



Statistical Tests for Multivariate Bioequivalence

Weizhen Wang; J. T. Gene Hwang; Anirban Dasgupta

Biometrika, Vol. 86, No. 2. (Jun., 1999), pp. 395-402.

Stable URL:

<http://links.jstor.org/sici?sici=0006-3444%28199906%2986%3A2%3C395%3ASTFMB%3E2.0.CO%3B2-Q>

Biometrika is currently published by Biometrika Trust.

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/bio.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.

Statistical tests for multivariate bioequivalence

BY WEIZHEN WANG

*Department of Mathematics and Statistics, Wright State University, Dayton, Ohio 45435,
U.S.A.*

wwang@math.wright.edu

J. T. GENE HWANG

Department of Statistical Science, Cornell University, Ithaca, New York 14853, U.S.A.

hwang@math.cornell.edu

AND ANIRBAN DASGUPTA

Department of Statistics, Purdue University, West Lafayette, Indiana 47907, U.S.A.

dasgupta@stat.purdue.edu

SUMMARY

Although the U.S. Food and Drug Administration (FDA, 1992) recommends testing bioequivalence of individual pharmacokinetic parameters one at a time, it seems reasonable and interesting to conduct a test simultaneously for all the parameters. In this paper, we discuss several ways to construct such tests. It is shown that the confidence set approach leads to a test which can be uniformly improved by the intersection of Schuirmann's two one-sided tests procedure. The latter test can further be improved upon noticeably by using the one-dimensional unbiased test of Brown, Hwang & Munk (1997). Numerical calculations of powers are given to support this claim.

Some key words: Confidence set; Intersection-union method; Likelihood-ratio test.

1. INTRODUCTION

In bioequivalence studies, one typically is interested in demonstrating that a new drug is similar in efficacy to a brand-name drug. The FDA (1992) recommends a 2×2 crossover design. Typically 24 subjects are randomly divided into two groups. One group will be given the brand-name drug and after a wash-out period the new drug. The other group is similarly treated except that the order of drugs is reversed. The blood samples are then collected from each subject at various times and a blood concentration curve against time of a certain ingredient is obtained.

The three most typical characteristics of the blood concentration curve considered are the area under the concentration curve, AUC, the maximum concentration, C_{\max} , and the time to reach the maximum concentration, T_{\max} . The FDA (1992) then recommends that one applies Schuirmann's (1987) two one-sided tests procedure on one of the above characteristics to test the hypotheses

$$H_0: |\theta| \geq \Delta \text{ versus } H_A: |\theta| < \Delta.$$

Here $\theta = \theta_T - \theta_R = \log(\mu_T/\mu_R)$, where \log is the natural logarithm function, and θ_T and θ_R represent respectively the means of the characteristic in a log scale corresponding to the new treatment and the reference, i.e. brand-name, treatment. The FDA's recommended cut-off number Δ is $\log(1.25)$ so that μ_T and μ_R stay within 80% of each other. If H_0 is rejected, then bioequivalence is declared.

Schuirman's (1987) test, however, is one-dimensional. It would seem reasonable and interesting to consider simultaneously the test involving all relevant characteristics. In this paper we shall consider testing the hypotheses

$$H_0: \max_{1 \leq i \leq p} |\theta^{(i)}| \geq \Delta \text{ versus } H_A: \max_{1 \leq i \leq p} |\theta^{(i)}| < \Delta, \quad (1.1)$$

instead. Here the $\theta^{(i)}$'s are, for example when $p = 3$, the differences of the means of the area under the concentration curve, C_{\max} , and T_{\max} , usually in logarithmic scales, corresponding to the treatment and the reference drugs. Statisticians have proposed various criteria in the area of multiparameter bioequivalence testing. Some have even proposed to test separately. Our hypotheses (1.1) can be generalised to different Δ 's as in (1.5) and allow the least interaction among the parameters. If H_0 is rejected, it ensures that each θ^i is bounded in absolute value by Δ or Δ_i .

Now consider the canonical form

$$X \sim N_p(\theta, \Sigma), \quad \hat{\Sigma} \sim W(\Sigma, d), \quad (1.2)$$

with the unknown parameters $\theta = (\theta^{(1)}, \dots, \theta^{(p)})'$ and $\Sigma = (\sigma_{ij})_{p \times p}$, and X and $\hat{\Sigma}$ are independent. Here $W(\Sigma, d)$ denotes a Wishart distribution with d degree of freedom (Anderson, 1984, p. 249) and p is the number of characteristics. This canonical form can be applied to a general linear model including the crossover design below with period effects and subject effects. Specifically, consider a standard 2×2 crossover design that compares a test drug formulation with a reference drug formulation, as in Chow & Liu (1992, p. 34):

for Sequence 1

$$Y_{iR1} = \mu + S_{i1} + F_R + P_1 + \varepsilon_{iR1}, \quad Y_{iT1} = \mu + S_{i1} + F_T + P_2 + \varepsilon_{iT1}, \quad (1.3a)$$

and for Sequence 2

$$Y_{iT2} = \mu + S_{i2} + F_T + P_1 + \varepsilon_{iT2}, \quad Y_{iR2} = \mu + S_{i2} + F_R + P_2 + \varepsilon_{iR2}, \quad (1.3b)$$

where Y_{ijk} , a p -dimensional random vector, is the response of the i th subject in the k th sequence for the j th formulation, in which $j = R, T$, $k = 1, 2$ and $i = 1, 2, \dots, n_k$; μ is the overall mean vector; F_j is the fixed effect for the j th formulation with $F_R + F_T = 0$; P_1 and P_2 are the fixed period effects with $P_1 + P_2 = 0$; S_{ik} is the random subject effect; ε_{ijk} is the intra-subject random error in observing Y_{ijk} . It is also assumed that, for $i = 1, \dots, n_k$ and $k = 1, 2$, $\varepsilon_{iT2} - \varepsilon_{iR2}$ are independently and identically distributed normal random vectors with mean zero and common unknown $p \times p$ covariance matrix Σ_0 . There are examples in which a multinormal model appears reasonable; see the Cornell University 1997 master's thesis by Xuan Ma. Let $\theta = F_T - F_R$, $V_{ik} = (Y_{iT2} - Y_{iR2})/2$ for all i and k and $\bar{V}_{.k} = \sum V_{ik}/n_k$. Then

$$X = \sum_{k=1}^2 \bar{V}_{.k}, \quad \hat{\Sigma} = \left(\frac{1}{n_1} + \frac{1}{n_2} \right) \sum_{k=1}^2 \sum_{i=1}^{n_k} (V_{ik} - \bar{V}_{.k})(V_{ik} - \bar{V}_{.k})' \quad (1.4)$$

satisfy (1.2) with $\Sigma = (1/n_1 + 1/n_2)\Sigma_0/4$ and $d = n_1 + n_2 - 2$.

Now return to model (1.2). In this paper, we shall derive tests for (1.1). Note that our

result applies to the hypotheses in which the equivalence limits are not equal for different characteristics. That is, consider

$$H_0: |\theta^{(i)}| \geq \Delta_i \quad \text{for some } i \quad (1.5a)$$

versus

$$H_A: |\theta^{(i)}| < \Delta_i \quad \text{for all } i, \quad (1.5b)$$

where the Δ_i 's, the threshold constants, may be different for different i . To see this, let A be a $p \times p$ diagonal matrix with i th element Δ/Δ_i , and define $X_{\text{new}} = AX$ and $\hat{\Sigma}_{\text{new}} = A\hat{\Sigma}A'$. Then (1.5) reduces to (1.1) with X_{new} and $\hat{\Sigma}_{\text{new}}$ as the test statistics.

We review some results for $p = 1$ first. Schuirmann (1987) provided a size- α test for (1.1) by using the intersection-union method with the rejection region

$$|X| < \Delta - t_d(\alpha)(\hat{\Sigma}/d)^{\frac{1}{2}}, \quad (1.6)$$

where $t_d(\alpha)$ is the upper α quantile of Student's t -distribution with d degrees of freedom. The size of a test is defined as the supremum of the type I error over the null hypothesis space. In contrast, a test is said to have level α if its size is no greater than α . Brown et al. (1997) obtained a size- α unbiased test with the rejection region

$$|X| < B(\hat{\Sigma}^{\frac{1}{2}}), \quad (1.7)$$

where B is some positive function which is never smaller than the right-hand side of (1.6). Therefore, (1.7), the test of Brown et al. (1997), uniformly improves Schuirmann's test (1.6) in power. This fact will be used in § 3.

For the multivariate case $p > 1$, Westlake (1988), Hauck et al. (1995) and a University of the Orange Free State technical report by R. Schall, P. C. N. Groenewald, L. A. Potgieter and H. G. Luus considered tests which reject H_0 if and only if, for each component, applying the two one-sided tests leads to the rejection of the nonequivalence hypothesis; i.e. reject H_0 in (1.1) if

$$|X^{(i)}| < \Delta - C(\hat{\Sigma}_{ii}/d)^{\frac{1}{2}} \quad (1 \leq i \leq p) \quad (1.8)$$

for some constant C , where $X^{(i)}$ is the i th component of X and $\hat{\Sigma}_{ii}$ is the i th diagonal element of $\hat{\Sigma}$. In particular, this test is called the repeated univariate test, based on Decision 2, in the report by Schall et al. Westlake (1988) and Hauck et al. (1995) recommended different choices of C and make a wrong suggestion or conjecture about the size of the test.

Is the test (1.8) reasonable? If so, what is the appropriate choice of C that would lead to the correct size? Is it possible to do better than (1.8)? These are the questions that we will address in this paper.

We try various approaches to derive tests. In § 2, first we show that the intersection-union method leads to (1.8). Secondly, the likelihood ratio approach is shown also to lead to (1.8). The result seems interesting even for $p = 1$, since this concludes that the two one-sided tests procedure can be justified by the likelihood ratio approach. As for the multivariate case, such a simple expression is not possible if we test again H_A as the null hypothesis. Thirdly, we take a confidence set approach which also leads to (1.8) with a larger choice of C and then yields a less powerful test than the intersection-union test. The latter one, however, can be improved uniformly as shown in § 3. By adapting the one-dimensional test of Brown et al. (1997), we can construct a multivariate test uniformly improving upon the intersection-union test (1.8). The improvement in power may be as

big as 0.11. As an interesting fact, the tests in §§ 2 and 3 are also valid as long as each characteristic follows a univariate normal distribution; they do not have to have a jointly multivariate normal distribution. Numerical studies are reported in § 4.

2. INTERSECTION-UNION TEST

Schuirmann (1987) proposed test (1.6) by using an approach called the intersection-union method in Berger (1982); see also Casella & Berger (1990, p. 356). In this section we will generalise Schuirmann's test to $p > 1$.

Let us consider p sets of hypotheses,

$$H_{0i}: |\theta^{(i)}| \geq \Delta \text{ versus } H_{Ai}: |\theta^{(i)}| < \Delta, \quad (2.1)$$

for $i = 1, \dots, p$. For each set of hypotheses, one has the size- α rejection region

$$R_i = \{ |X^{(i)}| < \Delta - t_d(\alpha)(\hat{\Sigma}_{ii}/d)^{\frac{1}{2}} \}$$

corresponding to the Schuirmann test (1.6). The rejection region

$$R^I = \bigcap_{i=1}^p R_i = \{ |X^{(i)}| < \Delta - t_d(\alpha)(\hat{\Sigma}_{ii}/d)^{\frac{1}{2}}, \text{ for all } i \} \quad (2.2)$$

defines a level- α test for (1.1) since $H_A = \bigcap_{i=1}^p H_{Ai}$. This method is called intersection-union since the rejection region is the intersection of several rejection regions and the null hypothesis is the union of several null hypotheses. The proof of the following result, established in Berger & Hsu (1996), is omitted.

THEOREM 1. *Consider the testing problem (1.1) under model (1.2). Then the test R^I for (1.1) has size α .*

It is worth pointing out that the intersection-union method depends on the marginal distributions of each $X^{(i)}$ only. As long as R_i defines a level- α test for (2.1), which is true if each characteristic is normally distributed, the intersection $\bigcap_{i=1}^p R_i$ then defines a level- α test for (1.1). Therefore, we do not need to assume in Theorem 1 above that $X^{(i)}$'s are jointly normal.

Secondly, note that the test (2.2) is basically a likelihood ratio test. Indeed, a direct calculation shows that the likelihood ratio $\lambda(X, \hat{\Sigma})$ equals

$$\lambda(X, \hat{\Sigma}) = \frac{\sup_{H_0} L(\theta, \Sigma; X, \hat{\Sigma})}{\sup_{H_0 \cup H_A} L(\theta, \Sigma; X, \hat{\Sigma})} = \left\{ \frac{1}{1 + \inf_{\theta \in H_0} (X - \theta)' \hat{\Sigma}^{-1} (X - \theta)} \right\}^{(d+1)/2},$$

where L is the likelihood function for model (1.2). Using this, we conclude the following.

Result 1. For $0 < K < 1$, $\lambda(X, \hat{\Sigma}) < K$ if and only if

$$|X^{(i)}| < \Delta - C(\hat{\Sigma}_{ii}/d)^{\frac{1}{2}} \quad (1 \leq i \leq p), \quad (2.3)$$

where $C^2 = d(K^{-2/(d+1)} - 1)$.

See the Appendix for a sketch of the proof. Putting Result 1 and Theorem 1 together, we conclude that the choice $C = t_d(\alpha)$ in (2.3) results in a test of size α , identical to R^I .

Remark. It is interesting to note that, in Result 1, $\lambda(X, \hat{\Sigma})$ has a simple expression because $\inf_{\theta \in H_0} (X - \theta)' \hat{\Sigma}^{-1} (X - \theta)$ has a simple one. There is no simple analytic expression, however, for $\inf_{\theta \in H_A} (X - \theta)' \hat{\Sigma}^{-1} (X - \theta)$, a problem relating to that of finding the maximum likelihood estimator of θ when θ is known to be in a cube.

Another way to obtain a test is the confidence set approach. Let $C(X, \hat{\Sigma})$ be a $1 - \alpha$ confidence set. Then a test for (1.1), which rejects the null hypothesis if and only if

$$C(X, \hat{\Sigma}) \subset H_A, \quad (2.4)$$

has level α . Specifically, consider the Hotelling's $T^2 = d(X - \theta)' \hat{\Sigma}^{-1}(X - \theta)$. Then $(d - p + 1)T^2/(dp)$ has an F distribution with p and $d - p + 1$ degrees of freedom. Let $F_{p, d-p+1}(\alpha)$ be its upper α quantile. The set

$$C(X, \hat{\Sigma}) = \{\theta : T^2 \leq C_1^2\}, \quad (2.5)$$

with $C_1^2 = F_{p, d-p+1}(\alpha)(dp)/(d - p + 1)$, has coverage probability $1 - \alpha$. Putting (2.4) and (2.5) together yields a test with the rejection region

$$|X^{(i)}| < \Delta - C_1(\hat{\Sigma}_{ii}/d)^{\frac{1}{2}} \quad (1 \leq i \leq p). \quad (2.6)$$

The test has size much less than α . In fact the actual size equals

$$\alpha_1 = \text{pr}(\chi_1^2/\chi_{d-p+1}^2 > q)/2,$$

where q is the α upper quantile of χ_p^2/χ_{d-p+1}^2 and χ_k^2 denotes a chi-squared random variable with k degrees of freedom. All the above chi-squared random variables are independent. In particular, if $p = 1$, then $\alpha_1 = \alpha/2$, a well-known result in the one-dimensional case. See, for example, Hsu et al. (1994). Table 1 shows that α_1 is much less than the nominal level $\alpha = 0.05$. The test (2.6) has a rejection region contained in R^I and hence is uniformly less powerful than R^I .

Table 1. *The actual size α_1 of the test derived from the confidence set approach when the test level is 0.05, the number of characteristics p varies and $d - p + 1$ is fixed at 23.*

	$p = 1$	$p = 2$	$p = 3$	$p = 4$	$p = 5$	$p = 10$
α_1	0.025	6.66×10^{-3}	2.14×10^{-3}	7.37×10^{-4}	2.62×10^{-4}	1.61×10^{-6}

3. IMPROVED TEST

Although the intersection-union test R^I was shown to be a reasonable test in the last section, it can be uniformly improved in power. In this section we will generalise the test of Brown et al. (1997), given in (1.7), to the case $p > 1$. Here, we will apply the intersection-union method again.

Consider the rejection region of an unbiased test

$$R_i^U = \{|X^{(i)}| < B(\hat{\Sigma}_{ii}^{\frac{1}{2}})\}$$

of Brown et al. (1997), given in (1.7) for the one-dimensional case. It is known that R_i^U is an unbiased level- α test for (2.1) which contains properly the rejection region of Schuirmann's (1987) test. Consequently, it has a uniformly larger power than Schuirmann's test. Consider the rejection region

$$R_U^I = \bigcap_{i=1}^p R_i^U$$

for testing (1.1). Although no detailed evaluation was given, a similar test of this kind was mentioned in Berger & Hsu (1996) except that R_i^U is replaced by the one-dimensional test

of Berger & Hsu which is analogous to Brown et al. (1997). Berger & Hsu's test is slightly less powerful, although having a smoother boundary. Obviously R_U^I uniformly improves R^I . Further, it is clear that R_U^I has size α for testing (1.1) under only a marginal normal assumption which is weaker than the multinormal assumption of $X^{(i)}$. Both R_U^I and R^I have the good property that their power functions are decreasing in each $|\theta^{(i)}|$ when the other $\theta^{(j)}$'s and Σ are fixed. It seems desirable to have maximum power at $\theta = 0$, or equivalently when the two drug effects have the same mean. Unlike the one-dimensional case, when $p > 1$ R_U^I is not unbiased. For $p > 1$, we shall show below that a nontrivial unbiased test with a power function decreasing in $\theta^{(i)}$ does not exist. Therefore, an unbiased test for (1.1), if it exists, may have a pathological rejection region and may be difficult to construct.

THEOREM 2. *Let $\Phi(X, \hat{\Sigma})$ be the critical function of a level- α unbiased test for (1.1) whose power function is unimodal with respect to each coordinate $\theta^{(i)}$. Then $\Phi(X, \hat{\Sigma})$ is the trivial test, that is $\Phi(X, \hat{\Sigma}) = \alpha$, almost surely.*

See the Appendix for the proof.

Even in the one-dimensional case, the issue of whether or not one should use the test of Brown et al. (1997) and a similar test proposed by Berger & Hsu remains somewhat controversial; see Berger & Hsu (1996) and the related discussion by various authors.

One objection against these improved tests, as also pointed out by the associate editor handling the paper, is the possibility of concluding bioequivalence even though, for example in the one-dimensional case, the point estimator X of θ is outside the bioequivalence range $(-\Delta, \Delta)$. There is a simple way to resolve the problem, however. One may simply modify the test of Brown et al. (1997) by further requiring to reject only if X is inside $(-\Delta, \Delta)$. The resulting test uniformly improves upon the two one-sided tests procedure and has a power similar to that of the original test; see Brown et al. (1997). For our multidimensional problem, we may similarly modify the test by further requiring to reject only if

$$\max_{1 \leq i \leq p} |X^{(i)}| < \Delta.$$

We would expect the power to be similar to that of the unmodified one.

4. NUMERICAL STUDIES OF POWER

We employ the model (1.3) and choose $n_1 = n_2 = 12$, that is $d = 22$, and $\Delta = \log(1.25)$ in (1.1), which means that the ratios of two population means for the characteristics must be within (0.8, 1.25) for bioequivalence to be asserted. All results are based on 100 000 simulations.

When $p = 3$, i.e. consider the area under the concentration curve, C_{\max} , and T_{\max} simultaneously, the power is a function of

$$\theta = \begin{pmatrix} \theta^{(1)} \\ \theta^{(2)} \\ \theta^{(3)} \end{pmatrix}, \quad \Sigma_0 = \begin{pmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \rho_{13}\sigma_1\sigma_3 \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 & \rho_{23}\sigma_2\sigma_3 \\ \rho_{13}\sigma_1\sigma_3 & \rho_{23}\sigma_2\sigma_3 & \sigma_3^2 \end{pmatrix},$$

where σ_i is the standard deviation of each variable and ρ_{ij} is the correlation coefficient of the (i, j) th variables. Therefore, $\Sigma = \Sigma_0/24$ by (1.4). In the simulation, $\theta^{(i)} = a$, $\sigma_i = b$ and $\rho_{ij} = c$ for all i, j . It is reasonable to assume that the ρ_{ij} 's are positive since the area under the concentration curve, C_{\max} , and T_{\max} are usually positively associated.

Table 2. The power simulations of the test R^I and the improved test R_U^I in §§ 2 and 3 when $p = 3, 2$, $a = 0$, $b = 0.2, 0.4$ or 0.6 and c varies.

c	$b = 0.2$		$b = 0.4$		$b = 0.6$	
	R^I	R_U^I	R^I	R_U^I	R^I	R_U^I
$p = 3$						
0	0.99911	0.99911	0.32068	0.32285	0.00253	0.01435
0.1	0.99915	0.99915	0.32116	0.32285	0.00271	0.01462
0.2	0.99907	0.99909	0.32749	0.32990	0.00304	0.01521
0.3	0.99930	0.99930	0.34180	0.34413	0.00361	0.01666
0.4	0.99930	0.99930	0.35490	0.35667	0.00444	0.01896
0.5	0.99916	0.99917	0.37449	0.37594	0.00619	0.02129
0.6	0.99923	0.99923	0.40116	0.40335	0.00826	0.02489
0.7	0.99922	0.99924	0.43944	0.44131	0.01221	0.03345
0.8	0.99936	0.99940	0.48392	0.48525	0.01965	0.04440
0.9	0.99942	0.99943	0.54011	0.54181	0.03744	0.07259
1.0	0.99971	0.99971	0.68450	0.68580	0.13766	0.24317
$p = 2$						
0	0.99943	0.99944	0.46706	0.46763	0.01923	0.05993
0.1	0.99947	0.99948	0.46802	0.46822	0.01932	0.05869
0.2	0.99945	0.99945	0.47310	0.47501	0.02045	0.06034
0.3	0.99943	0.99944	0.47657	0.47850	0.02137	0.06303
0.4	0.99950	0.99951	0.48763	0.48976	0.02421	0.06621
0.5	0.99937	0.99937	0.50121	0.50299	0.02658	0.06977
0.6	0.99955	0.99955	0.51687	0.51865	0.03052	0.07497
0.7	0.99938	0.99939	0.53662	0.53856	0.03636	0.08267
0.8	0.99941	0.99941	0.56142	0.56292	0.04674	0.09612
0.9	0.99970	0.99970	0.59507	0.59663	0.06331	0.12260
1.0	0.99972	0.99973	0.68422	0.68562	0.13720	0.24286

Table 2 gives for $p = 3$ and $p = 2$ the power simulations of the intersection-union test and the improved test in §§ 2 and 3 when $a = 0$, $b = 0.2, 0.4, 0.6$ and as c increases from 0 to 1 in ten steps. The improved test R_U^I , as expected, has a uniformly higher power than the test R^I . The improvement in power is larger when the standard deviation b and the correlation coefficient c become larger. In a somewhat extreme case, where $b = 0.6$ and $c = 1$, the power improvements in Table 2 are about 0.10. However, the observations have been log-transformed, and hence the standard deviation b is roughly the coefficient of variation for the observations in the original scale, which is usually less than or equal to 0.4 in bioequivalence studies. It is unfortunate that for such b 's the improvement is small. Drastically different methods have to be considered in order to improve upon R^I substantially. However, our result does demonstrate the existence of tests uniformly improving upon R^I .

APPENDIX

Proofs

Sketch of the proof of Result 1. Under model (1.2), one may show that

$$\lambda(X, \hat{\Sigma}) = \left\{ 1 + \inf_{\theta \in H_0} d(X, \theta) \right\}^{-(d+1)/2}, \quad (\text{A.1})$$

where $d(X, \theta) = (X - \theta)' \hat{\Sigma}^{-1} (X - \theta)$. Then $\lambda(X, \hat{\Sigma}) < K$ is equivalent to

$$\inf_{\theta \in H_0} d(X, \theta) > C^2/d. \quad (\text{A} \cdot 2)$$

If $X \in H_0$, both (2.3) and (A.2) fail. If $X \notin H_0$, then X stays inside the cube H_A . Let θ_{\min} be the minimiser of the left-hand side of (A.2). Let P_{ij} denote the $(p-1)$ -dimensional plane which consists of points whose i th coordinate is equal to $j\Delta$, where $j=1$ or $j=-1$, and let $d_{ij} = \min_{\theta \in P_{ij}} d(X, \theta)$. Since θ_{\min} is the point of the intersection between the cube H_A and the largest ellipsoid of the form $S = \{\theta : d(X, \theta) = c\}$ such that S is contained in the cube,

$$d(X, \theta_{\min}) := \inf_{\theta \in H_0} d(X, \theta) = \min_{i,j} d_{ij}. \quad (\text{A} \cdot 3)$$

Let $\eta = \hat{\Sigma}^{-\frac{1}{2}} \theta$, and write P_{ij} and $d(X, \theta)$ as

$$\eta' \hat{\Sigma}^{\frac{1}{2}} e_i = j\Delta, \quad (\hat{\Sigma}^{-\frac{1}{2}} X - \eta)' (\hat{\Sigma}^{-\frac{1}{2}} X - \eta), \quad (\text{A} \cdot 4)$$

respectively, where e_i is the i th coordinate vector. Then d_{ij} is the square of the distance between point $\hat{\Sigma}^{-\frac{1}{2}} X$ and plane P_{ij} in the η -space. Hence

$$d_{ij}^2 = \frac{|X^{(i)} - j\Delta|^2}{\hat{\Sigma}_{ii}}.$$

Using this, (A.3) and (A.2), we may establish (2.3).

Proof of Theorem 2. The normal distribution belongs to the exponential family and hence $E_{\theta, \Sigma} \Phi(X, \hat{\Sigma})$ is an analytic function of θ for each Σ . Now $E_{\theta, \Sigma} \Phi(X, \hat{\Sigma}) = \alpha$ if $\theta^{(1)} = \pm \Delta$, because of the unbiasedness of Φ . Let A be the set of θ such that $|\theta^{(1)}| < \Delta$. Since $E_{\theta, \Sigma} \Phi(X, \hat{\Sigma})$ is unimodal in $\theta^{(1)}$, it is not smaller than α on A . On the other hand $E_{\theta, \Sigma} \Phi(X, \hat{\Sigma})$ should be no larger than α on $A \cap H_0$ because of its unbiasedness. This implies that $E_{\theta, \Sigma} \Phi(X, \hat{\Sigma})$ is constant α on $A \cap H_0$, which contains an open set of θ . Therefore, by analyticity, $E_{\theta, \Sigma} \Phi(X, \hat{\Sigma}) = \alpha$ for all θ and Σ . We conclude that $\Phi(X, \hat{\Sigma}) = \alpha$ almost surely by completeness, and the proof of the theorem is complete.

REFERENCES

- ANDERSON, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*, 2nd ed. New York: Wiley.
- BERGER, R. L. (1982). Multiparameter hypothesis testing and acceptance sampling. *Technometrics* **24**, 295–300.
- BERGER, R. L. & HSU, J. (1996). Bioequivalence trials, intersection-union tests, and equivalence confidence sets (with Discussion). *Statist. Sci.* **11**, 283–319.
- BROWN, L. D., HWANG, J. T. G. & MUNK, A. (1997). An unbiased test for the bioequivalence problem. *Ann. Statist.* **6**, 2345–67.
- CASELLA, G. & BERGER, R. L. (1990). *Statistical Inference*. Belmont, CA: Duxbury.
- CHOW, S. C. & LIU, J. P. (1992). *Design and Analysis of Bioavailability and Bioequivalence Studies*. New York: Marcel Dekker.
- FDA (1992). *The FDA Guidance on Statistical Procedures for Bioequivalence Studies using a Standard Two-Treatment Crossover Design*. Rockville, MD: Division of Bioequivalence, Office of Generic Drugs, Center for Drug Evaluation and Research, U.S. Food and Drug Administration.
- HAUCK, W. W., HYSLOP, T., ANDERSON, S., BOIS, F. Y. & TOZER, T. N. (1995). Statistical and regulatory consideration for multiple measures in bioequivalence testing. *Clin. Res. Regulat. Affairs* **12**, 249–65.
- HSU, J. C., HWANG, J. T. G., LIU, H. K. & RUBERG, S. J. (1994). Confidence intervals associated with tests for bioequivalence. *Biometrika* **81**, 103–14.
- SCHUIRMANN, D. J. (1987). A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. *J. Pharmacokin. Biopharm.* **15**, 657–80.
- WESTLAKE, W. J. (1988). Bioavailability and bioequivalence of pharmaceutical formulations. In *Biopharmaceutical Statistics for Drug Development*, Ed. K. E. Peace, pp. 329–52. New York: Marcel Dekker.

[Received April 1997. Revised June 1998]