

A Comparative Study of Assessing Average Bioequivalence in 2x2 Crossover Design with Missing Observations¹⁾

Sang-Gue Park²⁾ · Jiyun Choi³⁾

Abstract

A modified Anderson and Hauck(1983) test for analyzing a two-sequence two-period crossover design in bioequivalence trials is proposed when some observations at the second period are missing. It is based on the maximum likelihood estimators of average bioequivalence model and designed for handling missing at random(MAR) situation. The performance of the proposed test is compared to other tests using Monte Carlo simulations.

Keywords : Average bioequivalence, Incomplete crossover design, Maximum likelihood estimation, Missing at random

1. Introduction

The bioequivalence problem is of practical importance because the approval of most generic drugs requires the establishment of bioequivalence between the brand-name drug and the proposed generic version. Recently there has been great interest in the problem of demonstrating bioequivalence of two drugs, especially in the pharmaceutical industry. Two different drugs or formulations of the same drug are called bioequivalent if they are absorbed into the blood and become available at the drug action site at about the same rate and concentration. Bioequivalence is usually studied by administering dosage to subjects and

1) This research was supported by the Chung-Ang Research Grants in 2005.

2) First Author : Professor, Department of Statistics, Chung-Ang University, Seoul, 156-756, Korea
E-mail : spark@cau.ac.kr

3) Ph.D Candidate, Department of Statistics, Chung-Ang University, Seoul, 156-756, Korea
E-mail : jiyun1023@hotmail.com

measuring concentration of the drug in the blood just before and at set times after the administration. These data are then used to determine if the drugs are absorbed at the same rate.

These concentration by time measurements are connected with a polygonal curve and several variables are measured. The common measurements are AUC(area under curve), CMAX(maximum concentration) and TMAX(time until maximum concentration). The two drugs are bioequivalent if the population means of AUC and CMAX are sufficiently close. Descriptive statistics of TMAX are usually provided, but formal tests are not required.

The experimental design for assessing the bioequivalence between two drugs is routinely crossover design. Crossover design is the design in which some or all subjects receive more than one treatment in a specific order or sequence. The crossover is an efficient study design, because much or all of the estimated treatment effect is based on a within-subject comparison. A common crossover design is the two-treatment, two-sequence and two-period design in which each subject receives both treatments but in different orders depending on sequence assignment.

In a bioequivalence trial, it is not unusual to observe a patient either withdrawing from the trial for an administrative reason or being dropped from the trial by the investigator because of one or more protocol violations such as poor drug compliance, unauthorized use of concomitant medication, etc. Sometimes a patient is dropped because of adverse experiences or unsatisfactory therapeutic response. For these and other reasons, values at one or both periods of the crossover design are missing.

A simple way to analyze a missing data set from the two-sequence two-period crossover design is to exclude the data from subjects who do not receive both drugs. This, however, may result in bias and a substantial loss in efficiency as the missing rate increases.

Patel (1985, 1991) has suggested a maximum likelihood estimator treating monotone incomplete data in a clinical trial setting. But the assumption of the model is somewhat different to the setting of the usual bioequivalence model. Patel has assumed that the variance of the response variables are defined by periods, not by the drug, which means the variance of the reference drug in the first sequence is different from the variance of the same reference drug in the second sequence. Kenward and Molenberghs(1998) also pointed out that Patel's method could not be used in missing at random situation because his precision estimator is based on the use of missing completely at random framework. Some modifications of Patel's procedure for assessing average bioequivalence are required when some observations of subjects are missing at random and recently, Lee et al.(2005) proposed a modified two one-sided test, which originally given by Schuirmann(1987), by the maximum likelihood estimation of the average bioequivalence model.

Anderson and Hauck(1983) proposed a statistical method for assessing the average bioequivalence between two drugs. It has been known that Anderson and Hack test could correct the poor power problem of the best-known Schuirmann's test in bioequivalence trials when the variabilities of drugs are relatively large. Since the recent tendency of bioequivalence assessment between drugs focuses to the robustness of drugs' variabilities, Anderson and Hauck test would be preferred over the well-known Schuirmann's test, especially in case of small samples with large variabilities. The motivation of this study is to extend Anderson and Hauck test for assessing the average bioequivalence when some of the data during the trial are missing at random.

In section 2, we present the maximum likelihood estimators of the average bioequivalence model under missing at random and discuss their statistical issues. A modified Anderson and Hauck(1983) test procedure for assessing the average bioequivalence in case of missing data is proposed in section 3 and an illustrative example using proposed method is presented in section 4. Finally in section 5, we present a simulation study of showing the efficiency of the proposed method.

2. Model and Maximum Likelihood Estimator

Two-sequence and two-period crossover design is hired to evaluate the average bioequivalence between two drugs. Suppose that n_k subjects in each sequence k ($k = 1, 2$) receive pre-assigned drug at each period j ($j = 1, 2$) and some of subjects drop the trial at the second period due to various reasons. The m_k subjects have no missing observations on both periods among n_k subjects in each sequence k and the $n_k - m_k$ subjects in each sequence have observations in period 1 only. If y_{ijk} stands for the response of the i th subject in sequence k at period j ($i = 1, \dots, n_k$; $j = 1, 2$; $k = 1, 2$), then $y_{m_k+1, 2k}, \dots, y_{n_k, 2k}$ in the second period of sequence k are missing. Let Y be an $m \times 2$ matrix representing complete pairs of both periods, and \underline{y}_1^* , an $(n - m) \times 1$ vector representing the period 1 only, be given as

$$Y = \begin{pmatrix} \mathcal{Y}_{11} & \mathcal{Y}_{21} \\ \mathcal{Y}_{12} & \mathcal{Y}_{22} \end{pmatrix}, \quad \underline{y}_1^* = \begin{pmatrix} \mathcal{Y}_{11}^* \\ \mathcal{Y}_{12}^* \end{pmatrix},$$

where $n = n_1 + n_2$, $m = m_1 + m_2$, $\mathcal{Y}'_{jk} = (y_{1jk}, \dots, y_{m_j, jk})$ and $\mathcal{Y}'_{1k} = (y_{m_k+1, 1k}, \dots, y_{n_k, 1k})$ for $j, k = 1, 2$. We assume that $(y_{i1k}, y_{i2k}) \sim IN((\mu_{1k}, \mu_{2k}), \Lambda)$ for $i = 1, \dots, m_k$ and $y_{i1k} \sim IN(\mu_{1k}, \lambda_{11})$ for

$i = m_k + 1, \dots, n_k$, where IN denotes independent normality and

$$A = \begin{pmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{12} & \lambda_{11} \end{pmatrix}.$$

The missing data mechanism for a response related dropout would be dependent on the data already observed, but not on the potential response measurements that could have been obtained subsequent to the time of dropout. Little and Rubin (2002) quote this kind of missing mechanism missing at random and we consider the likelihood based inference under this situation with assumption that the parameter of missing data process is distinct from the parameter of interest.

It is customary to employ lognormal models in bioequivalence studies of AUC and CMAX. For the log transformed y_{ijk} , μ_{jk} can be usually assumed as follows (Chow and Liu, 2000),

$$\mu_{jk} = \mu + P_j + F_{(j,k)} + C_{(j-1,k)}, \quad (j = 1, 2; k = 1, 2). \quad (1)$$

In model equation (1), P_j represents the period effect, $F_{(j,k)}$ the treatment effect, $C_{(j-1,k)}$ the carryover effect, and we can assume the following,

$$\begin{aligned} F_{(1,1)} = F_{(2,2)} = F_R; \quad F_{(2,1)} = F_{(1,2)} = F_T, \\ C_{(0,1)} = C_{(0,2)} = 0, \quad C_{(1,1)} = C_R, \quad C_{(1,2)} = C_T. \end{aligned}$$

Let

$$\begin{aligned} \bar{y}_{jk}^{(c)} &= \sum_{i=1}^c y_{ijk} / c, \quad (c = n_k \text{ or } m_k), \\ s_{uv} &= \sum_{k=1}^2 \sum_{i=1}^{m_k} (y_{iuk} - \bar{y}_{uk}^{(m_k)})(y_{ivk} - \bar{y}_{vk}^{(m_k)}) / m, \quad (u, v = 1, 2). \end{aligned}$$

The maximum likelihood estimators of the parameters are obtained from the likelihood function given by Lee et al(2005);

$$\hat{\mu}_{1k} = \bar{y}_{1k}^{(n_k)}, \quad \hat{\mu}_{2k} = \bar{y}_{2k}^{(m_k)} + \hat{\rho} \left(\hat{\mu}_{1k} - \bar{y}_{1k}^{(m_k)} \right), \quad (k = 1, 2).$$

The intraclass correlation ρ can be estimated by solving the following equation

$$[ms_{11} - (n-2)\hat{\lambda}_a]\hat{\rho}^3 + (n-m)s_{12}\hat{\rho}^2 + [(n-2)\hat{\lambda}_a - ns_{22} - (n+m)s_{11}]\hat{\rho} + (n+m)s_{12} = 0, \quad (2)$$

where

$$\hat{\lambda}_a = \sum_{k=1}^2 \sum_{i=1}^{n_k} (y_{i1k} - \bar{y}_{1k}^{(n_k)})^2 / (n-2).$$

The cubic equation (2) has exactly one root in $[-1, 1]$ having the same sign as s_{12} . This real root is the unique maximum likelihood estimator of ρ (Dahiya and Korwar, 1980). We can solve the cubic by the Newton-Raphson method with $\hat{\rho}_M = 2s_{12}/(s_{11} + s_{22})$ as the starting value (Morrison, 1973). And the estimator of the variance λ_{11} is

$$\hat{\lambda}_{11} = \frac{n\hat{\lambda}_a + m\hat{\lambda}_b/(1 - \hat{\rho}^2)}{n + m}$$

where

$$\hat{\lambda}_b = \sum_{k=1}^2 \sum_{i=1}^{m_k} \{ (y_{i2k} - \hat{\mu}_{2k}) - \hat{\rho}(y_{i1k} - \hat{\mu}_{1k}) \}^2 / (m - 2)$$

and

$$\hat{\lambda}_{12} = \hat{\rho} \hat{\lambda}_{11}.$$

The symmetric observed information matrix that relate to $\mu_{1k}, \mu_{2k}, \lambda_{11}, \rho$ ($k = 1, 2$) is :

$$\hat{i}_{Ok} = - \begin{bmatrix} \frac{n_k}{\hat{\lambda}_{11}} + \frac{m_k \hat{\rho}^2}{\hat{\lambda}_{11}(1 - \hat{\rho}^2)} & -\frac{m_k \hat{\rho}}{\hat{\lambda}_{11}(1 - \hat{\rho}^2)} & 0 & i_o(\hat{\mu}_{1k}, \hat{\rho}) \\ \frac{m_k}{\hat{\lambda}_{11}(1 - \hat{\rho}^2)} & 0 & 0 & i_o(\hat{\mu}_{2k}, \hat{\rho}) \\ & & -\frac{(n+m)}{2\hat{\lambda}_{11}^2} & i_o(\hat{\lambda}_{11}, \hat{\rho}) \\ & & & i_o(\hat{\rho}) \end{bmatrix},$$

where

$$i_o(\hat{\mu}_{1k}, \hat{\rho}) = -\frac{\hat{\rho} \sum_{i=1}^{m_k} (\hat{\mu}_{1k} - y_{i1k})}{\hat{\lambda}_{11}(1 - \hat{\rho}^2)},$$

$$i_o(\hat{\mu}_{2k}, \hat{\rho}) = \frac{\sum_{i=1}^{m_k} (\hat{\mu}_{1k} - y_{i1k})}{\hat{\lambda}_{11}(1 - \hat{\rho}^2)},$$

$$i_o(\hat{\lambda}_{11}, \hat{\rho}) = \frac{A}{\hat{\lambda}_{11}^2(1 - \hat{\rho}^2)} + \frac{(m-2)\hat{\rho}\hat{\lambda}_b}{\hat{\lambda}_{11}^2(1 - \hat{\rho}^2)^2},$$

$$i_o(\hat{\rho}) = \frac{m(1 + \hat{\rho}^2)}{(1 - \hat{\rho}^2)^2} - \frac{\sum_{k=1}^2 \sum_{i=1}^{m_k} (\hat{\mu}_{1k} - y_{i1k})^2}{\hat{\lambda}_{11}(1 - \hat{\rho}^2)} - \frac{(m-2)(1 + 3\hat{\rho}^2)\hat{\lambda}_b}{\hat{\lambda}_{11}(1 - \hat{\rho}^2)^3} - \frac{4\hat{\rho}A}{\hat{\lambda}_{11}(1 - \hat{\rho}^2)^2},$$

and

$$A = \sum_{k=1}^2 \sum_{i=1}^{m_k} A_{ik}(\hat{\mu}_{1k} - y_{i1k}), \text{ with } A_{ik} = y_{i2k} - \hat{\mu}_{2k} + \hat{\rho}(\hat{\mu}_{1k} - y_{i1k}).$$

The asymptotic variance covariance matrix of $(\hat{\mu}_{1k}, \hat{\mu}_{2k})$ can be obtained by simply inverting the above 4×4 observed information matrix and then taking the upper-left 2×2 matrix. This can be expressed as follows,

$$\hat{T}_{Ok} = \{\hat{i}_{Ok}^{-1}\}_{(2 \times 2)}. \quad (3)$$

The equation (3) can not be expressed in a closed form, but it is easily calculated when the data are obtained.

3. Assessing Average Bioequivalence

Most popular statistical tests for assessing bioequivalence in practice are Schuirmann(1987)'s two one-sided test(TOST) and Anderson and Hauck(1983) test(AHT). Later on Berger and Hsu(1996) discussed the statistical methods for assessing average bioequivalence and proposed the intersection-union test(IUT) by pointing out the shortcomings of TOST and AHT. However, Hauck and Anderson(1996) commented that both the TOST and the AHT are still comparable to IUT in practical situation in assessing bioequivalence, because FDA usually requires the proper sample size to guarantee the preassigned power such as 80%.

The poor power of the TOST has been known when the variabilities increase and the AHT has been known slightly 'liberal', that is, the real level may exceed the nominal level, but it is powerful than the TOST. Frick(1987) calculated exactly the level and power of the AHT and showed that it does not overestimate the given nominal level much. The practical sample size, for example $n=24$, the discrepancy between the actual level and the nominal level is just less than 0.005. Therefore, even though most FDAs recommend that the TOST is a standard statistical method in assessing bioequivalence, AHT should be considered as an alternative method to complement the TOST with a proper sample size.

The interval hypotheses for average bioequivalence can be formulated as

$$\begin{aligned} H_0: |F_R - F_T| &\geq \theta \\ H_1: |F_R - F_T| &< \theta \end{aligned}$$

where F_T and F_R are drug effects and θ is some clinically meaningful limits. Currently, the United Food and Drug Administration (1992) recommends $\theta = -\log(0.8) = \log(1.25)$ with log-transformed response variables.

The AHT is based on the statistics

$$T_{AH} = \frac{D}{\widehat{\sigma}_D}, \quad (4)$$

where

$$D = \widehat{F}_R - \widehat{F}_T = \bar{d}_{.1} - \bar{d}_{.2}$$

with

$$d_{ik} = \frac{1}{2}(Y_{2k} - Y_{1k}), \quad \bar{d}_{.k} = \frac{1}{n_k} \sum_{i=1}^{n_k} d_{ik}, \quad (5)$$

$$\widehat{\sigma}_D = \widehat{\sigma}_d \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}, \quad \widehat{\sigma}_d = \frac{1}{n_1 + n_2 - 2} \sum_{k=1}^2 \sum_{i=1}^{n_k} (d_{ik} - \bar{d}_{.k})^2. \quad (6)$$

T_{AH} is distributed as non-centrally t with $df = n_1 + n_2 - 2$ degrees of freedom and non-centrality parameter $\delta = \frac{(F_R - F_T)}{\sigma_D}$.

When missing observations in the second period occur, we can modify AHT by substituting (5) and (6) in test statistics (4) with the following;

$$D = \widehat{F}_R - \widehat{F}_T = (\widehat{\mu}_{11} + \widehat{\mu}_{22} - \widehat{\mu}_{21} - \widehat{\mu}_{12})/2, \quad (7)$$

$$\widehat{\sigma}_D = \sqrt{\sum_{k=1}^2 (\widehat{\Gamma}_{11,Ok} - 2\widehat{\Gamma}_{12,Ok} + \widehat{\Gamma}_{22,Ok})/4}, \quad (8)$$

where $\widehat{\Gamma}_{11,Ok}$ and $\widehat{\Gamma}_{22,Ok}$ are the variance estimators and $\widehat{\Gamma}_{12,Ok}$ is the covariance estimator obtained from $\widehat{\Gamma}_{Ok}$ in equation (3).

The modified AHT rejects H_0 at level α and declares two drugs bioequivalent if

$$F_{m-2} \left(|t_{AH}| - \frac{\theta}{\widehat{\sigma}_D} \right) - F_{m-2} \left(-|t_{AH}| - \frac{\theta}{\widehat{\sigma}_D} \right) \leq \alpha, \quad (9)$$

where t_{AH} is an observed value of T_{AH} obtained from the experiment and F_{m-2} is the cumulative t distribution with $df = m - 2$ degrees of freedom. It should be noted that Patel(1985) suggested to use $m - 2$ degrees of freedom for better approximation in this type of statistics.

Recently Lee et al.(2005) proposed an modified TOST when some of data are missing at random. The modified TOST is based on two test statistics

$$T_L = \frac{D + \theta}{\hat{\sigma}_D} \quad \text{and} \quad T_U = \frac{D - \theta}{\hat{\sigma}_D}. \quad (10)$$

The modified TOST rejects H_0 at level α and declares two drugs bioequivalent if both tests reject; that is, if

$$T_U < t_{\alpha, m-2} \quad \text{and} \quad T_L > -t_{\alpha, m-2}, \quad (11)$$

where $t_{\alpha, m-2}$ is the upper 100α percentile of a Student t distribution with $m-2$ degrees of freedom.

4. An Illustrative Example

A reanalysis of the 2×2 bioequivalence trial example given by Chow and Liu (2000, p.73) would help understand the motivation and usefulness offered by this paper. The original trial does not have any missing observation, so we delete the last two subjects of period 2 in both sequences for illustration purpose.

The simplest way to assess the average bioequivalence is to use the TOST with 20 complete pairs after deleting 4 incomplete pairs. This would definitely produce some amount of power loss and might lead to the incorrect conclusion. We present this result with our proposed method.

The treatment difference estimator and its variance covariance estimator are the followings ;

$$\begin{aligned} D &= - .005542, \\ \hat{\Gamma}_{o1} &= \begin{pmatrix} .003270 & .001769 \\ .001769 & .006560 \end{pmatrix}, \quad \hat{\Gamma}_{o2} = \begin{pmatrix} .003270 & .001769 \\ .001769 & .006561 \end{pmatrix}, \\ \hat{\sigma}_D &= .0561, \end{aligned}$$

with starting intraclass correlation value $\hat{\rho}_M = .496$.

$$\begin{aligned} &\text{Modified AHT p-value} \\ &: F_{18}(0.09879 - 3.98) - F_{18}(-0.09879 - 3.98) = 0.0002, \\ &\text{Modified TOST statistics} \\ &: T_U = -4.076, \quad T_L = 3.879. \end{aligned}$$

The modified AHT's p-value is compared to its nominal level 0.05 and modified TOST is compared to their values to 1.734, which is the upper 5% t-distribution percentile with degrees of freedom 18. From these results both tests conclude to be bioequivalent.

5. Simulation Results and Discussion

We conduct simulation studies to show how much the modified AHT and the modified TOST are effective compared to the current available methods in 2x2 crossover design. Without loss of generality we assume the zero period and carryover effect. We set the sample size of 12 at each sequence, which is best known sample size because some countries like India (Rani and Pargal, 2004) and Korea FDA(2002) require minimum sample size as 12 at each sequence. We also set 1 or 2 missing observations in the second periods at each sequence, because any bioequivalence study with more than 20% of the missing values at period 2 would be of little value in practice.

Bivariate lognormal random variates are generated in the following way;

Step 1 : Generate 12 observations from each of the pair of independent unit normal (U, V) in each sequence.

Step 2 : Obtain the bivariate normal (Y_1, Y_2) through the relationship

$$Y_1 = \sigma U + F_R, Y_2 = \sigma \rho U + \sigma \sqrt{1 - \rho^2} V + F_T.$$

Step 3 : Set $Y_1^* = \exp(Y_1)$ and $Y_2^* = \exp(Y_2)$. Then (Y_1^*, Y_2^*) has a bivariate lognormal distribution with correlation coefficient $\rho^* = [\exp(\rho \sigma^2) - 1] / [\exp(\sigma^2) - 1]$.

To make the mean and the coefficient of variation of the reference drug as nearly 100 and 0.2 in the original scale, respectively, we set $F_R = 4.6$, $\sigma = 0.2$ and $F_T = F_R + \Delta$, where Δ is ranged from -.25 to +.25. We also set intraclass correlation ρ between reference and test drug as 0.2, 0.5 and 0.8. Random numbers are generated in SAS IML program.

The simulation results are based on 10,000 repetitions on each of the following 54 categories:

- category 1: $n_1 = n_2 = 12$,
 $(m_1, m_2) = (10, 10), (11, 10)$.
- category 2: $\rho = .2, .5, .8$.
- category 3: $\Delta = -.25, -.2231, -.2, -.1, 0, .1, .2, .2231, .25$

Table 1 and 2 show simulation results for testing $H_0: |F_R - F_T| \geq \theta$ against $H_1: |F_R - F_T| < \theta$, with 5% of the significance level under various combinations of category 2 and 3. Table 1 represents the simulation results with 2 missing observations in each sequence, Table 2 shows with 1 missing observation in sequence 1 and 2 missing observations in sequence 2. Each entry in the table

denotes the average number of rejections of H_0 per 10,000 repetitions. We include 4 tests; i.e., "TOST after deletion (TOST)", "Anderson and Hauck test after deletion (AHT)", "modified TOST (MTOST)" and "modified AHT (MAHT)" in the table.

The simulation results show that the MTOST and the MAHT are reasonably well maintaining the given nominal level 0.05, which is the case of $\Delta = \pm .2231$ compared to other tests like TOST and AHT. We also notice that the MAHT consistently shows higher significance level than the MTOST as we expected. This indicates that this simulation results show the consistency of results in other simulation studies(Chow and Liu, 2000).

In the case of bioequivalence such as $-.2 \leq \Delta \leq +.2$, both two tables consistently show that the MAHT and the MTOST performs better than the usual TOST and AHT. And the power of the MAHT is the highest among others. As intraclass correlation ρ increases, the all the tests' rates of declaring bioequivalence also increase. In the case of bioinequivalence such as $\Delta = \pm .25$, both two tables consistently show that the MAHT and the MTOST performs better than the usual TOST and AHT as well.

From all these observations the proposed MAHT is worth to hire for analysis of the bioequivalence when some of the observations are missing at random. Even though KFDA and USFDA recommend to use the TOST for assessing bioequivalence, the MTOST looks reasonable to use with a proper sample size. Therefore, when the MTOST fails to prove the bioequivalence between two drugs or drugs show high variabilities, the MAHT would be a good candidate for assessing bioequivalence since KFDA and USFDA encourage to use more powerful statistical method as long as it is backed up with reasonable scientific results.

One might point out the limitation of simulation study, because it is based on the sample size of 12 with one or two missing observations in each sequence. But the simulation results show quite a consistency with other simulation study like Chow and Liu(2000) and it would be inferred that the proposed method MAHT should perform better than other TOST and AHT in other sample size case as well.

In a bioequivalence trial, it is not unusual to observe that a subject could not participate in the second period because of some adverse experiences or protocol violations. Some countries' FDAs do not specify how to handle such cases or recommend that the statistical analysis after just deleting the cases would be conducted. This paper discusses and proposes the statistical method of handling missing observations occurred at random in bioequivalence testing and confirms the validity through the simulation study. This research should be continued in various ways for setting up the guidance for testing average bioequivalence and need further study for testing population or individual bioequivalence.

<Table 1> Empirical bioequivalence testing results
($n_1 = n_2 = 12, m_1 = m_2 = 10$)

ρ	Δ	TOST	AHT	MTOST	MAHT
0.2	-0.25	117	118	111	113
	-0.2231	506	509	480	481
	-0.2	1314	1319	1331	1371
	-0.1	9040	9042	9208	9331
	0	9998	9998	9998	9999
	0.1	9069	9071	9213	9312
	0.2	1418	1425	1468	1499
	0.2231	527	531	511	512
	0.25	98	99	89	90
0.5	-0.25	64	64	57	57
	-0.2231	552	552	544	549
	-0.2	1814	1820	1821	1840
	-0.1	9847	9850	9877	9900
	0	10000	10000	10000	10000
	0.1	9835	9835	9857	9899
	0.2	1775	1781	1803	1811
	0.2231	487	492	499	500
	0.25	73	73	67	70
0.8	-0.25	8	8	11	10
	-0.2231	472	478	487	499
	-0.2	3019	3039	3080	3121
	-0.1	10000	10000	10000	10000
	0	10000	10000	10000	10000
	0.1	10000	10000	10000	10000
	0.2	3019	3038	3103	3219
	0.2231	520	524	525	524
	0.25	18	19	19	20

<Table 2> Empirical bioequivalence testing results
 ($n_1 = n_2 = 12$, $m_1 = 11$, $m_2 = 10$)

ρ	Δ	TOST	AHT	MTOST	MAHT
0.2	-0.25	88	88	91	92
	-0.2231	514	516	518	520
	-0.2	1421	1426	1429	1501
	-0.1	9203	9210	9326	9330
	0	9999	9999	10000	10000
	0.1	9206	9209	9306	9399
	0.2	1401	1406	1430	1470
	0.2231	491	494	472	488
	0.25	103	103	101	111
0.5	-0.25	71	71	65	66
	-0.2231	551	552	547	550
	-0.2	1809	1818	1845	1899
	-0.1	9849	9849	9870	9890
	0	10000	10000	10000	10000
	0.1	9886	9887	9896	9913
	0.2	1901	1912	1914	1935
	0.2231	476	479	475	488
	0.25	72	72	64	65
0.8	-0.25	17	17	18	19
	-0.2231	504	508	506	508
	-0.2	3131	3146	3169	3198
	-0.1	10000	10000	10000	10000
	0	10000	10000	10000	10000
	0.1	10000	10000	10000	10000
	0.2	3142	3162	3190	3200
	0.2231	502	506	505	500
	0.25	19	19	19	20

References

1. Anderson, S. and Hauck, W.W. (1983). A new procedure for testing equivalence in comparative bioavailability and other clinical trials. *Communications in Statistics*, A12, 2663-2692.
2. Berger, R.L. and Hsu, J.C.(1996). Bioequivalence trials, intersection-union tests and equivalence confidence sets. *Statistical Science*, 11, 283-319.
3. Chow, S. C., Liu, J. P. (2000). *Design and Analysis of Bioavailability and Bioequivalence Studies*. Marcel Dekker, New York.
4. Dahiya, R. C., Korwar, R. M. (1980), Maximum likelihood estimates for a

- bivariate normal distribution with missing data, *The Annals of Statistics* 8:687–692.
5. Food and Drug Administration (1992). *Guidance on Statistical Procedures for Bioequivalence Using a Standard Two-treatment Crossover Design*, Division of Bioequivalence, Office of Generic Drugs, Center for Drug Evaluation and Research.
 6. Hauck, W.W. and Anderson, S.(1996). Bioequivalence trials, intersection-union tests and equivalence confidence sets. *Comment. Statistical Science*, 11, 303.
 7. Kenward, M. G., Molenberghs, G. (1998). Likelihood based frequentist inference when data are missing at random, *Statistical Science* 13:236–247.
 8. Korea Food and Drug Administration (2002). *KFDA Publications 2002-61*.
 9. Lee, J., Kim, B. and Park, S.(2005). Average bioequivalence for two sequence two period crossover design with incomplete data. *Journal of Biopharmaceutical Statistics* 15, 857–867.
 10. Little, R. J., Rubin, D. B. (2002). *Statistical Analysis with Missing Data*, Wiley, New York.
 11. Morrison, D. F. (1973). A test for equality of means of correlated variates with missing data on one response, *Biometrika* 60:101–105.
 12. Patel, H. I. (1985). Analysis of incomplete data in a two-period crossover design with reference to clinical trials. *Biometrika* 72:411–418.
 13. Patel, H. I. (1991). Analysis of incomplete data from a clinical trial with repeated measurements. *Biometrika* 78:609–619.
 14. Rani, S. and Pargal, A. (2004). Bioequivalence : An overview of statistical concepts. *Indian Journal of Pharmacology*. 36, 209–216.
 15. 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.

[recieved date : Jan. 2006, accepted date : Feb. 2006]