Comparison of Outlier Detection Methods in Crossover Design Bioequivalence Studies

A. Rasheed[1,*], T. Ahmad[2,#] and J.S. Siddiqi[3,#]

[1]*Department of Research, Dow University of Health Sciences, Karachi, Pakistan*

[2]*Pharma Professional Services, Karachi, Pakistan*

[3]*Department of Statistics, University of Karachi, Pakistan*

Abstract: The significance of bioequivalence (BE) studies is rising due to large scale production and utilization of generic products all over the world. The correct identification of outlying data in BE studies is substantial for deciding two products either bioequivalence or bioinequivalent. For the detection of outliers in BE studies with the crossover designs different methods have been suggested in the literature. In the present work, we compared three outlier detection tests; (i) the Likelihood distance (LD) test (ii) the estimated distance (ED) test and the principal component analysis (PCA) test. In this work, the PCA test has been first time compared with the LD and ED test. For the purpose of comparison, we used two-way and three-way BE crossover data sets on linear and logarithmic scales. During the course of work it was found interesting and note-worthy that the performances of the ED and PCA tests in the sense of outlier detection are better than the LD test and this performance persists even for the log-transformed data. The results of our simulation study also indicated that the performance of the ED test for outliers' identification is better than the other two tests.

Keywords: Bioequivalence, Outliers, Likelihood distance, Estimated distance, Principal, Component.

1. INTRODUCTION

1.1. Bioequivalence

The term that considered as the rate and extension in which an active molecule is absorbed and becomes available at the drug action site is known as bioavailability. When comparison of pharmacokinetic parameters related to bioavailability is made between two formulations, the phenomena referred as bioequivalence. One of the two drugs is considered as the reference. The major pharmacokinetic parameters employed for the bioavailability assessment are: Area under the curve (AUC), Peak concentration (C_{max}), Time for achieving peak concentration (T_{max}). Bioequivalent drugs are pharmaceutical equivalents that, when administrated in the similar molar dose, in identical conditions, does not reveal significant statistical differences concerning bioavailability.

The most broadly used measure of bioequivalence is average bioequivalence. Average bioequivalence depends on the comparison of difference between formulations in the sense of mean and including the fact that the distributions of selected pharmacokinetic parameters may differ between two formulations in other distributional characteristics. According to Food and Drug Administration [1] guidelines, two drugs or formulations are declared to be bioequivalent if the 90% confidence interval of the ratio of geometric mean of pharmacokinetic parameters such as, AUC and C_{max}, lies within the pre-specified range (80%, 125%).

Sometime in regulatory submission, for evaluation of different formulations, the detection of outliers becomes mandatory. Correct identification and treatment of outlying data in BE studies is substantial for deciding two products either bioequivalence or bioinequivalent.

During the analysis of bioequivalence the logarithmic transformation of pharmacokinetic parameters is recommended by the FDA [1] under the fact that logarithmic transformation makes the distribution much closer to the normal.

1.2. Outlying Observations in Bioequivalence Studies

In bioequivalence study, data may have some outlying (extremely large or small) observations. These outlying observations may have some profound effect on the conclusion of bioequivalence studies. These extremely large or small values may be the results from different mechanisms, such as: (1) product failure (coated tablet broken; single tablet with drug dosage) (2) Adverse event affecting drug absorption (3) Laboratory error/data transcription error (4) unusual reaction of a single subject to one of the formulations (so-called subject-by-formulation interaction). Mechanism 1 to 3 can be viewed as outliers due to product or process failure and mechanism 4 can be

*Address corresponding to this author at the Department of Research, Dow University of Health Sciences, Karachi, Pakistan; Tel: +92 321-2598763; E-mail: abdur.rasheed@duhs.edu.pk

#Co-Author E-mails: tasneem_ahmad2001@yahoo.com, jssdr123@yahoo.com

ISSN: 2223-3806 / E-ISSN: 1927-5951/13

viewed as outliers due to subject-by-formulation interaction [2].

1.3. Outlying Observations in Crossover Bioequivalence Studies

In a non-replicated crossover design comparing f formulations of a drug, the model is given as

$$y_{ijl} = \mu + S_i + P_l + F_j + e_{ijl} \quad i = 1, ..., n \quad j, l = 1, ... f \qquad (1)$$

Where y_{ijl} is the response of the ith subject in lth period for jth formulation, μ is the overall mean, P_l is the fixed effect of the lth period with $\sum P_l = 0$, F_j is the fixed effect for the jth formulation with $\sum F_j = 0$, S_i is the random effect for the ith subject, and e_{ijl} is the random error in observing y_{ijl} . It is assumed that the $\{S_i\}$ and $\{e_{ijl}\}$ are independently and normally distributed with means zero and variances σ_s^2 and σ_e^2 , respectively.

For the model (1) we can differentiate two types of outliers

1. *Single data point outlier*: Unusual subjects who reveal extremely low or high bioavailability relative to reference treatment, this type of outlier also termed as within-subject outliers.

2. *Subject outlier*: Unusual subjects who reveal an extreme bioavailability for both test and reference treatments, this type of outliers termed as between-subject outliers.

FDA [1] advocates that product or process failure and subject-by-formulation interaction are two reasons of outliers in crossover bioequivalence studies. These mechanisms for outlier detection, the product or process failure and the subject-by-formulation interaction can show clearly as a single data point (within-subject) outlier. In crossover design for two treatments with two periods and two sequences, it is not possible to separate outliers occurred due to product or process failure from the outliers occurred due to subject-by-formulation interaction [2]. The crossover design for three treatments with three periods and three sequences has also the similar issue. Consequently regulatory authorities do not allow the exclusion of outliers from the statistical analysis of 2x2 crossover bioequivalence studies only based on statistical criteria, but if a crossover BE data set contain outlying observation then it might be of interest to present the results with and without outlying observations.

Lund [3] has developed a method for outlier detection in the linear model, use of this method has also been suggested by the FDA but Chow and Liu [4] pointed out and proved that this method no longer appropriate for the crossover design due to correlated pharmacokinetic responses from the same subject.

For detection of outliers in bioequivalence studies, two procedures based on Cook's likelihood distance and the estimated distance, were proposed by Chow and Tse [5]. Enachescu and Enachescu [6] used principal components to introduce a test for outlier detection in bioequivalence studies with crossover design.

Ramsay and Elkum [7] compared different outlier detection methods proposed by Chow and Tse and Liu and Weng [5, 8] and with the help of the simulation study, he presented that the estimated distance test performed better than other tests.

Enachescu and Enachescu [6] has initially used principal components for the identification of outlying observations in crossover BE studies. In this work we first compared the outliers' identification test based on principal components with other two tests based on Cook's likelihood distance and the estimated distance. We performed these three tests on non-replicated 2X2 and 3X3 crossover BE data sets and observed the numbers of subjects were detected as outliers. This work was carried on the linear and logarithmic scale as recommended by the FDA [1]. The performance of these test were also observed through a simulation study.

2. THREE OUTLIER DETECTION TEST FOR CROSSOVER DESIGN

2.1. Principal Component Analysis (PCA)

The objective of PCA is to discover or to reduce the dimensionality of the data set and identify new meaningful underlying variables. [6] has mentioned that for normally distributed observation U_i / λ_i are independent $\chi^2_{1,j}$ variables, where λ_i is called Eigen value denotes the variance of the i-th principal component. He also Considered $\sum_{i=1}^{p} \lambda_i \chi^2_{1,j}$ the weighted sum of square distance to zero of the

projected data into principal factorial plane, with $E\left(\sum_{i=1}^{p}\lambda_i\chi^2_{1,j}\right)=\sum_{i=1}^{p}\lambda_i=p$ and $Var\left(\sum_{i=1}^{p}\lambda_i\chi^2_{1,j}\right)=2\sum_{i=1}^{p}\lambda_i^2$. Now the Observations with a square distance greater than m (the rule of 2σ) may be considered as outliers where $m=p+2\sqrt{2\sum_{i=1}^{3}\lambda_i^2}$.

2.2. Likelihood Distance (LD) Test

Chow and Tse [5] introduced the likelihood distance test for identifying outlier in a bioequivalence study where the null hypothesis assumes that there are no period and formulation effects. Now the model becomes

$$y_{ij}=\mu+S_i+e_{ij} \qquad i=1,...,n \quad j=1,...f \tag{2}$$

The parameters of interest are μ,σ_s^2, and σ_e^2. Let $\theta=(\theta_1\theta_2\theta_3)^t$, where $\theta_1=\mu$, $\theta_2=\sigma_e^2$, and $\theta_3=\sigma_e^2+f\sigma_s^2$. The maximum likelihood function is

$$L(\theta)=\frac{-Nf}{2}\log 2\pi-\frac{N}{2}\log(\theta_2\theta_3{}^{f-1})-\frac{1}{2\theta_3}\sum_{i=1}^{N}\sum_{j=1}^{f}(Y_{ij}-\theta_1)^2 \tag{3}$$
$$=-\frac{f}{2}\left(\frac{1}{\theta_2}-\frac{1}{\theta_3}\right)\sum_{i=1}^{N}(\bar{Y}-\theta_1)^2$$

The maximum likelihood estimators of the parameters are then

$$\theta_1=\bar{Y}=\frac{1}{nf}\sum_i\sum_j Y_{ij},\ \theta_2=\frac{1}{n(f-1)}\sum_i\sum_j(Y_{ij}-\bar{Y}_i)^2$$
$$\theta_3=\frac{f}{n}\sum_i\sum_j(\bar{Y}_i-\bar{Y})^2$$

The likelihood distance statistic for the ith subject is twice the difference between the model log likelihood evaluated by using estimates from all the subjects and from estimates obtained after deleting the ith subject.

$$LD_i(\theta)=2[L(\theta)-L_i(\theta_{-i})$$

Where θ_{-i} is the maximum likelihood estimator of θ obtained by deleting the ith subject from the data. Asymptotically, as $n\to\infty$, $LD_i(\theta)$ is distributed as a chi-square statistic with three degrees of freedom. The ith subject is called an outlier if $LD_i(\theta)>\chi^2(3)$

2.3. Estimates Distance (ED) Test

Estimated distance statistics depends on the difference in the parameter estimates arising from the deletion of ith observation, rather than on the difference in the log-likelihood. The estimated distance statistic is

Table 1: 3X3 BE Crossover Data on Linear and Logarithmic Scale

		Period I	Period II	Period III	Period I	Period II	Period III
Sequence	**Subject no.**	**AUC on linear scale**			**AUC on logarithmic scale**		
ACB	2	4.88	4.88	3.54	1.585145	1.585145	1.264127
ACB	3	6.77	6.98	5.99	1.912501	1.943049	1.790091
ACB	8	2.7	3.85	2.43	0.993252	1.348073	0.887891
ACB	11	4.76	5.86	2.95	1.560248	1.76815	1.081805
ACB	13	4.76	4.4	2.84	1.560248	1.481605	1.043804
ACB	14	3.27	3.76	2.9	1.18479	1.324419	1.064711
ACB	20	8.2	11.64	6.2	2.104134	2.454447	1.824549
ACB	22	5.61	4.58	4.25	1.724551	1.521699	1.446919
BAC	6	2.53	4.91	3.74	0.928219	1.591274	1.319086
BAC	7	3	4.81	1.95	1.098612	1.570697	0.667829
BAC	9	1.96	3.67	2.37	0.672944	1.300192	0.86289
BAC	15	6.11	6.25	5.66	1.809927	1.832581	1.733424
BAC	18	7.34	6.67	8.47	1.993339	1.89762	2.136531
BAC	19	4.27	3.17	3.15	1.451614	1.153732	1.147402
BAC	21	6.41	4.54	3.74	1.857859	1.512927	1.319086
CBA	10	3.56	5.34	3.11	1.269761	1.675226	1.134623
CBA	12	2.71	2.53	2.07	0.996949	0.928219	0.727549
CBA	16	6.21	3.49	4.79	1.826161	1.249902	1.56653
CBA	17	5.71	3.81	6.58	1.742219	1.337629	1.884035
CBA	23	4.05	4.66	4.88	1.398717	1.539015	1.585145
CBA	24	7.66	4.59	6.05	2.036012	1.52388	1.800058

$$ED_i(\theta) = f^2 = (\theta - \theta_{-i})^t \Sigma^{-1} (\theta - \theta_{-i})$$

Where Σ^{-1} is the maximum likelihood estimator of the variance matrix

$$\Sigma = \begin{bmatrix} \theta_3 / n & 0 & 0 \\ 0 & 2\theta^2_2 / (n-1) & 0 \\ 0 & 0 & 2\theta^2_3 \end{bmatrix}$$

[5] proved that, $ED_i(\theta)$ is asymptotically distributed as a chi-square variable with three degrees of freedom. Hence, the estimated distance test declares the ith subject as an outlier if $ED_i(\theta) > \chi^2(3)$.

3. APPLICATION AND RESULTS

3.1. 3x3 Crossover Design

In order to apply above defined test procedures for detecting outliers in BE studies, we found a BE data set on FDA website, where it is mentioned as data set 8. In that BE study there were 3 formulations (A, B and C), 3 periods and 3 sequences (ACB, BAC and CBA); it was a crossover design with an equal number of sequence, periods and formulations. 1st Sequence contained 8 subjects, 2nd sequence contained 7 and 3rd had only 6 subjects. The data set on linear and logarithmic scales are given in Table **1**. As only 21 subjects out of 24 subjects completed the study, now for the sake of our convenience we coded them as subject number from 1 to 21.

In Figure **1**, the likelihood and estimated distances for the LD and ED tests and the squared distances for observations in the PCA test are presented for the linear scale data. For the LD and ED tests, subject is to be considered as outlier if $LD_i(\theta) > \chi^2(3) = 7.81473$ and $ED_i(\theta) > \chi^2(3) = 7.81473$ respectively and for the PCA test, subject with the squared distance greater than threshold value ($m = p + 2\sqrt{2\sum_{i=1}^{3} \lambda_i^2}$) 9.7934464 is

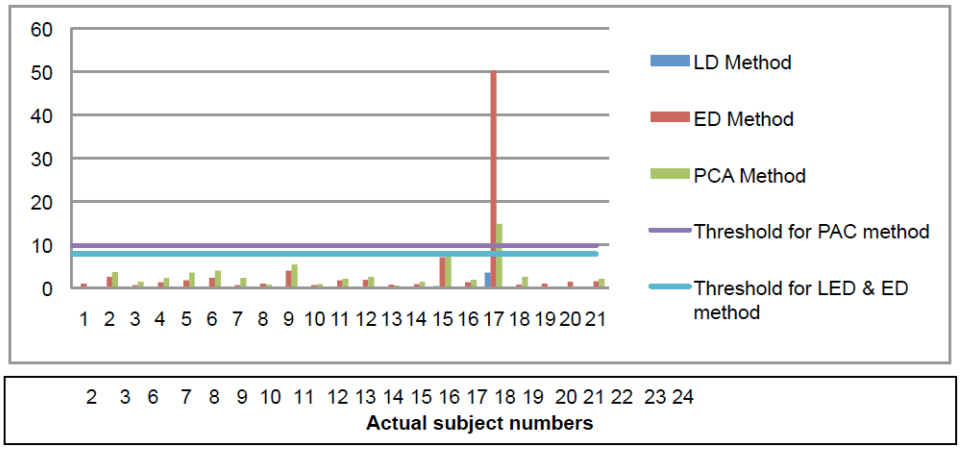

Figure 1: For 3x3 crossover BE data on linear scale, the likelihood and estimated distances for the LD and ED tests and the squared distances for observations in the PCA test respectively.

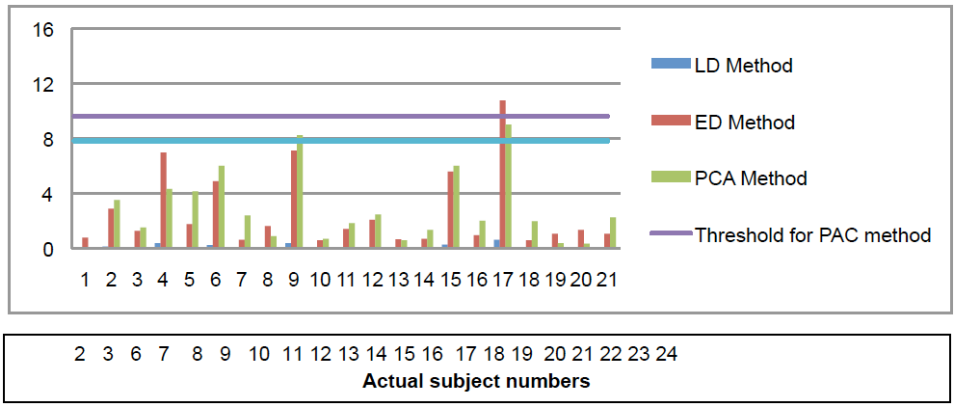

Figure 2: For 3x3 crossover BE data on the logarithmic scale, the likelihood and the estimated distances for the LD and ED tests and the squared distances for observations in the PCA test respectively.

Table 2: The Result of Outlier Detection from all Three Methods for 3x3 Crossover Data on the Linear and Logarithmic Scale

BE data on linear scale			BE data on logarithmic scale		
LD test	**ED test**	**PCA test**	**LD test**	**ED test**	**PCA test**
No outlier	Subj. 20 is outlier	Subj. 20 is outlier	No outlier	Subj. 20 is outlier	No outlier

considered as outlier. Figure **2** depicts the results on the logarithmic scale and for the PCA test threshold value was 9.5963537.

From above figures and table it is evident that none of the subject was detected as an outlier from the likelihood method, subject 20 was detected as an outlier from the estimated distance method on both scales whereas the principal component method detected it only on the linear scale.

3.2. 2x2 Crossover Design

We considered a BE data set from [4], this study was conducted with 24 healthy volunteers. The design of the BE study was 2x2 crossover i.e., there were two sequences (RT and TR) and two periods. Each of the two sequences contained 12 subjects. Five-50-mg tablets as test formulations or 5-mL of an oral suspension as a reference formulation were given to each of the subjects during each dosing period. The data set on linear and logarithmic scales are given in the Table **3**.

In Figure **3**, the likelihood and estimated distances for the LD and ED tests and the squared distances for observations in the PCA test are presented for the linear scale data. For the likelihood and estimated distance tests, the criterion for declaring a subject as an outlier is same as defined above and for the PCA test, subject with squared distance greater than threshold value ($m = p + 2\sqrt{2\sum_{i=1}^{3}\lambda_i^{\,2}}$) 6.7324548 is

Table 3: 2X2 BE Crossover Data on Linear and Logarithmic Scale

		Period I	**Period II**	**Period I**	**Period II**
Sequence	**Subject no.**	AUC on linear scale		AUC on logarithmic scale	
RT	1	74.675	73.675	4.313145	4.299664
RT	4	96.4	93.25	4.568506	4.535284
RT	5	101.95	102.125	4.624482	4.626198
RT	6	79.05	69.45	4.370081	4.240607
RT	11	79.05	69.025	4.370081	4.234469
RT	12	85.95	68.7	4.453766	4.229749
RT	15	69.725	59.425	4.244559	4.084715
RT	16	86.275	76.125	4.457540	4.332377
RT	19	112.675	114.875	4.724508	4.743845
RT	20	99.525	116.25	4.600409	4.755743
RT	23	89.425	64.175	4.493400	4.161614
RT	24	55.175	74.575	4.010510	4.311805
TR	2	74.825	37.35	4.315152	3.620333
TR	3	86.875	51.925	4.464470	3.949800
TR	7	81.675	72.175	4.402748	4.279094
TR	8	92.7	77.5	4.529368	4.350278
TR	9	50.45	71.875	3.920983	4.274928
TR	10	66.125	94.025	4.191547	4.543561
TR	13	122.45	124.975	4.807703	4.828114
TR	14	99.075	85.225	4.595877	4.445295
TR	17	86.35	95.925	4.458409	4.563567
TR	18	49.925	67.1	3.910522	4.206184
TR	21	42.7	59.425	3.754199	4.084715
TR	22	91.725	114.05	4.518795	4.736637

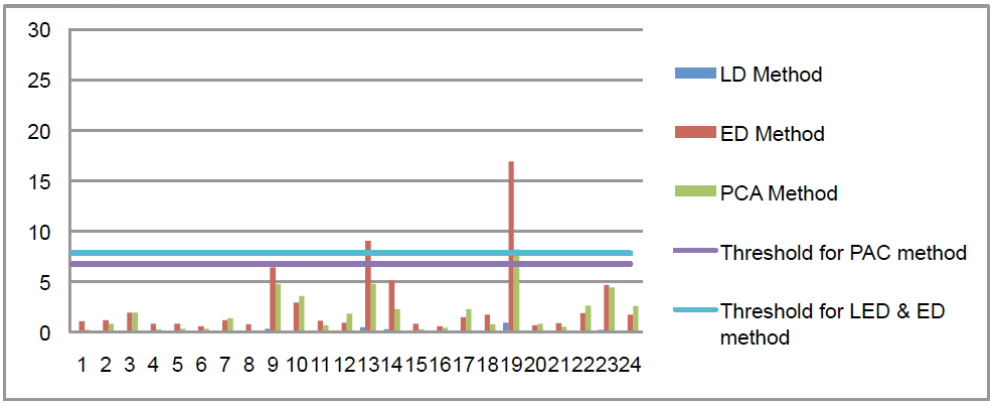

Figure 3: For 2x2 crossover BE data on the linear scale, the likelihood and the estimated distances for the LD and ED tests and the squared distances for observations in the PCA test respectively.

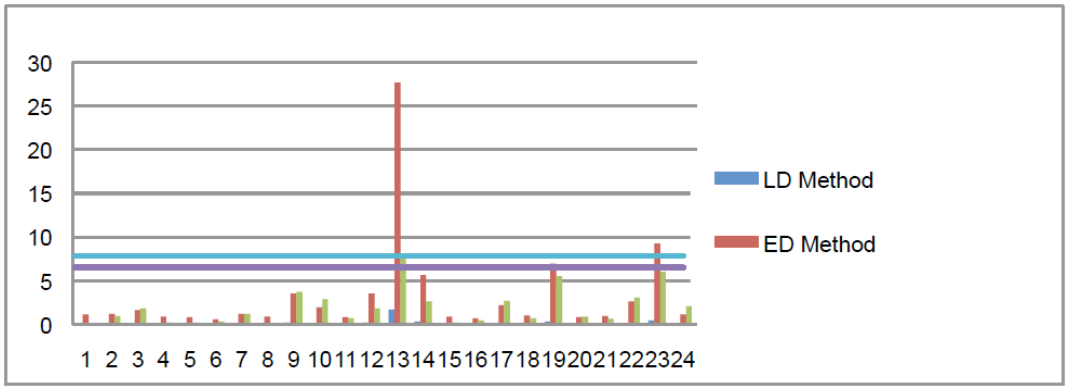

Figure 4: For 2x2 crossover BE data on the logarithmic scale, the likelihood and estimated distances for the LD and ED tests and the squared distances for observations in the PCA test respectively.

considered as outlier. Figure **4** shows the results on the logarithmic scale. For the PCA test threshold value was 6.5054357.

From above Figures **3**, **4** and Table **4** it is evident that none of the subject was detected as an outlier from the likelihood method, subjects 13 and 19 were detected as outliers from the estimated distance method on the linear scales and subjects 13 and 23 were detected as outlier on the logarithmic scale, moreover, on the logarithmic, subject 19 was not found as an outlier but it was very close to the threshold value. Whereas from the principal component analysis method subjects 19 and 13 were found as outliers on linear and logarithmic scales respectively. Subjects 19 and 23 were very close to the threshold value on the logarithmic but were not found as outliers.

4. SIMULATION STUDY

For each combination of sample size=18, 24 and intrasubject variabilities= 20, 40, a total of 100 data sets of AUC values were generated from the statistical model for 2X2 crossover design under the normality assumption. For the sake of our convenience we assumed that there were no period and carryover effects. The true test and reference means were both chosen to be 100. The results of simulations study are given in the Tables **5** and **6** on the linear and logarithmic scales respectively. From Tables **5** and **6** it is evident that the rate of outliers' identification of ED test is higher than LD and PCA tests, likewise the percentage of the simulation (in parenthesis) in which ED test rate was at least as good as each of the alternative is also high. For example in Table **5** for the linear scale, for sample size 24 and intrasubject

Table 4: The Result of Outlier Detection from all Three Methods for 2x2 Crossover Data on the Linear and Logarithmic Scale

BE data on linear scale			BE data on logarithmic scale		
LD test	**ED test**	**PCA test**	**LD test**	**ED test**	**PCA test**
No outlier	Subj.13 and 19 are outliers	Subj. is 19 outlier	No outlier	Subjs. 13 and 23 are outlier	Subj. 13 is outlier

Table 5: The Simulations Percentage where the Estimated Distance Test Performed Better than other Tests on the Linear Scale, in the Sense of Detecting Outliers More Frequently, and the Percentage (in Parenthesis) in which it was at Least as Good

Linear scale			
Sample size	**Intrasubject variabilities**	**ED>LD (ED≥LD)**	**ED>PCA (ED≥PCA)**
24	20	97(100)	82(100)
18		94(100)	71(100)
24	40	95(100)	77(100)
18		97(100)	78(100)

Table 6: The Simulations Percentage where the Estimated Distance Test Performed Better than other Tests on the Logarithmic Scale, in the Sense of Detecting Outliers More Frequently, and the Percentage (in Parenthesis) in which it was at Least as Good

Log scale			
Sample size	**Intrasubject variabilities**	**ED>LD (ED≥LD)**	**ED>PCA (ED≥PCA)**
24	20	90(100)	62(99)
18		86(100)	61(98)
24	40	80(100)	61(93)
18		84(100)	63(99)

variabilities 20, the estimated distance test performed better than the likelihood distance test in 97% of the simulations and better than the PCA test in 82% of the simulations. Under the same condition the estimated distance test performed at least as well as the likelihood distance test in 100% of the simulations and at least as well as the PCA test in 100% of the simulations. In Table **6** for the logarithmic scale, for sample size 24 and intrasubject variabilities 20, the estimated distance test performed better than the likelihood distance test in 90% of the simulations and better than the PCA test in 62% of the simulations. Under the same condition the estimated distance test performed at least as well as the likelihood distance test in 100% of the simulations and at least as well as the PCA test in 99% of the simulations.

5. CONCLUSION

Findings of our work propose that the estimated distance test is superior to the likelihood distance and the principal component analysis test. The performance of the estimated distance test persists even for log-transformed 2x2 and 3x3 crossover BE data sets. The principal component analysis test has been first time compared in this research with any of the outlier detection test. It is obvious from above findings that the principal component analysis test is much superior to the likelihood distance test as likelihood distance test did not detect any subject as an outlier on any of the scales. Moreover, the results of our simulation study advocate that, the performance of outliers' identification of the estimated distance test is superior to the likelihood distance test and the PCA test. These simulations results also support the performance of the estimated distance test over the likelihood distance test and the principal component analysis test when tested on real crossover BE data sets.

Exclusion of the outlying observations in BE studies are allowed only when they are caused by the product or process failure; when the reason of outlying observation is subject-by-formulation interaction, the exclusion of such outlying observation may not be allowed. In 2x2 or 3x3 crossover BE it is not possible to separate outliers appeared due to the product or process failure from the outliers appeared due to the subject-by-formulation interaction, only basis of statistical criteria. Although it is substantial to present the statistical results of crossover BE study with and without outlier observations.

REFERENCES

[1] FDA. Guidance for industry on statistical approaches to establishing bioequivalence.Center for Drug Evaluation and

Research, Food and Drug Administration, Rockville, Maryland 2001.

[2] Schall R, Endrenyi L, Ring A. Residuals and outliers in replicate design crossover studies. J Biopharm Stat 2010; 20(4): 835-49. http://dx.doi.org/10.1080/10543401003618876

[3] Lund RE. Tables for an approximate test for outliers in linear models. Technometrics 1975; 17: 473-76. http://dx.doi.org/10.1080/00401706.1975.10489374

[4] Chow S-C, Liu J-P. Design and analysis of bioavailability and bioequivalence studies (3rd ed.). New York: Dekker 2009. http://dx.doi.org/10.1002/sim.4780090508

[5] Chow SC, Tse SK. Outlier detection in bioavailability/ bioequivalence studies. Stat Med 1990; 9(5): 549-58.

[6] Enachescu D, Enachescu C. A new approach for outlying records in bioequivalence trials. Paper presented at the The XIII International Conference "Applied Stochastic Models and Data Analysis" 2009.

[7] Ramsay T, Elkum N. A comparison of four different methods for outlier detection in bioequivalence studies. J Biopharm Stat 2005; 15(1): 43-52. http://dx.doi.org/10.1081/BIP-200040815

[8] Liu JP, Weng CS. Detection of outlying data in bioavailability/bioequivalence studies. Statist Med 1991; 10(9): 1375-89. http://dx.doi.org/10.1002/sim.4780100906

Received on 15-04-2013 Accepted on 28-04-2013 Published on 30-04-2013

DOI: http://dx.doi.org/10.6000/1927-5951.2013.03.02.7