An Assessment on Handling of Outlier Subjects in Bioequivalence Study – A Review

Rajneesh Singh1, Kuldeep Kumar Namdev1 and Dr. Deepak Chilkoti2
1Pharmacokinetic and Biostatistics Department, Fortis Clinical Research Limited
2Clinical Pharmacology Unit, Fortis Clinical Research Limited

Abstract
The purpose of this article is to explicate the procedures for detecting outliers subject in bioequivalence study as per various regulatory agencies (e.g. USFDA, TPD, ANVISA and CDSCO). The existence of outlier subjects in bioequivalence studies possibly will have an effect on the results of the study. The present article provides the details regarding Studentized Residual Test and Lund Test for identification of outlier subjects in bioequivalence study. This article also represents the view of various drug regulatory agencies about re-testing / redosing of outlier subjects.

INTRODUCTION
Outlier data in bioequivalence studies are defined as subject data for one or more bioavailability measures that are discordant with corresponding data for that subject and/or for the rest of the subjects in a study. [1]

Comparative bioavailability studies are having small sample size in comparison to others clinical trials; one or two extreme value of subjects could have a large consequence on the inference to be made from these small studies. The usual parametric assumptions and estimation are not robust against extreme values. An outlier is defined as a residual (observed less predicted) data point that has large value - i.e. the model does not fit the data point well hence it as an observation with a large residual (what we observed is way different from what we predict). Outliers occur frequently in bioequivalence trials simply by chance. However, outlier detection is held to be potentially indicative of product failure or subgroup identification and is scrutinized carefully by regulatory authorities prior to approval. Statistical detection of an outlier is not sufficient reason to exclude a subject's observation(s). If data were to be excluded 'scientific evidence or explanations' (FDA, 1992) should be supplied. An example of an acceptable reason as a subject failed to swallow their medication or took too much. On a practical level, this essentially means in practice that there is no such thing as an outlier in a bioequivalence data set, and while we recommend that statisticians always check the assumptions of their model, in this context there is little utility in spending too much time worrying about outliers' impact on the findings. [2, 3]

Outlier can influence the biasness in a statistical model since it may give bias estimate of regression parameters. An outliers, is very different from all the others observations of data sets. Outlier’s subjects can be identifying by residual (looking for large difference in residual).Basically residual is the difference between the values of the outcome predicted by the model and the values of the outcome observed in the sample. If any cases stand out as having a large residual in statistical model, then it could be outliers. As well as testing for outliers by looking at the...
error in the model, it is also possible to look at whether certain case exerts unnecessary weight over the parameters of the models. So, if we were to delete a certain case, we can get the different regression coefficient. [4] There are others statistical method to detect the outlier’s subjects: The likelihood distance test (LD) is one of the tests used for determining outliers or influential observations in a bioequivalence study and was developed by Cook and Weisberg (1982) based on likelihood distances. Liu ve Weng (1991) suggested a procedure based on the order statistics of the two sample Hotelling T^2 (HT) statistics to identify possible outlying subjects. Likelihood distance, Estimates distance, Hotelling T^2, and Liu and Weng’s residuals method have been applied for bioequivalence study to detect the outlier’s subjects. The present article provides the details regarding Studentized Residual Test and Lund Test for identification of outlier subjects in bioequivalence study.

**STATISTICAL METHODOLOGY TO DETECT THE OUTLIER'S SUBJECTS**

After having computed the pharmacokinetic parameters, it shall be first manually/graphically inspected for the presence of any extreme (i.e. high or low) observations. And if any extreme observation is found during manual / graphical inspection then to ensure that it is a statistical outlier, this data will be subjected to statistical outliers test. In present article, we illustrate the some statistical outliers’ test which is applicable in bioequivalence study.

**Studentized Residual Test**

The difference between the observed value of the dependent variable \(y\) and the predicted value \(\hat{y}\) is called the residual \(e\). Each data point has one residual. Residual = Observed value - Predicted value.

Residuals are identifying problems with normality, constancy of variance and linearity of variables. The normal or unstandardized residual are measured in the same unit as the outcome variable and so are difficult to interpret across different model. However standardized residual are the residual converted to Z score, which means they are converted into standard deviation units (i.e. they are distributed around a mean of 0 with standard deviation 1). By converting residual into Z-score (standardized residual) we can compare residual from different model and use unacceptable values. So if the residuals, \(r_i\) have an approximate normal distribution with mean 0 and variance \(\sigma^2\) then \(\frac{(r_i - 0)}{\sigma_i} \sim N (0, 1)\). We don’t know the standard deviation of the residuals but we have an estimate of standard error of the \(i^{th}\) residual. It is estimate of the overall standard deviation (the square root of the Residual Mean Square) times a function of the associated hat matrix diagonal, i.e. S.E (\(r_i\)) = S* \(\sqrt{(1-h_i)}\). The standardized residuals are given by then \(r_i (s) = \frac{r_i}{S (i)} \times S.E (r_i)\), where S (i) is the residual standard deviation if the least squares procedure is run without the \(i^{th}\) observation. It is similar to the standardized residuals; however it has better theoretical properties than standardized residual. . The standardized residuals have a standard normal distribution, and then it is very easy to spot large residuals. Large residuals will have large Z-scores and hence will have low probabilities of being observed if these points truly came from a standard normal distribution.

Analysis of outliers usually focuses on deleted residuals, studentized residual is statistical method to test the residual at each X-points (independents variable), so statistically, assume that residual follows the normal distribution with unknown variance; however studentized residual follow the t-distribution with N-K-2 degree of freedom (N is sample size of data, K is numbers of independents variable). The method of calculating studentized residuals is also called ‘mean-shift’ outlier model. It might be preferable for small sample size (< 20) because it has followed a t-distribution. There will be a t - value for each residual, with degree of freedom n - k - 1. When t exceeds the critical value for a given alpha level then the case is considered an outlier.

If we consider a linear regression models with \(k\) regressor variables:

\[
Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_k X_{ik} + e_i;
\]

\[
Y = X\beta + E; \text{ Where } E \text{ follows } N (0, \sigma^2 I)\]
The studentized residual $E_i$ is distributed as $t_{n-k-2}$; we would expect that 5% of the studentized residuals would be beyond $t_{0.025} \pm 2$ by chance alone. It shows that Observations that have a studentized residual outside the $\pm 2$ range are considered statistically significant at the 95% level.

In a plot of studentized residuals versus ordinary residuals, one may draw lines at plus and minus two standard units to highlight cases outside the range where 95% of the cases normally lie; points substantially off the straight line are potential leverage problems.[5]

**Lund’s Statistical Procedure to Detect the Outliers**

If any extreme observation is found during manual / graphical inspection then to ensure that it is a statistical outlier the data will be subjected to Lund’s test. [6]

**Estimation of Outliers:** Following SAS code is used in the detection of outliers:

```
Proc glm data=new outstat=Lund; class Subject Form Seq Period; Model lnCmax=seq subject (seq) form period; lsmeans Form/stderr; Estimate 'test-ref' Form -1 1; Output out=Lund P=PRED R=RESID stderr=eresid STUDENT=STDRESID; Title4 'Studentised Residuals for Lntransformed Cmax (Test- Reference)'; Title5 'Lunds Test for XXXXXX Drug name -Ln-transformed Cmax';
```

On running the above SAS code, output containing Studentized Residual values for individual subjects are displayed. If the Studentized Residual value for any subject is found within the limit of Tabulated value obtained from “Studentized Residual Table” then that subject is considered to be a statistical outlier.

For example for $n=35$, $q = 1$ and $a = 0.05$ the tabulated value of Studentized Residual Value from the table is 3.03. If the Studentized Residual for any subject is outside the range of $\pm 3.03$ then it will be termed as a Statistical Outlier.

Where $n =$ number of subjects, $q =$ number of independent variables (including count for intercept if fitted) and $a =$ level of significance.

For the given, $n=35$ and using the model as there are two independent variables period and treatment which gives $q=34+1+1=36$ and $N=70$ rows (observations).

The residuals distribution has $DFE = N – q – 1 = 34$ degrees of freedom. And also $p=\alpha/N/2 = 2*0.05/35= 0.001429$ and $F_e$ is the tabulated value for $F$ distribution. Entering these values into (a) yields $d_e=3.033$ which is the value found in the Table of Studentized Residual Lund for $N=35$ and $q=1$. The Studentized Residual value from the table is obtained by having information on $n$, $q$ and $a$

For $n>100$ we cannot obtain the tabulated value for Studentized Residual from “Studentized Residual Table” [6].

**REGULATORY CONCERN REGARDING OUTLIERS SUBJECTS**

We illustrate the assessment of outliers’ subjects in bioequivalence study as per various regulatory implications like ANVISA - Brazil, CDSCO- India, EMEA - European Regions, Therapeutic Products Directorate (TPD) - Canada and USFDA - United States of America. Here, we are bringing up the concerns regarding the outliers subjects as per various international drug regulatory agencies as below.

**Agência Nacional de Vigilância Sanitária (ANVISA) [7]**

Considerations regarding outliers in the relative bioavailability/bioequivalence study with crossover design, the discrepant points are defined as those where some subjects (outliers) differ notably from the other subjects of the study when comparing test and reference product in the subject himself. The existence of an outlier without violation of the protocol may indicate one of the following situations:

A) Failure of the product: in this case, an abnormal response may be present both for the test product and the reference product;
B) Subpopulation: this may occur when an individual represents a population, in which the bioavailability of two products is notably different from the majority of the population. Due to these facts, in general, the exclusion of outliers is not recommended, mainly for designs that are not replicated.

**Central Drugs Standard Control Organization (CDSCO) [8]**

Post hoc exclusion of outliers is not recommended, a scientific explanation should be provided to justify the exclusion of a subject from the analysis.

**European Medicines Agency (EMEA) [9]**

European Medicines Agency (EMEA) is not suggested any statistical test for outlier’s subjects.

**Therapeutic Products Directorate (TPD) Health Canada [10]**

A strategy to identify and account for outliers should be part of the protocol. No more than 5% of the subjects may be considered to be outliers, unless there are 20 or fewer subjects, in which case only one subject may be removed. If a protocol for handling outliers is stated it must be followed before the results of the analysis are summarized into confidence intervals (i.e., regardless of whether results meet the standard, the outlier protocol should be followed).

The protocol for handling outliers should include the following:

A. The observation(s) should be identified by an outlier test. It is recommended that a simple outlier test such as a studentised residual being greater than 3 be used.

B. The observation(s) must be outside the range of all the other observations regardless of formulation. In other words, the procedure should only identify observations which are very different from all others collected.

C. The subject in question should be identified as an outlier for all parameters, for either the test or reference product, upon which the bioequivalence decision is to be based. Parameters of interest are usually an AUC and Cmax measure, but in some instances other parameters are required.

Note: Re-testing of subjects identified as outliers is not recommended.

**United States Food and Drug Administration (USFDA) [1]**

As per USFDA proposal, in the case of product failure, the unusual response could be present for either the tests or reference product. However, in the case of a subpopulation, even if the unusual response is observed on the reference product, there could still be concern for lack of interchangeability of the two products. For these reasons, deletion of outlier values is generally discouraged, particularly for nonreplicated designs. With replicated crossover designs, the retest character of these designs should indicate whether to delete an outlier value or not. Sponsors or applicants with these types of data sets may wish to review how to handle outliers with appropriate review staff.

The existence of a subject outlier with no protocol violations could indicate one of the following situations:

**A. Product Failure:** Product failure could occur, for example, when a subject exhibits an unusually high or low response to one or the other of the products because of a problem with the specific dosage unit administered. This could occur, for example, with a sustained and/or delayed-release dosage form exhibiting dose dumping or a dosage unit with a coating that inhibits dissolution.

**B. Subject-by-Formulation Interaction:** A subject-by-formulation interaction could occur when an individual is representative of subjects present in the general population in low numbers, for whom the relative BA of the two products is markedly different than for the majority of the population, and for whom the two products are not bioequivalent, even though they might be bioequivalent in the majority of the population.

BE studies are usually carried out as crossover studies, the most important type of subject outlier is the within-subject outlier, where one subject or a few subjects differ notably from the rest of the subjects with respect to a within-subject T-R comparison.
As FDA recommendation, an appropriate statistical outliers test can be apply for distinguish the outlier’s subjects in BE study. Re-testing / Re-dosing of outliers subjects will be recommended; again clinical experiment on these outlier’s subjects to test their validity for outlying value. Re-dosing studies allows for the verification of statistical outliers and is beneficial to the Agency in evaluating whether the statistical results obtained for a bioequivalence study indeed meet the BE limits of 80.00-125.00%, thus establishing the drug products are bioequivalent. [11]

Statistical Procedure of Re-dosing

Re-dosing studies allows for the verification of statistical outliers subjects. Once the subjects identified as statistically outlier then re-dosing study should be planned. The subjects (Outlier and control) used for re-dosing study must be from the previous pivotal study. The selection of subjects in the re-dosing study will be based upon the protocol. Minimum number of subjects for redosing will be 6 or 20% of total numbers of subjects of pivotal study or which is higher.

For each re-dosed subjects, a ratio (test/reference) should be calculated for AUCt, AUChat and Cmax. Each ratios should be compared to the acceptance range from the pivotal study calculated using the formula mean ±3 SD for the log transformed data.

The following conditions are possible after redosing study and based on these only the final conclusion.

Condition-I: The ratios of AUCt, AUChat and Cmax for the suspected outliers lie outside the corresponding acceptance range and all the ratios of AUCt, AUChat and Cmax for the control subject fall within the acceptance range. Then the suspected subjects will be confirmed as outlier’s subjects and should not be dropped from the bioequivalence evaluations.

Condition-II: The ratios of AUCt, AUChat and Cmax for the suspected outliers lie inside the corresponding acceptance range and all the ratios of AUCt, AUChat and Cmax for the control subject fall within the acceptance range, and then the values for suspected outliers from the original study will be confirmed as anomalous. Therefore, the results of original study will be recalculated without suspected outliers subjects and bioequivalence conclusions will be drawn.

Condition-III: The ratios of AUCt, AUChat and Cmax for one or more of the control subject lie outside the corresponding acceptance range, and then this re-dosing study will be inconclusive with regard to suspected outlier subjects.

Condition-IV: All the suspected outlier subjects dropped or withdrawn during the course of the study. The study will be terminated immediately and no analysis of samples or pharmacokinetic and statistical analysis will be done as it will be inconclusive. [1 and 11]

SUMMARY AND CONCLUSION

As the impact of outliers cannot be controlled after the study completes, the best way to deal with them is to recognize that they can occur at random and to keep the study's power for random appearance of outliers at the design stage. To do so, it is recommended that bioequivalence studies be powered at 90% confidence interval and that such trials have at least 80% power under potential inflation of the variability estimate and for potential changes in δ of up to 5% since two formulations (Test and Reference) to be considered as bioequivalent, that the limits of a 90% confidence interval for the ratio of the geometric means of the Test and Reference primary kinetic responses ( i.e. AUC & Cmax) should be within the interval of [80%-125%]. Usually, drug regulatory agencies does not accept the exclusion of outliers subjects in a bioequivalence study; the assessment and handling of outliers subjects should be predefined in the protocol by the biostatistician in association with clinicians. [12]

“Unusual values” of pharmacokinetic endpoints, either extremely large or extremely small observations, are frequently an issue in the analysis and review of bioequivalence studies. Such extreme values, or outliers, may arise through various mechanisms, including the following:
A. Product failure (coated tablet broken; single tablet with wrong drug dosage)
B. Adverse event affecting drug absorption (e.g. vomiting, diarrhoea)
C. Laboratory error / data transcription error
D. Unusual reaction of a single subject (or of a subset of subjects) to one of the formulations (so-called “subject-by-formulation interaction”).

Because of the different regulatory attitude to outliers caused by different mechanisms, it will be useful to distinguish between mechanisms A to C, to which we refer as outliers caused by “product or process failure”, and mechanism D, to which we refer as outliers caused by “subject-by-formulation interaction”. [13]

Outlier subjects possibly will effects the BE limit. Regulatory authorities might permit exclusion from analysis of outliers caused by product or process failure, while exclusion of outliers caused by subject-by-treatment interaction generally is not acceptable.

ACKNOWLEDGEMENT
We have taken efforts in this research paper. However, it would not have been possible without the kind support and help of Fortis Clinical Research LTD. I would like to extend my sincere thanks to all of them.

REFERENCES
[2] Dr. Scott Patterson; BIOEQUIVALENCE and Statistics in Clinical Pharmacology; Villanova University, Summer III, 2010
[8] European medicines agency, Committee for Medicinal Products for Human use (chmp) doc. Ref.: cpmp/qwp/ewp/1401/98 rev. 1
[12] Farid Kianifard and William H. Swallow; A comparison of some classical approaches to outlier detection In linear regression and an approach based on adaptively-ordered recursive residuals