Assessment of Average Bioequivalence in a 2x2 Crossover Design

Tasneem Ahmad, PhD
Chief Executive Officer
Pharma Professional Services, Karachi
A CRO for Bioequivalence Studies

Professor and Ex-Chairman of Pharmaceutics
Faculty of Pharmacy, University of Karachi

Sohaib Yahya
Head of Operations
Pharma Professional Services, Karachi, Pakistan

Publication Date: Mar 30 2015
ISBN/EAN13: 1483920046 / 9781483920047
Page Count: 202
Binding Type: US Trade Paper
Trim Size: 6" x 9"
Language: English
Color: Black and White
Related Categories: Medical / Biostatistics

www.createspace.com/4215274
This book examines the biopharmaceutics, pharmacokinetics and biostatistics involved in a 2x2 crossover bioequivalence study. It enlightens every topic in required detail with solved examples of biostatistical and mathematical analysis, data presentation and results with interpretation.

This book facilitates the reader to plan a bioequivalence study through sample size and power calculation methods, verify via checking the significance of confounding effects like Period, Sequence, Subject and Formulation effects using Analysis of variance. It also helps analyze the BE studies' data with the help of numerous statistical methods including parametric and non-parametric approaches used in the evaluation of bioequivalence.

The last chapter of the book reviews the issue of outliers in the BE Studies and attitude of regulatory authorities towards inclusion or exclusion of the data of the outlying subjects. It also includes the latest techniques to detect the outliers which may impact outcome of BE studies.
Basic Statistical Considerations in Bioequivalence

Till now we have done pharmacokinetics of the data which we had arranged in an organized format. Before going ahead for computation of bioequivalence, it is better to understand the basic principles of statistics. As this book is basically for pharmaceutical scientists and not for professional statisticians, it would be useful for the readers to comprehend these things before reading further; however, if you are a perfect statistician you may skip reading this chapter.

In following lines we are trying to make the whole BE-biostatistics easy for the understanding of the readers. In this chapter the whole discussion is divided into three parts. First of all we will select a data set of one PK parameter from the data discussed in previous chapter. Secondly we will study the basic statistical terms frequently used in bioequivalence. And finally on third count we will get acquaintance with the basic notations, so enabling us to understand all the equations in the forthcoming chapter of statistical computation of Bioequivalence.

A Data-Set Selected for Biostatistical Demonstration

After this chapter onward in Part II, computation of bioequivalence with regulatory requirements is discussed and additional biostatistics applied on PK or plasma level data is dealt in Part III. To make the whole biostatistics easy for understanding of the readers in these two parts, we have employed only one set of the data throughout the book is given in Table 3-1. This makes easy and smooth the explanation of the application of mathematical or statistical formulæ. This set of data consists on AUC_{0-t} obtained from raw data given in Table 1-1 (chapter# 1), it already have been appeared in Complete PK Metrics in Table 2-1.

Definitions of Statistical Terms

Here we are explaining the most commonly used following terms in bioequivalence-statistics which are:

- Additive Model (24), ANOVA (10), Average Bioequivalence (17), Confidence Interval (13), F statistics (12), Individual Bioequivalence (19), Inter and Intra Subject (7), Mean (4), Multiplicative Model (25), Non Significant (9), Normal distribution (14),
Standard Deviation(5), t-statistics(11), Variance(6), Null Hypothesis(16), Population Bioequivalence(19), Population(2), Power(15), Sample(1), Sample Size(3), Significant(8), Standard Deviation(5), t-statistics(11), Variance(6)

The terms are neither in alphabetical order nor in any other particular order instead they are selected in gradation of difficulty, from less to more complex. And understanding of each consecutive term depends upon the understanding of previous one. We do hope that a reading of all these definitions in the given order would be certainly beneficial for non-statisticians; those interested in finding any term alphabetically can find the number in parenthesis for their term of interest given in the list of terms in preceding paragraph.

1. **Sample**: A set of objects taken or selected either randomly or deliberately using a predefined criteria from the population which possesses certain characteristics that we want to study; is known as sample. In a BE study it is the human subject included in the study on the basis of Inclusion and Exclusion criteria set in the study protocol.

2. **Population**: A complete set of objects is called population, in most cases, this set is very large and complete enumeration of all the values in population is impractical or impossible.

3. **Sample Size**: The No of objects in a statistical sample is called sample size, the sample size should be greater than the minimum No. of objects that are required to attain a desired power of statistical test, for example if the desired power of the test is 80% and sample size required to attain this power is 18 so we must have to pick at least 18 objects from the population to make inference from the sample to the population. In a BE study it refers to number of subjects or volunteers participating in the study to give blood samples after taking a dose.

4. **Mean**: The mean has different definitions depending on their context in preceding text; we will discuss here two types of means arithmetic mean and geometric mean.

5. **The arithmetic mean** is sum of all values divided by the total number of values (total count). It is also mean or average value, and is denoted by \( \bar{x} \)
6. **The geometric mean** is equal to the nth root of the product of all values. i.e. $\sqrt[n]{\prod x_i}$ and is denoted by $\bar{x}$ this can also be calculated by taking exponential of arithmetic mean of log transformed values i.e. $\bar{x} = e^{\frac{\sum \ln x}{n}}$

7. **The arithmetic mean** is always greater or equal to the geometric mean, this equality holds only when both means are calculated from same sample.

8. **Standard Deviation**: The standard deviation is a measure of dispersion from the mean value of data set, which shows how much variation exists from the mean value in both positive and negative direction, larger standard deviation indicates that the data values are far from the mean and conversely smaller standard deviation shows that values are closer to the mean, this is denoted by $\sigma$. The standard deviation is calculated by taking the square root of mean square difference of each value from the mean i.e $\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}$

9. **Variance**: Variance is a measure of dispersion of data values from the mean value of data set, this is calculated by the mean square difference of each value from the mean value i.e $Var(x) = \frac{\sum (x_i - \bar{x})^2}{n}$, the standard deviation of the data is equal to the square root of its variance $\sigma = \sqrt{Var(x)}$ or $\sigma^2 = Var(x)$, therefore it is usually denoted by $\sigma^2$

10. **Inter and Intra Subject**: Inter means between and inter subject means between subjects, when we discuss the results we will notice Inter Subject effects which means the effects between subjects.

    While Intra means within and intra subject means within subject, in intra subject effects we will discuss the factors like period sequence or treatment affecting the test variable within a subject.

    In a 2x2 crossover study both reference and test drug products are tested in each subject; it is expected that in a properly conducted good study the intra subject variability would be minimum. To
evaluate the intra-subject variability residual sum of squares for each parameter is divided by degree of freedom.

11. **Significant**: A significance level is used for statistical hypothesis tests performed on a dataset like 0.05 (for 5% level of significance), if the probability of the test is less than the significance level, we reject our null hypothesis and accept the alternative hypothesis, which suggests that at 5% level of significance the results are significant.

12. **Non Significant**: The other way around if the probability of the test is greater than the significance level we accept our null hypothesis and reject the alternative hypothesis, which suggests that at defined level of significance the results are insignificant.

13. **ANOVA**: ANOVA is an abbreviation of *Analysis of variance* which is a statistical analysis tool that separates the total variability found within a data set into two components: random and systematic factors. The random factors do not have any statistical influence on the given data set, while the systematic factors do. The ANOVA test is used to determine the impact independent variables have on the dependent variable in a regression analysis.

14. **Computation of ANOVA**: Likewise the comparisons of means in t-test the comparison of variances of two treatments is performed in F-test. In the Analysis of Variance (ANOVA) F-test is applied to assess the significance of all components of the statistical model. The purpose of ANOVA in biostatistics for evaluation of a set of BE data is same as that of t-test; the assessment of significance of sequence, period and treatment on the model. The model assumes that these effects are insignificant. Again insignificant effects reveal the creditability of study design and management.

15. **t statistics**: Student t distribution, or t-test is the most commonly used test to check whether the two means are statistically different from each other for example when we give reference product and test product (generic or placebo) to a group of patients and check that the means of both tests and reference products are statistically different we use the t statistics.
The p-value reported by the t-test represents the probability of error in accepting our assumption or null hypothesis, that both means are statistically different.

16. **F statistics**: F statistics is used to check whether the two variances are statistically different from each other, the observed value of f-test is calculated by taking the ratio of one variance to the other variance, the more the ratio deviates from 1 the more the evidence for unequal variance, the p-value reported by the f-test represents the probability of error in accepting our research hypothesis that both variances are statistically different.

17. **Confidence Interval** (CI): The confidence interval gives us the estimated range being calculated from a given set of sample data, such that we can claim that 90, 95, 99 or whatever % (depending on our level of confidence) population values of unknown parameter will lie within this range. The wider is confidence interval; higher would be the uncertainty about the unknown parameter.

18. Regulatory agencies for a 2x2 crossover BE study require the Classical 90% Confidence Interval for AUC and $C_{\text{max}}$ based on log transformed data and for $t_{\text{max}}$ based on untransformed data. These limits are computed both for difference of means and the ratio between means. Confidence interval is the key parameter on which bioequivalence is established. Other procedures, parameters or approaches are mainly to establish the creditability of the data obtained in the study and study itself.

19. **Normal distribution**: If we plot the data against its frequency the plot could be spread out on left, on right, it can be all disorderly arranged or in most cases it is spread around a central value with no bias toward left or right, and shaped like a bell, this bell-shaped distribution having mean, median and mode equals to $\mu$ and variance equals to $\sigma^2$, it is symmetric about the center and about 50% values less than the mean and 50% values are greater than the mean.

20. **Power**: The statistical power of the test is the probability of rejecting the research null hypothesis when the hypothesis is wrong. That is probability of not committing an error (Type II
error i.e. failure to reject a false null hypothesis), the higher the power of a test the lesser the chances of occurring a Type II error. In simple words in a 2x2 crossover BE study it implies to the power of not committing an error of declaring two products bioequivalent when the two are actually not bioequivalent.

21. **Null Hypothesis**: In the significance testing approach, null hypothesis is potentially rejected or disproved on the basis of data that is significantly under its assumption. It should be chosen in such a way that it allows us to conclude whether the alternative hypothesis can either be accepted or stays undecided as it was before the test. There could be more than one null hypotheses which are written as $H_{01}, H_{02}$ etc.

22. **Average Bioequivalence**: The Statistical analysis for pharmacokinetic measures, such as area under the curve (AUC) and peak concentration (Cmax), be based on the two one-sided tests procedure to determine whether the average values for the pharmacokinetic measures determined after administration of the T and R products were comparable. This approach is termed average bioequivalence and involves the calculation of a 90% confidence interval for the ratio of the averages (population geometric means) of the measures for the T and R products. To establish BE, the calculated confidence interval should fall within a BE limit, usually 80-125% for the ratio of the product averages.

The **average** BE approach focuses only on the comparison of population averages of a BE measure of interest and not on the variances of the measure for the T and R products. The average BE method does not assess a subject-by-formulation interaction variance, that is, the variation in the average T and R difference among individuals.

In contrast, **population and individual** BE approaches include comparisons of both averages and variances of the measure. The population BE approach assesses total variability of the measure in the population. The individual BE approach assesses within-subject variability for the T and R products, as well as the subject-by-formulation interaction.
23. **Sequence, Period and Treatment Effect:** The comparison of means of two treatments provides the basis of t-test and thus the p-value is established. The p value of different elements of the model must be greater than 0.05 to establish non-significance at 5 percent level of probability. The effects of sequence, period and treatment also known as **Fixed Effects.** In a BE study these effects are assessed by calculating MVUE (minimum variance of unbiased estimator), V(Est), 95% CI, t-value and p-value. In an authentic, well designed and managed study the sequence and period effects must be insignificant.

24. **Additive Model:** Through statistical model we determine that any PK parameter is composed of mean of the population plus effects of all different variance contributing factors and error term. When on log scale we write this model it takes additive form of all actors i.e. $y_{ijk} = \mu + S_{ik} + P_j + F(j, k) + R(j, k) + \epsilon_{ijk}$

25. **Multiplicative Model:** When we write the statistical model for any PK parameter linear scale it becomes multiplicative as: $y_{ijk} = \mu s_{ik} p_j f(j, k) r(j, k) \epsilon_{ijk}$

**Statistical Notations**

<p>| $H_0$ | Null Hypothesis, in the significance testing approach, null hypothesis is potentially rejected or disproved on the basis of data that is significantly under its assumption. It should be chosen in such a way that it allows us to conclude whether the alternative hypothesis can either be accepted or stays undecided as it was before the test. There could be more than one null hypotheses which are written as $H_{01}$, $H_{02}$ etc. |
| $H_a$ | Alternative Hypothesis is potentially accepted or approved if null hypothesis is rejected or disproved on the basis of data. |
| $1SCV$ | Intra Subject Coefficient of Variation |
| $LD_{inf}$ | Lower bound for difference |
| $LD_{sup}$ | Upper bound for difference |
| $LR_{inf}$ | Lower bound for ratio |</p>
<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$LR_{sup}$</td>
<td>Upper bound for ratio</td>
</tr>
<tr>
<td>$n_1$</td>
<td>Number of subjects in sequence 1.</td>
</tr>
<tr>
<td>$n_2$</td>
<td>Number of subjects in sequence 2.</td>
</tr>
<tr>
<td>$SQ_{inter}$</td>
<td>Inter group or between group sum of squares</td>
</tr>
<tr>
<td>$SQ_{intra}$</td>
<td>Intra group or within group sum of squares</td>
</tr>
<tr>
<td>$SQ_{residual(carryover)}$</td>
<td>Sum of squares of sequence effect</td>
</tr>
<tr>
<td>$SQ_{drug}$</td>
<td>Sum of squares of drug effect</td>
</tr>
<tr>
<td>$SQ_{period}$</td>
<td>Sum of squares of period effect</td>
</tr>
<tr>
<td>$QM_{inter}$</td>
<td>Inter group or between group mean square</td>
</tr>
<tr>
<td>$QM_{intra}$</td>
<td>Intra group or within group mean square same as $MS_{intra}$</td>
</tr>
<tr>
<td>$QM_{residual(carryover)}$</td>
<td>Mean square of sequence effect</td>
</tr>
<tr>
<td>$QM_{period}$</td>
<td>Mean square of period effect</td>
</tr>
<tr>
<td>$MS_{intra}$</td>
<td>Intra group or within group mean square</td>
</tr>
<tr>
<td>$R_R$</td>
<td>Carryover effect of Reference product</td>
</tr>
<tr>
<td>$R_T$</td>
<td>Carryover effect of Test product</td>
</tr>
<tr>
<td>$S^2_{TR}$</td>
<td>Covariance of test and reference</td>
</tr>
<tr>
<td>$S^2_{TT}$</td>
<td>Variance of test</td>
</tr>
<tr>
<td>$S^2_{RR}$</td>
<td>Variance of reference</td>
</tr>
<tr>
<td>$T_{AH}$</td>
<td>T value calculated using Anderson and Hauk’s method and it follows a non-central t distribution</td>
</tr>
<tr>
<td>$T_L$</td>
<td>Lower value of calculated T-test from sample data.</td>
</tr>
<tr>
<td>$T_U$</td>
<td>Upper value of calculated T-test from sample data.</td>
</tr>
<tr>
<td>$Y_{ijk}$</td>
<td>Pharmacokinetic parameter of $i^{th}$ subject in $j^{th}$ period and $k^{th}$ sequence</td>
</tr>
<tr>
<td>$\mu_T$</td>
<td>Mean of Test values of any PK parameter such as $AUC, C_{Max}, T_{Max}, T_{1/2}$ etc,</td>
</tr>
<tr>
<td>Symbol</td>
<td>Description</td>
</tr>
<tr>
<td>--------</td>
<td>-------------</td>
</tr>
<tr>
<td>$\mu_R$</td>
<td>Mean of reference values of any PK parameter such as $AUC, C_{Max}, T_{Max}, T_{1/2}$ etc.</td>
</tr>
<tr>
<td>$\bar{Y}_T$</td>
<td>Least square means for the test formulation</td>
</tr>
<tr>
<td>$\bar{Y}_R$</td>
<td>Least square means for the reference formulation</td>
</tr>
<tr>
<td>$\hat{\sigma}_d$</td>
<td>Standard deviation calculated by taking square root of mean square error i.e. $\hat{\sigma}<em>d = \sqrt{\frac{\sigma_d^2}{2}} = \sqrt{\frac{MS</em>{\text{intra}}}{2}}$</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>Level of significance (probability of rejection)</td>
</tr>
<tr>
<td>$1 - 2\alpha$</td>
<td>Since $\alpha$ is the probability of rejection $1 - 2\alpha$ is the probability of acceptance in two tailed test.</td>
</tr>
<tr>
<td>$\theta_{\text{inf}}$</td>
<td>Observed Lower bound for difference</td>
</tr>
<tr>
<td>$\theta_{\text{sup}}$</td>
<td>Observed Upper bound for difference</td>
</tr>
<tr>
<td>$\delta_{\text{inf}}$</td>
<td>Observed Lower bound for ratio</td>
</tr>
<tr>
<td>$\delta_{\text{sup}}$</td>
<td>Observed upper bound for ratio</td>
</tr>
<tr>
<td>$\mu$</td>
<td>Arithmetic mean</td>
</tr>
<tr>
<td>$\varepsilon_{ijk}$</td>
<td>Residuals (Observed value minus estimated value)</td>
</tr>
<tr>
<td>$n_k$</td>
<td>Number of subjects in $k^{th}$ period</td>
</tr>
<tr>
<td>$t(\alpha, df)$</td>
<td>$T$ Tabulated value when probability (level of significance) = $\alpha$ and degree of freedom = $df$.</td>
</tr>
</tbody>
</table>
SUBJECT INDEX

68% Range, 119, 120
Calculations for AUC_{0-t}, 124

90% Confidence Interval for:
- Data on Log Transformed Scale, 70
- Difference of Means on Linear Scale, 68
- Difference of Means on Log Scale, 70
- Ratio between Means on Linear Scale, 69
- Ratio of means on Log Scale, 71

Additive Model, 62, 44, 57
Analysis of Variance (ANOVA)
- Computation of, 41
- Consolidated Tables, 95
- GLM tool for ANOVA, 89 of PK data, 79, 92

Anderson and Hauck's Test, 138
Andrews Curves, 166
Area Under the Curve up to Last sampling Time AUC_{0-t}, 25
Arithmetic mean, 39
Average Bioequivalence, 43
Bayesian method, 152
Bayesian Statistical Procedures, 154

Bioequivalence (BE)
- Computation Strategies, 57
- Computation, 49
- Limits for Difference of means, 120
- Limits for Ratio of means, 120
- Data Compilation, 3
- Data presentation, 4

Carryover Effect, 60
Clinical Rationale for Log Transformation, 54

Comparative Account of PK Parameters, 34
Profiles, 15
Comparison software, 3
Complete PK metrics, 22
Complete Raw Data of Plasma concentration, 6
Computation of ANOVA, 41
Confidence Interval (CI), 42
Consolidated ANOVA Tables, 95
Construction of Experimental Model, 58
Criterion of Bioequivalence in 2x2 Crossover Model, 62
Crossover Design, 57
Data Compilation, 3
Data presentation, 4
Decoding of analytical data, 3
Deviation in sampling-time, 7
Drug Effect (In Absence of Carryover Effect), 107
Drug Effect (In Presence of Carryover Effect), 110
Empirical power of a BE study, calculation, 98
Estimated Distance Test, 161
Estimation of AUC_{t-\infty}, 27
Evaluation of Significance, 59
Exclusion of Outlier Subjects, 158
Experimental Models, 57
Extent of Absorption, 25
F statistics, 42
Fieller’s Confidence Interval, 133
Fixed Effects, 44
Fixed Fieller’s Confidence Interval, 136
Geometric Means, 40, 55, 119
Calculations, 122
GLM tool for ANOVA, 89
Hodges-Lehmann Estimator, 146
Implications of Fixed Effect, 101
Inter and Intra Subject, 40
Inter Subject Variability, 80
Intra Subject Variability, 82, 97
Likelihood Distance Test, 160
Logarithmic Transformation of PK Data, 54
Mandallaz and Mau’s Method, 152
Mean, 39
Mean Comparative Profiles, 50
Multiplicative Model, 62, 44, 57
Non significant results, definition, 41
Nonparametric Tests, 141
Normal distribution, 42
 Normally Distributed Data, 60
Not normally distributed data, 62
Null Hypothesis, 43
Outlier Subjects
Analysis, 159
Definition 155
Detection through Andrew Curves, 167
Detection through principal components, 171
due to Product Failure, 158
due to Subject-by-Formulation Interaction, 158
in Bioequivalence Data, 155
in different Situations, 157
Purpose of Identifying, 155
Parametric Tests, 133
Peak Concentration; C_{max}, 23
Period Effect, 60 112
Pharmacokinetic Parameters, 21
Pharmacokinetic Rationale for Log Transformation, 54
PK Summary in BE Study, 28
Profiles, 9
Point estimator, 122
Population, 39
Population and individual BE approaches, 43
Power of a BE Study, 42, 98
Power of Study Evaluated on Untransformed Data, 100
Presentation of Raw Data, 3
Principal Component Analysis for Outliers, 169
Procedure for obtaining an empirical power of stdudy, 88
Proportion of AUC0-t in AUC0-\infty, 31
Rationale for 80-125% criteria, 56
Rationale for a Crossover Design, 57
Rodda And Davis Method, 150
Sample, 39
Sample Size, 39
SAS® script of GLM Procedure, 89
Schuirmann’s Two One Sided Test, 75
Selection of Terminal Points for Estimation of Ke, 28
Sequence-by-Period plots, 104
Significance defined, 41
Standard Deviation, 40
Statistical Considerations, 37
Statistical Models, 57
Statistical Notations, 44
Statistical Power of Study for different PK parameters, 42, 99
Statistical Terms, definitions, 37
Strategies for BE Computation, 60
Subject by Period Plots, 87
Subject by Treatment Plots, 36
Subject-Effect, 59
t statistics, 41
Testing of Carryover Effect, 84
Testing of Formulation Effect, 85
Testing of Period Effect, 85
Testing of Subject Effect, 83
Tests to Detect Outliers, 160
Time of Maximum Concentration; $T_{\text{max}}$, 25
Trapezoidal method, 25
Treatment-by-Period plots, 103
Types and Impact of Outliers, 156
Variance, 40
Variance Contribution of Different Components of the Model, 59
Wilcoxon Rank Sum Test, 141
Wilcoxon–Mann-Whitney Two One Sample Tests, 143

www.createspace.com/4215274