Sains Malaysiana 45(3)(2016): 499–506

Comparison of Outlier Detection Methods in Standard 2 × 2 Crossover Design

(Perbandingan Kaedah Pengesanan Titik Terpencil dalam Piawai 2 × 2 Reka Bentuk Pindah Silang)

FONG PENG LIM, IBRAHIM MOHAMED*, NOORIZAM DAUD & SIEW LI GOH

ABSTRACT

This paper discussed methods for outlier detection in standard 2 × 2 *crossover studies. Two outlier detection procedures were carried out based on residuals. Under a simplified model of* 2 × 2 *crossover design, we present the classical calculation of studentized residual (*SR1*) and propose a new studentized residual using median absolute deviation (*SR2*) to identify possible outlying subjects. The performances of both procedures in detecting subject outliers were compared. We show via simulation that a proposed procedure using* SR2 *is more powerful than that using* SR1 *for outlier detection. As an illustration, these procedures were applied to two real data sets from studies of bioavailability and kinesiology, respectively.*

Keywords: Crossover design; median absolute deviation; outlier; residual

ABSTRAK

Kertas kerja ini membincangkan kaedah untuk mengesan titik terpencil dalam piawai 2 × 2 *kajian pindah silang. Dua prosedur mengesan titik terpencil telah dijalankan berdasarkan sisa. Berdasarkan satu model* 2 × 2 *kajian pindah silang yang diringkaskan, kami menunjukkan pengiraan studentized sisa klasik (*SR1*) dan mencadangkan satu studentized sisa baharu yang menggunakan sisihan mutlak median (*SR2*) untuk mengesan subjek terpencil yang mungkin wujud. Prestasi kedua-dua prosedur dalam pengesanan subjek terpencil dibandingkan. Melalui simulasi, kami mendapati prosedur menggunakan* SR2 *yang dicadangkan memberikan prestasi yang lebih baik daripada prosedur menggunakan* SR1 *untuk pengesanan titik terpencil. Sebagai ilustrasi, prosedur tersebut digunakan untuk dua set data kajian sebenar dalam bidang bioketersediaan dan kinesiologi.*

Kata kunci: Reka bentuk pindah silang; sisa; sisihan mutlak median; titik terpencil

INTRODUCTION

In the standard 2 × 2 crossover design, we assume that there are two different groups of subjects. Each group receives the two treatments in a different order and the trial is to last for two treatment periods, with the order of treatments reversed in the second period. A common problem in crossover trials is the occurrence of extremely large or small observations. These extraordinary observations are called outliers and they may influence the conclusion drawn from the data set.

Chow and Tse (1990) proposed two procedures based on Cook's likelihood distance and the estimated distance for the detection of outliers in bioequivalence studies, where the crossover design is widely applied in their statistical analyses. Liu and Weng (1991) carried out procedures based on Hotelling T^2 statistics and residuals for the same purpose. Wang and Chow (2003) presented a general test procedure based on a mean-shift model. Furthermore, Ramsay and Elkum (2005) compared different outlier detection methods proposed by Chow and Tse (1990) and Liu and Weng (1991) via simulation study. They concluded that the estimated distance test performs better than other tests. Most recently, Karasoy and Daghan (2012) applied these existing methods to a real data set from a real bioequivalence analysis to investigate outliers. In bioequivalence studies with crossover design, Enachescu and Enachescu (2009) initially used principal components for the identification of outliers. Meanwhile, Singh et al. (2014) provided details regarding a studentized residual test and the Lund test for identification of outlier subjects.

It is therefore important that methods of identifying outliers in 2 × 2 crossover design were developed for proper handling of the data in studies. These methods are usually graphical or numerical.

In this paper, we carry out outlier detection based on residuals. Under a simplified model of 2 × 2 crossover design, we present a classical calculation of studentized residual (*SR*1) and propose a new studentized residual using median absolute deviation (*SR*2). Suitable outlier tests can then be applied to the resulting sets of studentized residuals in order to detect the possible within-subject outliers in the study. In the following section, we briefly describe the concept of outlier detection in standard 2 × 2 crossover design. We shall denote the procedure using by *SR*1 and *P*(*SR*1) that using *SR*2 by *P*(*SR*2). The next two sections discuss and in detail. Simulation studies were then carried out to obtain the percentage of times that the outlier was identified correctly and to investigate the performance of the procedures. We will then apply the procedures

to two real data sets from studies of bioavailability and kinesiology, respectively.

Outlier Detection for the 2 × 2 Crossover Design

Let Y_{kij} be the response of the th subject in period j under the ith treatment, where $i, j = 1, 2$; n_i is the size of group with treatment i and $k = 1, 2, \ldots, n_i$. We also consider here the crossover model used by Chow and Tse (1990):

$$Y_{kij} = \mu + S_k + F_i + P_j + e_{kij}, \tag{1}$$

where is the overall mean, F_i the fixed effect of the th treatment with $\Sigma_i F_i = 0$, P_j, the fixed effect of the jth period with $\Sigma_j P_i = 0$, S_k, the random effect of the kth subject and e_{kij} the random error. The variance components $\{S_k\}$ and $\{e_{kij}\}$ were assumed to be independent and normally distributed with mean 0 and variances σ_s^2 and σ_e^2, respectively. According to Liu and Weng (1991), when no period effect is assumed, the model (1) can be reduced as follows:

$$\begin{aligned} Y_{ki} &= \mu + S_k + F_i + e_{ki} \\ &= \alpha_i + S_k + e_{ki}, \end{aligned} \tag{2}$$

where $\alpha_i = \mu + F_i$ and $k = 1, 2, \ldots, n_i$ for $i = 1, 2$.

For model (1), we can distinguish two types of outlier: the between-subject outlier and the within-subject outlier. As stated by Chow and Liu (2009), the between-subject outliers were the unusual subjects who had extreme bioavailability to both treatments. Occurrence of the between-subject outliers may indicate that the underlying genetic mechanism for metabolism may be different from subjects to subjects. On the other hand, the within-subject outliers are the unusual subjects who exhibit extremely high or low bioavailability relative to the reference treatment. In other words, the within-subject outliers show unusual reaction to one of the treatments. Regulatory authorities generally do not allow the exclusion of outliers from the statistical analysis of 2 × 2 crossover design based solely on statistical criteria. However, if such a data set does contain outliers, then it might be of interest to present the results of analysis with and without outliers.

Outlier Detection using *SR*1

For the calculation of *SR*1, we refer to model (2). The repeated measurements on each subject are assumed to be values of independent, normally distributed random variables with equal variances. The residual is then given by:

$$r_{ki} = \left(1 - \frac{1}{n_i}\right) Y_{ki} - \left(\frac{1}{n_i}\right)\left[\left(\sum_{t=1}^{n_i} Y_{ti}\right) - Y_{ki}\right], \tag{3}$$

for each i, respectively.

The r_{ki} were estimators of the random error e_{ki} in model (2). These r_{ki} were normally distributed with mean zero and variance:

$$V(r_{ki}) = \left(1 - \frac{1}{n_i}\right)\sigma_e^2. \tag{4}$$

It was noticed that the r_{ki} were approximately independent with equal variances. Refer to Liu and Weng (1991) for the details of the derivation of residuals and their variance. Thus, the studentized residuals were:

$$SR1 = \frac{r_{ki}}{\left(\hat{V}(r_{ki})\right)^{1/2}}, \tag{5}$$

where $\hat{V}(r_{ki})$ is the estimated value of $V(r_{ki})$ obtained by replacing σ_e^2 by the mean square of the within-subject residual. The were treated as standard normal variables.

According to Jones and Kenward (1989), the response values corresponding to unusually large *SR*1 were called outliers or discordant values. The larger the residual, the more discordant is the corresponding response. To identify the outliers, the tabulated critical values, created by Lund (1975) can be used to determine whether the largest |*SR*1| is significantly large at the 5% level.

Outlier Detection using *SR*2

When the distribution of the residuals was far from normal, r_{ki} was not a good estimator of e_{ki}. A frequently used robust estimate of scale is the median absolute deviation (MAD), which is more resilient to outliers than the standard deviation and was defined by:

$$MAD = median\{|r_{ki} - median(r_{ki})|\}. \tag{6}$$

When sampling from a normal distribution, however, MAD estimates $Z_{0.75}e_{ki}$ rather than e_{ki}, where $Z_{0.75}$ is the 0.75 lower-tail probability of the standard normal distribution. Typically, MAD was rescaled so that it estimated e_{ki} when sampling from a normal distribution to give:

$$MADN = \frac{MAD}{Z_{0.75}} = \frac{MAD}{0.6745}. \tag{7}$$

In this study, was used as the robust estimate of scale and therefore the scaled residuals *SR*2 were given by:

$$SR2 = \frac{r_{ki} - median(r_{ki})}{MADN}. \tag{8}$$

Responses was labeled as outliers when $|SR2| > D$ in Table 1, where D is the critical value of the largest |*SR*2| at the significance level of 5%.

As presented in Table 1, we apply the parametric bootstrap technique to construct a table of critical values for the largest |*SR*2| under model (2). For each sample of size 20, 60 and 100, we calculate the *SR*2 and determine the largest |*SR*2|. The procedure of finding the largest was repeated 1000 times to generate 1000 values of the largest |*SR*2|. They were then sorted in ascending order and the 90, 95, 97.5 and 99% percentiles were obtained. These

percentiles approximate the critical values for significance levels of 0.10, 0.05, 0.025 and 0.01, respectively.

Simulation Study

In this section, we compare the power of the $P(SR2)$ in testing for subject outliers with that of the $P(SR2)$ in a standard 2 × 2 crossover design. Random samples were generated under a two-sequence, two-period crossover model based on the procedure used in Luzar-Stiffler and Stiffler (2005). Random samples Y_{ki} are first generated based on the following formula:

$$Y_{ki} = \gamma(z_{k0} + z_{ki})\,\mu_i, \tag{9}$$

where z_{k0} and z_{ki} is the i.i.d. standard normal ($i = 1, 2$; $k = 1, 2, \ldots, n_i$). Note that z_{k0} and z_{ki} were used to account for the between- and within-subject variations, respectively. Without loss of generality, the mean of treatment 1 (μ_1) was set to 100, while the mean of treatment 2 (μ_2) was set to 60, 80, 90, 100, 110, 125. The ability of and in detecting outlier then can be observed when the difference between μ_1 and μ_2 increased. The values of the constant γ were chosen to be 0.5, 5, 10 and 15, so that the coefficient of the intra-subject variation for the treatment 1 were 0.5, 5, 10 and 15%, respectively.

For simplicity, in generating the random samples Y_{ki}, we assumed the values of n_i were equal. Three types of n_i, 10, 30 and 50, were considered in the simulation. Let the total sample size, $N = \Sigma\, n_i$. The corresponding values N of therefore take the values 20, 60 and 100, respectively. The first subject was made into an outlier by multiplying the responses Y_{11} and Y_{12} by a constant p which varies from 10 to 200%. The process was repeated 200 times and the power of performance was assessed by computing the percentage of times that the outlier was identified correctly.

In order to conduct the power studies for ($SR1$), the procedure were summarized as follows:- Step 1. Using S-plus, random samples of z_{k0} and z_{ki} are generated from the standard normal distribution, where $i = 1, 2$ and $k = 1, 2, \ldots, n_i$. Step 2. Calculate the sum of each pair z_{k0} and z_{ki}. Each value of sum then multiplied with the chosen constant γ. Step 3. Set the mean of treatment 1 (μ_1) to 100. Repeat the Step 1 - 2 and calculate Y_{ki} for different mean of treatment 2 (μ_2). Step 4. The first subject was made into an outlier by multiplying the responses Y_{11} and Y_{12} by a constant p. The contaminated sample of was then obtained. Step 5. Calculate the residual in model (3) using the contaminated sample of Y_{ki}. Step 6. Conduct the analysis of variance for the contaminated sample of Y_{ki}. Then, estimate the variance of r_{ki}, $V(r_{ki})$ in model (4), by replacing σ_e^2 with the within-subjects residual mean squares. Step 7. Calculate the studentized residuals $SR1$ in model (5) using the r_{ki} and $V(r_{ki})$ in Steps 4 and 5. Step 8. Steps 1-7 were repeated (200) times and the times that the outlier was identified correctly were recorded.

Step 9. Calculate the power, where

$$\text{Power} = \frac{\text{Number of Times that the outlier is identified correctly}}{200} \times 100\%.$$

Step 10. Steps 1-9 were repeated for different constant γ ($\gamma = 0.5, 5, 10, 15$), mean of treatment 2 ($\mu_2 = 60, 80, 90, 100, 110, 125$) and constant p ($p = 10, 30, 50, 130, 150, 200$) and Step 11. The complete procedure (Steps 1-10) was repeated for different sizes of group with treatment i ($n_i = 10, 30, 50$). The corresponding total sample sizes N will be 20, 60 and 100, respectively.

TABLE 1. Critical values of the largest at significance level of 5%

N	μ_2	γ			
		0.5	5	10	15
20	60	0.704	0.973	1.271	1.640
	80	0.734	1.271	1.781	1.724
	90	0.794	1.781	1.693	1.641
	100	1.408	1.408	1.408	1.408
	110	0.794	1.781	1.693	1.641
	125	0.722	1.152	1.705	1.767
60	60	0.718	1.111	1.585	1.983
	80	0.762	1.585	2.100	2.139
	90	0.849	2.100	2.146	2.121
	100	1.947	1.947	1.947	1.947
	110	0.849	2.100	2.146	2.121
	125	0.744	1.380	2.014	2.149
100	60	0.724	1.165	1.702	2.104
	80	0.773	1.702	2.266	2.351
	90	0.871	2.266	2.356	2.300
	100	2.167	2.167	2.167	2.167
	110	0.871	2.266	2.356	2.300
	125	0.753	1.482	2.137	2.318

In order to conduct the power studies for $P(SR2)$, we use the same contaminated sample of in the power studies for $P(SR1)$, which was obtained using Steps 1-4 in the procedure above. We then calculate the scaled residuals $SR2$, as given by model (8). We repeat this calculation (200) times and record the times that the outlier was identified correctly. With the same combination of constant γ, mean of treatment 2 (μ_2) and constant p, the power of $P(SR2)$ can then be obtained. The complete procedure was repeated for the same sizes of group with treatment in the power studies for $P(SR1)$.

Tables 2-4 show the percentages of correctly detecting the designated outlier for sample sizes of 20, 60 and 100, respectively. For all sample sizes considered, the percentages of detection for both $P(SR1)$ and $P(SR2)$ were almost 100% when $\gamma = 0.5$ or 5. However, when or 15, the performance of $P(SR2)$ was always better than that of $P(SR1)$ since it has higher percentages of detection. These results showed that $P(SR2)$ was obviously more powerful than $P(SR1)$ for detecting outliers in a standard 2×2 crossover design.

Numerical Example

As an illustration, both procedures stated earlier were applied to two real data sets from studies of bioavailability and kinesiology, respectively. In the first real data set, we consider the blood concentration-time curve (AUC) data from two erythromycin formulations in a bioavailability study published by Clayton and Leslie (1981). In this study, a standard 2×2 crossover experiment was conducted with 18 subjects to compare a new erythromycin formulation

TABLE 2. Percentage of correctly identifying the designated outlier for sample size of 20

μ_2	p	$SR1$ γ				$SR2$ γ			
	(%)	0.5	5	10	15	0.5	5	10	15
60	10	100	100	100	74	100	100	100	99.5
	30	100	100	96.5	64.5	100	100	100	95
	50	100	100	96.5	74	100	100	98	88.5
	130	100	100	99	90	100	100	98	92.5
	150	100	100	99.5	92	100	100	100	94
	200	100	100	100	95.5	100	100	100	99
80	10	100	100	100	86	100	100	100	100
	30	100	100	98.5	53	100	100	100	100
	50	100	100	75.5	29	100	100	99	96
	130	100	99	74.5	43	100	100	93	95
	150	100	100	87.5	67.5	100	100	98.5	98.5
	200	100	100	100	94	100	100	100	99.5
90	10	100	100	100	92	100	100	100	100
	30	100	100	99	57	100	100	100	100
	50	100	100	77	24.5	100	100	100	98.5
	130	100	96.5	53	28.5	100	99.5	97.5	96.5
	150	100	100	85.5	60.5	100	100	99	98.5
	200	100	100	100	93.5	100	100	100	100
100	10	100	100	100	96	100	100	100	100
	30	100	100	99.5	67.5	100	100	100	100
	50	100	100	82.5	24.5	100	100	100	98.5
	130	100	95.5	45.5	22	100	100	98.5	96
	150	100	100	85.5	58.5	100	100	100	98.5
	200	100	100	100	96	100	100	100	100
110	10	100	100	100	97.5	100	100	100	100
	30	100	100	100	76.5	100	100	100	100
	50	100	100	85.5	34	100	100	100	99
	130	100	98.5	56	34.5	100	100	98	96
	150	100	100	91.5	63.5	100	100	100	98
	200	100	100	100	98	100	100	100	100
125	10	100	100	100	98.5	100	100	100	100
	30	100	100	100	85.5	100	100	100	100
	50	100	100	92.5	54.5	100	100	100	94.5
	130	100	100	86	61	100	100	95	93.5
	150	100	100	97.5	76.5	100	100	100	98.5
	200	100	100	100	99	100	100	100	100

TABLE 3. Percentage of correctly identifying the designated outlier for sample size of 60

μ_2	p	$SR1$				$SR2$			
		γ				γ			
	(%)	0.5	5	10	15	0.5	5	10	15
60	10	100	100	100	76	100	100	100	99
	30	100	100	100	68	100	100	100	91.5
	50	100	100	98	74.5	100	100	96	88
	130	100	100	100	83.5	100	100	98	95
	150	100	100	99.5	88.5	100	100	99.5	95.5
	200	100	100	100	97.5	100	100	100	99
80	10	100	100	100	87	100	100	100	100
	30	100	100	99.5	33.5	100	100	100	99
	50	100	100	68.5	20	100	100	100	94
	130	100	100	71	37	100	100	96	94.5
	150	100	100	86	57.5	100	100	98.5	97
	200	100	100	100	92.5	100	100	100	100
90	10	100	100	100	94.5	100	100	100	100
	30	100	100	100	39.5	100	100	100	100
	50	100	100	64.5	11.5	100	100	100	97.5
	130	100	95	42.5	23.5	100	100	96.5	94
	150	100	100	83	50	100	100	99	98
	200	100	100	100	92.5	100	100	100	100
100	10	100	100	100	98.5	100	100	100	100
	30	100	100	100	47.5	100	100	100	100
	50	100	100	71.5	7.5	100	100	100	98
	130	100	96	33.5	16	100	100	98.5	96.5
	150	100	100	82	46.5	100	100	99	98.5
	200	100	100	100	96	100	100	100	99.5
110	10	100	100	100	100	100	100	100	100
	30	100	100	100	65.5	100	100	100	100
	50	100	100	87	15	100	100	100	98.5
	130	100	99	49.5	25	100	100	98	94
	150	100	100	89.5	54	100	100	100	99
	200	100	100	100	98	100	100	100	100
125	10	100	100	100	100	100	100	100	100
	30	100	100	100	86	100	100	100	100
	50	100	100	97.5	41.5	100	100	100	98
	130	100	100	88	53	100	100	96.5	95.5
	150	100	100	98	77.5	100	100	100	96.5
	200	100	100	100	99	100	100	100	100

(i.e. erythromycin stearate) with a reference formulation (i.e. erythromycin base). As no sequence identification of each subject was provided in Clayton and Leslie (1981), we adapt the order of periods given in Weiner (1989) and assign subject 1 through 9 to sequence 1 and the remaining subjects to sequence 2. Using both procedures $P(SR1)$ and $P(SR2)$, subject 7 in group 1 was identified as an outlying subject. The similar result was showed by either the two-sample Hotelling T^2 or likelihood distance as stated by Chow and Liu (2009).

In the second real data set, we consider 77 measurements of peak oxygen consumption or VO_2 peak (in mL/kg/min) recorded from a 6 min Astrand submaximal cycling exercise test conducted at least one week apart. Figures 1 and 2 indicate the scatter plots of VO_2 peak for periods 1 and 2, respectively. A point (34th measurement) seems to be far from the others at the bottom of Figure 1. We fit the full data to model (2) and proceed to both procedures $P(SR1)$ and $P(SR2)$ for detecting the possible outliers in the data.

From Figures 3 and 4, it shows that at each plot, the 34th measurement gives the largest values of $|SR1|$ and $|SR2|$. These largest studentized residuals were then compared to their corresponding tabulated critical values. All of them are greater than the critical values. Figures 3 to 4 shows a dramatic outlier in subject 34. In order to compare the ability of both procedures in detecting the possible outliers in the data, we proceed to test the subsequent largest values of $|SR1|$ and $|SR2|$ in the same manner. The sequential testing procedure stops when only their subsequent largest values were less than their corresponding tabulated critical values. The results indicate

TABLE 4. Percentage of correctly identifying the designated outlier for sample size of 100

μ_2	p	$SR1$ γ				$SR2$ γ			
	(%)	0.5	5	10	15	0.5	5	10	15
60	10	100	100	100	76	100	100	100	100
	30	100	100	99	73.5	100	100	100	95
	50	100	100	99	81	100	100	96.5	94.5
	130	100	100	100	88	100	100	97	96
	150	100	100	100	90	100	100	99.5	97
	200	100	100	100	96	100	100	100	100
80	10	100	100	100	92.5	100	100	100	100
	30	100	100	99.5	35.5	100	100	100	99.5
	50	100	100	70.5	25.5	100	100	100	95
	130	100	100	66.5	38	100	100	95.5	95.5
	150	100	100	84	54	100	100	98	97
	200	100	100	99.5	93	100	100	100	99
90	10	100	100	100	97.5	100	100	100	100
	30	100	100	99.5	32.5	100	100	100	100
	50	100	100	66	11	100	100	100	97
	130	100	95	40.5	26.5	100	99.5	96	96
	150	100	100	77	42	100	100	99	97.5
	200	100	100	100	92	100	100	100	99
100	10	100	100	100	98	100	100	100	100
	30	100	100	100	43	100	100	100	100
	50	100	100	71	9.5	100	100	100	98.5
	130	100	91.5	31.5	16.5	100	100	96.5	96
	150	100	100	74	44.5	100	100	99.5	97
	200	100	100	100	92.5	100	100	100	100
110	10	100	100	100	99.5	100	100	100	100
	30	100	100	100	63.5	100	100	100	100
	50	100	100	86	15	100	100	100	99
	130	100	98.5	50	25	100	100	98.5	97
	150	100	100	83	56	100	100	100	97.5
	200	100	100	100	96.5	100	100	100	100
125	10	100	100	100	100	100	100	100	100
	30	100	100	100	90.5	100	100	100	100
	50	100	100	98	44.5	100	100	100	98
	130	100	100	87.5	53.5	100	100	98	96
	150	100	100	97.5	77	100	100	99.5	97.5
	200	100	100	100	99.5	100	100	100	100

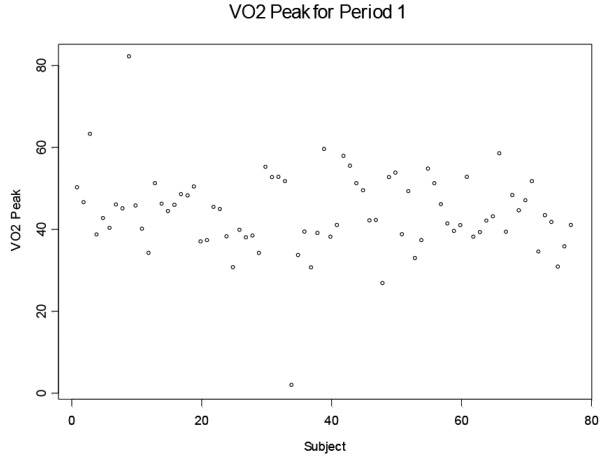

FIGURE 1. Scatter plot of peak for period 1

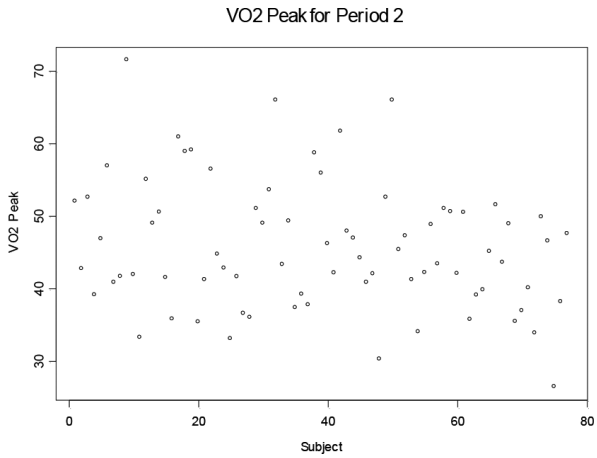

FIGURE 2. Scatter plot of peak for period 2

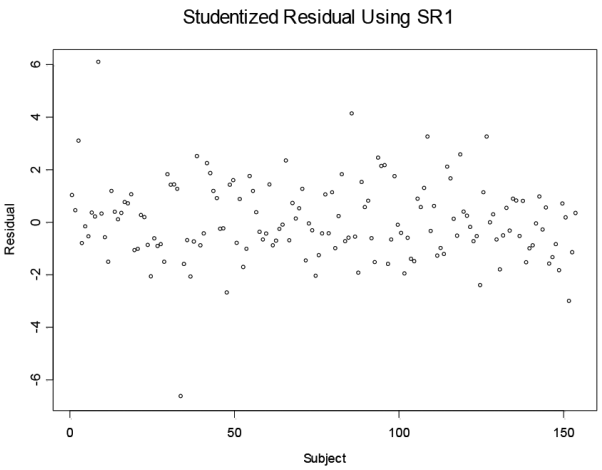

FIGURE 3. Studentized residual using

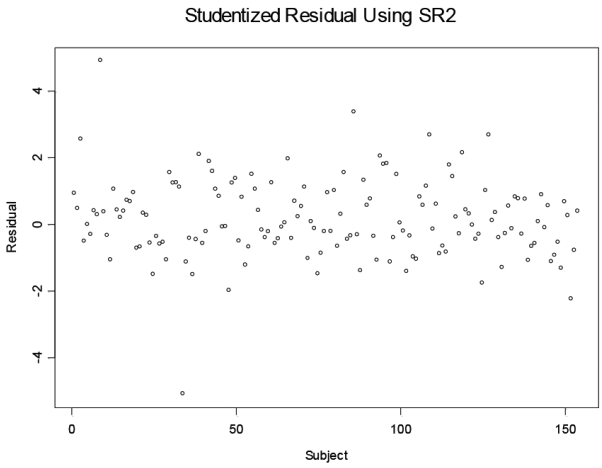

FIGURE 4. Studentized residual using

that was able to detect 7 possible outliers from the data, while $P(SR1)$ detects only 3 possible outliers. This suggests that $P(SR2)$ performs better than $P(SR1)$ in testing for possible outliers in a standard 2×2 crossover design.

CONCLUSION

In this paper, we investigated the detection of outliers based on residuals in a standard 2×2 crossover design. We calculated two types of studentized residual: $SR1$ using a classical procedure and $SR2$ using a new procedure based on median absolute deviation. Their performances in testing for within-subject outliers were compared. Based on a simulation study, we concluded that $P(SR2)$ is more powerful than $P(SR1)$. As an illustration, both procedures were applied to the AUC data and VO_2 peak data in the studies of bioavailability and kinesiology respectively and the superiority of our new procedure was confirmed.

ACKNOWLEDGEMENTS

We thank University of Malaya Medical Center (UMMC) Sport Medicine Clinic for the data provided. This research was partially funded by the UM-PPP (No. PG133-2012B), UM-BKP (BK 017 - 2012) and Universiti Putra Malaysia.

REFERENCES

Chow, S.C. & Liu, J.P. 2009. *Design and Analysis of Bioavailability and Bioequivalence Studies*. London: Chapman and Hall.

Chow, S.C. & Tse, S.K. 1990. Outliers detection in bioavailability/bioequivalence studies. *Statistics in Medicine* 9: 549-558.

Clayton, D. & Leslie, A. 1981. The bioavailability of erythromycin stearate versus enteric-coated erythromycin based when taken immediately before and after food. *Journal of International Medical Research* 9: 470-477.

Enachescu, D. & Enachescu, C. 2009. A new approach for outlying records in bioequivalence trials. Paper presented at *The XIII International Conference on Applied Stochastic Models and Data Analysis*.

Jones, B. & Kenward, M.G. 1989. *Design and Analysis of Cross-over Trials*. London: Chapman and Hall.

Karasoy, D. & Daghan, G. 2012. Examination of outliers in bioequivalence studies. *Bulletin of Clinical Psychopharmacology* 22(4): 307-312.

Liu, J.P. & Weng, C.S. 1991. Detection of outlying data in bioavailability/bioequivalence studies. *Statistics in Medicine* 10: 1375-1389.

Lund, R.E. 1975. Tables for an approximate test for outliers in linear models. *Technometrics* 17(4): 473-476.

Luzar-Stiffler, V. & Stiffler, C. 2005. Higher-order crossover design outlier detection. In *Proceedings of the 27th International Conference on Information Technology Interfaces*. Zagreb: University of Zagreb. pp. 650-655.

Ramsay, T. & Elkum, N. 2005. A comparison of four different methods for outlier detection in bioequivalence studies. *Journal of Biopharmaceutical Statistics* 15(1): 43-52.

Singh, R., Namdev, K.K. & Chilkoti, D. 2014. An assessment on handling of outlier subjects in bioequivalence study - A review. *Journal of Pharmaceutical and Bioanalytical Science* 3(1): 14-20.

Wang, W. & Chow, S.C. 2003. Examining outlying subjects and outlying records in bioequivalence trials. *Journal of Biopharmaceutical Statistics* 13(1): 43-56.

Weiner, D. 1989. *Bioavailability*. Notes on the training course for new clinical statisticians sponsored by the Biostatistics Subsections of Pharmaceutical Manufacture Association, March, 1989, Washington, DC.

Fong Peng Lim & Ibrahim Mohamed*
Institute of Mathematical Sciences
University of Malaya
50603 Kuala Lumpur, Federal Territory
Malaysia

Noorizam Daud
Faculty of Computer and Mathematical Sciences
Universiti Teknologi MARA
40450 Shah Alam, Selangor Darul Ehsan
Malaysia

Siew Li Goh
Sport Medicine Clinic, University of Malaya Hospital
50603 Kuala Lumpur, Federal Territory
Malaysia

*Corresponding author; email: imohamed@um.edu.my

Received: 19 January 2015
Accepted: 21 September 2015