
7	.16	.0018	.2444	.8169	.9342
8	.10	.0011	.3293	.8722	.9588
9	.04	.0007	.4154	.9115	.9740
10	0	.0004	.4968	.9382	.9838

while still controlling the size of the test. To obtain a power of .8, β should be at least 1.2 ($i = 5$, in Table 2); that is, $(\sigma_D^2 + \sigma_T^2)/\sigma_R^2 \leq 1/3$, when $\theta = 0$. It is no surprise that small subject-by-formulation interaction and within-subject variability are needed for the test formulation to obtain a large power. In fact, the reference formulation has a power of only .328 to establish individual bioequivalence to itself. For such a case, $(\sigma_D^2 + \sigma_T^2)/\sigma_R^2 = 1$. Table 3 presents a similar numerical study for $\gamma_m = 2$ with $k = .618$. Again, the type I error is controlled to be at most .05. A power larger than .8 can be obtained only when $\beta > .8$; that is, $(\sigma_D^2 + \sigma_T^2)/\sigma_R^2 < .75$.

I close this section with an attempt to improve (14) in power uniformly. Notice that the type I error of (14) is less than the test level, especially when β is large, so the rejection region of (14) may be enlarged at large $\hat{\beta}$. Define

$$\left| \frac{\hat{\theta}}{\hat{\sigma}/\sqrt{n}} \right| < T_1(\hat{\beta}) = \begin{cases} 0 & \text{if } k\hat{\beta} \leq 1/(2\gamma_m - .5) \\ T_0(2) & \text{if } k\hat{\beta} > k_1 \\ T_0(k\hat{\beta}) & \text{otherwise,} \end{cases} \quad (17)$$

where the constant k is the one in (14) and k_1 , less than 2, is a new constant to be determined. It is clear that the new test is uniformly more powerful than (14) if its size is still α , because it has a larger rejection region. When $\gamma_m = 1.5$ and $n = 24$, numerical calculations show that $k_1 = 1.09$. [See Fig. 2 for the rejection region of the new test, which equals the region of (14) plus two dashed areas.] Table 4 presents the simulations of the type I error and power of (17). The test level is still controlled at .05, and the power increments over (14) are substantial when θ/σ is close to 1. In the application, however, one may be cautious about using this test, because the boundary of the rejection region is not continuous at $\hat{\beta} = k_1/k$. It is of interest to obtain a more powerful test with a smooth rejection region.

4. EXAMPLES

Here I apply (14) for individual bioequivalence (5) with $\gamma_m = 1.5$ on two examples. The test level α here is .05.

Example 1 (Continued). From the discussion in the previous section, there is no need to use a 2×4 crossover design. One only needs a 2×3 crossover design in which

the test and reference formulations are administrated on each subject once and twice. For the component C of the MAO inhibitor, C_{\max} is obtained on two sequences: TRRT and RTTR. Therefore, for the purpose of illustration, I ignore the observations in the fourth period of sequence 1 and the third period of sequence 2 and pretend to have a 2×3 crossover design; Table 5 redisplay the data. Note that subjects 14 and 15 are not given. Thus the sample size is $n = 38$. As pointed out by a referee, in practice most studies are conducted as four-period designs. For such designs, a solution is to use the average of two responses of test formulation instead of a single response. Somewhat surprisingly, more parameters and more test statistics are involved, and the hypothesis spaces are regions in R^3 instead of R^2 . This concept is beyond the scope of this article, even though the idea of reparameterization can still be applied, and would be of interest for future research. Assume a three-period design; the analyses on C_{\max} proceeds as follows:

- Step 1. Obtain U_i and V_i on each subject using the transformations (8); for example, $V_3 = 12.247 - (11.574 + 10.988)/2$ and $U_3 = 11.574 - 10.988$. The results are given in Table 5.
- Step 2. Obtain the sample mean $\hat{\theta} = -.607$ and the sample standard deviation $\hat{\sigma} = 1.453$ of V_i 's and the sum of squares of U_i 's, $\sum U_i^2 = 289.231$. Then

$$\frac{\hat{\theta}}{\hat{\sigma}/\sqrt{n}} = -2.573, \quad \hat{\beta} = \frac{(\sum_{i=1}^n U_i^2)/(2n)}{\hat{\sigma}^2} = 1.803.$$

Table 3. The Type I Error at the Boundary of H_0 and the Power at $\theta/\sigma = 0, .4$, and 1 for the Test (14) When $\alpha = .05$, $n = 24$, $\gamma_m = 2$, and $\beta = 1/3.5 + (2 - 1/3.5)i/10$ for $i = 0, \dots, 10$

i	$(\sigma_D^2 + \sigma_T^2)/\sigma_R^2$	Type I error	$\theta/\sigma = 1$	$\theta/\sigma = .04$	$\theta/\sigma = 0$
0	3	.0500	5e-05	.0152	.0499
1	1.69	.0164	.0023	.1333	.2940
2	1.09	.0092	.0196	.3525	.5851
3	.75	.0057	.0731	.5737	.7842
4	.53	.0040	.1686	.7389	.8933
5	.38	.0026	.2938	.8467	.9479
6	.26	.0016	.4266	.9110	.9743
7	.17	.0010	.5521	.9487	.9870
8	.10	.0007	.6583	.9701	.9933
9	.05	.0003	.7440	.9825	.9965
10	0	.0001	.8100	.9896	.9981

Table 4. The Type I Error at the Boundary of H_0 and the Power at $\theta/\sigma = 0, .4$, and 1 for the Test (17) When $\alpha = .05$, $n = 24$, $\gamma_m = 1.5$, and $\beta = .4 + .16i$ for $i = 0, \dots, 10$

i	$(\sigma_D^2 + \sigma_T^2)/\sigma_R^2$	Type I error	$\theta/\sigma = 1$	$\theta/\sigma = .04$	$\theta/\sigma = 0$
0	3	.0499	.0001	.0152	.0500
1	1.69	.0228	.0023	.0783	.1957
2	1.09	.0221	.0142	.2012	.3965
3	.75	.0329	.0454	.3548	.5853
4	.53	.0443	.1024	.5069	.7295
5	.38	.0489	.1814	.6389	.8295
6	.26	.0447	.2763	.7415	.8931
7	.17	.0353	.3753	.8183	.9342
8	.10	.0249	.4700	.8735	.9589
9	.05	.0156	.5586	.9125	.9744
10	0	.0095	.6336	.9392	.9839

- Step 3. Define

$$x = \begin{cases} 0 & \text{if } k\hat{\beta} \leq .4 \\ k\hat{\beta} & \text{if } k\hat{\beta} > 2 \\ 2 & \text{otherwise} \end{cases}$$

where $k = .666$ from Table 1. Hence $x = 1.200$.

- Step 4. If $x = 0$, then fail to reject H_0 ; otherwise, solve $T_0(x)$ from

$$F(T_0(x)) - F(-T_0(x)) = \alpha,$$

where F is the cdf of the noncentral t distribution with $n - 1 = 37$ df and noncentrality parameter $\sqrt{38(2.5x - 1)}$. Here $T_0(x) = 6.677$.

- Step 5. Individual bioequivalence is declared at the .05 level if and only if $|\hat{\theta}/\hat{\sigma}/\sqrt{n}|$ is less than $T_0(x)$. Hence I declare individual bioequivalence based on this dataset.

Example 2. The antihypertensive patch is a generic drug used to treat hypertension. The same 2×4 crossover experiment as in Example 1 is conducted on $n = 37$ subjects, and the dataset is available on the same web site, labeled as drug 17. I analyze the AUC_{0-t} of component A. Again, I ignore the observations in the fourth period of the sequence TRRT and the third period of the sequence RTTR. Following steps 1 and 2,

$$\frac{\hat{\theta}}{\hat{\sigma}/\sqrt{n}} = -.510, \quad \hat{\beta} = \frac{(\sum_{i=1}^n U_i^2)/(2n)}{\hat{\sigma}^2} = .724.$$

Table 5. C_{max} for Test and Reference Formulations

Sequence	Subject	Period			V_i	U_i
		I	II	III or IV		
TRR	3	12.247	10.988	11.574	.966	.586
TRR	6	11.726	14.756	15.277	-3.291	.521
TRR	8	10.491	14.719	10.698	-2.218	-4.021
TRR	9	8.501	9.647	9.574	-1.110	-.073
TRR	10	9.490	10.658	9.833	-.756	-.825
TRR	11	11.246	12.900	12.816	-1.612	-.084
TRR	12	10.653	12.058	12.567	-1.660	.509
TRR	13	12.606	12.542	12.030	.320	-.512
TRR	17	10.977	13.961	10.386	-1.197	-3.575
TRR	18	14.115	14.691	13.009	.265	-1.682
TRR	19	11.857	13.688	11.862	-.918	-1.826
TRR	21	10.613	10.612	10.249	.183	-.363
TRR	24	11.532	11.526	11.700	-.081	.174
TRR	26	12.383	13.904	12.283	-.711	-1.621
TRR	29	10.474	10.662	10.192	.047	-.470
TRR	32	12.385	12.092	8.927	1.876	-3.165
TRR	34	11.776	13.936	12.205	-1.295	-1.731
TRR	37	10.927	12.298	10.922	-.683	-1.376
RTR	1	11.745	11.120	12.923	-1.214	1.178
RTR	2	12.949	11.389	12.054	-1.113	-.895
RTR	4	14.646	13.177	12.365	-.329	-2.281
RTR	5	14.389	12.710	13.548	-1.259	-.841
RTR	7	12.956	12.062	12.436	-.634	-.520
RTR	16	14.859	14.372	12.223	.831	-2.636
RTR	20	10.272	10.280	12.043	-.878	1.771
RTR	22	9.479	10.494	10.671	.419	1.192
RTR	23	11.841	11.823	11.376	.215	-.465
RTR	25	10.899	10.981	10.906	.079	.007
RTR	27	12.671	12.622	13.320	-.374	.649
RTR	28	14.310	12.266	10.227	-.003	-4.083
RTR	30	12.288	10.040	10.669	-1.439	-1.619
RTR	31	11.957	11.439	10.829	.046	-1.128
RTR	33	9.755	11.365	11.087	.944	1.332
RTR	35	10.684	11.526	10.635	.867	-.049
RTR	36	10.228	11.235	11.213	.515	.985
RTR	38	9.331	8.049	10.764	-1.999	1.433
RTR	39	28.235	14.757	14.597	-6.659	-13.638
RTR	40	13.288	14.886	14.869	.8075	1.581

Now $x = k\hat{\beta} = .479$, where $k = .662$. Solving the equation in step 4, $T_0(.479) = 1.055$, and the individual bioequivalence is established.

5. OTHER CONSIDERATIONS

The method of constructing the test (14) for (7) can be easily applied to the following hypotheses:

$$H_0: \left| \frac{\theta}{\sigma} \right| \geq H(\beta) \quad \text{versus} \quad H_A: \left| \frac{\theta}{\sigma} \right| < H(\beta), \quad (18)$$

where H is a given nonnegative nondecreasing function, with the test statistics $\sqrt{n}\hat{\theta}/\hat{\sigma}$ and $\hat{\beta}$. Equation (13) can be solved with H_m replaced by H , then a new T can be defined as in (14) with an appropriate choice of k .

Return to the hypotheses (6). The parameters σ_D^2 and σ_T^2 are not identifiable based on (8). Therefore, we are not able to deal with the cases of $c_1 \neq c_2$. But when $c_1 = c_2 = c$ and $a = 0$, then (6) can be written as

$$H_0: \left| \frac{\theta}{\sigma} \right| \geq H_g(\beta) \quad \text{versus} \quad H_A: \left| \frac{\theta}{\sigma} \right| < H_g(\beta), \quad (19)$$

where $H_g(\beta) = \sqrt{(\gamma_1 + 1.5c)\beta - c}$ for $\beta \in (c/(\gamma_1 + 1.5c), 2]$. Hence a test can be derived as shown in Section 3. The same test is also a valid test for (6) with $a > 0$. However, this test may be conservative when used for $a > 0$. This is because the null hypothesis of (6) for $a > 0$ is included in that for $a = 0$. The hypotheses (6) with $a = 0$ are recommended, because they are invariant under scale changes and also give higher credits to the test formulation with a small variability than the hypotheses with $a > 0$.

There is another application for the hypotheses (18). Schall (1995) proposed an individual bioequivalence based on probability that generalizes (2). He considered the following:

$$H_0: p_{TRI} \leq p_0 \quad \text{versus} \quad H_A: p_{TRI} > p_0, \quad (20)$$

where

$$p_{TRI} = \Pr(|Y_{iT} - Y_{iR}| \leq \gamma_p \sqrt{2}\sigma_R), \quad (21)$$

and $\gamma_p > \Phi^{-1}(p_0/2 + 1/2)/\sqrt{2}$ and p_0 are positive constants. Here Φ is the cdf of $N(0, 1)$. If the null hypothesis is rejected, then individual bioequivalence is declared. For example, if $\gamma_p = 1.96$ and $p_0 = .8$, then on more than 80% subjects, about 95% of the differences between the responses from the test and reference formulations are within 1.96 times $\sqrt{2}\sigma_R$, the standard deviation of the difference between two responses of the reference formulation. All one needs to do is write (20) in the form of (18).

Lemma 3. The hypotheses (20) can be written as

$$H_0: \left| \frac{\theta}{\sigma} \right| \geq H_p(\beta) \quad \text{versus} \quad H_A: \left| \frac{\theta}{\sigma} \right| < H_p(\beta), \quad (22)$$

where H_p is determined by

$$\Phi \left(\frac{\gamma_p \sqrt{2\beta} - H_p(\beta)}{\sqrt{1 + \beta/2}} \right) - \Phi \left(-\frac{\gamma_p \sqrt{2\beta} + H_p(\beta)}{\sqrt{1 + \beta/2}} \right) = p_0. \quad (23)$$

Here Φ is the cdf of $N(0, 1)$ and $\beta \in (\beta_0, 1]$, where

$$\beta_0 = \frac{2[\Phi^{-1}(p_0/2 + 1/2)]^2}{4\gamma_p^2 - [\Phi^{-1}(p_0/2 + 1/2)]^2}.$$

Moreover, H_p is a nondecreasing function.

In summary, a method of constructing tests for a general class of testing problems (18), including individual bioequivalence based on moment or probability, has been proposed. Higher-order crossover designs (more than 2×3) are not necessary for assessing bioequivalence. The simulation studies show that the proposed tests are, at least in a practical sense, exact α -level tests. Uniform improvement in power on the proposed tests is possible and of interest for future research.

APPENDIX: PROOFS

Proof of Lemma 1

Let $\chi_{n-1}^2 = (n-1)\hat{\sigma}^2/\sigma^2$. For given χ_{n-1}^2 , $\hat{\theta}/\hat{\sigma}/\sqrt{n}$ is independent of $\hat{\beta}$, and the conditional distribution of $\hat{\beta}$ depends on β only. Therefore, the conditional probability

$$\Pr \left(\left| \frac{\hat{\theta}}{\hat{\sigma}/\sqrt{n}} \right| < T(\hat{\beta}) | \chi_{n-1}^2 \right) = E \int_{\{|z + \theta/\sigma/\sqrt{n}| < \chi_{n-1} T(\hat{\beta})/\sqrt{n-1}\}} \varphi(z) dz,$$

where φ is the pdf of $N(0, 1)$, is unimodal and symmetric about 0 as a function of θ/σ , as is the unconditional probability

$$\Pr \left(\left| \frac{\hat{\theta}}{\hat{\sigma}/\sqrt{n}} \right| < T(\hat{\beta}) \right).$$

Proof of Lemma 2

It is well known that the noncentral t distribution family is strictly totally positive of order 3 to noncentrality parameter θ/σ (see Lehmann 1986). Then (12) defines a uniformly most powerful test for (7) based on the statistic $\hat{\theta}/\hat{\sigma}$ by problem 30 of Lehmann (1986, p. 120) for a fixed β . Because $\hat{\theta}/\hat{\sigma}$ is maximal invariant with respect to the group of scale change, (12) defines a uniformly most powerful invariant test.

Proof of Lemma 3

Because $Y_{iT} - Y_{iR}$ follows a normal distribution $N(\theta, (1 + \beta/2)\sigma^2)$,

$$p_{TRI} = \Phi \left(\frac{\gamma_p \sqrt{2\beta} - \theta/\sigma}{\sqrt{1 + \beta/2}} \right) - \Phi \left(-\frac{\gamma_p \sqrt{2\beta} + \theta/\sigma}{\sqrt{1 + \beta/2}} \right).$$

It is clear that p_{TRI} is decreasing in $|\theta/\sigma|$ for any fixed β . Thus H_p is determined by (5). If the alternative space H_A is not empty, then $\beta > \beta_0$. To show that H_p is nondecreasing, it only must be shown

that p_{TRI} is increasing in β . Note that p_{TRI} is an integration of a standard normal density function on an interval with a radius $\gamma_p \sqrt{2\beta} / \sqrt{1 + \beta/2}$ and a center at $-\theta / (\sigma \sqrt{1 + \beta/2})$. The radius increases and the center moves toward 0 as β gets large. Therefore, for a fixed θ/σ , the probability p_{TRI} is increasing in β , and the proof is complete.

[Received October 1997. Revised March 1999.]

REFERENCES

- Anderson, S., and Hauck, W. W. (1983), "A New Procedure for Testing Equivalence in Comparative Bioavailability and Other Clinical Trials," *Communications in Statistics, Part A—Theory and Methods*, 12, 2663–2692.
- (1990), "Consideration of Individual Bioequivalence," *Journal of Pharmacokinetics and Biopharmaceutics*, 18, 259–273.
- (1996), "Comment," *Statistical Science*, 11, 303.
- Brown, L. D., Hwang, J. T. G., and Munk, A. (1997), "An Unbiased Test for the Bioequivalence Problem," *The Annals of Statistics*, 26, 2345–2367.
- Chen, M. L. (1997), "Individual Bioequivalence—A Regulatory Update," *Journal of Biopharmaceutical Statistics* 7, 5–11.
- Chow, S. C., and Liu, J. P. (1992), *Design and Analysis of Bioavailability and Bioequivalence Studies*, New York: Marcel Dekker.
- Hwang, J. T. G., and Wang, W. (1997), "The Validity of the Test of Individual Equivalence Ratios," *Biometrika*, 84, 893–900.
- Lehmann, E. T. (1986), *Testing Statistical Hypotheses* (2nd ed.), New York: Wiley.
- Liu, J. P. (1995), "Use of the Repeated Crossover Design in Assessing Bioequivalence," *Statistics in Medicine*, 14, 1067–1078.
- Liu, J. P., and Chow, S. C. (1992), "On Assessment of Bioequivalence in Variability of Bioavailability Studies," *Communications in Statistics, Part A—Theory and Methods*, 21, 2591–2607.
- (1996), "Comment," *Statistical Science*, 11, 306–312.
- Schall, R. (1995), "Assessing of Individual and Population Bioequivalence Using the Probability that Bioavailabilities are Similar," *Biometrics*, 51, 615–626.
- Schall, R., and Luus, H. G. (1993), "On Population and Individual Bioequivalence," *Statistics in Medicine*, 12, 1109–1124.
- Schuirmann, D. J. (1987), "A Comparison of the Two One-Sided Tests Procedure and the Power Approach for Assessing the Equivalence of Average Bioavailability," *Journal of Pharmacokinetics and Biopharmaceutics*, 15, 657–680.
- U.S. Food and Drug Administration (1997), *Guidance for Industry: In Vivo Bioequivalence Studies Based in Population and Individual Bioequivalence Approaches*, Rockville, MD: Author.
- Wang, W. (1995), "On Assessment of Bioequivalence," Ph.D. thesis, Cornell University.
- (1997a), "Optimal Unbiased Tests for Equivalence in Intra-Subject Variability," *Journal of the American Statistical Association*, 92, 1163–1170.
- (1997b), "Assessing Individual Bioequivalence Using 2×2 Crossover Design," technical report, Wright State University, Dept. of Mathematics and Statistics.
- Wang, W., and Hwang, J. T. G. (1997), "A Nearly Unbiased Test for Individual Bioequivalence," technical report, Cornell University, Statistics Center.