Optimal Unbiased Tests for Equivalence in Intrasubject Variability

Weizhen Wang

Journal of the American Statistical Association, Vol. 92, No. 439. (Sep., 1997), pp. 1163-1170.

Stable URL:
http://links.jstor.org/sici?sici=0162-1459%28199709%2992%3A439%3C1163%3AOUTFEI%3E2.0.CO%3B2-C

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

Optimal Unbiased Tests for Equivalence in Intrasubject Variability

Weizhen WANG

The equivalence in average bioavailability between formulations may not be sufficient for assessment of bioequivalence. The difference in intrasubject variability between formulations should also be considered. This article presents an unbiased test procedure for equivalence in intrasubject variability of bioavailability that is uniformly more powerful than the two one-sided tests procedure proposed by Liu and Chow. Under a stronger condition, a uniformly most powerful invariant test for this problem is proposed. Some numerical comparisons and an example are also presented.

KEY WORDS: Cross over design; Linear regression; Two one-sided tests procedure; Uniformly most powerful invariant test.

1. INTRODUCTION

In a traditional testing approach, one tries to establish that two treatments are different or one treatment is more powerful than the other. Assuming normality, the uniformly most powerful unbiased (UMPU) tests exist in many situations. In bioequivalence studies, however, the aim is to establish the similarity of two treatments. Thus the problem becomes two-sided. For example, a new drug can obtain U.S. Food and Drug Administration (FDA) approval by establishing its equivalence to a well-established drug, called the reference drug.

The statistical question is then how to establish the similarity of two treatments. Currently, the FDA (1992) requires evidence only of equivalence in average bioavailability between formulations, formulated by comparing two population means, for assessing bioequivalence. If these means are within a predetermined tolerance limit (e.g., 20% by FDA), then the two populations are considered similar. Many researchers have worked on this problem. Schuirmann (1987) proposed a two one-sided tests procedure that is computationally easy and is recommended in the FDA guidance (1992). This is an exact α-level test (i.e., the supremum of its type I error equals α), but it is biased. Recently Brown, Hwang, and Munk (1997) proposed an unbiased test that is uniformly more powerful than Schuirmann's test. But Anderson and Hauck (1990) and Liu and Chow (1992) argued that not only the averages but also the variabilities of two populations should be compared to ensure that two distributions are similar. This is especially true with the normality assumption, because a normal distribution is uniquely determined by its mean and variance. To address this problem, Liu and Chow (1992) proposed a two one-sided tests procedure that can detect equivalence in variability. This article presents an unbiased test that is more powerful than Liu and Chow's. The technique is based on the result of Brown et al. (1997). Additionally, for a simpler case in which there is no subject effect, a stronger result—a uniformly most powerful invariant (UMPI) test—is provided.

The model is a standard 2×2 crossover design that compares a test drug formulation with a reference drug formulation. It is expressed as follows:

$$\text{Sequence 1} \quad \begin{aligned} Y_{i\mathrm{R}1} &= \mu + S_{i1} + F_{\mathrm{R}} + P_1 + \varepsilon_{i\mathrm{R}1} \\ Y_{i\mathrm{T}1} &= \mu + S_{i1} + F_{\mathrm{T}} + P_2 + C_1 + \varepsilon_{i\mathrm{T}1} \end{aligned}$$

and

$$\text{Sequence 2} \quad \begin{aligned} Y_{i\mathrm{T}2} &= \mu + S_{i2} + F_{\mathrm{T}} + P_1 + \varepsilon_{i\mathrm{T}2} \\ Y_{i\mathrm{R}2} &= \mu + S_{i2} + F_{\mathrm{R}} + P_2 + C_2 + \varepsilon_{i\mathrm{R}2}, \end{aligned} \quad (1)$$

where Y_{ijk} is the response of the ith subject in the kth sequence for the jth formulation, in which $j = \mathrm{R}, \mathrm{T}$, $k = 1, 2$, and $i = 1, 2, \ldots, n_k$; μ is the overall mean; F_j is the fixed effect for the jth formulation with $F_{\mathrm{R}} + F_{\mathrm{T}} = 0$; P_1 and P_2 are the fixed-period effects with $P_1 + P_2 = 0$; C_1 and C_2 are the first-order carry-over effects with $C_1 + C_2 = 0$; S_{ik} is the random subject effect; and ε_{ijk} is the intrasubject random error in observing Y_{ijk}. It is also assumed that S_{ik} and ε_{ijk} are mutually independent, and that $S_{ik} \overset{\text{iid}}{\sim} N(0, \sigma_{\mathrm{S}}^2)$, $\varepsilon_{i\mathrm{T}k} \overset{\text{iid}}{\sim} N(0, \sigma_{\mathrm{T}}^2)$ and $\varepsilon_{i\mathrm{R}k} \overset{\text{iid}}{\sim} N(0, \sigma_{\mathrm{R}}^2)$. The intrasubject variability is represented by σ_j^2.

The difference between F_{T} and F_{R} represents the average bioavailability, and the ratio of σ_{T}^2 and σ_{R}^2 represents the equivalence in intrasubject variability. Let $\lambda = \sigma_{\mathrm{T}}^2/\sigma_{\mathrm{R}}^2$ and $\lambda_1 < \lambda_2$ be two positive numbers. The hypotheses considered are

$$H_0: \lambda \le \lambda_1 \quad \text{or} \quad \lambda \ge \lambda_2 \quad \text{vs.} \quad H_A: \lambda_1 < \lambda < \lambda_2 \quad (2)$$

If we reject H_0, then equivalence in intrasubject variability is confirmed. In practice, we usually select $\lambda_1 < 1 < \lambda_2$ with $\lambda_1 \lambda_2 = 1$, because then $\lambda = 1$ implies that the two populations have the same variability. But these assumptions are not needed in the following theoretical development, in which assessment of equivalence in variability does not need equivalence in mean and vice versa. In other words, the average bioavailability and the variability in bioavailability are two independent issues. Therefore, to ensure the switchability, the suggestion is to simultaneously test the similarity of means and of variances. This article focuses on how to develop tests for (2).

Weizhen Wang is Assistant Professor, Department of Mathematics and Statistics, Wright State University, Dayton, OH 45435. This work was partially supported by National Science Foundation grant DMS-9307499. The author wishes to express heartfelt thanks to his advisor, Professor J. T. Gene Hwang, for showing this interesting problem and for many constructive suggestions. He also thanks Professor Lawrence D. Brown and the referees for helpful comments and Professor Daniel Voss for editing the article.

Brown et al. (1997) derived an α-level unbiased test for

$$H_0: |\theta| \geq \Delta \quad \text{vs.} \quad H_A: |\theta| < \Delta \qquad (\Delta > 0) \tag{3}$$

based on two independent statistics D and S, where $D \sim N(\theta, \sigma^2)$ and $S^2 \sim \sigma^2 \chi^2_n$. Applying to (1), θ is the difference between F_{T} and F_{R}. When there is no carry-over effect (i.e., when the C's are all 0s and W_{ik} is defined as $(Y_{i\mathrm{T}k} - Y_{i\mathrm{R}k})/2$), then D is equal to $\sum_k \bar{W}_{\cdot k}$ and S^2 is a known multiple of $\sum_i \sum_k (W_{ik} - \bar{W}_{\cdot k})^2$. Brown et al. (1997) were also able to prove that their test has a larger rejection region than Schuirmann's test. These are used to derive an α-level unbiased test for (2).

In this article, three cases are considered:

a. When there is no subject effect, no period effect and no carry-over effect (i.e., $S_{ik} = 0$, $P_{\cdot} = 0$, $C_{\cdot} = 0$), or equivalently, when a two-group parallel design is used;
b. when there is a subject effect but no period effect and no carry-over effect (i.e. $S_{ik} \neq 0$, $P_{\cdot} = 0$, $C_{\cdot} = 0$);
c. when the full model (1) is assumed.

Each of the cases a and b is a special case of the case below it; case c is a general case that includes virtually all practical situations. Section 2 gives a UMPI test for case a. Section 3 discusses case b, because the subject effect tends to be more significant and the period effect and the carry-over effect tend to be negligible. An unbiased test for (2) is provided. Section 4 considers case c. The same technique as in Section 3 is used to obtain another unbiased test for (2). However, the former test for case b is more powerful when case b is true. Section 5 discusses the power of the proposed tests in Sections 2 and 3. Some numerical results describe what settings, in terms of the level of test, sample size, and alternative hypothesis (λ_1, λ_2), would provide reasonable power. Section 6 gives an example to demonstrate how to use the proposed unbiased tests, and Section 7 presents conclusions.

2. UNIFORMLY MOST POWERFUL INVARIANT TEST

This section focuses on case a in which no subject effect exists. A UMPI test is derived. To discuss an invariant test, first the group needs to be specified. The group consists of all linear transformations with a positive scale. More precisely, define

$$\mathbf{Y}_{\mathrm{R}} = (Y_{1\mathrm{R}1}, \ldots, Y_{n_1\mathrm{R}1}, Y_{1\mathrm{R}2}, \ldots, Y_{n_2\mathrm{R}2})'$$

and

$$\mathbf{Y}_{\mathrm{T}} = (Y_{1\mathrm{T}1}, \ldots, Y_{n_1\mathrm{T}1}, Y_{1\mathrm{T}2}, \ldots, Y_{n_2\mathrm{T}2})'.$$

Let G consist of transformations g on $R^{2(n_1+n_2)}$ such that

$$g(\mathbf{Y}_{\mathrm{R}}, \mathbf{Y}_{\mathrm{T}}) = (c\mathbf{Y}_{\mathrm{R}} + \theta_1, c\mathbf{Y}_{\mathrm{T}} + \theta_2)$$

for some constants $c > 0$, θ_1, $\theta_2 \in R^1$. It can be seen that G forms a group under the composition of transformations and that (2) is invariant under the group G.

The following notations are also necessary throughout this article: For any pair of vectors $\mathbf{X} = (X_1, \ldots, X_n)'$ and $\mathbf{Y} = (Y_1, \ldots, Y_n)'$, define $S_{\mathbf{XX}} = \sum_{i=1}^n (X_i - \bar{X})^2$ and $S_{\mathbf{XY}} = \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})$.

For case a, $\mathbf{Y}_{\mathrm{R}}$ and $\mathbf{Y}_{\mathrm{T}}$ are independent, because all S_{ik}'s vanish. The full model (1) reduces to

$$Y_{ijk} = \mu + F_j + \varepsilon_{ijk}. \tag{4}$$

Now σ_j^2 also represents the between-subject variability because of no subject effect. Then $S_{\mathbf{Y_TY_T}}/\sigma_{\mathrm{T}}^2$ and $S_{\mathbf{Y_RY_R}}/\sigma_{\mathrm{R}}^2$ are independent, and both have the $\chi^2_{n_1+n_2-1}$ distribution. Let $\eta = \log\lambda$ and $\eta_i = \log\lambda_i$, for $i = 1, 2$, and write (2) as

$$H_0: \eta \leq \eta_1 \quad \text{or} \quad \eta \geq \eta_2 \quad \text{vs.} \quad H_A: \eta_1 < \eta < \eta_2. \tag{5}$$

Let Z be $\log(S_{\mathbf{Y_TY_T}}/S_{\mathbf{Y_RY_R}})$. The distribution of Z involves only η, the new parameter of interest, and has nothing to do with any nuisance parameters, such as μ or F_{T}. The testing problem (2) thus reduces to a one-parameter problem. The statistic Z (or, equivalently, $S_{\mathbf{Y_TY_T}}/S_{\mathbf{Y_RY_R}}$) can be used to develop a uniformly most powerful (UMP) test among the tests based on Z for (5). First, note that the distribution family of Z is strictly totally positive of order 3 (STP_3; Lehmann 1986).

Lemma 2.1. The family of pdf of Z is STP_3 with respect to η.

Theorem 2.1. Under the model (4), among the tests based on the observed data $(\mathbf{Y}_{\mathrm{R}}, \mathbf{Y}_{\mathrm{T}})$ through $S_{\mathbf{Y_TY_T}}/S_{\mathbf{Y_RY_R}}$, there exists a UMP test Φ_{I} for (4) given by

$$\Phi_{\mathrm{I}}(Z) = \begin{cases} 1 & c_1 < Z = \log(S_{\mathbf{Y_TY_T}}/S_{\mathbf{Y_RY_R}}) < c_2 \\ 0 & \text{otherwise,} \end{cases}$$

where c_1 and c_2 are determined by $E_{\eta_i}\Phi_{\mathrm{I}}(Z) = \alpha$ for $i = 1, 2$. Moreover, $\Phi_{\mathrm{I}}(Z)$ is a UMPI test based on the observed data $(\mathbf{Y}_{\mathrm{R}}, \mathbf{Y}_{\mathrm{T}})$ with respect to the group G.

Remark 2.1. One can use a UMPU test to check the existence of a subject effect. The sample correlation coefficient of $\mathbf{Y}_{\mathrm{T}}$ and $\mathbf{Y}_{\mathrm{R}}$ is used to test the independence of $\mathbf{Y}_{\mathrm{T}}$ and $\mathbf{Y}_{\mathrm{R}}$, which implies no subject effect (See Lehmann 1986, p. 249).

Remark 2.2. The test Φ_{I} can also be applied to the noncrossover setting; that is, a randomized trial with parallel treatment groups. Because the different sequences, indexed by k, have no effect on the model (4), this setting is equivalent to a two-group parallel design.

When the period effect and the carry-over effect, but no subject effect, are present, the test in Theorem 2.1 may be adapted by using the test statistic $(S_{\mathbf{Y_{T1}Y_{T1}}} + S_{\mathbf{Y_{T2}Y_{T2}}})/(S_{\mathbf{Y_{R1}Y_{R1}}} + S_{\mathbf{Y_{R2}Y_{R2}}})$, where $\mathbf{Y}_{\mathrm{R}k} = (Y_{1\mathrm{R}k}, \ldots, Y_{n_k\mathrm{R}k})'$ and $\mathbf{Y}_{\mathrm{T}k} = (Y_{1\mathrm{T}k}, \ldots, Y_{n_k\mathrm{T}k})'$ for $k = 1, 2$. For different j and k, $\mathbf{Y}_{jk}$ has a different mean vector. If each Y_{ik} is adjusted by its own sample mean, then $(S_{\mathbf{Y_{T1}Y_{T1}}} + S_{\mathbf{Y_{T2}Y_{T2}}})/\sigma_{\mathrm{T}}^2$ and $(S_{\mathbf{Y_{R1}Y_{R1}}} + S_{\mathbf{Y_{R2}Y_{R2}}})/\sigma_{\mathrm{R}}^2$ are independent and both have the $\chi^2_{n_1+n_2-2}$ distribution. Omitting a similar proof, the following applies.

Proposition 2.1. Suppose that there is no subject effect in (1); that is, $S_{ik} = 0$, for all i, k. There exists a UMPI test

for (5) given by

$$\Phi_{\mathrm{I}1}(Z) = \begin{cases} 1 & c_1 < Z = \log((S_{\mathbf{Y}_{\mathrm{T}1}\mathbf{Y}_{\mathrm{T}1}} + S_{\mathbf{Y}_{\mathrm{T}2}\mathbf{Y}_{\mathrm{T}2}}) \\ & \div (S_{\mathbf{Y}_{\mathrm{R}1}\mathbf{Y}_{\mathrm{R}1}} + S_{\mathbf{Y}_{\mathrm{R}2}\mathbf{Y}_{\mathrm{R}2}})) < c_2 \\ 0 & \text{otherwise,} \end{cases}$$

where c_1 and c_2 are determined by $E_{\eta_i}\Phi_{\mathrm{I}1}(Z) = \alpha$ for $i = 1, 2$.

3. UNBIASED TEST

In Section 2 all results were based on the assumption of no subject effect. When the subject effect cannot be ignored, the problem is more difficult to resolve due to dependence of $\mathbf{Y}_{\mathrm{T}}$ and $\mathbf{Y}_{\mathrm{R}}$. It can no longer be reduced to a one-parameter problem, because σ_{S}^2, the variance of S_{ik}, is involved. However, it is possible to formulate it to be a model with only two parameters. One happens to be an increasing function of λ, and the other is a functionally independent nuisance parameter. In this section assume that

$$Y_{ijk} = \mu + S_{ik} + F_j + \varepsilon_{ijk}. \tag{6}$$

To test (2), transform and reparameterize the problem so that it becomes the testing problem of (3), where θ is the slope in a regression line.

To transform the data, for any $\delta \neq -1$, define

$$\mathbf{Z}_{ik}(\delta) = \begin{pmatrix} v_{ik} \\ u_{ik}(\delta) \end{pmatrix} \overset{\text{def}}{=} \begin{pmatrix} Y_{i\mathrm{T}k} - Y_{i\mathrm{R}k} \\ Y_{i\mathrm{T}k} + \delta Y_{i\mathrm{R}k} \end{pmatrix}. \tag{7}$$

Then $Z_{ik}(\delta) \overset{\text{iid}}{\sim} N_2(\boldsymbol{\theta}_Z(\delta), \boldsymbol{\Sigma}_Z(\delta))$, where

$$\boldsymbol{\theta}_Z(\delta) = \begin{pmatrix} F_{\mathrm{T}} - F_{\mathrm{R}} \\ (1+\delta)\mu + F_{\mathrm{R}}(\delta - 1) \end{pmatrix}$$

and

$$\boldsymbol{\Sigma}_Z(\delta) = \begin{pmatrix} \sigma_{\mathrm{R}}^2 + \sigma_{\mathrm{T}}^2 & \sigma_{\mathrm{T}}^2 - \delta\sigma_{\mathrm{R}}^2 \\ \sigma_{\mathrm{T}}^2 - \delta\sigma_{\mathrm{R}}^2 & (1+\delta)^2\sigma_{\mathrm{S}}^2 + \sigma_{\mathrm{T}}^2 + \delta^2\sigma_{\mathrm{R}}^2 \end{pmatrix}.$$

Different δ's correspond to different transformations. As is shown later, an unbiased test for (2) will be derived for each transformation. Although it appears that there are a great number of unbiased tests, through a complicated algebraic calculation it can be shown that these tests are identical to each other. Therefore, there is indeed only one test. One version, with a specified δ_0 that simplifies the problem statement, is developed later in this section.

To regress $u_{ik}(\delta)$ against v_{ik}, write

$$\begin{aligned} u_{ik}(\delta) &= [\mathrm{cov}(u_{ik}(\delta), v_{ik})/\mathrm{var}(v_{ik})]v_{ik} \\ &\quad + [u_{ik}(\delta) - (\mathrm{cov}(u_{ik}(\delta), v_{ik})/\mathrm{var}(v_{ik}))v_{ik}] \\ &= [(\lambda - \delta)/(\lambda + 1)]v_{ik} + \tilde{e}_{ik}(\delta), \end{aligned}$$

where $\tilde{e}_{ik}(\delta) = u_{ik}(\delta) - [(\sigma_{\mathrm{T}}^2 - \delta\sigma_{\mathrm{R}}^2)/(\sigma_{\mathrm{T}}^2 + \sigma_{\mathrm{R}}^2)]v_{ik}$. Let $e_{ik}(\delta) = \tilde{e}_{ik}(\delta) - E\tilde{e}_{ik}(\delta)$. Then

$$u_{ik}(\delta) = \alpha(\lambda, \delta) + \beta(\lambda, \delta)v_{ik} + e_{ik}(\delta), \tag{8}$$

where $\alpha(\lambda, \delta) = E\tilde{e}_{ik}(\delta) = (1+\delta)[\mu + F_{\mathrm{R}}(\lambda - 1)/(\lambda + 1)]$ and

$$\beta(\lambda, \delta) = (\lambda - \delta)/(\lambda + 1). \tag{9}$$

These $\tilde{e}_{ik}$'s are iid and normally distributed, because the Z_{ik}'s are iid and normally distributed and the transformation is linear. Their common mean and variance are $\alpha(\lambda, \delta)$ and $(1+\delta)^2[\sigma_{\mathrm{S}}^2 + \sigma_{\mathrm{R}}^2\lambda/(1+\lambda)]$, respectively. Most important, v_{ik} and $e_{ik}(\delta)$ are independent, because their covariance is equal to 0. If the simple linear regression is modeled with $u_{ik}(\delta)$ as regressor and v_{ik} as predictor, then the only difference from standard linear regression is that the predictor v_{ik} is a random variable. However, due to independence of v_{ik} and e_{ik}, one may condition on v_{ik} first and reduce it to a standard case. I summarize the discussion as follows.

Lemma 3.1. Let $\sigma^2(\delta) = (1+\delta)^2[\sigma_{\mathrm{S}}^2 + \sigma_{\mathrm{R}}^2\lambda/(1+\lambda)]$. Then (8) holds, $e_{ik}(\delta) \overset{\text{iid}}{\sim} N(0, \sigma^2(\delta))$, $v_{ik} \overset{\text{iid}}{\sim} N(F_{\mathrm{T}} - F_{\mathrm{R}}, \sigma_{\mathrm{T}}^2 + \sigma_{\mathrm{R}}^2)$, and $e_{ik}(\delta)$ and v_{ik} are independent.

The following lemma is a classical result, so the proof is omitted.

Lemma 3.2. Let $\mathbf{V} = (v_{11}, \ldots, v_{n_1 1}, v_{12}, \ldots, v_{n_2 2})'$, let $\hat{\alpha}(\delta)$ and $\hat{\beta}(\delta)$ be the least squares estimators for $\alpha(\lambda, \delta)$ and $\beta(\lambda, \delta)$, and let $S^2(\delta) = \sum_{ik} \|u_{ik}(\delta) - \hat{\alpha}(\delta) - \hat{\beta}(\delta)v_{ik}\|^2/S_{\mathbf{VV}}$. Then for given $S_{\mathbf{VV}}$, $\hat{\beta}(\delta)$ and $S^2(\delta)$ are independent and $\hat{\beta}(\delta) \sim N(\beta(\lambda, \delta), \sigma^2(\delta)/S_{\mathbf{VV}})$, and $S^2(\delta) \sim (\sigma^2(\delta)/S_{\mathbf{VV}})\chi^2_{n_1+n_2-2}$. Moreover, $S_{\mathbf{VV}} \sim (1+\lambda)\sigma_{\mathrm{R}}^2\chi^2_{n_1+n_2-1}$.

Return to the choice of δ. From (9), one can see that $\beta(\lambda, \delta)$, the new parameter of interest, is strictly increasing in λ for any fixed $\delta > -1$ and that (2) is equivalent to

$$\begin{aligned} &H_0\colon \beta(\lambda, \delta) \notin (\beta(\lambda_1, \delta), \beta(\lambda_2, \delta)) \\ &\qquad \text{vs.} \quad H_A\colon \beta(\lambda, \delta) \in (\beta(\lambda_1, \delta), \beta(\lambda_2, \delta)). \end{aligned} \tag{10}$$

This relationship reduces the problem of testing (2) to testing (10), to which the result of Brown et al. (1997) can be readily applied due to Lemma 3.2. To obtain an interval centered at 0, $\delta_0 = (2\lambda_1\lambda_2 + \lambda_1 + \lambda_2)/(\lambda_1 + \lambda_2 + 2)(> 0)$ should be chosen. In the case of $\lambda_1\lambda_2 = 1$, $\delta_0 = 1$. Let $\beta = \beta(\lambda, \delta_0)$. Then (2) turns out to be

$$H_0\colon |\beta| \geq \Delta \quad \text{vs.} \quad H_A\colon |\beta| < \Delta, \tag{11}$$

where $\Delta = (\lambda_2 - \lambda_1)/(\lambda_1 + \lambda_2 + 2) > 0$. Now an unbiased test for (11) can be derived when working on this specified δ_0. Let $D_0 = \hat{\beta}(\delta_0)$ and $S_0 = S(\delta_0)$ and use them as the test statistics. Combined with $S_{\mathbf{VV}}$, three statistics emerge, and $S_{\mathbf{VV}}$ is hidden in the conditional variance of D_0. These satisfy all of the conditions in Lemma 3.2 in which $\delta = \delta_0$. For any given $S_{\mathbf{VV}}$, the problem becomes the average bioavailability problem (3). If $\mathbf{U}(\delta) = (u_{11}(\delta), \ldots, u_{n_1 1}(\delta)$ and $u_{12}(\delta), \ldots, u_{n_2 2}(\delta))'$, then

$$D_0 = S_{\mathbf{U}(\delta_0)\mathbf{V}}/S_{\mathbf{VV}}$$

and

$$S_0 = (S_{\mathbf{U}(\delta_0)\mathbf{U}(\delta_0)}/S_{\mathbf{VV}} - S^2_{\mathbf{U}(\delta_0)\mathbf{V}}/S^2_{\mathbf{VV}})^{1/2}. \tag{12}$$

The following lemma of Brown et al. (1997) deals with the hypotheses (3).

Lemma 3.3. Let $D \sim N(\theta, \sigma^2)$, $S^2 \sim \sigma^2\chi^2_n$, where θ and σ are unknown parameters and D and S are independent. Then for any $\alpha \in (\alpha^*, 1/2)$, there is an α-level unbiased test $\Phi_{\text{B-H-M}}$ for (3) with $\Delta = (\lambda_2 - \lambda_1)/(\lambda_1 + \lambda_2 + 2)$, so that

$$\Phi_{\text{B-H-M}}(D, S) = \begin{cases} 1 & \text{if } |D| < T(S) \\ 0 & \text{otherwise} \end{cases} \tag{13}$$

for some positive function T, where

$$\alpha^* = \int_{3\pi/4}^{\pi} \sin^{n-1}(x)\,dx \Big/ \int_0^{\pi} \sin^{n-1}(x)\,dx. \tag{14}$$

Moreover, $\Phi_{\text{B-H-M}}$ is uniformly more powerful than Schuirmann's test Φ_{S}, where

$$\Phi_{\text{S}}(D, S) = \begin{cases} 1 & \text{if } |D| < \Delta - t_{\alpha,n}S/n^{1/2} \\ 0 & \text{otherwise} \end{cases} \tag{15}$$

and $t_{\alpha,n}$ is the upper α quantile of the Student's t distribution with n df.

As mentioned by Brown et al. (1997), α^* is always smaller than .05 unless n is less than 5. Thus the constraint does not cause a problem. The next theorem is the basic result for Section 3. It defines an unbiased test for (2).

Theorem 3.1. For any $\alpha \in (\alpha^*, 1/2)$ and $n = n_1 + n_2 - 2$, $\Phi_{\text{U}} = \Phi_{\text{B-H-M}}(D_0, S_0)$ defines an unbiased test for (11), then for (2). Let $\Phi_{\text{L-C}}$ be the two one-sided tests procedure for (2) proposed by Liu and Chow (1992); that is, $\Phi_{\text{L-C}} = 1$ if

$$t_1 = r_1[n/(1 - r_1^2)]^{1/2} > t_{\alpha,n}$$

and

$$t_2 = r_2[n/(1 - r_2^2)]^{1/2} < -t_{\alpha,n}, \tag{16}$$

where r_l is the sample Pearson correlation coefficient between v_{ik} and $u_{ik}(\lambda_l)$ for $l = 1, 2$, and $\Phi_{\text{L-C}} = 0$ otherwise. Then Φ_{U} is uniformly more powerful than $\Phi_{\text{L-C}}$.

Remark 3.1. In the proof of Theorem 3.1 in the Appendix, (16) is equivalent to

$$|D_0| < \Delta - t_{\alpha,n}S_0/n^{1/2}, \tag{17}$$

which has a similar mathematical form to Schuirmann's test in (15). The rejection region of Liu and Chow's test seems to depend on δ_0 by (17); however, (16) indicates that this is not so. This may suggest that the unbiased test does not depend on δ_0 either.

Actually, Theorem 3.1 is just one application of a general strategy: Any valid test based on D and S for (3) can provide a test for (2) replacing D and S by D_0 and S_0. Lemma 3.2 connects (3) to (2). Through this connection, the problems of average bioavailability and bioavailability in variability become one problem.

The tests based on D_0 and S_0 are all invariant with respect to the group G. Consequently, the UMPI test Φ_{I}, as defined in Theorem 2.1, is uniformly more powerful than Φ_{U} when there is no subject effect. However, when there is a subject effect, Φ_I is inappropriate and may have size greater than α.

4. GENERALIZATION

Up to this point, the results considered do not assume existence of the period effect and the carry-over effect. But the example in Section 6 does consider such effects. Here I generalize the results to the full model (1), including these effects. In short, the only modification in the testing procedure is the reduced degrees of freedom. The results here parallel those in Section 3, and the proofs are omitted.

The same data transformation, (7), is used. But the $Z_{ik}(\delta)$'s are no longer iid. They have the same covariance matrix as before but a different mean vector for different sequences (indexed by k). Fortunately, $u_{ik}(\delta)$ can still be regressed on v_{ik} within a sequence (i.e., for each k). The following equality holds:

$$u_{ik}(\delta) = \alpha_k(\lambda, \delta) + \beta(\lambda, \delta)v_{ik} + e_{ik}(\delta), \tag{18}$$

where $\alpha_k(\lambda, \delta) = (1 + \delta)[\mu + (F_{\text{R}} + P_k)(\lambda - 1)/(\lambda + 1) + C_k/(\lambda + 1)]$, for $k = 1, 2$, and the other definitions remain the same. This gives two regression lines with different intercepts but the same slope. Independence of v_{ik} and $e_{ik}(\delta)$ is still true. This leads to the following

Lemma 4.1. Equation (18) holds, and $e_{ik}(\delta) \overset{\text{iid}}{\sim} N(0, \sigma^2(\delta))$ is independent of v_{ik}.

The means for different sequences are different. After adjustment by its own sample mean, each corresponding statistic has one less degree of freedom.

Lemma 4.2. Let $\mathbf{V}_k = (v_{1k}, \ldots, v_{n_k k})'$, for $k = 1, 2$, let $\hat{\alpha}_k(\delta)$ and $\hat{\beta}_1(\delta)$ be the least squares estimators for $\alpha_k(\lambda, \delta)$ and $\beta(\lambda, \delta)$, and let $S_1^2(\delta) = \sum_{ik} \|u_{ik}(\delta) - \hat{\alpha}_k(\delta) - \hat{\beta}_1(\delta)v_{ik}\|^2 / \sum_k S_{\mathbf{V}_k\mathbf{V}_k}$. Then for given $\sum_k S_{\mathbf{V}_k\mathbf{V}_k}$, $\hat{\beta}_1(\delta)$ and $S_1^2(\delta)$ are independent and $\hat{\beta}_1(\delta) \sim N(\beta(\lambda, \delta), \sigma^2(\delta)/\sum_k S_{\mathbf{V}_k\mathbf{V}_k})$, and $S_1^2(\delta) \sim (\sigma^2(\delta)/\sum_k S_{\mathbf{V}_k\mathbf{V}_k})\chi^2_{n_1+n_2-3}$. Moreover, $\sum_k S_{\mathbf{V}_k\mathbf{V}_k} \sim (1 + \lambda)\sigma_{\text{R}}^2\chi^2_{n_1+n_2-2}$.

Theorem 4.1. Define $\mathbf{U}_k(\delta) = (u_{1k}(\delta), \ldots, u_{n_k k}(\delta))'$ for $k = 1, 2$, $D_{01} = \hat{\beta}_1(\delta_0)$, and $S_{01} = S_1(\delta_0)$. Let $\Phi_{\text{B-H-M},1}(D, S)$ be the α-level Brown et al.'s test for $n = n_1 + n_2 - 3$, where $\alpha \in (\alpha^*, 1/2)$ and α^* is defined in (14). Then $\Phi_{\text{U1}} = \Phi_{\text{B-H-M},1}(D_{01}, S_{01})$ defines an unbiased test for (11) under the model (1). Moreover, there is a correspondence to Liu and Chow's test for (11), denoted by $\Phi_{\text{L-C},1}$, and $\Phi_{\text{L-C},1} = 1$ if

$$t_{11} = r_{11}[n/(1 - r_{11}^2)]^{1/2} > t_{\alpha,n}$$

and

$$t_{21} = r_2[n/(1 - r_{21}^2)]^{1/2} < -t_{\alpha,n}, \tag{19}$$

where $r_{l1} = \sum_k S_{\mathbf{U}_k(\lambda_l)\mathbf{V}_k} / (\sum_k S_{\mathbf{U}_k(\lambda_l)\mathbf{U}_k(\lambda_l)} \sum_k S_{\mathbf{V}_k\mathbf{V}_k})^{1/2}$, for $l = 1, 2$, and $\Phi_{\text{L-C},1} = 0$ otherwise. Then Φ_{U1} is uniformly more powerful than $\Phi_{\text{L-C},1}$.

Remark 4.1. Equation (19) is equivalent to

$$|D_{01}| < \Delta - t_{\alpha,n} S_{01}/n^{1/2} \tag{20}$$

and

$$D_{01} = \frac{\sum_k S_{\mathbf{U}_k(\delta_0)\mathbf{V}_k}}{\sum_k S_{\mathbf{V}_k\mathbf{V}_k}}$$

and

$$S_{01} = \left[\frac{\sum_k S_{\mathbf{U}_k(\delta_0)\mathbf{U}_k(\delta_0)}}{\sum_k S_{\mathbf{V}_k\mathbf{V}_k}} - \frac{(\sum_k S_{\mathbf{U}_k(\delta_0)\mathbf{V}_k})^2}{(\sum_k S_{\mathbf{V}_k\mathbf{V}_k})^2}\right]^{1/2}. \tag{21}$$

5. NUMERICAL STUDIES OF POWER

This section presents power comparisons between the UMPI test Φ_{I} and the unbiased test Φ_{U}, and between the unbiased test Φ_{U} and Liu and Chow's test $\Phi_{\mathrm{L-C}}$ of Sections 2 and 3 based on numerical integrations using the Gauss program. It focuses on cases a and b, because similar results are expected for case c. The results show that the improvement of Φ_{U} over $\Phi_{\mathrm{L-C}}$ is noticeable; compared to the bioequivalence test of means, a larger sample size is needed to achieve a reasonable power. For simplicity, assume that $\lambda_1\lambda_2 = 1$ in this section. It implies that $\delta_0 = 1$.

In the following three figures, the alternative hypothesis given in (2) is (.5, 2), the sample size $n_1 + n_2$ is 41, and the test level is .1.

Figure 1 plots the powers of Φ_{I} and Φ_{U} against λ under the reduced model (4) for case a. It is expected that Φ_{I} is uniformly more powerful than Φ_{U} by Theorem 2.1. The maximal improvement of power can be 8% (from .55 to .63), which occurs at $\lambda = 1$. Because $\lambda_1\lambda_2 = 1$, $c_2 = -c_1$, and c_2 is solved by $P_{\log(\lambda_1)}(|Z| < c_2) = .1$ from Theorem 2.1. The UMPI test Φ_{I} has a rejection region between two straight lines $S_{\mathbf{Y_TY_T}} = \exp(c_i) S_{\mathbf{Y_RY_R}}$ which are symmetric about $S_{\mathbf{Y_TY_T}} = S_{\mathbf{Y_RY_R}}$ in the $(S_{\mathbf{Y_RY_R}}, S_{\mathbf{Y_TY_T}})$ plane. Therefore, whenever there is evidence of no subject effect, the UMPI test is preferred due to its power and simplicity.

When there is a subject effect (i.e., case b), the only valid tests are our unbiased test Φ_{U} and Liu and Chow's test $\Phi_{\mathrm{L-C}}$, defined in Theorem 3.1. Figure 2 plots their rejection regions in terms of D_0 and S_0. It is clear that Φ_{U} has a larger rejection region than $\Phi_{\mathrm{L-C}}$, so it is uniformly more powerful. (See Brown et al. (1997) for how to determine the critical region of Φ_{U}.) Also, $T(S_0)$, defined in Theorem 3.1, is always positive for any S_0, and it decreases when S_0 is small and increases when S_0 is large. One issue raised is whether T defines a sensibly shaped rejection region, because it should be harder to reject the null hypothesis if a larger S_0 is observed. Brown et al. (1997) suggested that one adjust T to be a decreasing function or even truncate T at the point when it starts to increase, if necessary.

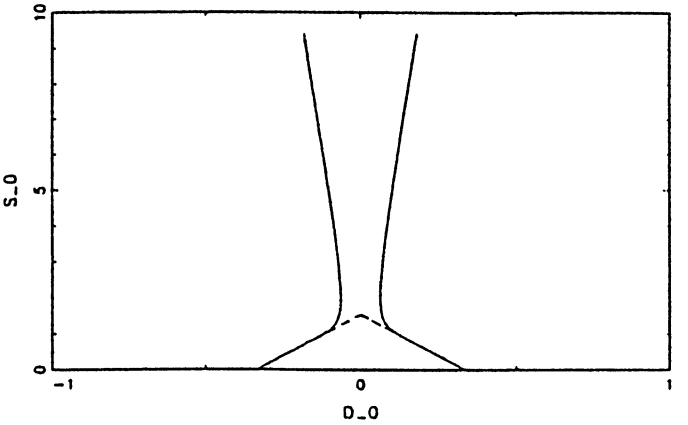

Figure 2. Rejection Regions of Φ_U and Φ_{L-C}. The area between the solid lines is the rejection region for Φ_I; the dashed triangle is the rejection region for Φ_{L-C}.

Figure 3 plots the powers of Φ_{U} and $\Phi_{\mathrm{L-C}}$ at $\lambda = 1$ against $\sigma_0 = \sigma_S^2/\sigma_{\mathrm{R}}^2$, the ratio of variances of S_{ik} and $\varepsilon_{i\mathrm{R}k}$ in (6). For fixed λ, the power is a function of σ_0. Note that the power of $\Phi_{\mathrm{L-C}}$ drops to 0 quickly when σ_0 becomes large, whereas the power of Φ_{U} is always above .1 due to its unbiasedness. The power of $\Phi_{\mathrm{L-C}}$ at $\sigma_0 = .625$ falls below the test level .1, whereas Φ_{U} has power .2439, which is 15% larger. When $\sigma_0 \approx 2$, which happens in the example of the next section, $\Phi_{\mathrm{L-C}}$ has power less than .001, whereas the power of Φ_{U} is approximately .144.

One may notice that the maximal power of the unbiased test from Figure 3 is only .55, which is attained at $\lambda = 1$ and $\sigma_{\mathrm{S}}^2 = 0$ and is much less than 1. When $\sigma_{\mathrm{S}}^2 = 0$ (i.e., case a), Φ_I has a larger power than Φ_{U}. But, its maximal power

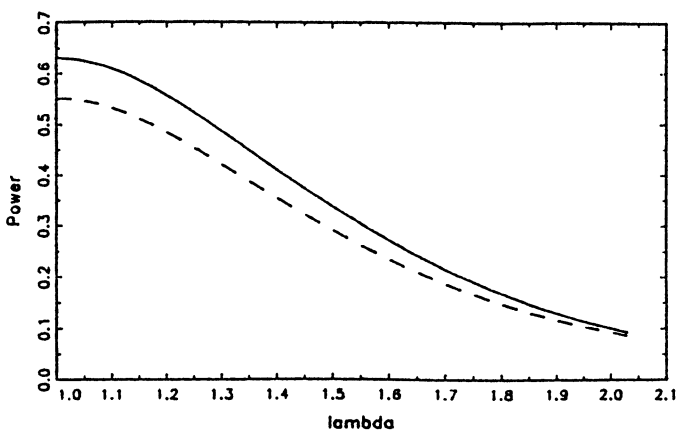

Figure 1. Power Comparison Between Φ_I and Φ_U When λ Varies. ——, the power of Φ_I; – – –, the power of Φ_U.

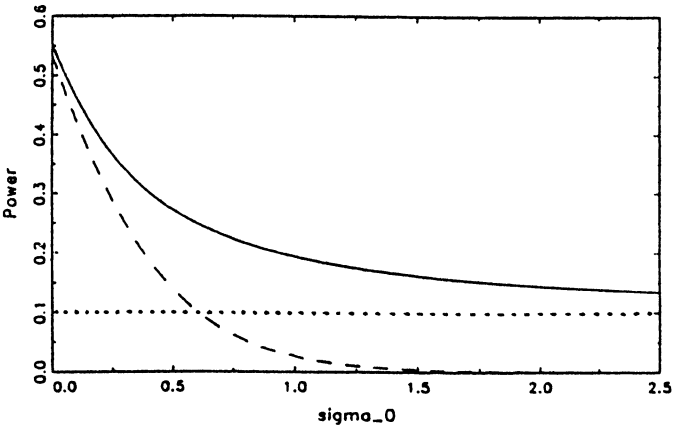

Figure 3. Power Comparison Between Φ_U and Φ_{L-C} at λ = 1 When $\sigma_0 = \sigma_S^2/\sigma_R^2$ Varies. ——, the power for Φ_U, – – –, the power for Φ_{L-C}, and the power for the .1 level.

Table 1. The Maximum Powers for the UMPI test (at $\lambda = 1$) With $\alpha = .05$ and $\alpha = .1$ When the Sample Size and the Alternative Hypothesis Vary

		(λ_1, λ_2)			
	$n_1 + n_2$	$(1.25^{-1}, 1.25)$	$(1.5^{-1}, 1.5)$	$(1.75^{-1}, 1.75)$	$(2^{-1}, 2)$
$\alpha = .05$	21	.0566	.075	.108	.160
	41	.0644	.112	.225	.423
	61	.0725	.167	.413	.693
	81	.0820	.243	.604	.849
$\alpha = .1$	21	.113	.150	.212	.306
	41	.128	.221	.408	.631
	61	.145	.318	.621	.834
	81	.163	.434	.773	.928

achieved at $\lambda = 1$ cannot be 1 either, which does not happen for detecting the average bioavailability. In that case, if the variance σ^2 goes to 0, then the power of $\Phi_{\mathrm{B-H-M}}$ goes to one; in our case, however, the corresponding parameter $\sigma^2(\delta_0)/(\sigma_{\mathrm{T}}^2 + \sigma_{\mathrm{R}}^2)$ has a positive lower bound and is not able to go to 0. Thus to increase the power, one needs to increase the sample size, the test level, or the alternative region.

Table 1 presents the maximal powers of the UMPI test Φ_{I} of case a for different settings. Enlarging the range of equivalence intervals turns out to be the most efficient way to increase the power. If (λ_1, λ_2) is too short, such as (.8, 1.25), then the maximum power is only .163 even for 81 observations and .1 level. If the interval (λ_1, λ_2) is expanded, then the power is improved. It seems reasonable to consider the interval (.5, 2), which requires one standard deviation to be within approximately 1.4 times the other. The power increments from such a choice are more than .2 over the column next on the left when the sample size exceeds 41. Enlarging the type I error also helps increase the power. For $\alpha = .05$ and $(\lambda_1, \lambda_2) = (.5, 2)$, the power is .423 for a sample size of 41. The power increases to .631, more than .21 higher, if the α-level is increased to .1. Usually the power increases by .15 in this situation. To increase the power to at least .8, the sample size must be at least 60. Additionally, if there is a subject effect, then the sample size must be even larger. The current choice of 12–24 for the bioequivalence test in mean is clearly too small. If one insists on using $\alpha = .05$ and $\lambda_1 = .8$, even for case a where Φ_{I} is valid, then the power reaches .8 or higher only when the sample size is around 700 or higher, as shown in Table 2. Requiring so many observations seems striking.

I close this section with Table 3, which contains the powers of the unbiased test for different sample sizes when $\sigma_S^2 = 2\sigma_{\mathrm{R}}^2$ and $\sigma_{\mathrm{R}}^2 = \sigma_{\mathrm{T}}^2$ under case b.

6. AN EXAMPLE

In this section the example presented by Liu and Chow (1992) is used to illustrate the proposed unbiased test for assessment of equivalence in variability. The results are compared to Liu and Chow's test. The purpose of the example was to compare the bioavailability between two formulations of a drug product. A standard 2×2 crossover design was carried out with 24 healthy volunteers. Each volunteer accepted either five 50 ml tablets (the test formulation) or 5 ml (50 mg/ml) of an oral suspension (the reference formulation). Blood samples were collected at 0 hours prior to dosing and at various times after dosing. The area under the blood concentration curve (AUC) values from 0–32 hours were calculated using the trapezoidal method and are given in Table 4.

Columns 5 and 6 of Table 4 present the period effect and the carry-over effect. It is not clear whether these effects are significant. The full model (1), which includes all effects, is then assumed to be on the safe side, and the unbiased test Φ_{U1} and Liu and Chow's test $\Phi_{\mathrm{L-C,1}}$ described in Section 4 are used in this section. Also, assume that $\lambda_1\lambda_2 = 1$.

From Table 4, we obtain $D_{01} = .0687$ and $S_{01} = 2.098$ by using (21). In fact, D_{01} is the least squares estimator of the slope of (18) with $\delta_0 = 1$, and $S_{01}^2/21$ is the unbiased estimator of the variance of D_{01}. Also, the point estimates for σ_{T}^2, σ_{R}^2, and σ_{S}^2 are calculated to be 14.89, 12.98, and 23.73. Therefore, λ is estimated to be 1.148. If the level α is .05, then among the choices of the alternative hypothesis in Table 1, (.5, 2) gives the largest $\Delta = 1/3$ (by (11)). With this Δ, (20) does not hold, because its right side is negative with $S_{01} = 2.098$. Thus $\Phi_{\mathrm{L-C,1}}$ fails to establish equivalence in variability for all choices of (λ_1, λ_2) due to the monotonicity of the right side of (20) in Δ. When the level α is .1, the same things happen: $\Phi_{\mathrm{L-C,1}}$ fails to establish equivalence in all cases. The situation is different for the unbiased test Φ_{U1}. The null hypothesis is rejected if $|D_{01}|$ is smaller than $T(S_{01})$ by Theorem 4.1. While $T(S_{01})$ is always positive, some chance (maybe small) does exist for a small $|D_{01}|$ to reject the null hypothesis. Also, T depends on Δ. Table 5 contains the $T(S_{01})$'s corresponding to different test levels and different choices of (λ_1, λ_2).

Table 2. The Maximum Power of Φ_{I} for Large Sample Sizes, Where the Level $\alpha = .05$ and $(\lambda_1, \lambda_2) = (.8, 1.25)$

	Sample Size				
	100	300	500	700	900
Power of Φ_{I}	.093	.297	.607	.809	.911

Table 3. The Power of Φ_{U} When the Sample Size Varies, for Fixed Level $\alpha = .1$, $(\lambda_1, \lambda_2) = (.5, 2)$, $\sigma_0 = 2$, and $\sigma_T^2 = \sigma_R^2$

	Sample size			
	21	41	61	81
Power of Φ_{U}	.1246	.1444	.1508	.1479

Table 4. AUC (0–32) for Test and Reference Formulations

Sequence	Subject number	Period I	Period II	Subject total	PD*	CD*
1						
RT	1	74.675	73.675	148.350	−1.000	−1.000
RT	4	96.400	93.250	189.659	−3.150	−3.150
RT	5	101.950	102.125	204.075	.175	.175
RT	6	79.050	69.450	148.500	−9.600	−9.600
RT	11	79.050	69.025	148.075	−10.025	−10.025
RT	12	85.950	68.700	154.650	−17.250	−17.250
RT	15	69.725	59.425	129.150	−10.300	−10.300
RT	16	86.275	76.125	162.400	−10.150	−10.150
RT	19	112.675	114.875	227.550	2.200	2.200
RT	20	99.525	116.250	215.775	16.725	16.725
RT	23	89.425	64.175	153.600	−24.250	−24.250
RT	24	54.175	74.575	129.750	19.400	19.400
2						
TR	2	74.825	37.350	112.175	−37.475	−37.475
TR	3	86.875	51.925	138.800	−34.950	34.950
TR	7	81.675	72.175	153.850	−9.500	9.500
TR	8	92.700	77.500	170.200	−14.200	14.200
TR	9	50.450	71.875	122.325	21.425	−21.425
TR	10	66.125	94.025	160.150	27.900	−27.900
TR	13	122.450	124.975	247.425	2.525	−2.525
TR	14	99.075	86.225	184.300	−13.850	13.850
TR	17	86.350	95.925	182.275	9.575	−9.575
TR	18	49.925	67.100	117.025	17.175	−17.175
TR	21	42.700	59.425	102.125	16.725	−16.725
TR	22	91.725	114.050	205.775	22.325	−22.325

PD* = 2 × (period difference).
CD* = 2 × (crossover difference).

In all cases considered here, $\Phi_{\mathrm{L-C},1}$ fails to establish equivalence in variability. So does $\Phi_{\mathrm{U},1}$, except for the last case, where $\alpha = .1$ and $(\lambda_1, \lambda_2) = (.5, 2)$. In this case, $|D_{01}| = .0687$ is smaller than $T(S_{01}) = .0748$, then $\Phi_{\mathrm{U},1}$ rejects the null hypothesis, and equivalence in variability is established.

When $(\lambda_1, \lambda_2) = (.8, 1.25)$, numerical calculations show that the p value for Φ_{U1} is .112, much smaller than the p value for $\Phi_{\mathrm{L-C},1}$, .4635. All of the evidence suggests that the unbiased tests do a much better job in establishing equivalence in variability.

7. SUMMARY

As has been shown, it is hard to detect the accuracy, such as (.8, 1.25), with .05 type I error and sample size 12–24. With certain alternative hypothesis and test level, we may need a large sample size to obtain a reasonable power. In the absence of a subject effect, the UMPI test Φ_{I} in Theorem 2.1 (or Φ_{I1}, depending on the existence of the period effect and the carry-over effect) is recommended. The power increment of the UMPI tests over the unbiased tests is substantial, and these tests are very easy to determine. Their critical values, c_i, are based on the F distribution. When a subject effect is present, the proposed unbiased test Φ_{U} in Theorem 3.1 (or Φ_{U1}) should be used. The main advantage of Liu and Chow's test $\Phi_{\mathrm{L-C}}$ (or $\Phi_{\mathrm{L-C},1}$) is its ease of use. But it has almost no power to detect the equivalence in many cases. On the other hand, the unbiased test Φ_{U} (or Φ_{U1}) always has higher power to detect equivalence. Also, the boundary of its rejection region (i.e., the function T) can be calculated by computer (see Brown et al. 1997). A Gauss program, for example, running on a Pentium computer can solve it within 5 minutes and is available from the author. An alternative formulation of equivalence in variability discussed by Wang (1995) allows the maximum power to be 1. Through linear regression, the problem of equivalence in variability reduces to the average bioavailability problem. As pointed out by one of the referees, the proposed unbiased tests are for intrasubject variability. Nonetheless, it can be shown that they also define exact α-level tests to detect the similarity of variabilities of the responses from two formulations; that is, the variances of $Y_{i\mathrm{T}k}$ and $Y_{i\mathrm{R}k}$.

Table 5. The Boundary $T(S_{01})$'s for $S_{01} = 2.098$

	(λ_1, λ_2), Δ			
	(.8, 1.25), 1/9	(1.5^{-1}, 1.5), 1/5	(1.75^{-1}, 1.75), 3/11	(.5, 2), 1/3
α = .05	.0298	.0317	.0342	.0371
α = .1	.0597	.0635	.0686	.0748

APPENDIX: PROOFS

Proof of Lemma 2.1

Let f_η be the pdf of Z and $n = n_1 + n_2 - 1$, $h(z)$ be $f_{\eta'}(z)/f_\eta(z)$ for $\eta < \eta'$. Then $f_\eta(z) = \exp(n(z - \eta)/2)/[\beta(n/2, n/2)(1 + \exp(z - \eta))^n]$ and

$$\frac{d \log h(z)}{dz} = ne^z \frac{e^{-\eta} - e^{-\eta'}}{(1 + e^{z-\eta})(1 + e^{z-\eta'})} > 0. \qquad \text{(A.1)}$$

Thus $h(z)$ is strictly increasing in z, and the distribution family of Z is monotone likelihood ratio. So that this distribution family

is STP_3, by problem 29 of Lehmann (1986, p. 119), it is sufficient to show that the following holds:

For $\eta < \eta' < \eta''$ and $K_1, K_2, K_3 > 0$, let $g(z) = K_1 f_\eta(z) - K_2 f_{\eta'}(z) + K_3 f_{\eta''}(z)$. If $g(z_1) = g(z_3) = 0$, then g is positive outside the interval (z_1, z_3) and negative inside.

$$g(z) > 0 \Leftrightarrow K_2 f_{\eta'}(z) > K_1 f_\eta(z) + K_3 f_{\eta''}(z)$$

$$\Leftrightarrow c > c_1 \left(\frac{1+e^{z-\eta'}}{1+e^{z-\eta}}\right)^n + \left(\frac{1+e^{z-\eta'}}{1+e^{z-\eta''}}\right)^n$$

(for some $c, c_1 > 0$).

Let $l(z)$ be the right side of the last line,

$$l'(z) = n\frac{(1+e^{z-\eta'})^{n-1}e^z}{(1+e^{z-\eta})^{n+1}}\left[-d_1 + d_2\left(\frac{1+e^{z-\eta}}{1+e^{z-\eta''}}\right)^{n+1}\right],$$

where $d_1 = -c_1(\exp(-\eta') - \exp(-\eta)) > 0$ and $d_2 = \exp(-\eta') - \exp(-\eta'') > 0$. Because $(1 + \exp(z - \eta))/(1 + \exp(z - \eta''))$ is strictly increasing in z, $l'(z) = 0$ at most one time. If $l'(z_0) = 0$ for some z_0, then $l(z)$ is strictly decreasing on $(-\infty, z_0)$ and increasing on $(z_0, +\infty)$. Thus if $g(z_1) = g(z_3) = 0$, which implies that there is $z_0 \in (z_1, z_3)$ so that $l'(z_0) = 0$, then $g(z)$ is positive outside (z_1, z_3) and negative inside.

Proof of Theorem 2.1

Because $f_\eta(z)$ is continuous in z for each η and is STP_3, $\Phi_I(Z)$ is the UMP test for (5) among the tests based on the data through $S_{\mathbf{Y}_T\mathbf{Y}_T}/S_{\mathbf{Y}_R\mathbf{Y}_R}$, by problem 30 of Lehmann (1986, p. 120). Also, $(\bar{Y}_T, \bar{Y}_R, S_{\mathbf{Y}_T\mathbf{Y}_T}, S_{\mathbf{Y}_R\mathbf{Y}_R})$ are sufficient statistics and $S_{\mathbf{Y}_T\mathbf{Y}_T}/S_{\mathbf{Y}_R\mathbf{Y}_R}$ is maximal invariant with respect to G; then $\Phi_I(Z)$ defines a UMPI test based on data through $(\bar{Y}_T, \bar{Y}_R, S_{\mathbf{Y}_T\mathbf{Y}_T}, S_{\mathbf{Y}_R\mathbf{Y}_R})$. Thus Φ_I is UMPI by theorem 6 of Lehmann (1986, p. 301).

Proof of Lemma 3.3

See theorem 1 of Brown et al. (1997).

Proof of Theorem 3.1

Now the parameter space $\Omega = \{\omega = (\beta, \mu, F_R, \sigma_R, \sigma_S) | \beta \in (-\delta_0, +1), \mu \in R, F_R \in R, \sigma_R > 0, \sigma_S > 0\}$. To see the unbiasedness of Φ_U, it is sufficient to show that for any given $S_{\mathbf{VV}}$,

$$E_\omega(\Phi_U | S_{\mathbf{VV}}) = \begin{cases} \geq \alpha & \text{if } |\beta| \leq \Delta \\ \leq \alpha & \text{if } |\beta| \geq \Delta; \end{cases} \tag{A.2}$$

then

$$E_\omega \Phi_U = E_\omega E_\omega(\Phi_U | S_{\mathbf{VV}}) = \begin{cases} \geq \alpha & \text{if } |\beta| \leq \Delta \\ \leq \alpha & \text{if } |\beta| \geq \Delta \end{cases}.$$

Applying Lemma 3.3 on D_0 and S_0 for any $S_{\mathbf{VV}}$, (A.2) holds. Thus Φ_U is unbiased.

By definition, $\Phi_U = 1$ iff $|D_0| < T(S_0)$, where D_0 and S_0 are defined in (12). Let $a = S_{\mathbf{Y}_R\mathbf{V}}/S_{\mathbf{VV}}$, $b = S_{\mathbf{Y}_R\mathbf{Y}_R}/S_{\mathbf{VV}}$. Then $D_0 = (\delta_0 + 1)a + 1$, $S_0 = (\delta_0 + 1)(b - a^2)^{1/2}$, and

$$t_1 = \sqrt{n}\left[S_{\mathbf{U}(\lambda_1)\mathbf{V}} \Big/ \sqrt{S_{\mathbf{U}(\lambda_1)\mathbf{U}(\lambda_1)} S_{\mathbf{VV}}}\right] \div \sqrt{1 - S^2_{\mathbf{U}(\lambda_1)\mathbf{V}}/(S_{\mathbf{U}(\lambda_1)\mathbf{U}(\lambda_1)} S_{\mathbf{VV}})}$$

$$= \sqrt{n}[1/(\lambda_1 + 1) + a] \Big/ \sqrt{b - a^2} = \sqrt{n}(D_0 + \Delta)/S_0.$$

Similarly $t_2 = n^{1/2}(D_0 - \Delta)/S_0$. Thus (16) is equivalent to (17). Hence for any given S_{VV}, Φ_{L-C} can be thought of as Φ_S, Schuirmann's test. Then that Φ_U is uniformly more powerful than Φ_{L-C} follows from Φ_{B-H-M} being uniformly more powerful than Φ_S.

[Received February 1995. Revised September 1996.]

REFERENCES

Anderson, S., and Hauck, W. W. (1990), "Consideration of Individual Bioequivalence," *Journal of Pharmacokinetics and Biopharmaceutics*, 18, 259–273.

Brown, L. D., Hwang, J. T. G., and Munk, A. (1997), "An Unbiased Test for the Bioequivalence Problem," submitted to *Annals of Statistics*.

Lehmann, E. T. (1986), *Testing Statistical Hypotheses* (2nd ed.), New York: Wiley.

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.

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 (1992), "FDA Guidance on Statistical Procedures for Bioequivalence Studies Using a Standard Two-Treatment Crossover Design," *Division of Bioequivalence, Office of Generic Drugs, Center for Drug Evaluation and Research, U. S. Food and Drug Administration, Rockville, MD.*

Wang, W. (1995), "On Assessment of Bioequivalence," Ph. D. thesis, Cornell University.