Sains Malaysiana 52(5)(2023): 1581-1593
http://doi.org/10.17576/jsm-2023-5205-19

Identifying Outlier Subjects in Bioavailability Trials Using Generalized Studentized Residuals

(Pengenalpastian Subjek Outlier dalam Ujian Ketersediaan Biologi Menggunakan Residu Terstuden)

F.P. Lim[1,*], L.L. Wong[2], H.K. Yap[3] & K.S. Yow[4]

[1,2,4]Department of Mathematics and Statistics, Faculty of Science, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor Darul Ehsan, Malaysia
[3]Department of Mathematical and Actuarial Sciences, Lee Kong Chian Faculty of Engineering and Science, Universiti Tunku Abdul Rahman, Sungai Long Campus, Bandar Sungai Long, Cheras, 43000 Kajang, Selangor Darul Ehsan, Malaysia
Received: 19 August 2022/Accepted: 10 May 2023

ABSTRACT

This paper discusses several outlier detection methods for bioavailability trials, particularly based on residuals. By considering a simplified model of standard crossover model, which is commonly used in bioavailability trials, we propose an outlier detection procedure based on the generalized studentized residuals (SR3) and compare its ability of detecting the possible outlying subjects with two existing procedures, which are carried out based on the classical studentized residual (SR1) and studentized residual using median absolute deviation (SR2). The performances of these procedures in detecting outlying subject are presented via an extensive simulation study. The results show that the proposed procedure SR3 performs more powerful than that using SR1, and as well as the procedure using SR2 for outlier detection. As an illustration, these procedures are implemented on a real dataset from bioavailability study, namely, the area under the curve (AUC) dataset for two erythromycin formulations.

Keywords: Bioavailability; crossover design; generalized studentized residuals; outlier; residual

ABSTRAK

Kertas ini membincangkan beberapa kaedah pengesanan titik terpencil untuk ujian bioketersediaan, terutamanya berdasarkan residu. Dengan mempertimbangkan satu model silang piawai yang biasa digunakan dalam ujian bioketersediaan, kami mencadangkan satu prosedur pengesanan titik terpencil berdasarkan residu terstuden teritlak (SR3) dan membandingkan keupayaannya untuk mengesan kemungkinan subjek terpencil dengan dua prosedur sedia ada, iaitu dijalankan berdasarkan residu terstuden klasik (SR1) dan residu terstuden yang menggunakan sisihan mutlak median (SR2). Prestasi prosedur berkenaan dalam mengesan subjek terpencil dibentangkan melalui kajian simulasi yang ekstensif. Keputusan menunjukkan bahawa prosedur yang dicadangkan SR3 berprestasi lebih baik daripada prosedur yang menggunakan SR1, dan juga prosedur menggunakan SR2 bagi pengesanan titik terpencil. Sebagai ilustrasi, prosedur tersebut dilaksanakan pada satu set data sebenar daripada kajian bioketersediaan, iaitu, luas di bawah lengkungan (AUC) set data untuk dua formulasi eritromisin.

Kata kunci: Bioketersediaan; pencilan; reka bentuk silang; residu terstuden teritlak; residu

Introduction

A comparative bioavailability trial is usually carried out for testing the bioequivalence of different formulations of a drug in terms of the rate and extent absorption, which is mostly measured by the area under the blood or plasma concentration-time curve (AUC) and the maximum concentration (C_{max}), respectively. Two formulations of the same drug or two drugs are claimed as bioequivalent when they provide the therapeutic effect or that they are therapeutically equivalent (Chow 2014). It is essential to

understand the performance of dosage forms of a drug in appraising the efficacy and quality of a new drug as the bioavailability trial demonstrates the comparable safety of the new drug.

However, the presence of extraordinary observations, which most frequently known as outliers, may influence the conclusion in assessment of equivalence between different formulations with regard to the rate and extent of absorption of the drug (Chow & Tse 1990; Liu & Weng 1991; Metzler & Huang 1983; Rodda 1986). As stated by Chow and Liu (2009), we can distinguish two types of outliers in bioavailability trials: The between-subject outlier and the within-subject outlier. The between-subject outliers are the unusual subjects who had extreme bioavailability to both formulations. 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 or test formulation. In other words, the within-subject outliers show unusual reaction to one of the formulations. Consequently, Sandulovici et al. (2020) analyzes the potential outliers in the bioanalytical and clinical part of a bioequivalence study, and investigates the effect on bioequivalence decisions whether it is appropriate to eliminate them from the statistical evaluation of bioequivalence. In summary, it is crucial to solve the problem of outliers in bioavailability trials to ensure the accuracy of analyses.

In literature, Chow and Tse (1990) proposed both procedures based on Cook's likelihood distance and the estimated distance to solve the problem of outlier in bioequivalence studies, wherein the crossover design is widely used in their statistical analyses. For the same purpose, Liu and Weng (1991) introduced Hotelling T^2 statistics and residuals, while Wang and Chow (2003) suggested a general test procedure based on a mean-shift model. Using simulation study, Ki et al. (1995) highlighted that it may had a masking effect in the intra-subject variability when doing outlier detection. Moreover, Ramsay and Elkum (2005) evaluated different outlier detection methods presented by Chow and Tse (1990), Liu and Weng (1991) and Wang and Chow (2003) through simulation studies. They pointed out that the superiority of the estimated distance test compared to other tests. Furthermore, Karasoy and Daghan (2012) had investigated these existing outlier detection methods by using a real dataset. On the other hand, Enachescu and Enachescu (2009) carried out the principal components analysis and projection pursuit for identifying possible outliers in bioequivalence studies. The details of a studentized residual test and the Lund test also are provided by Singh, Namdev and Chilkoti (2014) for determining the outlying subjects. Recently, Lim (2016) had considered the studentized residual based on median absolute deviation as a robust outlier detection method for standard crossover design. Lim et al. (2019) also suggested the solutions for the same problem in Bayesian framework. Besides, El-Kelany and Ahmed (2020) also compare the performance of both principal component analysis and Cook's Distance for detecting possible outliers in the real data which obtained from Central Administration for Pharmaceutical Affairs (CAPA), in Egypt.

This paper considers the procedure using generalized studentized residual (SR3) in standard crossover design as an alternative to deal with outliers in bioequivalence studies. As we all know, Imon (2005) proposes this group-deletion-based method for handling the outliers in the fixed dimensional problems. This procedure has been widely used in identifying multiple influential points (Baba et al. 2021; Silalahi et al. 2020; Zhao et al. 2019). Hence, it is worth to further investigate the implement of this procedure in standard crossover design. The rest of this article is organized as follows: Next section presents the details about the SR3 and another two existing procedures, which are carried out based on the classical studentized residual (SR1) and studentized residual using median absolute deviation (SR2). Subsequent section discusses the main settings and results of investigating the power of performance through simulation study. Finally, last section illustrates the implementation of the SR3 on a real dataset from bioavailability study.

CROSSOVER MODEL

In standard crossover design, the responses of the *kth* subject in the period j under treatment i is denoted as Y_{kij} where $i, j = 1, 2$ and $k = 1, 2, ..., n_i$. n_i is the size of the group with treatment i. Consider the model presented by Chow and Tse (1990), we have the general crossover design model is as follow:

$$Y_{kij} = \mu + S_k + F_i + P_j + e_{kij} \quad (1)$$

where μ is the overall mean; S_k is the random effect of

kth subject; F_i is the fixed effect of the *ith* treatment with $\sum_i F_i = 0$, P_j is the fixed effect of the *jth* period with $\sum_i P_j = 0$, and e_{kij} is the random error. Besides, the variance components $\{S_k\}$ and $\{e_{kij}\}$ are assumed to be independent and normally distributed with zero mean and variance σ_s^2 and σ_e^2 respectively.

As stated by Chow and Tse (1990), the response variables of primary interest in bioavailability trials are often the extent of absorption and the rate of absorption. The former can be obtained by considering the area under the plasma concentration-time curve (AUC), while the latter is interpreted in term of peak concentration (C_{max}) and time to peak concentration (t_{max}). Since the distribution of the response variable, that is AUC, is usually skewed, hence a log transformation is applied on AUC to adjust the skewness and further analyses then can be performed on the transformed AUC under the model (1).

In the following sections, we present three outlier detection procedures with use of a simplified model (1), which is suggested by Liu and Weng (1991). In order to extend the ideas to more general model, it is assumed that there is no period effect, which is $P_j = 0$, then the model (1) can be reduced to the following model:

$$Y_{ki} = \mu + S_k + F_i + e_{ki} \tag{2}$$

where $k = 1,2,\ldots, n_i$ for $i = 1,2$.

The following sections briefly discusses the concept of three types of studentized residual: classical studentized residual (SR1), studentized residual using median absolute deviation (SR2), and generalized studentized residual (SR3).

CLASSICAL STUDENTIZED RESIDUAL (SR1)

Consider the model (2), assume that the repeated measurements on each subject are assumed to be independent and normally distributed random variables with equal variances, the residual can be obtained such as

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

for each *i*, respectively, and *N* is the total summation of two treatments; $N = \sum n_i$, as mentioned by Lim et al. (2016). The r_{ki} are the estimators of the random error e_{ki} and are normally distributed with zero mean and variance as below,

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

where σ_e^2 is the mean square value of the within-subject residual. Thus, the *SR*1 becomes

$$SR1 = \frac{r_{ki}}{\sqrt{\hat{V}(r_{ki})}}. \tag{5}$$

The response value corresponding to extraordinary large SR1 is known as outlier (Jones & Kenward 1989). Lund (1975) also suggests that a response value is considered as an outlier when the corresponding |SR1| g greater than value 3.

STUDENTIZED RESIDUAL USING MEDIAN ABSOLUTE DEVIATION (SR2)

As r_{ki} is not a good estimator of e_{ki} when the residuals are far from the normal distribution, Lim et al. (2016) proposed a robust estimate of scale, which is median absolute deviation (MAD) instead of the classical studentized residual. If the samples come from normal distribution,

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

can be used to estimate $Z_{0.75}\, e_{ki}$ rather than e_{ki}, where $Z_{0.75}$ is the 0.75 quantile of the standard normal distribution. The scaled MAD (MADN) then is defined as

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

Refer to Lim (2010), Maarof, Peng and Ibrahim (2010), and Wilcox (2011) for the details of MAD. Hence, the *SR*2 is

$$SR2 = \frac{r_{ki} - median(r_{ki})}{MADN}. \tag{6}$$

If the largest |*SR*2| is greater than the critical value, *D* in Table 1, then, the corresponding response can be considered as an outlier. It is noted that the parametric bootstrap technique is used for constructing the Table 1 under model (2). *SR*2 is calculated for each size of group 20, 60 and 100 and then largest *SR*2 can be obtained. See Lim et al. (2016) for the details of the simulation procedures for constructing the Table 1.

GENERALIZED STUDENTIZED RESIDUAL (SR3)

In this study, we adapt the generalized studentized residual, which can identify the single influential observation in the linear regression, into a standard crossover trial. Let us consider a linear regression model

$$Y = X\beta + \varepsilon \tag{7}$$

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

N	μ_2	γ					
		0.5	5	10	15	30	40
20	60	0.705058	0.980175	1.285861	1.654673	1.709177	1.706658
	80	0.735627	1.285861	1.793335	1.709177	1.654034	1.593723
	90	0.796764	1.793335	1.706658	1.654034	1.542514	1.512168
	100	1.411959	1.411959	1.411959	1.411959	1.411959	1.411959
	110	0.796764	1.793335	1.706658	1.654034	1.542514	1.512168
	125	0.723400	1.163587	1.718965	1.738969	1.686018	1.634269
60	60	0.718478	1.114369	1.606454	1.994404	2.145909	2.154915
	80	0.762466	1.606454	2.117405	2.145909	2.091540	2.058979
	90	0.850442	2.117405	2.154915	2.091540	2.025157	1.988339
	100	1.906702	1.906702	1.906702	1.906702	1.906702	1.906702
	110	0.850442	2.117405	2.154915	2.091540	2.025157	1.988339
	125	0.744871	1.382749	2.010072	2.126348	2.128574	2.089173
100	60	0.723172	1.161311	1.696741	2.100141	2.317864	2.333515
	80	0.771854	1.696741	2.265473	2.317864	2.291519	2.245437
	90	0.869218	2.265473	2.333515	2.291519	2.212292	2.197347
	100	2.122826	2.122826	2.122826	2.122826	2.122826	2.122826
	110	0.869218	2.265473	2.333515	2.291519	2.212292	2.197347
	125	0.752381	1.479371	2.153228	2.307614	2.318592	2.282197

where Y is a $N \times 1$ matrix vector of responses variable; X is a $N \times m$ ($N > m$) matrix of predictor including the constant predictor; β is a $m \times 1$ vector of parameters to be estimated and ε is a $N \times 1$ vector of random error. In accordance with the model (2), we may recognise the X_1 and X_2 as subject effect, S_k and treatment effect, F_i, respectively. Model (7) then can be re-written as

$$y_t = x_t\beta + \varepsilon_t, \quad t = 1,2, \dots, N$$

where y_t is the t^{th} responses and x_t is the $1 \times m$ vector of predictor.

To estimate the regression parameters, the ordinary least square (OLS) technique is used. Let $\hat{\beta} = (X'X)^{-1}X'Y$, then t^{th} residual is

$$\hat{\varepsilon} = Y - X\hat{\beta}\,.$$

We fix these residuals into a weight matrix W, namely as leverage matrix, then $\hat{\varepsilon} = (I - W)Y$

where $W = X(X' X)^{-1}) X'$. As the high leverage point is caused by a set of influential X-values, hence the diagonal element of W is known as leverage value and denoted by

$$w_{tt} = x_t(X'X)^{-1}x_t' .$$

An observation is considered as high leverage point when w_{tt} is greater than the twice-the-mean-rule, $2m/N$ (Hoaglin & Welsch 1978).

As an extension, Hadi (1992) presented a single case deleted measure of leverage, which is commonly known as potentials, p_{tt}, such as

$$p_{tt} = x_t(X_{(t)}^t X_{(t)})^{-1}x_t'$$

where $X_{(t)}$ is the matrix X with t^{th} row is deleted. Alternatively, these p_{tt} can also be obtained as follows,

$$p_{tt} = w_{tt} / (1\text{-}w_{tt}).$$

A cut-off point for p_{tt} then can be calculated using median and median absolute deviation (MAD), that is

$$p_{tt} > median(p_{tt}) + 3 \text{ MAD } (p_{tt})$$

where $MAD\ (p_{tt}) = median\{|p_{tt}\text{-}median(p_{tt})|\}/0.6745$.

Similar with the calculation of σ^2 in Crossover model section, we can define the estimator of variance, $\hat{\sigma}^2$ based on the data set with t^{th} observation deleted as

$$\hat{\sigma}_{(t)}^2 = \frac{1}{(N-m-1)}\sum_j \left(y_j - x_j\hat{\beta}^{(-t)}\right)^2 . \qquad (8)$$

Consequently, the external studentized residual for a single case influence, SR3* becomes

$$SR3^* = \frac{y_j - x_j\hat{\beta}^{(-t)}}{\hat{\sigma}_{(t)}^2\sqrt{1-w_{tt}}}$$

The SR3* follows Student's t distribution with N-m-1 degree of freedom and the $|SR3^*|$ is considered as an outlier when greater than 2.5 (Ellenberg 1976).

Nevertheless, Imon (2005) found that the outlier detection approach of single case deleted proposed by Hadi (1992), ineffectively detects multiple influential observations due to the masking and swamping problems. The former means that the second outlier in the data set can be considered as an outlier by itself with the absence of the first outlier, while the latter occurs when the observation in the data set only can be considered as an outlier with the presence of the first outlier. Hence, a generalized version of studentized residual had been proposed to overcome these problems.

As suggested by Imon (2005), let a data set consist of two groups of observations, namely 'remaining' set (R) and 'deleted' set (D), respectively. Consider R contains (N - d) cases after the $d < (N - m)$ cases of D are deleted and all observations are assumed to be the last d rows of X and Y without loss of generality. Define $w_{tt(R)}$ as the t^{th} diagonal element of the $X(X_R' X_R)^{-1} X'$ when a group of observations D is deleted. The $w_{tt(R)}$ then can be calculated as

$$w_{tt(R)} = x_t(X_R'X_R)^{-1}x_t' .$$

and the generalized potentials is

$$p_{tt}^* = \begin{cases} \dfrac{w_{tt(R)}}{1 - w_{tt(R)}} & for\ t \in R \\ w_{tt(R)} & for\ t \in D \end{cases}$$

where D is any arbitrary deleted set of points (Imon 1996). The p_{tt}^* is considered large if

$$p_{tt}^* > median(p_{tt}^*) + 3\ MAD\ (p_{tt}^*)$$

where $MAD\ (p_{tt}^*) = median\{|p_{tt}^* - median(p_{tt}^*)|\}/0.6745$.

After deleting a group of size observations, the estimated parameters, $\hat{\beta}_{(R)}$ is

$$\hat{\beta}_{(R)} = (X_R'X_R)^{-1}X_R'Y_R ,$$

and the t^{th} deletion residual is

$$\hat{\varepsilon}_{t(R)} = y_t - x_t\hat{\beta}_{(R)} .$$

Thus, the generalized externally studentized residual for $t \in R$ becomes

$$SR3 = \frac{y_t - x_t\hat{\beta}_{(R)}}{\hat{\sigma}_{R-t}^2\sqrt{1-w_{tt(R)}}} = \frac{\hat{\varepsilon}_{t(R)}}{\hat{\sigma}_{R-t}^2\sqrt{1-w_{tt(R)}}} .$$

where $\hat{\sigma}_{R-t}$ can be derived by replacing the t in equation (8) with (R - t). Similarly, for case with $t \notin R$, we implement

$$w_{tt(R+t)} = x_t(X_R'X_R + x_t'x_t)^{-1}x_t' = \frac{w_{tt(R)}}{1 + w_{tt(R)}}$$

and

$$\hat{\beta}_{(R+t)} = (X_R'X_R + x_t'x_t)^{-1}(X_R'Y_R + x_t'y_t)$$

$$= \hat{\beta}_R + \frac{(X_R'X_R)^{-1}x_t'}{1 + w_{tt(R)}}\hat{\varepsilon}_{t(R)}$$

to compute the generalized external studentized residual for $t \notin R$ such as

$$SR3 = \frac{y_t - x_t\hat{\beta}_{(R+t)}}{\hat{\sigma}_R^2\sqrt{1 - w_{tt(R+t)}}} = \frac{\hat{\varepsilon}_{t(R)}}{\hat{\sigma}_R^2\sqrt{1 + w_{tt(R)}}} \ .$$

In summary, the generalized studentized residuals for a data set are as below:

$$SR3 = \begin{cases} \frac{\hat{\varepsilon}_{t(R)}}{\hat{\sigma}^2_{R-t}\sqrt{1-w_{tt(R)}}} & for\ t \in R \\ \frac{\hat{\varepsilon}_{t(R)}}{\hat{\sigma}^2_{R}\sqrt{1+w_{tt(R)}}} & for\ t \notin R \end{cases} \quad (9)$$

The rule for deciding an observation as outlier is similar to the rule used for cases using $SR3^*$.

POWER OF PERFORMANCE

An extensive simulation study is conducted to investigate the performance of three outlier detection methods for bioavailability trials: procedures using *SR*1, *SR*2 and *SR*3, respectively. As stated by Lim et al. (2016), random samples are generated based on the model suggested by Luzar-Stiffler and Stiffler (2005) such as

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

where the between-subject, z_{k0}, and within-subject, z_{ki}, variations are the independent and identically standard normal where $i = 1, 2$ and $k = 1, 2, \ldots, n_i$. For simplicity, we assume the values of n_i are equal and the total sample size, $N = \sum n_i$. Consider the n_i values of 10, 30 and 50 in this simulation, the corresponding values N then become 20, 60 and 100, respectively. Without the loss of generality, the treatment 1 mean, μ_1 is set as 100 while the treatment 2 mean, μ_2 takes the values of 60, 80, 90, 100, 110, and 125, respectively. Changes in the difference between two treatment means may exam the capability of outlier detection methods. Furthermore, the constant value, γ is assigned to be 0.5, 5, 10, 15, 30, and 40, which represents the coefficient of the intra-subject variation of 0.5%, 5%, 10%, 15%, 30%, and 40% for the treatment 1, respectively. The first subject is assigned as outlier by multiplying the responses Y_{11} and Y_{12} with a constant p which varies from 10% to 200%.

With the contaminated samples presented, we proceed to the power studies for procedure using *SR*1. We calculate the residual in equation (3) and its variance in equation (4) using the contaminated samples of Y_{ki}, so that studentized residuals *SR*1 in equation (5) can be computed. The similar calculations are repeated 200 times and the frequency of correctly identifying the assigned outlier are recorded. Thus, the power of performance (Q), also known as the percentage of correctly identifying the assigned outlier, can be computed as follows,

$$Q = \frac{g}{200} \times 100\% \ ,$$

where g is the frequency of correctly identifying the assigned the assigned outlier. For the power studies for both procedures using *SR*2 and *SR*3, we use the same contaminated samples of Y_{ki} in the power studies above to count the studentized residual using median absolute deviation, *SR*2 in equation (6) and the generalized studentized residuals in equation (9). We then repeat these calculations (200) times and record the corresponding frequency of correctly identifying the assigned outlier. With the same combination of mean treatment 2 (μ_2), same sizes of group with treatment (n_i), constant γ and constant p, finally the power of performance for procedures *SR*2 and *SR*3 can be calculated accordingly.

Tables 2-4 present the percentage of correctly identifying the assigned outlier for all procedures considered with sample sizes of 20, 60, and 100, respectively. For all sample sizes we consider, the results show that percentages of detection for all procedures using *SR*1, *SR*2 and *SR*3 are almost 100% when γ = 0.5 and 5. There is a decreasing in percentages of detection when γ increases from 10 to 40, especially for procedure using *SR*1. The dramatically drop in percentages of detection indicates that the procedure using *SR*1 is not a significant outlier detection method for bioavailability trials. On the other hand, the procedure using *SR*2 remains as the most powerful than others as the decreasing in percentages of detection is much slightly. However, the procedure using *SR*3 can be considered as an alternative outlier detection method for bioavailability trials due to its high percentages of detection when sample size is large.

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

μ_2	p (%)	SR1 γ						SR2 γ						SR3 γ					
		0.5	5	10	15	30	40	0.5	5	10	15	30	40	0.5	5	10	15	30	40
60	10	100	100	100	89	15	6	100	100	100	99.5	88	89.5	100	100	100	100	71.5	55
	30	100	100	95.5	62	13	6	100	100	100	92	90	95	100	100	100	99	58.5	52
	50	100	100	85	54	13	6	100	100	98.5	92	95.5	96	100	100	99.5	88	55	51.5
	130	100	100	94.5	74.5	33	21.5	100	100	97	93.5	97.5	97	100	83	67.5	65.5	61.5	62
	150	100	100	99	87	46	31.5	100	100	100	98	97.5	97.5	100	99.5	89.5	78	69	66.5
	200	100	100	100	99	76.5	61.5	100	100	100	100	98.5	99	100	100	99.5	98	83.5	79
80	10	100	100	100	99.5	10.5	3.5	100	100	100	100	94.5	93.5	100	100	100	100	66.5	53.5
	30	100	100	99.5	65	5	2.5	100	100	100	99.5	91.5	95	100	100	100	99	56.5	51
	50	100	100	76	21.5	3.5	2.5	100	100	97.5	90.5	94.5	96.5	100	100	99.5	83.5	52	50.5
	130	100	99.5	77.5	49.5	20	13.5	100	100	95.5	98	96	98	100	94.5	76	70.5	64	62
	150	100	100	97.5	78	35	26.5	100	100	99	99	96	98.5	100	100	93.5	82.5	71.5	67.5
	200	100	100	100	99	74	55	100	100	100	100	99.5	100	100	100	100	98.5	85.5	80.5
90	10	100	100	100	100	13.5	3.5	100	100	100	100	98.5	97	100	100	100	100	64	53.5
	30	100	100	80.5	100	4.5	2	100	100	100	99.5	95	96	100	100	100	98.5	55.5	50
	50	100	100	29.5	85.5	3	2	100	100	99.5	97.5	95	96	100	100	99.5	82	50.5	49
	130	100	99.5	38	65	16.5	11	100	100	99.5	96.5	98	99	100	98.5	80	71.5	62.5	62.5
	150	100	100	70.5	94.5	32	25	100	100	100	99.5	98	97.5	100	100	96	85	71	65.5
	200	100	100	98.5	100	72	52	100	100	100	100	100	99.5	100	100	100	99.5	87.5	81.5
100	10	100	100	100	100	16	2.5	100	100	100	100	98.5	96.5	100	100	100	100	64	53
	30	100	100	100	88	4	1	100	100	100	100	97.5	96	100	100	100	97.5	54	50
	50	100	100	94.5	41	1	0.5	100	100	100	99.5	96	96	100	100	98.5	78.5	50	48.5
	130	100	98.5	51	27	11	8.5	100	100	99.5	99	98	97.5	100	99	82.5	75	62	62
	150	100	100	89.5	60.5	27.5	21.5	100	100	100	99.5	98.5	98.5	100	100	97.5	87.5	74	66
	200	100	100	100	98.5	65.5	51	100	100	100	100	100	99	100	100	100	99.5	89	81
110	10	100	100	100	100	24.5	4.5	100	100	100	100	99.5	98	100	100	100	100	62	54
	30	100	100	100	94.5	7.5	2.5	100	100	100	100	96.5	95.5	100	100	100	96	53	51
	50	100	100	98	55	2.5	1.5	100	100	100	98.5	96	96	100	100	98	77	52	49.5
	130	100	93.5	42	23.5	10.5	9	100	99.5	99	97.5	97.5	97	100	99.5	87	76	62.5	61.5
	150	100	100	83	55	27	21.5	100	100	99.5	100	98	98.5	100	100	99	91	73.5	67.5
	200	100	100	100	97	64.5	51	100	100	100	100	100	99.5	100	100	100	99.5	89	82
125	10	100	100	100	100	33	9.5	100	100	100	100	99	95	100	100	100	100	59.5	53.5
	30	100	100	100	98	13.5	4.5	100	100	100	100	95.5	93	100	100	100	98	52	52.5
	50	100	100	99	77.5	7	3.5	100	100	100	98.5	93.5	93.5	100	100	94.5	71.5	53.5	52.5
	130	100	97.5	54.5	29.5	11	9.5	100	100	94	95.5	97.5	97.5	100	100	87.5	75.5	63	61.5
	150	100	100	82	54.5	25.5	21	100	100	99.5	97.5	97.5	98	100	100	98.5	90.5	73.5	67.5
	200	100	100	100	96.5	60	49	100	100	100	99.5	99	98.5	100	100	100	100	91.5	83

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

μ_2	p (%)	SR1 γ						SR2 γ						SR3 γ					
		0.5	5	10	15	30	40	0.5	5	10	15	30	40	0.5	5	10	15	30	40
60	10	100	100	100	99	47	32	100	100	100	100	92	93.5	100	100	100	100	80	70.5
	30	100	100	100	89	47	32	100	100	95	90.5	93	93.5	100	100	100	100	75.5	72
	50	100	100	99.5	88.5	47	32	100	100	89	91	93.5	94.5	100	100	100	94.5	75	72
	130	100	100	99.5	93	57.5	40.5	100	100	97.5	95	95	93.5	100	87.5	84.5	78.5	76	76.5
	150	100	100	100	97	70	50	100	100	99.5	97	96.5	95.5	100	100	92	87.5	80	79.5
	200	100	100	100	100	83.5	72	100	100	100	100	97	97	100	100	100	99	92	88
80	10	100	100	100	100	19.5	16.5	100	100	100	100	93.5	93.5	100	100	100	100	73.5	68.5
	30	100	100	100	76	19.5	16.5	100	100	100	99	93.5	95	100	100	100	100	70	69
	50	100	100	88	49	19.5	16.5	100	100	98.5	94.5	95	95	100	100	100	92.5	69.5	69.5
	130	100	100	88.5	70	31.5	24.5	100	100	97	96.5	95.5	95	100	98	86	81.5	78.5	75.5
	150	100	100	98	83	44	37	100	100	99.5	97	97	95	100	100	97	89	82.5	79
	200	100	100	100	99	72.5	65.5	100	100	100	100	98.5	97.5	100	100	100	99	90.5	88
90	10	100	100	100	100	14	11.5	100	100	100	100	93	92.5	100	100	100	100	69.5	68
	30	100	100	100	80.5	14	11.5	100	100	100	99.5	91.5	93	100	100	100	100	69.5	69
	50	100	100	87	30.5	14	11.5	100	100	99.5	97.5	92	93.5	100	100	100	89	70	69.5
	130	100	99.5	72.5	46.5	24	19.5	100	100	97.5	97	94.5	94	100	99.5	87	83.5	77	75.5
	150	100	100	95	70.5	38.5	29	100	100	100	98.5	95.5	95.5	100	100	98	90	84	80
	200	100	100	100	96.5	70.5	61	100	100	100	100	98	97.5	100	100	100	99.5	91	89
100	10	100	100	100	100	9.5	8	100	100	100	100	96.5	92	100	100	100	100	69.5	67
	30	100	100	100	91	8	8	100	100	100	100	94	92	100	100	100	100	68	69.5
	50	100	100	96	33.5	9	8	100	100	100	98	92.5	92	100	100	100	84.5	70.5	70
	130	100	97.5	54	32	17	16	100	100	98	97	94.5	93.5	100	100	88.5	83	77	75.5
	150	100	100	89	63.5	32.5	25	100	100	100	98	96.5	95.5	100	100	99.5	91.5	83	79.5
	200	100	100	100	96.5	66.5	55	100	100	100	100	98	97	100	100	100	100	91.5	89
110	10	100	100	100	100	16.5	12.5	100	100	100	100	97	92.5	100	100	100	100	68	67.5
	30	100	100	100	98	14.5	12.5	100	100	100	100	93.5	92.5	100	100	100	99.5	69.5	69.5
	50	100	100	99	56.5	14.5	12.5	100	100	100	99.5	92	93	100	100	99.5	84	71	70
	130	100	98	59.5	35.5	22.5	19.5	100	99.5	95	96	94	94	100	100	90	84.5	77	76
	150	100	100	86	64	35.5	28	100	100	98.5	97.5	95.5	95.5	100	100	99.5	92	83	80
	200	100	100	100	95	68.5	58	100	100	100	100	98	97.5	100	100	100	100	91.5	89
125	10	100	100	100	100	30	18.5	100	100	100	100	97	94	100	100	100	100	68	68.5
	30	100	100	100	99	26.5	18.5	100	100	100	100	95.5	94	100	100	100	99	69	69.5
	50	100	100	100	86.5	25	18.5	100	100	100	99.5	94.5	95	100	100	98.5	79	70	71.5
	130	100	99.5	89.5	63.5	30.5	21.5	100	99	96	95.5	94.5	94.5	100	100	92.5	87	80.5	77.5
	150	100	100	94.5	79	39.5	31.5	100	100	98	98	95.5	95.5	100	100	98.5	92.5	85.5	81.5
	200	100	100	100	98	70	58	100	100	100	99.5	97.5	98	100	100	100	100	92	89

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

μ_2	p (%)	SR1 γ						SR2 γ						SR3 γ					
		0.5	5	10	15	30	40	0.5	5	10	15	30	40	0.5	5	10	15	30	40
60	10	100	100	100	99.5	69.5	53	100	100	100	99	92	90	100	100	100	100	87	82.5
	30	100	100	100	97	69.5	53	100	100	98.5	91	92.5	90	100	100	100	100	85	83
	50	100	100	100	97	69.5	53	100	100	93.5	91.5	93	91	100	100	100	96.5	85	84
	130	100	100	100	97.5	73	56	100	100	96	94	93	92	100	96.5	90.5	89	84.5	85.5
	150	100	100	100	99	78	61	100	100	99.5	97	94	93	100	100	96.5	94	86.5	85
	200	100	100	100	99.5	92	81	100	100	100	99.5	97	96	100	100	100	99.5	94	91
80	10	100	100	100	100	39	32	100	100	100	100	90.5	92.5	100	100	100	100	84	82.5
	30	100	100	100	82.5	39	32	100	100	100	99	90.5	92.5	100	100	100	100	83	83
	50	100	100	94	70	39	32	100	100	97	93	91.5	93.5	100	100	100	95.5	83	83.5
	130	100	100	96	79	44	34.5	100	100	96	94	93.5	94.5	100	99.5	91.5	86	85.5	84.5
	150	100	100	99.5	91	53.5	43	100	100	99.5	97	94.5	96	100	100	98	93.5	86.5	86
	200	100	100	100	98.5	82.5	69	100	100	100	99.5	97.5	98	100	100	100	99	95	93
90	10	100	100	100	100	28.5	27.5	100	100	100	100	92.5	92	100	100	100	100	84.5	81.5
	30	100	100	100	88.5	28.5	27.5	100	100	100	100	92.5	92	100	100	100	100	81.5	82.5
	50	100	100	93	48	28.5	27.5	100	100	99.5	93	93	93.5	100	100	100	94	82	82.5
	130	100	100	80	53.5	33.5	30	100	100	96	95.5	93.5	94	100	99	91.5	87	85	83
	150	100	100	95.5	81.5	43.5	39.5	100	100	99.5	97.5	94.5	94.5	100	100	98	94	87.5	85
	200	100	100	100	98	77.5	66	100	100	100	99.5	97.5	97	100	100	100	99	94.5	93
100	10	100	100	100	100	21	20.5	100	100	100	100	94.5	93	100	100	100	100	81	81
	30	100	100	100	97	20.5	20.5	100	100	100	100	93	92.5	100	100	100	100	81.5	82
	50	100	100	98	39	20.5	20.5	100	100	100	99	93	92.5	100	100	100	92	82	82.5
	130	100	97.5	56	34	24.5	23.5	100	99.5	96.5	95	94.5	94.5	100	99.5	93	87.5	84	83.5
	150	100	100	91.5	67	34.5	32.5	100	100	99	98	94.5	95	100	100	99.5	94.5	87	85.5
	200	100	100	100	97.5	74	61.5	100	100	100	99	98	97.5	100	100	100	99	94	92.5
110	10	100	100	100	100	30	28	100	100	100	100	95.5	92	100	100	100	100	80.5	79
	30	100	100	100	98.5	29.5	28	100	100	100	100	93	92	100	100	100	100	79	81
	50	100	100	100	63	29.5	28	100	100	100	98	93	93	100	100	100	90	79.5	81
	130	100	98	64	45	32	31	100	99	93.5	94	93.5	93.5	100	100	94.5	88	84	84
	150	100	100	90	68	41.5	37.5	100	100	98.5	97.5	94.5	94.5	100	100	99.5	95	86.5	85
	200	100	100	100	98	75	62	100	100	100	99.5	98	96.5	100	100	100	99.5	94	93
125	10	100	100	100	100	44.5	36.5	100	100	100	100	93	93	100	100	100	100	78.5	79.5
	30	100	100	100	100	44	36.5	100	100	100	100	92	93	100	100	100	99.5	79.5	80
	50	100	100	100	92.5	43.5	36.5	100	100	100	98	91.5	93.5	100	100	99.5	86	80	81
	130	100	100	96	79	46	39	100	96	93.5	92	92	94	100	100	96	89	84.5	83.5
	150	100	100	98.5	85	51.5	43	100	100	96	94	92.5	94	100	100	99.5	96	88	85.5
	200	100	100	100	97	77.5	64.5	100	100	100	100	96.5	96	100	100	100	99.5	94.5	93

APPLICATION ON BIOAVAILABILITY DATA

In this section, a real dataset from bioavailability study is used for illustration purpose. We consider the plasma concentration-time curve (AUC) dataset from two erythromycin formulations, which was published by Clayton and Leslie in 1981 (Table 5). With the participation of 18 subjects, a crossover trial was carried out and it aimed to evaluate a new erythromycin formulation (i.e., erythromycin stearate) with a reference formulation (i.e., erythromycin base). Since there is no sequence identification of each subject is provided by Clayton and Leslie (1981), thus, we adapt the order of periods given in Weiner (1989) and assign subjects 1 through 9 as sequence 1 while the remaining subjects as sequence 2.

Initially, the existence of outlying subjects in the dataset can be illustrated through a scatter plot as Figure 1, wherein the subject 7 in group 1 is obviously far from most observations. As per Table 6, we may conclude that there are three outliers are detected in this AUC data by the procedure using SR1 as the mean of both periods is quite high. Besides, the procedure using SR2 detects five outliers detected in this AUC data as presented in Table 7, while the procedure using SR3 able to identify four outliers as shown in Table 8. In summary, the results prove that the subject 7 in group 1 is significantly recognized as outlying subject by using all procedures considered in this study. These results are in accordance with the results presented by Chow and Liu (2009), whereas the used methods are two-sample Hotelling T^2 and likelihood distance.

TABLE 5. Plasma concentration-time curve (AUC) for two erythromycin formulations

Subject	Sequence	AUC ($mcg \cdot h \cdot mL^{-1}$)	
		Treatment 1 (Base)	Treatment 2 (Stearate)
1	BS	5.47	2.52
2	BS	4.84	8.87
3	BS	2.25	0.79
4	BS	1.82	1.68
5	BS	7.87	6.95
6	BS	3.25	1.05
7	BS	12.39	0.99
8	BS	4.77	5.60
9	BS	1.88	3.16
10	SB	4.98	3.19
11	SB	7.14	9.83
12	SB	1.81	2.91
13	SB	7.34	4.58
14	SB	4.25	7.05
15	SB	6.66	3.41
16	SB	4.76	2.49
17	SB	7.16	6.18
18	SB	5.52	2.85

Table adapted from Clayton and Leslie (1981)

TABLE 6. Outlier detected by the procedure using SR1

Number of outliers detected	Subject detected as outlier	AUC ($mcg \cdot h \cdot mL^{-1}$)	
		Period 1	Period 2
1	Subject 7	12.39	0.99
2	Subject 11	9.83	7.14
3	Subject 2	4.84	8.87
	Mean	9.02	5.67

TABLE 7. Outlier detected by the procedure using SR2

Number of outliers detected	Subject detected as outlier	AUC ($mcg \cdot h \cdot mL^{-1}$)	
		Period 1	Period 2
1	Subject 7	12.39	0.99
2	Subject 11	9.83	7.14
3	Subject 2	4.84	8.87
4	Subject 5	7.87	6.95
5	Subject 14	7.05	4.25
	Mean	8.40	5.64

TABLE 8. Outlier detected by the procedure using SR3

Number of outliers detected	Subject detected as outlier	AUC ($mcg \cdot h \cdot mL^{-1}$)	
		Period 1	Period 2
1	Subject 7	12.39	0.99
2	Subject 2	4.84	8.87
3	Subject 11	9.83	7.14
4	Subject 5	7.87	6.95
	Mean	8.73	5.99

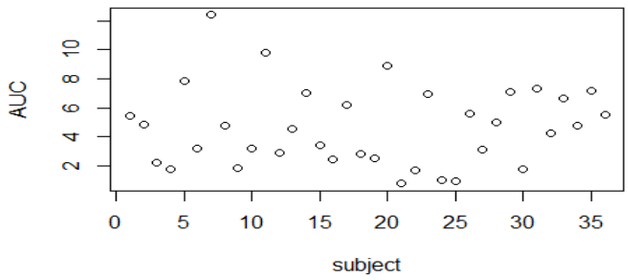

FIGURE 1. Scatter plot of AUC data

CONCLUSION

In this paper, we have considered the problem of detecting outlier in bioavailability trial based on residual. We propose the procedure using generalized studentized residual (*SR3*) and investigate its performance with two existing procedures that using classical studentized residual (*SR1*) and studentized residual based on median absolute deviation (*SR2*). Their performance in detecting the outlying subject is compared via a simulation study. It is observed that the percentage of correctly identifying the assigned outlier (Q) for the procedure using *SR*3 always higher than that using *SR*1, and it tends to provide the similar results as the procedure using *SR*2 (Tables 2 - 4). Hence, we may conclude that the procedure using *SR*3 performs more powerful than that using *SR*1, and comparable with the procedure *SR*2. As an illustration, all procedures are applied to the AUC dataset and the capability of our proposed procedure in identifying the possible outlying subjects in bioavailability trial is confirmed.

ACKNOWLEDGEMENTS

The authors would like to thank the Ministry of Higher Education, Malaysia (Grant no: FRGS/1/2019/STG06/UPM/02/18) and Universiti Putra Malaysia (Grant no: GP-IPS/2019/9676000) for funding this study.

REFERENCES

Baba, A.M., Midi, M., Adam, M.B. & Rahman, N.H.A. 2021. Detection of influential observations in spatial regression model based on outliers and bad leverage classification. *Symmetry* 13(11): 2030.

Chow, S.C. 2014. Bioavailability and bioequivalence in drug development. *Wiley Interdisciplinary Reviews: Computational Statistics* 6(4): 304-312. doi:10.1002/wics.1310

Chow, S.C. & Liu, J.P. 2009. *Design and Analysis of Bioavailability and Bioequivalence Studies*. 3rd ed. London: Chapman and Hall.

Chow, S.C. & Tse, S.K. 1990. Outlier detection in bioavailability/bioequivalence studies. *Statistics in Medicine 9*(5): 549-558.

Clayton, D. & Leslie, A. 1981. The bioavailability of erythromycin stearate versus enteric-coated erythromycin base when taken immediately before and after food. *Journal of International Medical Research* 9(6): 470-477.

El-Kelany, G.A. & Ahmed, S.M. 2020. Detecting outliers in bioequivalence studies for crossover design. *Asian Journal of Applied Science and Technology (AJAST).* 4(1): 30-49.

Ellenberg, J.H. 1976. Testing for a single outlier from a general linear regression. *Biometrics* 32(3): 637-645.

Enachescu, D. & Enachescu, C. 2009. A new approach for outlying records in bioequivalence trials. In *Proceedings of the 13th International Conference on Applied Stochastic Models and Data Analysis, Vilnius, Lithuania*. pp. 250-257.

Hadi, A.S. 1992. A new measure of overall potential influence in linear regression. *Computational Statistics & Data Analysis* 14(1): 1-27.

Hoaglin, D.C. & Welsch, R.E. 1978. The hat matrix in regression and ANOVA. *The American Statistician* 32(1): 17-22.

Imon, A.H.M.R. 2005. Identifying multiple influential observations in linear regression. *Journal of Applied Statistics* 32(9): 929-946.

Imon, A.H.M.R. 1996. Subsample methods in regression residual prediction and diagnostics. PhD thesis, School of Mathematics and Statistics, University of Birmingham, UK (Unpublished).

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. *Klinik Psikofarmakoloji Bülteni-Bulletin of Clinical Psychopharmacology* 22(4): 307-312.

Ki, F.Y., Liu, J.P., Wang, W. & Chow, S.C. 1995. The impact of outlying subjects on deciding of bioequivalence. *Journal of Biopharmaceutical Statistics* 5(1): 71-94.

Lim, F.P. 2016. Statistical appraisal in solving some medical problems. Doctoral Dissertation, University of Malaya, Kuala Lumpur, Malaysia (Unpublished).

Lim, F.P. 2010. Goodness-of-fit test for standard logistics distribution with outliers. Doctoral Dissertation, Universiti Putra Malaysia (Unpublished).

Lim, F.P., Mohamed, I., Ibrahim, A.I.N., Goh, S.L., Mohamed, N.A. & Rahman, A. 2019. Outlier detection in 2 × 2 crossover design using Bayesian framework. *Sains Malaysiana* 48(4): 893-899.

Lim, F.P., Mohamed, I., Daud, N. & Goh, S.L. 2016. Comparison of outlier detection methods in standard 2×2 crossover design. *Sains Malaysiana* 45(3): 499-506.

Liu, J.P. & Weng, C.S. 1991. Detection of outlying data in bioavailability/bioequivalence studies. *Statistics in Medicine* 10(9): 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.

Rodda, B.E. 1986. *Bioequivalence of Solid Oral Dosage Forms, Presentation to U. S. Food and Drug Administration Hearing on Bioequivalence of Solid Dosage Forms*. 29 September to 1 October. Pharmaceutical Manufacturers Association.

Maarof, F., Peng, L.F. & Ibrahim, N.A. 2010. Bootstrapping the confidence intervals of for samples from contaminated standard logistic distribution. *Pertanika Journal of Science & Technology* 18(1): 209-221.

Metzler, C.M. & Huang, D.C. 1983. Statistical methods for bioavailability and bioequivalence. *Clinical Research Practices and Drug Regulatory Affairs* 1: 109-132.

Sandulovici, R., Mircioiu, I., Aboul-Enein, H.Y., Manolache, M., Mircioiu, C., Voicu, V. & Anuta, V. 2020. Sources of outlier data in the bioanalytical and clinical part of a piroxicam bioequivalence study. *International Journal of Clinical Pharmacology and Therapeutics* 58(11): 652-663.

Silalahi, D.D., Midi, H., Arasan, J., Mustafa, M.S. & Caliman, J.P. 2020. Kernel partial diagnostic robust potential to handle high-dimensional and irregular data space on near infrared spectral data. *Heliyon* 6(1): e03176.

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.

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.

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

Wilcox, R.R. 2011. *Introduction to Robust Estimation and Hypothesis Testing*. Massachusetts: Academic Press.

Zhao, J.L., Liu, C., Niu, L. & Leng, C.L. 2019. Multiple influential point detection in high dimensional regression spaces. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 81(2): 385-408.

*Corresponding author; email: fongpeng@upm.edu.my