*Sociedad Española de Estadística e Investigación Operativa*

# Test

**Volume 11, Number 1. June 2002**

# Optimal confidence sets for testing average bioequivalence

**Yu-Ling Tseng**
*Department of Applied Math*
*Dong Hwa University, Taiwan*

*Sociedad de Estadística e Investigación Operativa*
*Test* (2002) *Vol.* **11**, *No.* 1, pp. 127–141

*Sociedad de Estadística e Investigación Operativa*
*Test* (2002) *Vol.* **11**, *No.* 1, pp. 127–141

# Optimal confidence sets for testing average bioequivalence

**Yu-Ling Tseng***
*Department of Applied Math*
*Dong Hwa University, Taiwan*

**Abstract**

Under the multivariate normal setup, the bioequivalence problem is studied, using the confidence approach. First a confidence set which has smallest expected effective length at the origin is proposed. For the known variance case, its induced test can be shown to have constant level $\alpha$. Also, it is unbiased and uniformly most powerful among equivariant tests. For the unknown variance case, an approximated confidence set is proposed. The induced test enjoys similar good properties. Simulation shows that our test substantially outperforms some existing tests, in general.

**Key Words:**Average bioequivalence, Confidence set, Effective length, Multivariate bioequivalence testing, Optimal bioequivalence test.

**AMS subject classification:** 62F03, 62F25.

## 1 Introduction

Assessment of bioequivalence is important in drug development. In statistical terms, the problem of bioequivalence is to decide if the difference of two parameters $\theta = \mu - \mu_0$ is close to zero. Formally, it is to test

$$H_0 \ : \theta \notin \Delta \quad \text{vs} \quad H_1 \ : \theta \in \Delta, \tag{1.1}$$

where $\Delta$ is a (small) set containing the origin. Often the set $\Delta$ is of the type $\{\theta \in R^p \ : \ ||\theta|| \leq \delta\}$ for some (small) $\delta > 0$. See Metzler (1972), Westlake (1972, 1976, 1979) and Anderson and Hauck (1983) for a detailed discussion.

*Correspondence to: Yu-Ling Tseng, Department of Applied Math Dong Hwa University, Hualien 974, Taiwan. Email: yltseng@mail.ndhu.edu.tw

This research was supported in part by NSC, Taiwan, grant No. 84-2121-M-001-031.

Received: May 2000; Accepted: October 2001

Regarding the formulation, existing results have been focusing on the univariate case, i.e., $p = 1$. For example, Anderson and Hauck (1983), Patel and Gupta (1984), Rocke (1984), Schuirmann (1987), Hsu and Ruberg (1992), Hwang and Liu (1992), Hsu et al. (1994), Brown et al. (1995) and Berger and Hsu (1996) have proposed tests for this problem in univariate setup. Also see Bofinger et al. (1993) and Giani and Finner (1991) for other related results. Besides the theoretical generality, a multivariate formulation of the problem is also of interest in practice. For example, Chow and Liu (1992) address the issue of design of study, cf. their page 290 for details. Despite of this, very few results for multivariate formulations are available in the literature in contrast to those for univariate formulations. The multivariate bioequivalence problem is certainly an important yet rather understudied terrain for applied and theoretical statisticians. Recently, Wang et al. (1999) advanced along this direction. The present study is another new and interesting addition.

The confidence approach is a popular way to construct tests for the bioequivalence problem. It generalizes that in Westlake (1972, 1976). In this approach, $H_0$ is rejected and bioequivalence declared if $C(X) \subset \Delta$ where $C(X)$ is a confidence set for $\theta$ and the distribution of the observation vector $X$ depends on $\theta$. This approach is justified by Theorem 2.3 in Hsu et al. (1994) which applies to univariate and multivariate setups. In short, if the confidence level of $C(X)$ is at least $1-\alpha$ then the size of the induced test is at most $\alpha$.

In this article, the confidence approach under the multivariate normal setup is considered, i.e., $X$ follows a $N_p(\theta, \sigma^2 I)$ distribution and $\sigma^2$ is either known or can be estimated by an independent estimator $S^2$ such that $mS^2/\sigma^2$ has a $\chi^2_m$ distribution. In terms of testing bioequivalence, the effective length of the confidence set for $\theta$ is a crucial quantity. The effective length of a confidence set $C(X)$, denoted by $ef\ell(C(X))$, is defined as twice the supremum of the distances of points in $C(X)$ from the origin;

$$ef\ell(C(X)) \equiv 2 \sup_{y \in C(X)} ||y||, \tag{1.2}$$

where $||y||$ is the usual Euclidean norm of $y$. Specifically, a straightforward generalization of Theorem 2.4 in Hsu et al. (1994) to the multivariate setup shows that confidence set with shorter effective length induces more powerful test through the confidence approach.

In light of this, we aim at finding confidence sets with shortest effective length. These sets in turn induce better bioequivalence tests. In Section 2 we consider the case when $\sigma^2$ is known. The proposed confidence set is shown to minimize the expected effective length at the origin. Its associated test has level exactly $\alpha$, is unbiased, and is the uniformly most powerful (UMP) for (1.1) among tests based on $||X||$ (a reasonable reduced class of tests). The comparison between our test and that proposed in Brown et al. (1995) is provided. It is noticed that although their confidence set minimizes the expected volume at the origin, its associated test may not be unbiased for some $\Delta$.

We propose an approximate confidence set for the unknown $\sigma^2$ case in Section 3. By simulation, its coverage probabilities and the power comparison among its induced test, the two one-sided test proposed by Schuirmann (1987) and the test induced by $R^I$ in Wang et al. (1999) are provided. The results suggest that our test based on the approximate confidence set generally substantially outperform those of Schuirmann (1987) and Wang et al. (1999).

## 2 Known Variance

Since $\sigma$ is known, without loss of generality we assume that $\sigma = 1$. Let $\lambda = ||\theta||^2$. The propose confidence set is

$$C^*(X) = \left\{\theta : ||X||^2 \geq c(\lambda)\right\}, \tag{2.1}$$

where $c(\lambda)$ is the $\alpha$ upper percentile of a non-central chi-squared distribution, denoted by $\chi^2_p(\lambda)$, with $p$ degrees of freedom and noncentrality parameter $\lambda$; that is $P_\theta\left\{||X||^2 \geq c(\lambda)\right\} = 1 - \alpha$, for all $\theta$.

Note that

$$C^*(X) \subset \Delta = \{\theta : ||\theta|| \leq \delta\} \iff ||X||^2 \leq c(\delta^2).$$

Hence, for testing (1.1) the associated test of $C^*(X)$ has rejection region $\{x : ||x||^2 \leq c(\delta^2)\}$, and, by the definition of $c(\delta^2)$, power function satisfying $P_\theta\left(||X||^2 \leq c(\delta^2)\right) \geq \alpha$, for all $\theta$ with $||\theta|| \leq \delta$; which means that the test is unbiased. Note that when $\Delta = \{\theta : ||\theta|| \leq \delta\}$, problem (1.1) is invariant under $G = \{g : g \text{ is an } p \times p \text{ orthogonal matrix}\}$ and $||X||$ is a maximal invariant statistic under $G$. Intuitively, a reasonable procedure should make

it harder to declare bioequivalence as the observed $||x||$ increases and this means that we should focus on invariant tests under $G$, i.e., those depending on $X$ only through $||X||$. With this reduction, problem (1.1) becomes that of testing

$$H_0 \ : \lambda > \delta^2 \quad \text{vs} \quad H_1 \ : \lambda \leq \delta^2, \tag{2.2}$$

from an observation $Y = ||X||^2$ following a $\chi^2_p(\lambda)$ distribution ($\lambda = ||\theta||^2$). As a consequence of the monotone likelihood ratio property of the non-central $\chi^2$ distribution, the UMP size $\alpha$ test for (2.2) has rejection region $\{y : y \leq c(\delta^2)\}$. Therefore, the associated test of $C^*(X)$ is the UMP invariant size $\alpha$ test for (1.1).

It is interesting to notice from the above mentioned properties of the associated test of $C^*(X)$ that by choosing an appropriate confidence set, the confidence approach to testing bioequivalence can lead to the known optimal solution of testing standard hypotheses (2.2). As mentioned previously, however, for testing (1.1) confidence sets with shorter effective length induce more powerful bioequivalence tests. In the following theorem, as in the works of Brown et al. (1995) and Tseng and Brown (1997), the Ghosh-Pratt identity (Ghosh (1961), Pratt (1961)) is utilized to establish the optimality of $C^*(X)$ in terms of effective length.

**Theorem 2.1.** *Let $p \geq 1$. Among all $1-\alpha$ confidence sets whose effective length depends on $X$ only through $||X||$, $C^*(X)$ minimizes the expected effective length at the origin.*

*Proof.* Let $C(X)$ be any $1-\alpha$ confidence set for $\theta$ and $K(X) = K(||X||) = ef\ell(C(X))$. Note for all $\theta$, $P_\theta[\theta \in C'(X)] \geq P_\theta[\theta \in C(X)] \geq 1-\alpha$, where $C'(X) = \left\{\theta : ||\theta|| \leq \frac{1}{2}K(||X||)\right\}$. Also, $ef\ell(C'(x)) = ef\ell(C(x))$ for all $x$. It suffices to consider, therefore, only $1-\alpha$ confidence sets of the form $C(X) = \{\theta : ||\theta|| \leq k(||X||)\}$, for some function $k(||X||)$.
Note that by the Ghosh-Pratt identity

$$\begin{aligned}
E_0\{ef\ell(C(X))\} &= E_0\{2k(||X||)\} \\
&= 2\int_{R^p}\int_{I\{||\theta||:||\theta||\leq k(||x||)\}} d||\theta||(\tfrac{1}{2\pi})^{p/2}e^{-\frac{1}{2}||x||^2}dx \\
&= 2\int_0^\infty \int_{I\{x:||\theta||\leq k(||x||)\}} (\tfrac{1}{2\pi})^{p/2}e^{-\frac{1}{2}||x||^2}dx\,d||\theta|| \\
&= 2\int P_0\{||\theta|| \leq k(||X||)\}\, d||\theta||
\end{aligned}$$

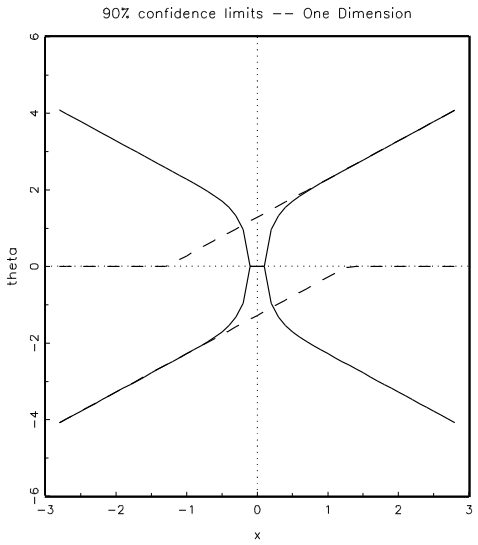


*Figure 1*: *Comparison of $C^*(x)$ (solid lines) with $C^{bch}(x)$ (dashed lines).*

$$\begin{aligned} &\geq 2\int P_0\left\{||X||^2 \geq c(||\theta||^2)\right\} d||\theta|| \\ &= E_0\left\{ef\ell(C^*(X))\right\} \end{aligned}$$

where the inequality is implied by the Neyman-Pearson fundamental lemma and the monotone likelihood property of the non-central chi-squared distribution family. □

As a result, the associated tests are more powerful than the related tests in the literature. For example, in comparing to the usual confidence set $C^0(X)$ and $C^{bch}(X)$, proposed in Brown et al. (1995), which minimizes the expected volume at the origin among all $1-\alpha$ confidence sets for $\theta$ (see Brown et al. (1995) for details), we have

**Corollary 2.1.** *For $p \geq 1$, $C^*(X)$ induces more powerful bioequivalence test than both $C^0(X)$ and $C^{bch}(X)$.*

*Proof.* It is easy to see that $ef\ell(C^{bch}(x)) = 2(||x|| + z_\alpha)$ and $ef\ell(C^0(x)) = 2(||x|| + k)$, where $k^2$ is the $1-\alpha$ cutoff point of a chi-square distribution with $p$ degrees of freedom. The proof is complete since as tests for (1.1), they are both based on $||X||$. □

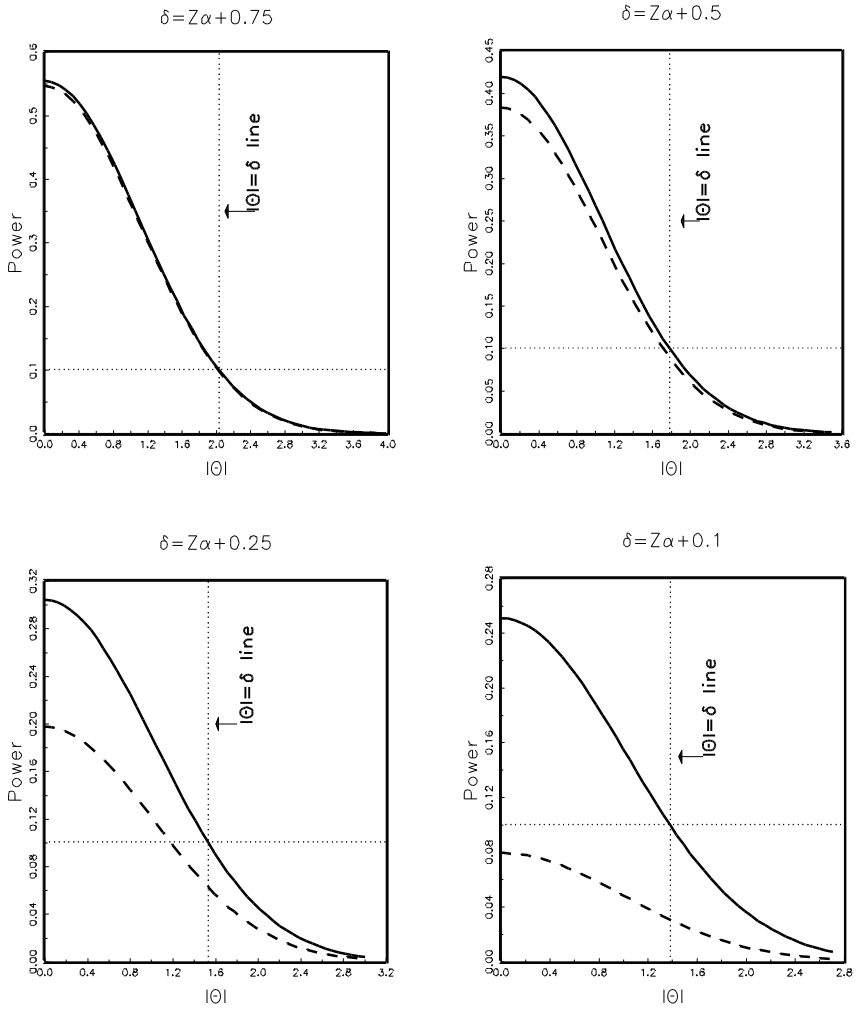


*Figure 2*: *Power Functions of $C^*(x)$ (solid lines) and $C^{bch}(x)$ (dashed lines) for testing bioequivalence when $p = 1$, $\alpha = 0.1$ and $z_\alpha = 1.28$ with values of $\delta$ as indicated. The horizontal dotted line denotes Power $= \alpha$ line.*

Figure 1 compares the one-dimensional 90% confidence interval $C^*(X)$ with $C^{bch}(X)$ which, when $p = 1$, is

$$C^{bch}(X) = \{\theta \,:\, min(0, X - z_\alpha) \leq \theta \leq max(0, X + z_\alpha)\}\,.$$

For an observation $x_0$, the limits of $C^*(x_0)$ ($C^{bch}(x_0)$) are given by the

intersection of the solid lines (dashed lines) and the line $x = x_0$. It is obvious from the graph that for very small values of $x$, $C^*(x)$ is narrower than $C^{bch}(x)$. For larger values of $x$, however, it is not surprising to see that $C^*(x)$ is wider. Nevertheless, we can see from the picture that for most of the $x$ values $ef\ell(C^*(x)) \leq ef\ell(C^{bch}(x))$. That is $C^*(x)$ is better than $C^{bch}(x)$ in terms of effective length. In fact, Corollary 2.1 proves that bioequivalence test induced by $C^*(X)$ is more powerful than that associated with $C^{bch}(X)$. In Figure 2, the power functions of the associated tests of $C^*(X)$ and $C^{bch}(X)$ when $p = 1$, $\alpha = 0.1$ and $\Delta = [-\delta, \delta]$ are graphed for various values of $\delta$. Note that the power function corresponding to $C^{bch}(X)$ is $P_\theta(|X| + z_\alpha \leq \delta)$, and that corresponding to $C^*(X)$ is $P_\theta\left(|X|^2 \leq c(\delta^2)\right)$. It is noticed that the improvement can be substantial for small $|\theta|$ and, especially, for small $\delta$. In fact, when $\delta \leq z_\alpha$, $C^{bch}(X)$ will never be entirely contained in $\Delta$ and, hence, in this case the test using $C^{bch}(X)$ has power zero for all $\theta$. As a result, using $C^{bch}(X)$ does not even give an unbiased bioequivalence test. While, as we noted previously, using $C^*(X)$ always induce a unbiased bioequivalence test, regardless of the value of $\delta$. One remark before we leave this section:

**Remark 2.1.** From Figure 1, we see that sometimes $C^*(x)$ is an empty set, an ill conditional property as a confidence set. However, a better confidence set *per se* is not our major concern. Instead, we aim at finding better tests for (1.1) induced by confidence sets through the confidence approach. The fact that $C^*(x)$ is an empty set when $||x||^2 < c(0)$ implies that the induced test will declare average bioequivalence whenever the observation is very near the origin. As we have seen, the induced test rejects $H_0$, i.e., declares bioequivalence, when $||x||^2 \leq c(\delta^2)$ which properly contain the region where $||x||^2 < c(0)$. Hence, the possibility that $C^*(x)$ being empty does not seems to have a detrimental effect in terms of test.

## 3 Unknown Variance

Here, we first consider in Section 3.1 the case when $\Delta$ is specified in terms of the parameter $\eta = \theta/\sigma$. In other words, drug products are declared to be bioequivalent if their mean differences after "normalizing" by the common variance lie within the pre specified set. There are many instances in practice, especially in engineering science, where the "signal-to-noise ratio", $\eta$, is of interest and for such cases this formulation is naturally

called for.

However, this is not what is currently suggested by the FDA (see FDA 1992 for details) for drug companies' conducting bioequivalence trials. Hence, in Section 3.2 we propose an approximate confidence set, generalized from (3.1), for the usual FDA formulation and present simulation results in supporting its use. However, it is our strong belief that the "signal-to-noise ratio" formulation should also be considered in the assessment of bioequivalence for drug products. Of course, different settings may lead to distinct optimal confidence sets.

### 3.1 Equivalence based on $\theta/\sigma$

Here, assume that $\Delta = \{\eta : ||\eta|| \leq \delta\}$. To test (1.1) through the confidence approach, our proposed confidence set for $\eta$ is

$$C^*(X, S^2) = \left\{\eta : \frac{||X||^2}{pS^2} \geq c^*(||\eta||^2)\right\} \tag{3.1}$$

where $c^*(||\eta||^2)$ is the $\alpha$ cutoff point of a noncentral $F$ distribution with degrees of freedom $p$ and $m$ and the noncentrality parameter $||\eta||^2$.

The following results are very similar to those in Section 2, hence their proofs are omitted.

**Theorem 3.1.** *Let $p \geq 1$. Suppose that $\Delta = \{\eta : ||\eta|| \leq \delta\}$ with a constant $\delta > 0$. Then, the test induced by $C^*(X, S^2)$ is uniformly most powerful for (1.1) among tests based on $||X||/S$, regardless of $\delta$.*

**Theorem 3.2.** *Among all $\eta$'s $1-\alpha$ confidence sets whose effective length depends on $X$, $S^2$ only through $||X||/S$ , $C^*(X, S^2)$ minimizes the expected effective length at $\theta = 0$ for every $\sigma^2 > 0$.*

**Remark 3.1.** Brown et al. (1995) also propose a $1-\alpha$ confidence interval for $\eta$ and their interval is proved to minimize the expected volume at $\theta = 0$ and any $\sigma > 0$. While, the set $C^*(X, S^2)$ presented here minimizes the expected effective length at $\theta = 0$ and any $\sigma > 0$. Since the effective length is the right measure for testing bioequivalence, the associated test of $C^*(X, S^2)$ is more powerful. Comparisons in power between these two tests are expected to be similar as those we have in Figure 2, hence are not provided.

### 3.2 Equivalence based on $\theta$

We now present the approximated confidence region for testing average bioequivalence under the FDA formulation. Simulations on its coverage probabilities and the power comparison between its associated test and that proposed by Schuirmann (1987) and $R^I$ in Wang et al. (1999) are provided, respectively. All results are based on 10000 simulations.

Modifying the set in Section 3.1, we propose an approximated confidence region, denoted by $C^a(X, S^2)$, for bioequivalence testing. Define

$$C^a(X, S^2) = \left\{ \theta : \frac{||X||^2}{pS^2} \geq c^*(||\theta||^2/S^2) \right\}.$$

Note that $C^a(X, S^2)$ is an approximated confidence set for $\theta$ since its coverage probability is now no longer a constant $1 - \alpha$.

Table 3.2 provides the simulated coverage probabilities of $C^a(X, S^2)$ when $p = 1$ with $\alpha = 0.1$ and 0.05. Note that the coverage probabilities are not very far off the nominal one $1 - \alpha$ except possibly for $|\theta|$ near the value $\delta$. We also notice that for small $\sigma$, $C^a(X, S^2)$ seems to have coverage probabilities larger than $1 - \alpha$.

Figure 3 gives the simulated power function of $C^a(X, S^2)$ and the numerically calculated power function of Schuirmann's test for $\delta = 1$, $m = 20$, $\alpha = 0.1$ or 0.05 and $\sigma = 0.3$, 0.55, or 0.8. Since the power functions of $C^a(X, S^2)$ are simulated curves, due to possible simulation errors solid lines in Figure 3 are not as smooth as those dashed lines which are calculated by numerical integrations. Nevertheless, it is clear from these power curves that $C^a(X, S^2)$ improves upon Schuirmann's two one-sided test. In particular, when $\sigma = 0.55$ or 0.8, for smaller values of $\theta$ we see a noticeable improvement in power of our test over the two one-sided tests procedure. This range of $\sigma$ is not unusual in testing bioequivalence, as best explained in Brown et al. (1997). We also notice that Schuirmann's test can have very poor power for large $\sigma$; while for $|\theta| \leq \delta$, $C^a(X, S^2)'s$ power seems to remain being above $\alpha$. We also did simulations corresponding to $\sigma < 0.3$. The powers of the two tests are quite similar in these cases, hence no figures are reported for $\sigma < 0.3$. Simulations we had for the case when $m = 10$ show similar comparison results, hence are not reported, either.

In Wang et al. (1999) Schuirmann's test is generalized for multivariate

bioequivalence. More precisely, for testing the hypotheses

$$H_0 : \max_{1\le i\le p} |\theta_i| \ge \Delta_0 \quad \text{vs} \quad H_1 : \max_{1\le i\le p} |\theta_i| < \Delta_0 \tag{3.2}$$

the generalized test has the rejection region

$$R^I = \bigcap_{i=1}^{p} R_i = \{|X_i| < \Delta_0 - t_m(\alpha)S, \text{ for all } i = 1, \ldots, p\},$$

where $t_m(\alpha)$ is the upper $\alpha$ quantile of Student's $t-$distribution with $m$ degrees of freedom. Note that the hypotheses (3.2) differ from (1.1) in that they have different regions for bioequivalence. Table 2 provides simulated powers at $\theta = 0$ of $C^a(X, S^2)$ and $R^I$ for cases where $\delta = \Delta_0$, $\delta = \sqrt{p}\Delta_0$, or volume of $\Delta = (2\Delta_0)^p$ volume of the region for bioequivalence in (3.2) with various values of $\Delta_0$, and $p = 2$, 3. It is quite evident that for most cases $C^a(X, S^2)$ has a higher power than $R^I$. In many cases the power improvements in Table 2 are larger or much larger than 0.10. Exceptions are the cases when $\delta = \Delta_0 = 1$ where $R^I$ has a higher power. But, when $\delta = \Delta_0 = 1$ the region for bioequivalence in (3.2) is much larger in size than that in (1.1), as a result, such way of power comparison is in favor of $R^I$. We also notice from Table 2 that $R^I$ is biased for several cases, while the power of $C^a(X, S^2)$ are all larger than $\alpha$. Based on our simulation, the associated test of $C^a(X, S^2)$ seems to be unbiased. If so, unlike $R^I$, its power function cannot be decreasing with respect to each $|\theta_i|$. Otherwise, it would be a trivial test as implied by Theorem 2 of Wang et al. Straightforward derivations show that the power function of our test is decreasing with respect to $||\theta||$, instead. Both tests associated with $C^a(X, S^2)$ and $R^I$ have maximum power at $\theta = 0$ – a desirable property. Seemingly, better unbiased tests for either (1.1) or (3.2) can be obtained when considering tests whose power is a decreasing function of $||\theta||$, rather than $|\theta_i|$ for all $i$.

As a closing note: Wang et al. (1999) compares the simulated power of the tests based on $R^I$ and $R^I_U$, a generalization based on Wang et al. (1999). With few exceptions in extreme cases, the powers are comparable. They hence conclude *"Drastically different methods have to be considered in order to improve upon $R^I$ substantially."* (cf. page 401, Wang et al.) Interestingly, our study suggests otherwise: Based on the confidence approach (still), the simulation shows our tests have substantially better power performance than theirs – maybe a suitable choice of the test statistic will just do the trick.

$\alpha = 0.1/\alpha = 0.05$

| | $m = 10$ | | | $m = 20$ | | |
|---|---|---|---|---|---|---|
| $\lvert\theta\rvert$ | $\sigma = 0.3$ | $\sigma = 0.55$ | $\sigma = 0.8$ | $\sigma = 0.3$ | $\sigma = 0.55$ | $\sigma = 0.8$ |
| 0 | .899/.950 | .903/.948 | .901/.947 | .900/.950 | .899/.946 | .902/.950 |
| 0.1 | .894/.945 | .905/.945 | .900/.950 | .900/.949 | .902/.949 | .897/.948 |
| 0.2 | .892/.949 | .903/.949 | .895/.948 | .900/.950 | .903/.945 | .904/.950 |
| 0.3 | .889/.946 | .898/.948 | .897/.948 | .897/.948 | .899/.946 | .899/.950 |
| 0.4 | .886/.939 | .896/.949 | .894/.948 | .894/.944 | .901/.948 | .899/.950 |
| 0.5 | .881/.935 | .899/.947 | .895/.947 | .895/.937 | .895/.948 | .899/.945 |
| 0.6 | .885/.936 | .885/.944 | .892/.946 | .893/.942 | .884/.949 | .893/.948 |
| 0.7 | .894/.937 | .878/.941 | .893/.947 | .896/.940 | .892/.948 | .895/.948 |
| 0.8 | .892/.937 | .876/.933 | .886/.942 | .895/.946 | .893/.946 | .899/.950 |
| 0.9 | .897/.942 | .877/.931 | .889/.943 | .905/.943 | .886/.946 | .889/.945 |
| 1 | .904/.947 | .885/.935 | .887/.937 | .903/.946 | .891/.940 | .893/.943 |
| 1.1 | .908/.948 | .882/.932 | .884/.937 | .902/.947 | .887/.940 | .888/.946 |
| 1.2 | .915/.958 | .891/.933 | .883/.936 | .907/.948 | .894/.943 | .892/.940 |
| 1.3 | .919/.960 | .885/.933 | .885/.932 | .909/.955 | .892/.939 | .893/.940 |
| 1.4 | .931/.963 | .892/.938 | .883/.932 | .919/.960 | .895/.946 | .892/.942 |
| 1.5 | .931/.971 | .893/.935 | .880/.931 | .920/.957 | .898/.943 | .886/.942 |
| 1.6 | .941/.972 | .898/.944 | .881/.930 | .923/.966 | .898/.940 | .889/.944 |
| 1.7 | .944/.977 | .901/.945 | .891/.931 | .927/.965 | .905/.942 | .894/.941 |
| 1.8 | .953/.980 | .903/.949 | .889/.938 | .931/.971 | .901/.947 | .896/.942 |
| 1.9 | .954/.980 | .906/.947 | .890/.937 | .937/.971 | .901/.949 | .896/.940 |

*Table 1*: *Simulated coverage probabilities of* $C^a(X, S^2)$

## Acknowledgments

The author is grateful to the referee and the associated editor for helpful remarks and excellent suggestions which led to much improvement.

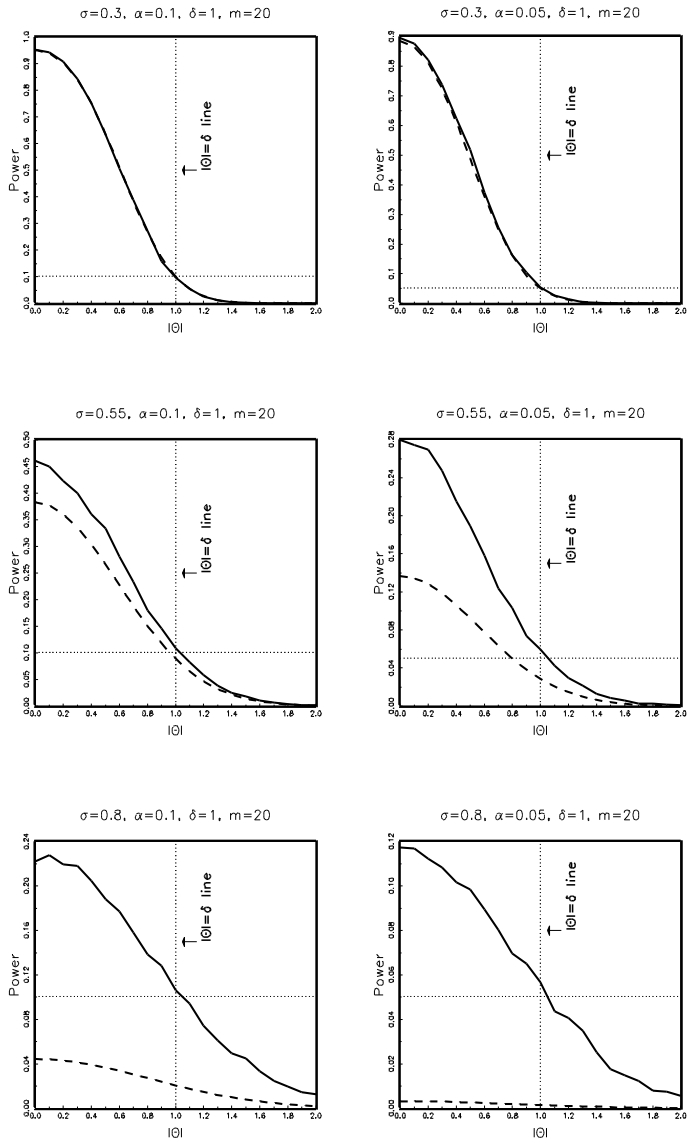


*Figure 3*: *Power Functions of* $C^a(X, S^2)$ *(solid line) and Schuirmann's test (dashed line) when* $m = 20$ *and* $\delta = 1$ *with* $\alpha = 0.1$ *or* $0.05$ *and various values of* $\sigma$.

| $\delta = \Delta_0 = \ln(1.25)$ | | | | | | |
|---|---|---|---|---|---|---|
| $p = 2$ | | | | | | |
| | $\alpha = 0.05$ | | | $\alpha = 0.1$ | | |
| $\sigma =$ | 0.2 | 0.4 | 0.6 | 0.2 | 0.4 | 0.6 |
| $C^a(X, S^2)$ | 0.0944 | 0.0564 | 0.0563 | 0.1863 | 0.1210 | 0.1022 |
| $R^I$ | 0.0911 | 0.0196 | 0.0119 | 0.1886 | 0.0530 | 0.0341 |
| $p = 3$ | | | | | | |
| | $\alpha = 0.05$ | | | $\alpha = 0.1$ | | |
| $\sigma =$ | 0.2 | 0.4 | 0.6 | 0.2 | 0.4 | 0.6 |
| $C^a(X, S^2)$ | 0.0872 | 0.0583 | 0.0513 | 0.1673 | 0.1131 | 0.1029 |
| $R^I$ | 0.0295 | 0.0040 | 0.0012 | 0.0848 | 0.0146 | 0.0068 |

| $\delta = \Delta_0 = 1$, $\alpha = 0.1$ | | | | | | |
|---|---|---|---|---|---|---|
| | $p = 2$ | | | $p = 3$ | | |
| $\sigma =$ | 0.2 | 0.4 | 0.6 | 0.2 | 0.4 | 0.6 |
| $C^a(X, S^2)$ | 0.9989 | 0.6670 | 0.3239 | 0.9975 | 0.5973 | 0.2861 |
| $R^I$ | 0.9994 | 0.7719 | 0.4001 | 0.9994 | 0.6845 | 0.2665 |

| $\delta = \sqrt{p}\Delta_0$, $\Delta_0 = 1$, $\alpha = 0.1$ | | | | | | |
|---|---|---|---|---|---|---|
| | $p = 2$ | | | $p = 3$ | | |
| $\sigma =$ | 0.2 | 0.4 | 0.6 | 0.2 | 0.4 | 0.6 |
| $C^a(X, S^2)$ | 1.0000 | 0.9423 | 0.6175 | 1.0000 | 0.9841 | 0.7438 |
| $R^I$ | 0.9996 | 0.7736 | 0.4040 | 0.9997 | 0.6852 | 0.2797 |

| Volume of $\Delta = (2\Delta_0)^p$, $\Delta_0 = 1$, $\alpha = 0.1$ | | | | | | |
|---|---|---|---|---|---|---|
| | $p = 2$ | | | $p = 3$ | | |
| $\sigma =$ | 0.2 | 0.4 | 0.6 | 0.2 | 0.4 | 0.6 |
| $C^a(X, S^2)$ | 0.9998 | 0.7802 | 0.4116 | 1.0000 | 0.8073 | 0.4222 |
| $R^I$ | 0.9995 | 0.7671 | 0.4143 | 0.9992 | 0.6840 | 0.2708 |

*Table 2*: *The simulated power at $\theta = 0$ of the test induced from $C^a(X, S^2)$ and the test $R^I$ in Wang et al. (1999) when $m = 20$, $p = 2,\ 3$, $\alpha = 0.05,\ 0.1$ and $\sigma = 0.2,\ 0.4,\ 0.6$*

## References

Anderson, S. and Hauck, W. W. (1983). A new procedure for testing equivalence in comparative bioavailability and other clinical trials. *Communications in Statististics Theory and Methods*, 12:2663–2692.

Berger, R. L. and Hsu, J. C. (1996). Bioequivalence trials, intersection-union tests and equivalence confidence sets. *Statistical Science*, 11(4):283–319.

Bofinger, E., Hayter, A. J., and Liu, W. (1993). The construction of upper confidence bounds on the range of several location parameters. *Journal of American Statistical Association*, 88:906–911.

Brown, L. D., Casella, G., and Hwang, J. T. G. (1995). Optimal confidence sets, bioequivalence, and the limacon of Pascal. *Journal of the American Statistical Association*, 90(431):880–889.

Brown, L. D., Hwang, J. T. G., and Munk, A. (1997). An unbiased test for the bioequivalence problem. *Annals of Statistics*, 25(6):2345–2367.

Chow, S. C. and Liu, J. P. (1992). *Design and Analysis of Bioavailability and Bioequivalence Studies*. Marcel Dekker, Inc., New York.

Ghosh, J. K. (1961). On the relation among shortest confidence intervals of different types. *Calcutta Statistical Association Bulletin*, pp. 147–152.

Giani, G. and Finner, H. (1991). Some general results on least favorable parameter configurations with special reference to equivalence testing and the range statistic. *Journal of Statistical Planning and Inference*, 28:33–47.

Hsu, J. C., Hwang, J. T. G., Liu, H. K., and Ruberg, S. J. (1994). Equivalence confidence intervals in bioequivalence trials. *Biometrika*, 81:103–114.

Hsu, J. C. and Ruberg, S. J. (1992). Equivalence confidence intervals in bioequivalence trials. Technical report, Ohio State University, Columbus, Ohio.

Hwang, J. T. G. and Liu, H. K. (1992). Symmetric confidence intervals associated with tests for bioequivalence. Technical report, Cornell University, Ithaca, New York.

Metzler, C. M. (1972). Bioavailability - a problem in equivalence. *Biometrics*, 30:309–317.

Patel, H. L. and Gupta, G. (1984). A problem of equivalence in clinical trials. *Biomedical Journal*, 26:471–474.

Pratt, J. W. (1961). Length of confidence intervals. *Journal of the American Statistical Association*, 56:541–567.

Rocke, D. M. (1984). On testing for bioequivalence. *Biometrics*, 40:225–230.

Schuirmann, D. J. (1987). A comparison of the two one-sided test procedure and the power approach for assessing the equivalence of average bioavailability. *Journal of Pharmokinetics and Biopharmaceutics*, 15:657–680.

Tseng, Y. L. and Brown, L. D. (1997). Good exact confidence sets for a multivariate normal mean. *Annals of Statistics*, 25(5):2228–2258.

Wang, W., Hwang, J. T. G., and Das Gupta, A. (1999). Statistical tests for multivariate bioequivalence. *Biometrika*, 86(2):395–402.

Westlake, W. J. (1972). Use of confidence intervals in analysis of comparative biovailability trials. *Journal of Pharmaceutical Science*, 61:1340–1341.

Westlake, W. J. (1976). Symmetrical confidence intervals for bioequivalence trials. *Biometrics*, 32:741–744.

Westlake, W. J. (1979). Statistical aspects of comparative bioavailability trials. *Biometrics*, 32:273–280.