2009

NONLINEAR MODELS IN MULTIVARIATE POPULATION BIOEQUIVALENCE TESTING

Bassam Dahman
Virginia Commonwealth University

School of Medicine
Virginia Commonwealth University

This is to certify that the dissertation prepared by Bassam A Dahman entitled "NONLINEAR MODELS IN MULTIVARIATE POPULATION BIOEQUIVALENCE TESTING" has been approved by his or her committee as satisfactory completion of the thesis or dissertation requirement for the degree of
DOCTOR OF PHILOSOPHY

Ramakrishnan, Viswanathan PhD, Director, School of Medicine

Elswick, Ronald K. Jr. Ph.D., Co-Director, School of Medicine

Barr, William H. Ph.D, School of Pharmacy

Masho, Saba W. Ph.D., School of Medicine

Mcclish, Donna K. Ph.D., School of Medicine

Jerome F. Strauss, III, M.D., Ph.D., Dean, School of Medicine

Dr. F. Douglas Boudinot, Dean of the School of Graduate Studies

November, 17, 2009

"NONLINEAR MODELS IN MULTIVARIATE POPULATION BIOEQUIVALENCE

TESTING"

A Dissertation submitted in partial fulfillment of the requirements for the degree of DOCTOR OF PHILOSOPHY at Virginia Commonwealth University.

by

BASSAM A. DAHMAN
MS, Virginia Commonwealth University, 2007
BSc, Kuwait University, Kuwait, 1982

Director: RAMAKRISHNAN, VISWANATHAN, PH.D.
ASSOCIATE PROFESSOR, DEPARTMENT OF BIOSTATISTICS

Director: ELSWICK,RONALDK.JR., PH.D.
ASSOCIATE PROFESSOR, DEPARTMENT OF BIOSTATISTICS

Virginia Commonwealth University
Richmond, Virginia
November 2009

Acknowledgements

To my wife Elham: Thank you for all the love, help and sacrifices. I love you.

Special thanks to the members of my graduation committee. Drs Ramesh, Elswick, McClish, Barr and Masho, you have always been there when I needed.

Thanks for Dr. Bradley and Dr. Smith for all the support to reach this goal.

Table of Contents

Page

List of Tables

Page

List of Figures

Page

Abstract

NONLINEAR MODELS IN MULTIVARIATE POPULATION BIOEQUIVALENCE TESTING

By Bassam A Dahman, MS

A Dissertation submitted in partial fulfillment of the requirements for the degree of PhD at Virginia Commonwealth University.

Virginia Commonwealth University, 2009

Major Director: Ramakrishnan, Viswanathan Ph.D.
Associate Professor, Department of Biostatistics

In this dissertation a methodology is proposed for simultaneously evaluating the population bioequivalence (PBE) of a generic drug to a pre-licensed drug, or the bioequivalence of two formulations of a drug using multiple correlated pharmacokinetic metrics. The univariate criterion that is accepted by the food and drug administration (FDA) for testing population bioequivalence is generalized.

Very few approaches for testing multivariate extensions of PBE have appeared in the literature. One method uses the trace of the covariance matrix as a measure of total variability, and another uses a pooled variance instead of the reference variance. The former ignores the correlation between the measurements while the later is not equivalent

to the criterion proposed by the FDA in the univariate case, unless the variances of the test and reference are identical, which reduces the PBE to the average bioequivalence.

The confidence interval approach is used to test the multivariate population bioequivalence by using a parametric bootstrap method to evaluate the $(1-\alpha)100\%$ confidence interval. The performance of the multivariate criterion is evaluated by a simulation study. The size and power of testing for bioequivalence using this multivariate criterion are evaluated in a simulation study by altering the mean differences, the variances, correlations between pharmacokinetic variables and sample size. A comparison between the two published approaches and the proposed criterion is demonstrated. Using nonlinear models and nonlinear mixed effects models, the multivariate population bioequivalence is examined. Finally, the proposed methods are illustrated by simultaneously testing the population bioequivalence for *AUC* and $C_{\max}$ in two datasets.

1 Introduction

Bioequivalence (BE) studies are an essential component of the applications for approval of generic drugs or new formulations of previously licensed drugs submitted to the regulatory agencies. Any two drugs are deemed to have the same therapeutic effect if they have the same rate of absorption, the same maximum concentration or level of the pharmacologically active material at the site of action, and the same total amount available before the drug is completely excreted. This is considered fundamental for bioequivalence and is sometimes referred to as the fundamental assumption for bioequivalence (Chow and Liu, 2009).

Bioequivalence is closely related to bioavailability (BA) in drug testing. Both are required by the Food and Drug Administration (FDA) for the approval of drugs, and therefore essential in investigational new drug applications (INDs), new drug applications (NDAs), abbreviated new drug applications (ANDAs), and their supplements. The FDA regulates BE studies and the regulations governing these studies are provided in part 320 of 21 CFR (FDA, 2000).

Bioavailability is defined by the FDA in 21 CRF 320.1 as:

"The rate and extent to which the active ingredient or active moiety is absorbed from a drug product and becomes available at the site of action. For drug products that are not intended to be absorbed into the bloodstream, bioavailability may be assessed by measurements intended to reflect the rate and extent to which the active ingredient or active moiety becomes available at the site of action."

Bioequivalence is defined by the FDA in 21 CRF 320.1 as:

"The absence of a significant difference in the rate and extent to which the active ingredient or active moiety in pharmaceutical equivalents or pharmaceutical alternatives becomes available at the site of drug action when administered at the same molar dose under similar conditions in an appropriately designed study."

Establishing bioavailability (BA) of any drug is a benchmark effort with comparisons between formulations and routes of admission such as oral solution, oral suspension, or an intravenous formulation. Whereas, demonstrating BE is usually formal and use comparative statistical tests that uses specific criteria for comparisons (FDA, 2000). To license a generic drug it is essential to demonstrate that the newly proposed drug or formulation contains the same active pharmaceutical moiety with the same dose and strength and it has the same route of administration. The producer of the test (generic) drug should show that the release of an active substance from the test drug product and the subsequent absorption into the systemic circulation is similar to the release and absorption of the reference drug. There is a need to demonstrate that the bioavailability of the proposed drug is similar to that of the approved and listed drug, the reference drug. To test for the similarity of the bioavailability of the two drugs, bioequivalence testing is required.

1.1. Types of Bioequivalence Measures:

1.1.1 Average Bioequivalence

Average bioequivalence (ABE) is the most widely used measure of BE in the pharmaceutical industry and research. It compares between the means or averages of the test and reference drug distributions. Bioequivalence is concluded if the confidence

interval of difference between the means of the reference and the test means falls within a predefined range.

This measure ignores the difference in the variance between the test and reference drug distributions. Ignoring the differences in the variance does not guarantee that the two drugs, reference and test, could be used interchangeably in terms of safety and efficacy.

1.1.2 Population Bioequivalence

Population bioequivalence (PBE) is another measure of bioequivalence that was proposed to evaluate prescribability of the drug. Prescribability of a drug is defined as the ability to get the same effect by prescribing the brand-name drug or its generic drug to a new patient (Chow and Liu, 2009). As mentioned in section 1.1.1 the ABE focuses only on the comparison of population averages of the rates or extent of absorption and not on the variances of these measures. In contrast, PBE includes comparisons of both the means and variances of the measures. Therefore, the PBE approach assesses total variability of the measure in the population (Hauk and Anderson, 1992, FDA 1997).

1.1.3 Individual Bioequivalence (IBE)

Individual bioequivalence was proposed to evaluate switchability of two drugs. Switchability (Anderson, 1993; Liu and Chow, 1995) is defined as the ability to switch from the brand-name drug to a generic drug while guaranteeing the same efficacy and safety to the patient who was using the brand-name drug. It is recommended to assess bioequivalence within individual subjects to assess switchability. Intra-subject variances are included in the comparison between the test and reference drugs when assessing IBE.

1.2. The Hypothesis of bioequivalence:

Let δ be a BE measure of interest, usually, in the case of ABE, the difference between the means of the pharmacokinetic parameters (PK) of the two drugs being compared. Let θ_1 and θ_2 be two pre-defined bioequivalent limits. Then the two-sided hypotheses to assess bioequivalence are:

$$H_0 : \delta \le \theta_1 \text{ or } \delta \ge \theta_2 \quad vs \quad H_a : \theta_1 < \delta < \theta_2$$

These hypotheses can also be rewritten as two one-sided hypotheses as

$$H_{01} : \delta \le \theta_1 \qquad vs \qquad H_{a1} : \delta > \theta_1$$

and

$$H_{02} : \delta \ge \theta_2 \qquad vs \qquad H_{a2} : \delta < \theta_2 .$$

If both null-hypotheses are rejected at level α, there is evidence of bioequivalence at 100(1-α)% significance. If δ is the difference between the average PK parameters of two drugs, the 100(1 – 2α)% confidence interval for δ could be constructed as

$$\left((\hat{\mu}_T - \hat{\mu}_R) - z_\alpha \hat{\sigma}_{\hat{\mu}_T - \hat{\mu}_R}, (\hat{\mu}_T - \hat{\mu}_R) + z_\alpha \hat{\sigma}_{\hat{\mu}_T - \hat{\mu}_R} \right) \qquad (1)$$

where $\hat{\mu}_T - \hat{\mu}_R$ is the difference between the two estimated means of the PK parameter for drugs T and R, and $\hat{\sigma}_{\hat{\mu}_T - \hat{\mu}_R}$ is the estimated standard deviation of the difference between the means. Then testing the two hypotheses simultaneously is equivalent to comparing the confidence interval to the bioequivalence acceptance region for $(\theta_1, \theta_2) = (-0.2\hat{\mu}_R, 0.2\hat{\mu}_R)$, where $\hat{\mu}_R$ is the estimated mean of the PK of the reference drug.

The estimates of the differences in the means and the $100(1 - 2\alpha)\%$ confidence intervals, and the estimate of the reference mean could be obtained from any experimental design such as parallel or cross-over trials.

1.3. Assessing Bioequivalence

The goal of bioequivalence studies is to test the hypothesis that two or more drugs are bioequivalent in terms of PK parameters. Experiments designed to assess bioequivalence of drugs take a wide range of measurements of the levels of the drug in the blood or plasma over a period of time. The key parameters in bioequivalence testing are shown as part of a typical plasma concentration time profile in Figure 1 (Mehrotra 2007). The figure shows also the minimum effective concentration (MEC) which is the minimum concentration to produce the desired pharmacological effect; and the maximum tolerable concentration (MTC) beyond which toxic and adverse events are intolerable. From such concentration-time data or curves several PK parameters such as the rate of absorption (k_a), and (T_{max}) the time until the maximum concentration (C_{max}) is reached, and total available dose (area under the blood level-time curve (AUC)) are either measured or estimated.

For example, the AUC resultant from a single dose of a drug formulation is commonly assessed with the linear trapezoidal method (Berger RL, 1996, Gibaldi, 1982). The average of two subsequent plasma-concentrations (C_i and C_{i-1}) is calculated, then it is multiplied by the difference between the consecutive time points (t_i and t_{i-1}). The partial areas are then summed to produce the AUC

$$AUC_{0-t} = \sum_{i=1}^{t} \left(\frac{C_i + C_{i-1}}{2} \right) (t_i - t_{i-1}). \quad (2)$$

The total area under the curve would be estimated as

$$AUC_{0-\infty} = AUC_{0-t} + \frac{C_{last}}{k_e}, \quad (3)$$

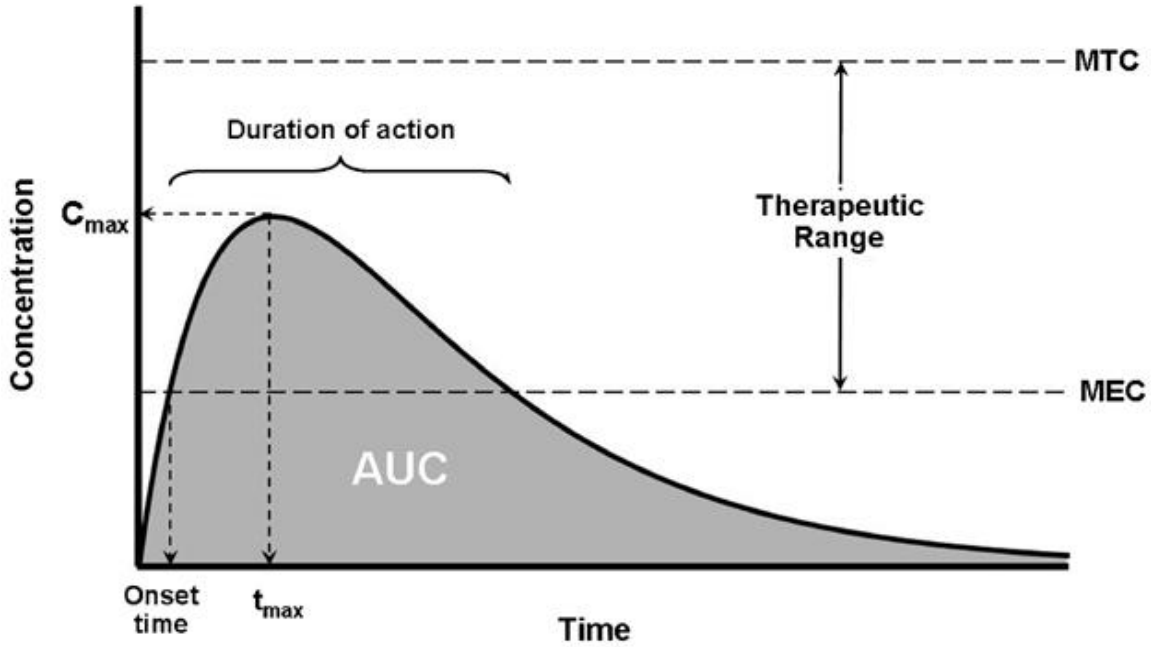
where k_e is the elimination constant, which describes the rate of reduction of the log plasma-concentration per unit time. This constant could be estimated as the value of the slope of the reduction of the log concentration by time. Thus it could be calculated using the elimination half-life ($t_{1/2}$) which is the time it takes for the concentration of the drug to fall to half its concentration. Suppose the concentration of the drug dropped for C_1 measured at time t_1 to its half C_2 at time t_2, then the time difference $t_2 - t_1$ is noted as $t_{1/2}$ which is known as the half time. Then the elimination constant could be calculated as rate of this drop as:

$$k_e = \frac{\log C2 - \log C1}{t_{1/2}} = \frac{\log(C2/C1)}{t_{1/2}} = \frac{\log(C2/C1)}{t_{1/2}} = \frac{\log(1/2)}{t_{1/2}} = \frac{0.693}{t_{1/2}}$$

The trapezoidal formula used for AUC is an approximation of the total area under the curve. The further the distance between the time points when the blood concentrations are measured, the larger the inaccuracy of the calculated AUC. Depending on the original profile of blood concentration curve, this could be an underestimation in some cases and an overestimation in other cases.

$C_{\max}$ is measured as the highest observed concentration. Although this measure

Figure 1 Typical plasma concentration time profile after oral administration



Cmax, maximum concentration; tmax, time to Cmax; AUC, area under the curve; MEC, minimum effective concentration; MTC, maximum tolerated concentration.

rarely coincides with the true $C_{\max}$ (the estimate is biased downward), this measure is widely used in bioequivalence determinations. It is not unusual for plasma concentration profiles that reach a peak then the concentration drops, only for the concentration to peak again. The second peak may be higher or lower than the first peak. In these situations, the $C_{\max}$ is usually estimated as the concentration of the highest peak in profile. However, the first peak may be used as the estimate of $C_{\max}$ when used as a measure of

absorption. The $T_{\max}$ is defined as the time when $C_{\max}$ is observed, and similar to $C_{\max}$, it is rarely accurate.

The rate of absorption could be measured in two ways. The first method is based on the linear fit of the first few points (at least three points) from beginning of the concentration profile to the first peak. This absorption constant, usually noted as k_0, is calculated as the slope of that linear fit. The number of points chosen for this fit is based on the R-squares of the fits. Other methods uses nonlinear models to estimate the absorption rate constant denoted as k_a. These estimates of the bioequivalence parameters are non-model based calculations and they cannot account for the uncertainty in measuring the drug concentration. Alternatively, these parameters could also be estimated by fitting mechanistically meaningful non-linear models.

By assumption, if the difference between the new test drug and the reference drug in terms of the means of these pharmacokinetic parameters are within a pre-defined acceptable magnitude then the drugs are deemed bioequivalent.

1.4. Extensions considered in this dissertation

Most of the pharmacokinetic parameters are derived from the same blood concentration-time profile. This makes them correlated and therefore individually testing each measure for BE is not optimal. Several approaches to extend the ABE methods to multivariate situation have been proposed (Brown, 1995; Berger and Hsu, 1996; Brown, 1997; Munk and Pflujer, 1999; Wang, 1999; and Tamhane and Logan, 2004). However, for the population BE only one method has been suggested and investigated in the literature for the multivariate bioequivalence (Chervoneva, 2007). However, this

approach is based only on the trace of the variance covariance matrix. Since the trace is the sum of the diagonal elements of the matrix, this approach essentially ignores the correlation between the pharmacokinetic measures and therefore could not be considered an extension of the univariate approach to the multivariate situation. Another method was suggested by Dragalin et al. (2003), in which the Kullback–Leibler divergence (KLD) was used to evaluate the multivariate case of IBE. They proposed an analogous method to be applied to evaluate the PBE. Their method was not studied, and was not accepted by the FDA.

Another aspect of the bioequivalence methods that is often ignored in the literature is the fact that many of the pharmacokinetic measures are derived from the concentration-time curves and therefore there is uncertainty in the estimates. When these measures are based on a single compartment non-linear fit of the data it is possible to estimate this uncertainty and incorporate it in the BE tests. One such method has been suggested in the literature but it considers each PK measure individually (Panhardt 2007). Multivariate extensions are considered in this dissertation.

In Chapter 2, a review of the literature on the topics of BE are presented. The chapter will discuss the methods of BE testing, types of BE tests, the PK used in testing for bioequivalence. Methods of estimating these PK are also presented and compared.

Chapter 3 will present the development of a methodology to simultaneously evaluate population BE using multiple PK. A multivariate extension of the FDA approved PBE criterion will be derived. A method to implement the proposed multivariate criterion will be presented. Also in this chapter a method to simultaneously

estimate the PK parameters from a nonlinear mixed effects models and test for BE using the proposed multivariate method will be presented.

In Chapter 4 the design and results of a simulation study to evaluate the multivariate criterion in testing for PBE are presented. The size and power of testing BE using this criterion are evaluated. The definition of the acceptable BE regions and regulatory limits are discussed, especially with the introduction of the covariance as a new factor in defining these regions.

In Chapters 5 and 6, the material from chapter 3 and 4 are consolidated in the form of two journal articles. The first paper (chapter 5) will introduce the multivariate extension to the PBE that accounts for the correlation between the pharmacokinetic variables. This paper will also include an illustration of the method using an existing data and a comparisons to other methods available in the literature. The second paper (Chapter 6) will focus on the nonlinear methods.

In Chapter 7, summary and conclusions of the finding of this research are presented along with the limitations and future work. Several appendices are supplementing this study. Appendix I provides mathematical presentation of PBE as a distance measure. The SAS programs used for this dissertation and tables of the data used in the examples. Since the chapters are written in the form of journal manuscripts, the mathematical derivations of the PBE criterion are presented in Appendix II. The distribution histograms, and blood concentration profiles for all subjects in this study are also displayed in the appendices C, E and F.

2 Background

2.1. Introduction

Bioequivalence studies are used in the development of generic drugs and the development of new formulations of drugs that were previously approved. Developing a new drug and obtaining approval from the Food and Drug Administration (FDA) requires multiple clinical trials to document the toxicity and the efficacy of the pharmacologically active ingredients of the new drug. A generic formulation of an approved compound is not subject to the multiple clinical trial requirement of a new compound because it is assumed that the active ingredients of the generic drug have the same toxicity and therapeutic efficacy as the approved drug. Thus, a generic drug needs only to demonstrate bioequivalence to the approved drug; once bioequivalence is demonstrated, it is also assumed that the therapeutic efficacy is similar between the approved and generic drugs. Thus the bioequivalence studies are designed to establish this expected similarity of the generic drugs to the approved drug having the same active ingredients.

Experiments are designed to measure the concentration of the active ingredient of both the test and the reference drugs in the blood, or in the biological site of action, at appropriate time intervals. A profile of the concentration of the drug over time is then generated. Several pharmacokinetic parameters are estimated from the concentration by time profiles and are used to quantify bioavailability. In general the PK parameters of interest are, the maximum absorbed drug (AUC), the time ($T_{\max}$) at which the highest

concentration in the blood ($C_{\max}$) occurs, rates of absorption (k_0 and k_a), rate of elimination(k_e), and blood or plasma half lives($t_{0.5}$) are calculated.

2.2. BE a function of distance

For any given metric two drugs are defined to be bioequivalent if

$$\delta^2 \le \theta, \tag{4}$$

where θ is a predefined constant, and for any given metric, δ is a critical value obtained from the distribution of a distance function of the new and the reference drugs. The upper limit (θ) is often prescribed by regulatory agencies. The δ in general is the (1 - α)*th* percentile of the distribution of the distance function for a given confidence level α. For instance, if the 90% confidence interval of the distance function falls completely within the interval $[-\theta, \theta]$, bioequivalence is concluded. This procedure is equivalent to testing two one-sided hypotheses such as those mentioned in section 1.2 each at level α using an analogous test (Schuirmann, 1987).

The ABE test focuses on the differences in the means of the pharmacokinetic parameters.

$$\delta = \mu_T - \mu_R, \tag{5}$$

The US FDA's guideline suggests comparing this distance measure with a BE predefined limit (θ) that is equal to 20% of the reference mean. BE is concluded when the confidence interval of the distance is within the BE acceptance region, i.e. $(-0.2\hat{\mu}_R, 0.2\hat{\mu}_R)$, where $\hat{\mu}_R$ is the estimated mean of the PK of the reference drug. The distribution of the PKs used in bioequivalence testing like $C_{\max}$ and AUC are known to

be lognormal, so the log-transformed parameters are often used in evaluating bioequivalence. The confidence interval of the distance measure δ is estimated as

$$\left((\hat{\mu}_T - \hat{\mu}_R) - z_\alpha \hat{\sigma}_{\hat{\mu}_T - \hat{\mu}_R}, (\hat{\mu}_T - \hat{\mu}_R) + z_\alpha \hat{\sigma}_{\hat{\mu}_T - \hat{\mu}_R}\right) \quad (6)$$

where $\hat{\mu}_T - \hat{\mu}_R$ is the difference between the two estimated means of the PK parameter for drug T and R, and $\hat{\sigma}_{\hat{\mu}_T - \hat{\mu}_R}$ is the estimated standard deviation of the difference between the means.

This method of evaluating bioequivalence, does not account for differences in the variability between the reference and test drugs. PBE was proposed to evaluate prescribability of the drug. Prescribability of a drug is defined as the ability to get the same effect by prescribing the brand-name drug or its generic drug to a new patient (Chow and Liu, 1992). In contrast to average BE, the PBE includes comparisons of the means and the total variability of the pharmacokinetic measures between the reference and test drugs (Hauk and Anderson, 1992).

The PBE was introduced by FDA in 1997 as an alternative method of testing BE. The PBE is a scaled distance between the test and reference distributions with respect to the first two moments while the ABE is simply the difference between the first moments only. The PBE may be thought of as the ratio of two expected squared distances where the numerator is the expected squared distance between the reference and the test and the denominator is the expected squared distance between two reference observations. Bioequivalence, then is determined by the ratio of the two expected squared differences is within a predefined distance, θ, from unity. That is,

$$\frac{E\left[\left(y_T - y_R\right)^2\right]}{E\left[\left(y_R - y_{R'}\right)^2\right]} - 1 \le \theta \tag{7}$$

where y_T is a random variable denoting the test PK metrics, y_R and $y_{R'}$ are two realizations of the reference random variable and *E* represents the expectation.

The univariate PBE criterion in (7), by substituting the unit ratio of the denominator term for the 1, could be redefined as (Sheiner 1992, Schall and Luus 1993),

$$\frac{E\left[\left(y_T - y_R\right)^2\right] - E\left[\left(y_R - y_{R'}\right)^2\right]}{E\left[\left(y_R - y_{R'}\right)^2\right] \Big/ 2} \le \theta . \tag{8}$$

Rewriting equation (8) in terms of the population mean and variance, it reduces to,

$$C = \frac{\left(\mu_T - \mu_R\right)^2 + \sigma_T^2 - \sigma_R^2}{\sigma_R^2} \le \theta . \tag{9}$$

where μ_R and μ_T are the means of the reference and the test random variables respectively, and σ_R^2 and σ_T^2 are the population variances of the reference and test pharmacokinetics respectively. Thus, the δ^2 from the original inequality in (4) is a function both of a distance metric of the means as well as the variances. The hypothesis test form of PBE uses the hypotheses $H_0 : C > \theta$ vs $H_a : C \le \theta$. Bioequivalence is concluded with $100\left(1-\alpha\right)\%$ confidence if $\hat{C}_{(1-\alpha)} \le \theta$, where $\hat{C}_{(1-\alpha)}$ is the estimate of the upper limit of the one-sided $100(1-\alpha)th$ confidence interval of the PBE criterion defined in (9) using the maximum likelihood estimates (mle's) of the means and variances.

Extending this to more than one metric requires accommodation of the correlation. For example, suppose that the blood absorption coefficient (K_a), and the time (T_{max}) until the maximum concentration (C_{max}) of the blood concentration is reached, and the area under the blood concentration curve (AUC), are all calculated from the same blood concentration-time profile. In this case, the assumption of independence in testing bioequivalence using multiple tests for each of the four parameters is not justifiable. Clearly, the correlations among these variables should be incorporated in the multivariate tests of bioequivalence.

2.3. Multivariate extensions of BE assessment

Multiple multivariate extensions for the average BE (Brown, 1995; Berger and Hsu, 1996; Brown, 1997; Munk and Pflujer, 1999; Wang, 1999; and Tamhane and Logan, 2004) have been proposed in the literature. However, there are only a couple that deal with the multivariate PBE. The first notable exception is Dragalin et al. (2003), in which the Kullback–Leibler divergence (KLD) is used as a measure of discrepancy between the distributions of the two formulations. They propose a generalization of average and PBE measures, and generalize it to the multivariate situation. Their multivariate method could be summarized as follows. Consider a multivariate random variable $\mathbf{Y}$ representing a set of PK metrics. Suppose $\mathbf{Y}$ is distributed as normal with mean vector $\boldsymbol{\mu}$ and variance covariance matrix $\boldsymbol{\Sigma}$. Let T and R represent test and reference treatments, respectively. The criterion proposed by Dragalin et al. (2003) is based on the following inequality

$$\frac{1}{2}trace\left\{\left[\left(\boldsymbol{\mu}_T-\boldsymbol{\mu}_R\right)\left(\boldsymbol{\mu}_T-\boldsymbol{\mu}_R\right)'+\boldsymbol{\Sigma}_T+\boldsymbol{\Sigma}_R\right]\left[\boldsymbol{\Sigma}_T^{-1}+\boldsymbol{\Sigma}_R^{-1}\right]\right\}-2p<\theta \qquad (10)$$

Here, the left hand side (LHS) of the equation is the KLD. Two formulations are declared bioequivalent if the upper bound of a level-α confidence interval for the KLD is less than a given specific value, θ. This criterion does not reduce to the univariate criterion proposed by the FDA in equation(9); instead it reduces to

$$\frac{1}{2}\left[\frac{(\mu_T-\mu_R)^2+\sigma_T^2-\sigma_R^2}{\sigma_R^2}+\frac{(\mu_T-\mu_R)^2+\sigma_R^2-\sigma_T^2}{\sigma_T^2}\right] \qquad (11)$$

Thus, this criterion may be seen as the average of two terms where the first term is the same measure of distance scaled by the reference variance proposed by FDA. The second term is similar except that it is scaled by the variance of the test. This criterion is equivalent to the FDA proposed criterion only if the reference and test variances are equal, in which case it is only a measure of the squared mean distances. Dragolin et al. argue that the measure proposed by the FDA is not a well defined distance measure, while the LHS of equation (10) is. However, the purpose of scaling the measure only with respect to the reference variance attributes more weight to the well established drug.

The second notable exception is Chervoneva et al. (2007) who propose a criterion using the trace of the variance-covariance matrices. Although this criterion reduces to the univariate PBE when *p*, the number of variables, is one, it does not incorporate the correlations. The trace of a matrix being the sum of the diagonal elements alone ignores the off diagonal elements which represent the covariances.

The bioequivalence rule proposed by Chervenova et. al. (2007) is,

$$B_p=\frac{(\boldsymbol{\mu}_T-\boldsymbol{\mu}_R)'(\boldsymbol{\mu}_T-\boldsymbol{\mu}_R)+tr(\boldsymbol{\Sigma}_T)-tr(\boldsymbol{\Sigma}_R)}{tr(\boldsymbol{\Sigma}_R)}\le\theta \quad . \qquad (12)$$

They linearize this inequality by writing,

$$(\boldsymbol{\mu}_T - \boldsymbol{\mu}_R)'(\boldsymbol{\mu}_T - \boldsymbol{\mu}_R) + tr(\boldsymbol{\Sigma}_T) - tr(\boldsymbol{\Sigma}_R) \le \theta tr(\boldsymbol{\Sigma}_R), \qquad (13)$$

which reduces to,

$$(\boldsymbol{\mu}_T - \boldsymbol{\mu}_R)'(\boldsymbol{\mu}_T - \boldsymbol{\mu}_R) + tr(\boldsymbol{\Sigma}_T) - (1+\theta) tr(\boldsymbol{\Sigma}_R) \le 0. \qquad (14)$$

They then construct estimates of the confidence interval for the LHS of the above inequality by developing confidence intervals of the traces and the quadratic term. They calculate the predefined θ using the same limits of differences in means and variances defined by the FDA in the univariate case. They concluded BE when the upper limit of the 90% confidence interval is negative. For p = 1 this rule reduces to the univariate rule in equation (9) .

In chapter 4 the implementation of the criterion suggested in equation (8) will be discussed and the two measures presented here will be compared to the criterion proposed in the chapter 3. One of the main issues in the two methods presented here and the one that is proposed in chapter 3 is regarding the specification of the upper limit, θ. Next, this is discussed briefly.

2.4. Upper bounds of PBE defined by FDA

In the univariate case, θ is defined according to predetermined limits determined by the FDA. The maximum difference between the variances of the test and the reference $(\sigma_T^2 - \sigma_R^2)$ allowed by FDA (1997) is 0.02, and the minimum allowed variance of the reference (σ_R^2) is 0.04. This minimum variance was motivated by the population difference ratio (PDR) and the corresponding criterion for ABE. The PDR is defined as

the square root of the ratio of the expected squared difference of the pharmacokinetic measure of the test and the reference to the expected squared difference of the same under replicated administration of the reference drug. The FDA defined 1.25 as the maximum allowable value of PDR to consider the two drugs bioequivalent. Notice that PDR will reduce to a function of the PBE criterion C in equation(9). That is, $PDR = \sqrt{\frac{C}{2}+1}$. The FDA also sets the upper limit of $(\mu_T - \mu_R)$ to the natural log of 1.25 to accept bioequivalence according to the '80/125' rule, where the ratio between the test and reference means should lie within the [80%, 125%] range. Using these facts and assuming that $\sigma_T^2 = \sigma_R^2$, the minimum value of σ_R^2 that fulfills FDA's maximum allowable value of PDR is about 0.04, and the maximum value of θ that determines PBE is 1.75 (Appendix C).

The FDA proposed the limits for the PBE upper bounds for the univariate case only. No guidance was provided for the multivariate case. Chervenova et al. (2007), used the same limits for each of the variables in the multivariate method they suggested. Their method ignored the correlations between the variable. The correlations should be accounted for and their effects need to be studied. This important issue will be further considered in chapters 3 and 4.

3 A multivariate criterion for testing PBE

In the previous chapter two multivariate criteria for testing BE were discussed. It was noted that these criteria are not appropriate analogs of the univariate FDA approved criterion. In this chapter, a multivariate criterion that is based on the motivation of the univariate FDA criterion is derived, and a method to implement it is presented.

3.1. Development of the multivariate bioequivalence criterion C_p

The univariate PBE criterion is expressed as

$$\frac{E\left[\left(Y_T - Y_R\right)^2\right] - E\left[\left(Y_R - Y_{R'}\right)^2\right]}{E\left[\left(Y_R - Y_{R'}\right)^2\right]/2}. \tag{15}$$

To develop the multivariate equivalent for the criterion in (15), let $\mathbf{Y}_T$ and $\mathbf{Y}_R$ be p-variate random variables denoting the test and reference PK metrics. Assume, $\mathbf{Y}_T$ is distributed as a p-variate normal with a mean $\boldsymbol{\mu}_T$ and a variance covariance matrix $\boldsymbol{\Sigma}_T$ and let $\mathbf{Y}_R$ and $\mathbf{Y}_{R'}$ be two realizations of the p-variate normally distributed random variables with mean vector $\boldsymbol{\mu}_R$ and a variance covariance matrix $\boldsymbol{\Sigma}_R$.

Then the multivariate equivalent of the denominator in(15) is

$$\frac{1}{2}E\left[\left(\mathbf{Y}_R - \mathbf{Y}_{R'}\right)\left(\mathbf{Y}_R - \mathbf{Y}_{R'}\right)'\right] = \boldsymbol{\Sigma}_R. \tag{16}$$

Then the multivariate criterion that is equivalent to (9) could be written,

$$C_p = E\left[\left(\mathbf{Y}_T - \mathbf{Y}_R\right)' \boldsymbol{\Sigma}_R^{-1}\left(\mathbf{Y}_T - \mathbf{Y}_R\right)\right] - E\left[\left(\mathbf{Y}_R - \mathbf{Y}_{R'}\right)' \boldsymbol{\Sigma}_R^{-1}\left(\mathbf{Y}_R - \mathbf{Y}_{R'}\right)\right] \tag{17}$$

To prove this, let $\mathbf{Z} = \boldsymbol{\Sigma}_R^{-1/2}(\mathbf{Y}_T - \mathbf{Y}_R)$ and let $\mathbf{K} = \boldsymbol{\Sigma}_R^{-1/2}(\mathbf{Y}_R - \mathbf{Y}_{R'})$, then by substituting $\mathbf{Z}$ and $\mathbf{K}$ appropriately in (17), the multivariate criterion could be expressed as

$$C_p = E[\mathbf{Z}'\mathbf{Z}] - E[\mathbf{K}'\mathbf{K}] \tag{18}$$

Note that

$$\begin{aligned} E[\mathbf{Z}'\mathbf{Z}] &= E\left[\sum_{i=1}^{p} z_i^2\right] = \sum_{i=1}^{p} E\left[z_i^2\right], \\ &= \sum_{i=1}^{p}\left(\sigma_{z_i}^2 + E[z_i]^2\right), \\ &= \sum_{i=1}^{p}\left(\sigma_{z_i}^2 + E[z_i]^2\right), \\ &= \sum_{i=1}^{p}\sigma_{z_i}^2 + \sum_{i=1}^{p} E[z_i]^2, \\ &= trace(\boldsymbol{\Sigma}_{\mathbf{Z}}) + E[\mathbf{Z}]' E[\mathbf{Z}]. \end{aligned} \tag{19}$$

Similarly, $E[\mathbf{K}\mathbf{K}'] = trace(\boldsymbol{\Sigma}_{\mathbf{K}}) - E[\mathbf{K}]' E[\mathbf{K}]$. Substituting these in (19), the multivariate PBE criterion reduces to,

$$C_p = trace(\boldsymbol{\Sigma}_{\mathbf{Z}}) + E[\mathbf{Z}]' E[\mathbf{Z}] - trace(\boldsymbol{\Sigma}_{\mathbf{K}}) - E[\mathbf{K}]' E[\mathbf{K}]. \tag{20}$$

The expected value of $\mathbf{Z}$ is

$$\begin{aligned} E[\mathbf{Z}] &= E\left[\boldsymbol{\Sigma}_R^{-1/2}(\mathbf{Y}_T - \mathbf{Y}_R)\right], \\ &= \boldsymbol{\Sigma}_R^{-1/2} E\left[(\mathbf{Y}_T - \mathbf{Y}_R)\right], \\ &= \boldsymbol{\Sigma}_R^{-1/2}(\boldsymbol{\mu}_T - \boldsymbol{\mu}_R). \end{aligned} \tag{21}$$

Therefore the second term of the right hand side of (20) is

$$
\begin{aligned}
E[\mathbf{Z}]' E[\mathbf{Z}] &= (\boldsymbol{\mu}_T - \boldsymbol{\mu}_R)' \boldsymbol{\Sigma}_R^{-1/2} \boldsymbol{\Sigma}_R^{-1/2} (\boldsymbol{\mu}_T - \boldsymbol{\mu}_R), \\
&= (\boldsymbol{\mu}_T - \boldsymbol{\mu}_R)' \boldsymbol{\Sigma}_R^{-1} (\boldsymbol{\mu}_T - \boldsymbol{\mu}_R).
\end{aligned} \tag{22}
$$

The expected value of $\mathbf{K}$ is,

$$
\begin{aligned}
E[\mathbf{K}] &= E\left[\boldsymbol{\Sigma}_R^{-1/2} (\mathbf{Y}_R - \mathbf{Y}_{R'}) \right], \\
&= \boldsymbol{\Sigma}_R^{-1/2} E\left[(\mathbf{Y}_R - \mathbf{Y}_{R'}) \right], \\
&= \boldsymbol{\Sigma}_R^{-1/2} (\boldsymbol{\mu}_R - \boldsymbol{\mu}_R), \\
&= \underline{\mathbf{0}}.
\end{aligned} \tag{23}
$$

Note that the variance covariance matrix of $\mathbf{Z}$ is:

$$
\begin{aligned}
\boldsymbol{\Sigma}_{\mathbf{Z}} &= V(\mathbf{Z}), \\
&= V\left(\boldsymbol{\Sigma}_R^{-1/2} (\mathbf{Y}_T - \mathbf{Y}_R) \right), \\
&= \boldsymbol{\Sigma}_R^{-1/2} Cov(\mathbf{Y}_T - \mathbf{Y}_R) \boldsymbol{\Sigma}_R^{-1/2}, \\
&= \boldsymbol{\Sigma}_R^{-1/2} \left(Cov(\mathbf{Y}_T) + Cov(\mathbf{Y}_R) \right) \boldsymbol{\Sigma}_R^{-1/2}, \\
&= \boldsymbol{\Sigma}_R^{-1/2} (\boldsymbol{\Sigma}_T + \boldsymbol{\Sigma}_R) \boldsymbol{\Sigma}_R^{-1/2}, \\
&= \boldsymbol{\Sigma}_R^{-1/2} \boldsymbol{\Sigma}_T \boldsymbol{\Sigma}_R^{-1/2} + \boldsymbol{\Sigma}_R^{-1/2} \boldsymbol{\Sigma}_R \boldsymbol{\Sigma}_R^{-1/2}.
\end{aligned} \tag{24}
$$

The last term of the above equation reduces to a p x p identity matrix, $\mathbf{I}$, and thus the variance-covariance matrix of $\mathbf{Z}$ reduces to $\boldsymbol{\Sigma}_R^{-1/2} \boldsymbol{\Sigma}_T \boldsymbol{\Sigma}_R^{-1/2} + \mathbf{I}$.

Similarly the variance covariance matrix of $\mathbf{K}$

$$
\begin{aligned}
\boldsymbol{\Sigma}_K &= V(K), \\
&= V\left(\boldsymbol{\Sigma}_R^{-1/2} (\mathbf{Y}_R - \mathbf{Y}_{R'}) \right), \\
&= \boldsymbol{\Sigma}_R^{-1/2} \left(Cov(\mathbf{Y}_R) + Cov(\mathbf{Y}_{R'}) \right) \boldsymbol{\Sigma}_R^{-1/2}, \\
&= \boldsymbol{\Sigma}_R^{-1/2} (\boldsymbol{\Sigma}_R + \boldsymbol{\Sigma}_R) \boldsymbol{\Sigma}_R^{-1/2}, \\
&= 2\mathbf{I}.
\end{aligned} \tag{25}
$$

Substituting these into (20) and using the cyclical properties of the trace, the multivariate criterion could be expressed as

$$\begin{aligned} C_p &= trace(\boldsymbol{\Sigma}_Z) + E[\mathbf{Z}]' E[\mathbf{Z}] - trace(\boldsymbol{\Sigma}_K) - E[\mathbf{K}]' E[\mathbf{K}], \\ &= trace\left(\boldsymbol{\Sigma}_R^{-1/2}\boldsymbol{\Sigma}_T\boldsymbol{\Sigma}_R^{-1/2} + \mathbf{I}\right) + (\boldsymbol{\mu}_T - \boldsymbol{\mu}_R)' \boldsymbol{\Sigma}_R^{-1}(\boldsymbol{\mu}_T - \boldsymbol{\mu}_R) - trace(2\mathbf{I}), \\ &= trace\left(\boldsymbol{\Sigma}_T\boldsymbol{\Sigma}_R^{-1}\right) + (\boldsymbol{\mu}_T - \boldsymbol{\mu}_R)' \boldsymbol{\Sigma}_R^{-1}(\boldsymbol{\mu}_T - \boldsymbol{\mu}_R) - p. \end{aligned}$$

It is simple to show that this multivariate criterion reduces to the univariate criterion (9) when p = 1. It also accounts for the total variability and the correlations among the PK metrics used in evaluating bioequivalence.

Using the invariance property, the maximum likelihood estimator (mle) of the multivariate PBE criterion could be obtained from the data as:

$$\hat{C}_p = trace\left(\hat{\boldsymbol{\Sigma}}_T\hat{\boldsymbol{\Sigma}}_R^{-1}\right) + (\hat{\boldsymbol{\mu}}_T - \hat{\boldsymbol{\mu}}_R)' \hat{\boldsymbol{\Sigma}}_R^{-1}(\hat{\boldsymbol{\mu}}_T - \hat{\boldsymbol{\mu}}_R) - p, \quad (26)$$

where $\hat{\boldsymbol{\mu}}_T$ and $\hat{\boldsymbol{\mu}}_R$ are the mle's of the population means, and $\hat{\boldsymbol{\Sigma}}_T$ and $\hat{\boldsymbol{\Sigma}}_R$ are the mle's of the variance covariance matrices of the test and reference variables.

3.2. Hypothesis Testing of Multivariate PBE

Using the proposed multivariate criterion (C_p), the hypotheses for multivariate PBE are

$$H_0 : C_p > \theta \text{ vs } H_a : C_p \le \theta, \quad (27)$$

where C_p is the p-variable PBE criterion and θ is the constant predetermined by the regulators as the upper acceptable value for the acceptance region. One could define the test statistic based on the mle of the C_p. However, the exact distribution of the test

statistic is not tractable. Therefore, numerical methods such as the Bootstrap algorithm have to be used to determine the distribution. In general, a size α test could be defined as,

$$\phi(\underline{y}_R, \underline{y}_T) = \begin{cases} 1 \text{ if } \quad \hat{C}_p(1-\alpha) > \theta, \\ 0 \text{ if } \quad \hat{C}_p(1-\alpha) \le \theta. \end{cases}, \qquad (28)$$

where $\underline{y}_R$ and $\underline{y}_T$ are the sample observations, and $\hat{C}_p(1-\alpha)$ is the $(1-\alpha)100th$ percentile of the distribution of the mle of the PBE criterion. This test rejects the null hypothesis of no BE when the test statistic is 1. Equivalently, the multivariate bioequivalence will be concluded at significance level α if the upper bound of the confidence interval for C_p, namely $\hat{C}_p(1-\alpha)$ is less than θ.

3.2.1. Constructing the $100(1-\alpha)th$ confidence interval of C_p

The exact distribution of the MV criterion of the mle $\hat{C}_p$ is not tractable. Therefore, a parametric bootstrap method (Efron & Tibshirani, 1993), as recommended by the FDA, is proposed. The steps of the parametric bootstrap method are:

1. Obtain the mle's of the population parameters $\boldsymbol{\mu}_T$ and $\boldsymbol{\Sigma}_T$ of the test and $\boldsymbol{\mu}_R$ and $\boldsymbol{\Sigma}_R$ of the reference metrics. Calculate the MV PBE criterion using (26).
2. Generate *B* pairs of bootstrap random samples. Each pair is made up of two random samples of size n_T (the number test drug samples) and n_R (the number of reference drug samples), selected from a multivariate normal distribution with

mean vector $\hat{\mu}_T$ and covariance matrix $\hat{\Sigma}_T$ for the test drug; and mean vector $\hat{\mu}_R$ and covariance matrix $\hat{\Sigma}_R$ for the reference drug.

3. For each pair of the *b*-th sample calculate the mle's of the means, $\hat{\mu}_{Tb}$ and $\hat{\mu}_{Rb}$ and the mle's of the variance covariance matrices, $\hat{\Sigma}_{Tb}$ and $\hat{\Sigma}_{Rb}$, for $b = 1,...,B$. Calculate the bootstrap estimate of the MV population criterion $\hat{C}_{pb}$ for each bootstrap sample.
4. Determine, $\hat{C}_{pb}(1-\alpha)$, the $100(1-\alpha)th$ percentile of the distribution of C_{pb} based on the *B* bootstrap samples.
5. PBE will be concluded if $\hat{C}_{pb}(1-\alpha) < \theta$.

3.3. Specifying the upper limit, θ, of BE

In Chapter 2, the limits and rational of BE acceptance used in calculating the univariate θ determined by the FDA were presented. These limits are extended to the multivariate criterion, by setting the maximum difference between the means of the test and reference pharmacokinetic measures as the natural logarithm of 1.25; the maximum difference between the test and reference variances as 0.02, and the lowest variances as 0.04. Since there is no analogous guideline for incorporating the correlations, different combinations of correlations among the test and reference variables will be used. In the case of independence of the two measures (i.e., $\rho_T = 0$ and $\rho_R = 0$), θ reduces to *p*-multiples of the univariate θ, where p is the number of variables. That is, $\theta_p = p\theta = 1.75p$, leading to a rectangular region of BE rather than an elliptical region.

Using the proposed multivariable limits, values of θ could be calculated for the case where $p = 2$ with the BE limits of means and variance differences as defined by FDA. To account for the correlations, since there are no FDA specifications, the θ could be computed for a range of values of ρ_R and ρ_T.

Figure 2 shows the change in θ in the positive range of the correlations. It is unlikely to find negative correlations between the AUC and $C_{\max}$. The horizontal reference line in the graph crosses the y-axis at 3.49 which is the value of θ when the reference variables are independent (i.e., $\rho_R = 0$). This value is noted as θ_0. Also, when the reference variables are independent, (i.e., $\rho_R = 0$), the value of θ is a constant (equal to 3.49) regardless of the correlation between the test variables.

The plot also illustrates, the value of θ is always lower than θ_0 when the correlation between the reference variables is less than or equal to 0.4. For reference correlations greater than 0.4, the value of θ is smaller than or greater than θ_0 depending on the values of ρ_R and ρ_T. Since C_p is scaled by the reference variance covariance matrix, θ is more sensitive as the difference between the reference correlation and the test correlations increase. The (second) figure shows the plot of θ versus correlation when the two correlations are assumed equal. Notice that in the positive range of values the plot is close to the horizontal line representing independence. However, the plot is consistently below the horizontal line. That is, the value of θ is always smaller than θ_0, calculated ignoring the correlation. As expected, this suggests, the acceptance region of BE will be

smaller when accounting for the correlations. (Most of the PK variables are generally positively correlated.)

In summary, these plots show, accounting for the correlations is necessary if the correlation between the variables in the reference and the test are expected to be vastly different. Further, when the correlations are vastly different guidance from the regulatory bodies is needed to determine the right upper bound. In the meanwhile, the BE should be tested for the most conservative bound. This problem becomes more complicated for p greater than two. Table 1 lists the values of θ for the case where the correlations in the reference and the test are assumed equal, in the case of p=3. In this case there are three correlations, the first between the first and second PK, the second between the first and third PK, and the third correlation between the second and third PK. the value θ when all the correlations are equal decreases as the correlation increases. When fixing the correlations between any two PK at low or medium values the value of θ decreases by increasing the third correlation. This pattern is not the same when fixing two correlations at high value (like 0.8), the value of θ increases from 2.39 when the third correlation is zero to 7.72 when the third correlations is at medium value, but θ drops to 2.94 when the correlation is high at 0.8. The table also demonstrates that the values of θ are higher in the case of p=3 than they are in the case of p=2.

Table 1 comparison of θ by different correlations among PK when p=3

ρ_{12}	ρ_{13}	ρ_{23}	θ
0	0	0	5.23
0	0	0.3	4.66
0	0	0.8	4.13
0	0.3	0.3	4.23

0	0.3	0.8	4.04
0	0.8	0.8	2.39
0.3	0.3	0.3	3.83
0.3	0.3	0.8	3.50
0.3	0.8	0.8	7.72
0.8	0.8	0.8	2.94

Figure 2: Effect of ρ_R and ρ_T on the rule θ

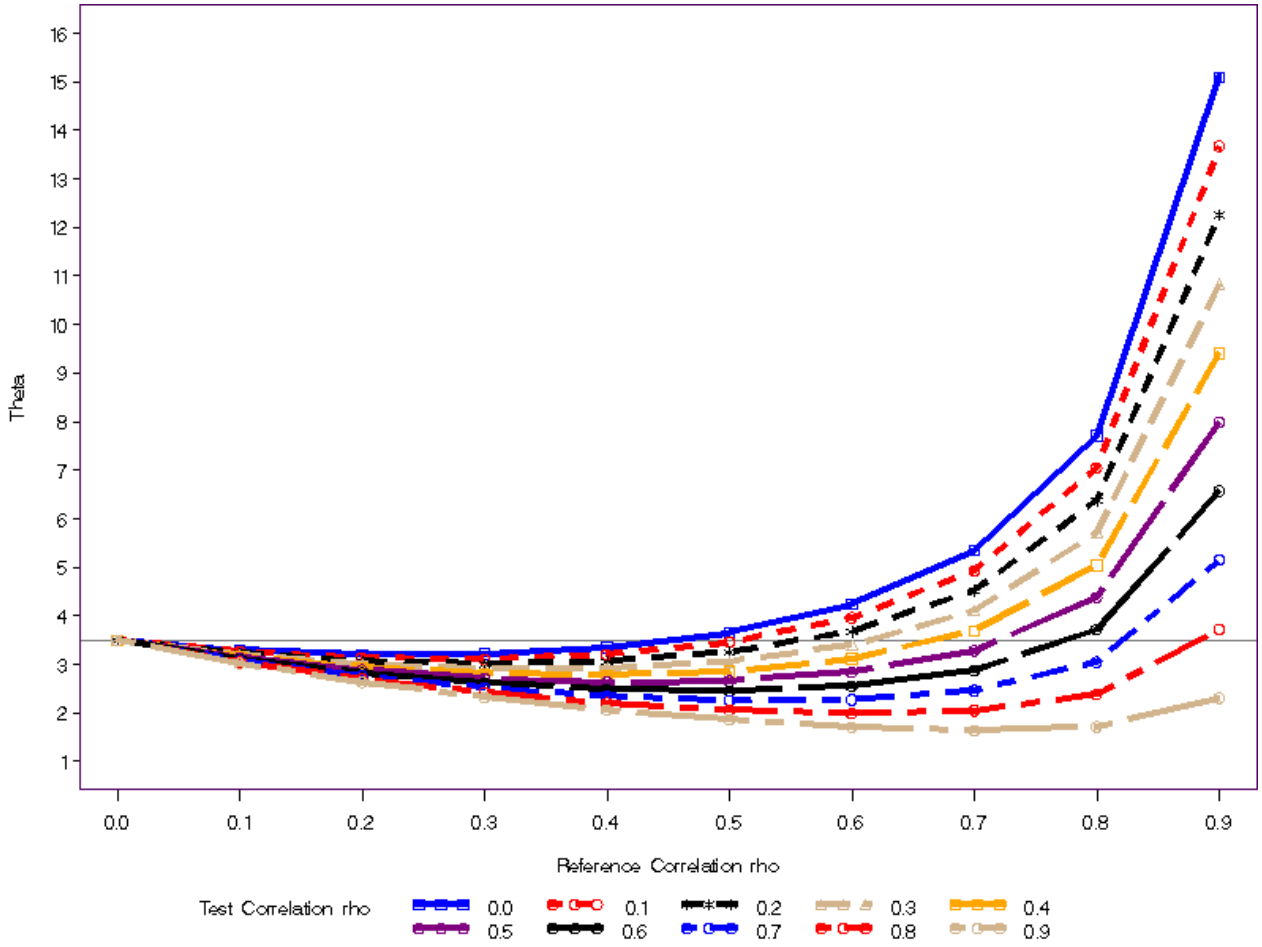
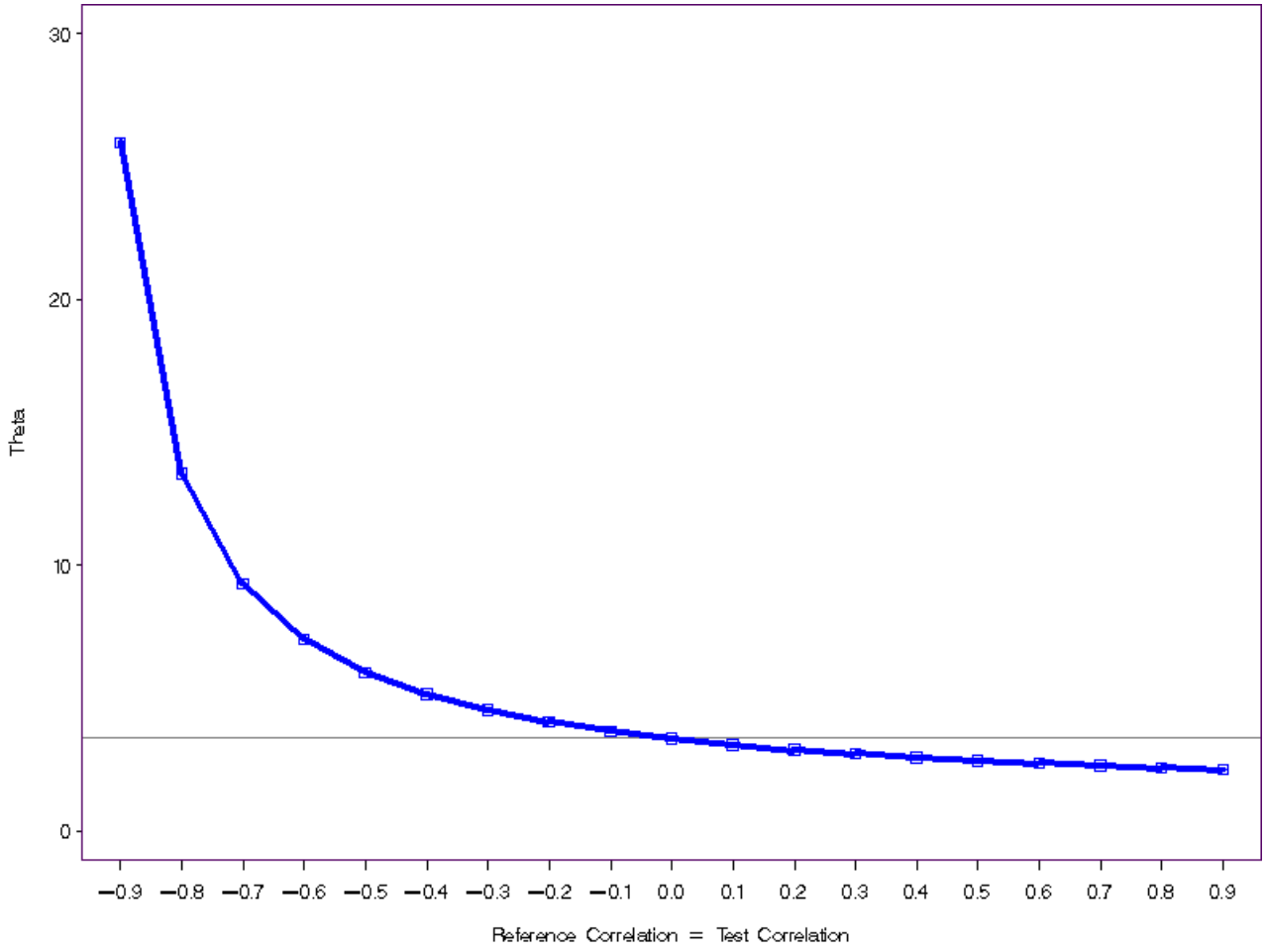


Figure 3 θ with the BE limits of mu and var differences, and equal ρ_R and ρ_T



3.4. Determination of PK parameters using model based estimation

It is known in PK that the drugs and other chemicals are absorbed, metabolized, and eliminated from the body according to mechanisms that are specific to each drug or groups of drugs and their physio-chemical metabolic pathways, in the anatomical compartments they are distributed through. These mechanisms could be represented mathematically through differential equations from which non-linear functional forms of blood/serum concentration profiles could be derived. In the area of pharmaco-kinetics these are known as one, two or higher order compartmental models. These models assume the existence of multiple separate but connected compartments in the body where the drug will be absorbed and eliminated from. In each compartment the constants of absorption and elimination are unique for that compartment. The absorption of each drug could be assessed using non-compartmental methods like zero-order models as discussed in chapter 1. As mentioned in chapter 1, the AUC resultant from a single dose of a drug formulation is commonly assessed with the linear trapezoidal method (Berger RL, 1996, Gibaldi, 1982), ignoring the mechanistic non-linear nature of the concentration profile. This is mainly because in the early phases of drug development these models might not be known. The non-linear functional form of the PK mechanisms of approved drugs are always studied extensively during and after the approval of the drug. However, after the approval of the drug, and before the development of new formulations or generic drugs the characteristics of the non-linear function would be studied from which they could be well specified and all the characteristics of the plasma-concentration curves could be determined. Acquiring this knowledge and the availability of advance analytic software

makes it logical to use compartmental or model-based methods for estimating the metrics used in evaluating bioequivalence.

Non-linear models have been used in pharmacology to study the PKs of drugs for a long time. These models could be based on theoretical models describing the underlying mechanism that produces the data. As a consequence, the non-linear model parameters have a better physical interpretation (Adams, 2002). However, these models are not generally used in drug testing, except for very limited tasks, like estimating the constants of absorption and elimination. Even in situations where non-linear models are used to estimate the other PK metrics such as AUC and $C_{\max}$ only point estimates are used in the bioequivalence testing. The uncertainties in the estimation are ignored (FDA 1992-2001, Chow SC, Liu JP 2000).

3.5. Analysis of pharmacological functions using nonlinear models

Consider the one-compartment pharmacological model that determines the drug concentration in the plasma or blood at any time point according to this function:

$$C = \frac{k_a k_e D}{Cl\left(k_a - k_e\right)}\left[e^{-k_e t} - e^{-k_a t}\right], \tag{29}$$

where C is the plasma concentration, D is the dose, Cl is the clearance, t is the time of the measurement, k_a is the constant of absorption, and k_e is the constant of elimination. Note that the clearance rate of the drug is $Cl = k_e V$, where V is the volume of the active compartment. The area under the curve AUC, could be estimated by integrating the plasma concentration with respect to time of the concentration function in (29). That is,

$$AUC = \int C\,dt = \int_0^{\infty} \frac{k_a k_e D}{Cl\left(k_a - k_e\right)}\left[e^{-k_e t} - e^{-k_a t}\right] dt, \qquad (30)$$

which yields $AUC = \frac{D}{Cl}$. If one is interested in AUC alone, the function could be re-parameterized in terms of AUC by substituting AUC for the ratio of the dose (D) to the clearance (Cl). That is, the model could be rewritten

$$C = \frac{AUC * k_a k_e}{\left(k_a - k_e\right)}\left[e^{-k_e t} - e^{-k_a t}\right]. \qquad (31)$$

Similarly $C_{\max}$ could be calculated by differentiating (29) with respect to *t* and equating it to zero. This yields the equation,

$$\frac{\partial C}{\partial t} = \left(\frac{k_a k_e D}{Cl\left(k_a - k_e\right)}\left[e^{-k_e t} - e^{-k_a t}\right]\right)' = 0 \qquad (32)$$

Solving the above equation yields, $C_{\max} = \frac{k_a k_e D}{Cl\left(k_a - k_e\right)}\left[e^{-k_e T_{\max}} - e^{-k_a T_{\max}}\right]$, where $T_{\max}$ is calculated as $\frac{\ln\left(k_e\right) - \ln\left(k_a\right)}{\left(k_e - k_a\right)}$. Other PK parameters such as the first order rate of absorption k_a, and the rate of drug elimination, k_e, could also be determined from these models through appropriate mathematical manipulations.

3.6. Modeling BE experiments using Nonlinear mixed effects models

Consider a bioequivalence study comparing a new test drug to a reference drug. In such studies, which are usually designed as cross-over studies, each subject receives both treatments, and he/she might receive each treatment multiple times. So these correlated repeated measures need to be accounted for when estimating the fixed effects. Suppose, the blood-concentration by time profile of the reference drug could be represented by a

known non-linear function *f* linking concentrations to sampling times of all the subjects with subject specific PK parameters, such as absorption (k_a), elimination (k_e)rate constants, and clearance half-life (Cl). An example of this function is the one compartment model function in (29)

Suppose several subjects are observed over a time interval at different occasions (of periods) on different treatments. At time point, t_{ijpk}, let C_{ijpk} represent the blood-concentration of the k^{th} treatment given to the i^{th} subject at the j^{th} time point, in the p^{th} period. Here i = 1, 2, …, n, k = T or R, p= 1, 2, …, P, and j = 1, 2, …, t_{ipk}, for, n subjects, 2 treatments, P periods and t_{ipk} time points. Assume that the sampling times are fixed and identical, for each treatment, period, and for all subjects, as often is the case in cross-over trials. Then for all $i, j, p,$ and k, the time t_{ijpk} could be simplified to t_j. Let $\boldsymbol{\lambda}_{ipk}$ be the vector of the PK parameters of the subject i for treatment k in period p. Then the nonlinear model for the concentration profile is,

$$C_{ijpk} = f\left(t_j, \boldsymbol{\lambda}_{ipk}\right) + \varepsilon_{ijpk}, \tag{33}$$

where ε_{ijpk} is the measurement error. It is also assumed that ε_{ijpk} are independent of $\boldsymbol{\lambda}_{ipk}$, and they are normally distributed with mean zero and variance σ^2_{ijpk}.

Assume that the parameters $\boldsymbol{\lambda}_{ipk}$ are random vectors that could be decomposed for each period and treatment as

$$\boldsymbol{\lambda}_{ipk} = \boldsymbol{\mu}_{\cdot} + \boldsymbol{\beta}_k + \boldsymbol{\gamma}_p + \mathbf{u}_{ik} \tag{34}$$

where $\boldsymbol{\mu}_{\cdot}$ is the overall mean, $\boldsymbol{\beta}_k$ is the fixed effect of the treatment, $\boldsymbol{\gamma}_p$ is the fixed effect of the period, and $\mathbf{u}_{ik}$ is the random effect of subject i for treatment k, it is also assumed that $\mathbf{u}_{ik}$ is distributed as a multivariate normal with mean zero vector and a variance covariance matrix $\boldsymbol{\Psi}_k$.

To ensure that the estimates are always positive, $\boldsymbol{\lambda}_{ipk}$ elements are the natural logarithms of the original PK parameters in the function *f*. The elements of $\boldsymbol{\lambda}_{ipk}$ are $\left[\log\left(Cl_{ipk}\right),\ \log\left(k_{a_{ipk}}\right),\ \log\left(k_{e_{ipk}}\right)\right]$.

The mle's of the original PK parameters: k_a , k_e, and Cl could be estimated using this nonlinear mixed effects model. Since $C_{\max}$ and AUC or other metrics are functions of these PK parameters, then the mle's of these metrics for each treatment group could be estimated as functions of the mle's of the PK. The asymptotic approximation of the variance covariance matrices of these metrics could be estimated using the Taylor series expansion theorem. Where the second partial derivatives are derived and the mles are estimated. This method is known as the delta method. When closed form derivatives are available, the estimation of the information matrix is easy. Otherwise other methods are utilized.

Using these estimates of the means and variance covariance matrices of the test and the reference drugs, the multivariate PBE criterion (C_p) that was proposed in section 3.1 is estimated. Then the 90% confidence interval is constructed around this estimate using the parametric bootstrap method as suggested by the FDA and as shown in our first paper. Two thousand samples are randomly generated from a multivariate

distribution with means and variances equal to the mle's obtained from the NLMEM. The upper limit of the resultant confidence interval is compared to the predefined limit of bioequivalence θ described earlier in the chapter. Bioequivalence is concluded if the upper limit of the 90% confidence interval is smaller than the predefined limit θ.

3.6.1 Modeling cross-over experimental design

The cross-over design is the most recommended design for bioequivalence studies. In this design subjects are randomized to receive one of the sequences of treatments. Each of these sequences contains both the test and the reference drugs at different periods. For example in the two period cross-over design, one sequence is TR, and the other treatment is RT. A washout period should follow any treatment period to make sure no effect of the drug from the previous period will affect the consecutive period. In multiple period designs, the order in which the two drugs are given is usually selected at random using a block randomization method to ensure the balance within subjects.

At the beginning of each period, baseline blood concentration data are collected prior to the administration of the drug to evaluate the washout period. The pharmacological baseline measurement is supposed to be zero if the wash out period is long enough. After the drug is administered, the specified pharmacological measurement (level of specific active material in the blood) is obtained over a period of time at fixed time intervals predesigned according to the investigators knowledge and expectations about the pharmacodynamics of the tested drug.

In this design, each subject receives both treatments, and one or more administrations of each treatment according the sequence assigned to him/her. Under this design, all

measures within each subject are correlated. The goal was to determine if the two drugs are bioequivalent. The NLMEM presented above in equation (34) is a suitable model for this design because the administrations are repeated in nature on each subject, and due to the expected missing data in such experiments. Using the same notation above the parameter vector could be decomposed as (34).

3.6.2 Modeling parallel experimental design

Although parallel designs are not the recommended designs for testing BE, in practice some drugs cannot be tested using other designs. In this design patients are randomized to either receive the test or the reference treatment. Similar to the nonlinear mixed effects model (NLMEM) described above. A fixed effect model could be constructed by excluding the period effect, and the random effect of the subject. This could be used to model the one-compartment pharmacological model that fits such data. Using the same notations above, the vector $\boldsymbol{\lambda}_{ik}$ is a fixed effects vector that can be decomposition as

$$\boldsymbol{\lambda}_k = \boldsymbol{\mu}_{\cdot} + \boldsymbol{\beta}_k \tag{35}$$

where $\boldsymbol{\mu}_{\cdot}$ is the mean value of all treatments, $\boldsymbol{\beta}$ is the coefficient of the fixed effect According to this model, the vector $\boldsymbol{\lambda}_k$, could be defined as $\left[\log(Cl),\ \log(k_a),\ \log(k_e)\right]$ $\boldsymbol{\mu}_{\cdot} = \left[\mu_1, \mu_2, \mu_3\right]$, and $\boldsymbol{\beta}_k = \left[\beta_{1k}, \beta_{2k}, \beta_{3k}\right]$. The pharmacological function in (29) could be fit in the nonlinear model by substituting the PK parameters Cl, k_a, and k_e by the exponentials of the members of the fixed effect vector $\boldsymbol{\lambda}_{ik}$; i.e. $(\mu_1 + \beta_{1k})$, $(\mu_2 + \beta_{2k})$, and $(\mu_3 + \beta_{3k})$ respectively.

3.7. Summary of the method

The estimates of the PK parameters for each treatment and the variance covariance matrices could be estimated as in section 3.6. The MV criterion would be estimated and the upper limit of the confidence interval, constructed using the bootstrap method, would be compared to the predefined limit θ to evaluate PBE.

4 Simulation Study

In this chapter, properties of the multivariate PBE criterion proposed in the previous chapter are studied using Monte-Carlo simulation methods. The simulation study was designed to evaluate the distribution of the proposed multivariate criterion C_p under different combinations of sample size (number of subjects in the trial), differences in the averages and variances of the pharmacokinetic (PK) parameters between the reference and test drugs, and under different correlations among the PK parameters within each treatment group. This study was mainly designed to guide in the selection of a method to construct the confidence interval for the proposed criterion. Another simulation study was designed to study the size and power of the hypothesis tests, and to compare the multivariate criterion versus the multiple testing using the univariate criteria. These studies were limited to equal size samples of reference and test drugs. The effect of different sample sizes, missing values and dependence between the treatments drugs should be examined in future studies.

4.1. Generating *p*-variate normally distributed data

It has been shown in many studies that the log-transformed pharmacokinetic parameters have a normal distribution, and that data extracted from the same concentration time profiles for each subject are correlated. To create samples that preserve these properties samples for the simulations were drawn from multivariate normal distribution. Two samples $\underline{Y}_T$ and $\underline{Y}_R$ of N sets of pairs of variables were generated from a bivariate normal distribution to represent the test and reference datasets.

The first dataset includes variable y_{R1i} and y_{R2i} hat represent the log-transformed pharmacokinetic parameters of the reference drug, namely $\log(C_{\max})$ and $\log(AUC)$. The second dataset includes y_{T1i} and y_{T2i} that represent the log-transformed data of the test drug, where i represents the i^{th} subject, for $i = 1,...,N$. The corresponding data matrices are

$$\mathbf{Y}_T = \begin{bmatrix} y_{T11} & y_{T12} & \cdots & y_{T1N} \\ y_{T21} & y_{T22} & \cdots & y_{T2N} \end{bmatrix},$$

$$\mathbf{Y}_R = \begin{bmatrix} y_{R11} & y_{R12} & \cdots & y_{R1N} \\ y_{R21} & y_{R22} & \cdots & y_{R2N} \end{bmatrix}.$$

have samples of bivariate normal distributions, which we will denote by

$$\mathbf{Y}_k \sim N_p(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k),$$

where the subscript k represent the treatment, $k = \mathrm{T}, \mathrm{R}$. For treatment k the population mean vector is

$$\boldsymbol{\mu}_k = [\mu_{k1}, \mu_{k2}],$$

and the variance covariance matrix is

$$\boldsymbol{\Sigma}_k = \begin{bmatrix} \sigma_{k1}^2 & \rho_k \sigma_{k1} \sigma_{k2} \\ \rho_k \sigma_{k1} \sigma_{k2} & \sigma_{k2}^2 \end{bmatrix}.$$

In the case of more than two PK, this covariance structure allows for different correlations between the pharmacokinetic variables within each of the two drugs. For the simulation a parallel design is assumed, so that the observations from the two treatments are assumed independent. This assumption is widely accepted in the case of the univariate evaluation of PBE.

SAS IML code calling the function (VNORMAL) was used in a macro to generate the samples of bivariate normal data. All simulations and analyses were done using SAS 9.2, SAS Institute, Cary, NC. Code is presented in Appendix K.

4.1.1 Evaluation of the distribution of the multivariate BE criterion.

A preliminary Monte-Carlo experiment was performed to study the distributions of the multivariate criterion. The results in Appendix C show that the distributions are skewed to the right, especially in smaller sample sizes. As the sample size increases the distribution gets closer to a symmetric normal distribution. The distribution under equal correlations, and three sample sizes (25, 50, and 100), were examined. Distribution histograms in Appendix C show that with a small sample size (n = 25) the distribution of the estimate of PBE criterion, C_p, is far from normality. As the sample size increases, the distribution approaches normality. However, if the distribution was investigated by the difference between the covariances of reference and the test, then this study demonstrates that these difference specific distributions are far from the normal distribution even with the larger sample sizes (n = 100).

The results of this Monte-Carlo experiment (presented in Appendix C) supported FDA's recommendation of using the bootstrap method to construct the confidence interval for the univariate PBE criterion rather than applying a normal approximation. It could be concluded that similar to the univariate case, using the parametric confidence intervals which assume normality for the multivariate PBE criterion C_p would not be appropriate. In order not to assume any distribution the parametric bootstrap method to

evaluate the percentile confidence interval was used as suggested by the FDA. The parametric bootstrap method was used because there is enough evidence that the log-transformed pharmacokinetic measures like *AUC*, $C_{\max}$ and $T_{\max}$ follow a normal distribution. It was convenient to resample (for the bootstrap) from a *p*-variate normal distribution in which means and variance covariance matrix are equal to those estimated from the sample.

4.2. Constructing confidence intervals for the PBE Criterion C_p

The non-parametric confidence interval was constructed using the percentiles from the parametric bootstrap method. From each replication of the experimental settings, the means and the variance covariance matrices were estimated. Then 2000 bootstrap samples were generated by randomly selecting from multivariate normal distribution with mean and variance equal to those estimated from the replication. Then estimates of the multivariate PBE criterion C_p were calculated for each of these 2000 bootstrap samples. The 95th bootstrap percentile of C_p was determined. This is the upper limit of the one sided 95% confidence interval for C_p that would be used in the hypothesis testing.

4.3. Simulation Configurations

The simulation study was run with 500 replicates. Each replicate represented a bioequivalence trial with N subjects in each treatment (test or reference). For each subject two measures representing the log-transformed AUC and $C_{\max}$, were selected from the bivariate normal distribution as described earlier. To estimate the nonparametric

confidence interval in each simulation, 2000 bootstrap samples were randomly selected from the bivariate distributions. Efron (1982) demonstrated that in general 2000 bootstrap samples would be large enough to obtain unbiased confidence intervals.

Following FDA guidelines for the univariate case of PBE, and based on the information from published bioequivalence trials the following factors and their levels were considered:

1. Sample size (The number of subjects in each sample (N)): It is important to evaluate the performance of any new statistic or test under a variety of sample sizes. The usual number of subjects in most of the bioequivalence drug trials varies between 20 and 100 per group. So it was decided to choose the values 25, 50 and 100.
2. Difference in the means: As discussed earlier the maximum allowable difference between the (log transformed) means of the test and the reference parameters is the natural logarithm of 1.25. Five values for the difference between the means were selected: -2log(1.25), -log(1.25)/2, zero, log(1.25), 2log(1.25).
3. The standard deviations: The FDA recommends a minimum value for the reference standard deviation of 0.2, so it selected as the value for the variance of the reference pharmacokinetics. Hence, σ_{R1}^2 & σ_{R2}^2 were set to be 0.04.
4. Difference in the variance: The FDA defines bioequivalence if the difference between the variances of the reference and the test pharmacokinetic is within 0.02. We selected the values for the variance of the test PK to be 0.04, 0.06 and 0.1.

5. Correlation between the pharmacokinetic variables: The correlations for variables are not used in the definition of bioequivalence. The correlations are ignored in the multiple univariate testing and in the MV criterion suggested by Chervoneva (2007). To evaluate the performance of the proposed criterion, a wide range of correlations that allow for all possible values including rare cases like negative correlations were selected. The correlations between -0.2 and 0.8 in increments of 0.2 were selected to evaluate the effect of the difference between the correlations on the power and the size of the tests for bioequivalence.

The size of testing the hypotheses of BE was evaluated under conditions that represent the maximum allowable values of variances and differences between the means of the reference and test drugs. The power of the test was evaluated under conditions that fulfill the BE with respect to the variances and the means. The size and power were evaluated under different sample sizes and variable correlations between the reference and between the test variables.

4.4. Description of the simulation steps

In summary, the simulations was performed using the following steps for each of the combinations of sample size, difference in means, difference in variance and correlations between the reference and the correlations between the test PK measures:

1. Generate 500 samples (replicates) of four random variables of size *N* from two bivariate normal distribution with reference means μ_{R1} and μ_{R2} and test means μ_{T1} and μ_{T2}, and correlations ρ_R and ρ_T and variances σ^2_{R1}, σ^2_{R2}, σ^2_{T1}, and σ^2_{T2} that define each configuration or setting.

2. Estimate the means vector and variances covariance matrix for each replication.
3. Bootstrap each replicate 2000 times using parametric bootstrap method by selecting 2000 random samples from multivariate normal distribution with the means and variances equal to those estimated in step 2. Calculate the estimated MV criterion $\hat{C}_{pb}$ for each bootstrap sample. The subscript b denotes bootstrap.
4. Calculate the 95th percentile for the 2000 calculated bootstrap estimates of $\hat{C}_{pb}$ for each simulation configuration. Let this percentile be denoted $\hat{C}_{p(.95)}$ which is the upper limit of the one-side confidence interval of C_p.
5. For each replication, reject the null hypothesis if the upper bound of the one-sided 95% confidence interval, $\hat{C}_{p(.95)}$, was less than the predefined θ. Calculate the number and percent of times the result was correct in agreement with the setting, and do the same using the univariate criterions for each of the variables.
6. Calculate the size and power of the test by calculating the percentage of times the null hypothesis was rejected among the 500 replicates, if the null hypothesis was true or was not true respectively.

4.5. Simulation Results

The results of the simulation study are tabulated by sample size, the difference between the averages of each of the variables and the difference in the variances, the correlations of the reference and the correlations of the test drugs in the Appendices C and D. The results show that the size of the test is affected by the sample size, and the correlations.

4.5.1 Evaluation of size of the test

The size of the test could be evaluated as the highest probability of rejecting the null hypothesis given the null hypothesis is true. That is, using the definition

$$\alpha = \sup\left\{P_{Ho}\left(\phi(\underline{y}_R, \underline{y}_T) = 1\right)\right\}. \quad (36)$$

In the example of bioequivalence this is the highest probability of determining two drugs as BE when they are actually not BE.

The size of the test was evaluated using random selection from distributions that have the maximum accepted values of the differences between the means and between the variances. These values define the boundaries of the multivariate PBE acceptance region in all dimensions. These are the values that define the true value of the multivariate PBE criterion as equal to the predefined θ. The table compares between two tests, the first accounts for the correlation in the predefined θ, the second one ignores the correlations when defining θ. The probability of type I error is very conservative in the case of smallest tested sample sizes of 25 and it increases by increasing the sample size. These errors exceed the 0.05 level only when the reference correlations are higher than 0.2, and the θ is calculated under the assumption of independence. The largest probability of type I errors are observed when the correlations of either the reference or the test are very high (0.8), and when the differences between the correlations of the reference and the correlations among the test are large.

Probability of the type I errors are also sensitive to the magnitude of the correlation used in calculating the θ (Table 3). Testing the bioequivalence hypothesis using negative correlations, on samples drawn from positively correlated random

variables, causes the highest increase in type I errors. There is a minimal increase in probability of Type I error when using a positive correlation that is not equal to the true correlation except for testing with lower correlations when the true correlation is high 0.8. This would result in rejecting the null hypothesis, and concluding PBE in more non-BE cases.

4.5.2 Evaluation of power of accepting BE

Table 4 compares the power of the proposed multivariate PBE test as a function of the sample size, the correlations and the differences in the means and the difference between the variances. As expected the power is highest when there is no difference in the true means of the reference and test variables. The power drops gradually as the difference between the means increases in either direction. Accounting for the correlation in the upper limit, θ, of PBE results in a test with less power than ignoring the correlation. The power increases as the absolute difference between the test and reference means increases. The test that ignores the correlation achieves the highest power, while the intersection of two separate univariate tests has the lowest power. Data in Table 5 presents the power as a function of the difference in the variances of the reference and test variables. The power increases as the sample size increases and as the difference between the variances decreases.

Each simulation sample was classified as bioequivalent or not according to the proposed multivariate PBE test accounting for correlations, ignoring the correlations, and using the intersection of two univariate tests of BE with and without Bonferoni correction. Then according to each scenario the simulation samples were classified into correctly classified and incorrectly classified. Table 6 compares the correct classification

of all BE and non-BE scenarios using the four methods. The sensitivity of testing for BE is higher when ignoring the correlation than all other tests. This means that more scenarios are classifying correctly as BE when the truth is BE. However this higher sensitivity is at the cost of having the highest false positive rate or the lowest specificity, i.e. the highest proportion of misclassification (type I error or incorrect classification) when the truth is not BE.

The power function was evaluated as a function of the test means under fixed values of the reference means and variance covariance matrix and under equal correlations between the test variables and between the reference variables. The x-axis in Figure 4 presents the value of the means of the test variables, assuming equal means for both PK. The acceptance region of Bioequivalence under these conditions is the area bounded by the two vertical lines in the graph which represent differences of (log 1.25) from the reference means. The figure shows that the power of rejecting the null hypothesis of "no BE" is higher than 90% when the reference means are equal to the test means. This power decreases as the difference between the means increases. After crossing the acceptance boundaries, the vertical reference lines in the plot, the probability of rejecting 'not BE' drops to below 0.05. The figure shows that this trend is common among all three test: i) comparing the multivariate criterion to a rule θ that does not account for the correlation, ii) comparing the multivariate criterion to a rule θ that does account for the correlation , iii) the intersection of two univariate comparisons of the univariate criterion to a univariate rule θ .

Figure 5 presents the drop in the power of testing the null hypothesis of 'no BE', under fixed variance of the reference, as the variance of the test increases. This power, as expected, increases as the sample size increases. The power is above 0.7 with sample size of 100 when the variance is 0.06, which is equivalent to a difference of 0.02 between the variances. As the variance of the test variable increases the probability of rejecting the null hypothesis of no BE decreases. When the difference between the variances exceeds 0.02 this probability represents the probability of type I error, and it is always 0.05 for the smallest sample size of 25. For larges sample sizes the probability of type I error is lower than 0.05 only when the difference between variances is greater than 0.05.

Table 2. Size* of the test,

		Ignoring correlation in θ, $P\left(C_p < \theta_0\right)$			Accounting for correlation in θ, $P\left(C_p < \theta\right)$		
ρ_R	ρ_T	n=25	n=50	n=100	n=25	n=50	n=100
-0.2	-0.2	0.0100	0.0000	0.0000	0.0133	0.0033	0.0133
	0	0.0000	0.0067	0.0000	0.0000	0.0067	0.0267
	0.2	0.0067	0.0000	0.0000	0.0067	0.0133	0.0400
	0.4	0.0033	0.0000	0.0033	0.0167	0.0133	0.0233
	0.8	0.0000	0.0067	0.0100	0.0200	0.0300	0.0267
0	-0.2	0.0033	0.0133	0.0133	0.0033	0.0133	0.0133
	0	0.0033	0.0200	0.0100	0.0033	0.0200	0.0100
	0.2	0.0033	0.0133	0.0333	0.0033	0.0133	0.0333
	0.4	0.0000	0.0133	0.0100	0.0000	0.0133	0.0100
	0.8	0.0100	0.0167	0.0300	0.0100	0.0167	0.0300
0.2	-0.2	0.0100	0.0000	0.0133	0.0067	0.0000	0.0100
	0	0.0000	0.0467	**0.0567**	0.0000	0.0167	0.0233
	0.2	0.0133	0.0267	0.0467	0.0100	0.0133	0.0067
	0.4	0.0133	0.0467	***0.0767***	0.0067	0.0267	0.0233
	0.8	0.0500	**0.0667**	***0.1433***	0.0233	0.0133	0.0433
0.4	-0.2	0.0033	0.0000	0.0100	0.0033	0.0033	0.0133
	0	0.0000	0.0067	0.0333	0.0000	0.0000	0.0200
	0.2	0.0067	0.0333	**0.0667**	0.0033	0.0200	0.0167
	0.4	0.0100	0.0467	***0.1300***	0.0067	0.0133	0.0067
	0.8	**0.0533**	***0.1700***	***0.3700***	0.0100	0.0233	0.0167
0.8	-0.2	0.0000	0.0000	0.0000	0.0067	0.0333	0.0333
	0	0.0000	0.0000	0.0000	0.0067	0.0100	0.0200
	0.2	0.0000	0.0000	0.0000	0.0067	0.0133	0.0033
	0.4	0.0000	0.0000	0.0000	0.0033	0.0067	0.0333
	0.8	0.0133	**0.1100**	***0.2867***	0.0000	0.0000	0.0200

- *the probability of rejecting the null hypothesis of "not bioequivalent" at the boundary of the BE region.
- ρ_R and ρ_T are correlations among reference measures and among test measures.
- ***Bold italic***: 95% CI does not include 0.05; **Bold**: 95% CI includes 0.05.

Table 3: Effect of misclassification of the rule theta on Type I error

Sample Size	rhoR	rhoT	$\theta_{-.2}$	θ_0	$\theta_{.2}$	$\theta_{.4}$	$\theta_{.8}$
25	-0.2	-0.2	0.0063	0.0063	0.0025	0.0000	0.0000
	0	0	0.0163	0.0063	0.0013	0.0013	0.0013
	0.2	0.2	0.0263	0.0063	0.0063	0.0013	0.0000
	0.4	0.4	0.0350	0.0050	0.0088	0.0050	0.0013
	0.8	0.8	0.0638	0.0025	0.0125	0.0088	0.0025
50	-0.2	-0.2	0.0125	0.0125	0.0000	0.0000	0.0000
	0	0	0.0475	0.0138	0.0088	0.0038	0.0000
	0.2	0.2	0.0875	0.0125	0.0125	0.0050	0.0000
	0.4	0.4	0.1213	0.0113	0.0238	0.0113	0.0038
	0.8	0.8	0.2388	0.0113	0.0588	0.0388	0.0113
100	-0.2	-0.2	0.0188	0.0188	0.0000	0.0000	0.0000
	0	0	0.0838	0.0163	0.0050	0.0025	0.0000
	0.2	0.2	0.2100	0.0113	0.0113	0.0025	0.0000
	0.4	0.4	0.3488	0.0088	0.0375	0.0088	0.0025
	0.8	0.8	0.5288	0.0238	0.1313	0.0663	0.0238

$\mu_{1T} - \mu_{1R} = \mu_{2T} - \mu_{2R} = \ln(1.25); \sigma^2_{1R} = \sigma^2_{2R} = 0.04; \sigma^2_{1T} = \sigma^2_{2T} = 0.06$

Table 4. Power of the test

		Ignoring correlation in θ, $P\left(C_p < \theta_0\right)$			Accounting for correlation in θ, $P\left(C_p < \theta\right)$		
rhoR	muT-muR	n=25	n=50	n=100	n=25	n=50	n=100
-0.2	-0.335	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	-0.223	0.0033	0.0100	0.0167	0.0133	0.0283	0.0567
	-0.112	0.1600	0.5100	0.9017	0.2683	0.6950	0.9800
	0	0.4183	0.9017	1.0000	0.5917	0.9633	1.0000
	0.112	0.1567	0.5483	0.8967	0.2567	0.7450	0.9717
	0.223	0.0050	0.0050	0.0167	0.0067	0.0283	0.0600
	0.335	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0	-0.335	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
	-0.223	0.0250	0.0467	0.0833	0.0250	0.0467	0.0833
	-0.112	0.1933	0.6283	0.9350	0.1933	0.6283	0.9350
	0	0.4333	0.9200	0.9983	0.4333	0.9200	0.9983
	0.112	0.2250	0.5850	0.9417	0.2250	0.5850	0.9417
	0.223	0.0183	0.0383	0.0817	0.0183	0.0383	0.0817
	0.335	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.2	-0.335	0.0000	0.0017	0.0000	0.0000	0.0017	0.0000
	-0.223	0.0300	0.0817	0.1850	0.0183	0.0350	0.0783
	-0.112	0.2167	0.6683	0.9583	0.1467	0.5017	0.8983
	0	0.4367	0.9133	1.0000	0.3383	0.8367	0.9917
	0.112	0.2267	0.6767	0.9650	0.1600	0.5383	0.9083
	0.223	0.0383	0.1150	0.2133	0.0250	0.0533	0.0850
	0.335	0.0017	0.0000	0.0000	0.0017	0.0000	0.0000
0.4	-0.335	0.0000	0.0017	0.0017	0.0000	0.0000	0.0000
	-0.223	0.0350	0.1383	0.3367	0.0150	0.0283	0.0867
	-0.112	0.2333	0.7167	0.9717	0.1233	0.4817	0.8567
	0	0.4350	0.9283	1.0000	0.2467	0.7500	0.9800
	0.112	0.2567	0.7667	0.9683	0.1333	0.4733	0.8383
	0.223	0.0350	0.1500	0.3100	0.0150	0.0383	0.0767
	0.335	0.0000	0.0033	0.0017	0.0000	0.0000	0.0000
0.8	-0.335	0.0117	0.0117	0.0133	0.0017	0.0017	0.0000
	-0.223	0.0483	0.2467	0.5850	0.0067	0.0600	0.0917
	-0.112	0.2817	0.7533	0.9783	0.0817	0.3633	0.7633
	0	0.4517	0.9133	1.0000	0.1650	0.6250	0.9300

0.112	0.2833	0.7500	0.9833	0.0750	0.3517	0.7700
0.223	0.0683	0.2250	0.5600	0.0150	0.0317	0.0883
0.335	0.0033	0.0150	0.0200	0.0000	0.0000	0.0000

- $\sigma_{R1}^2 = \sigma_{R2}^2 = 0.04$, $\sigma_{T1}^2 = \sigma_{T2}^2 = 0.05$, $\rho_R = \rho_T$ and $\mu_R = \mu_T$
- power of test = the probability of rejecting the null hypothesis "not bioequivalent" within the population BE region

Table 5. Power of the test as a function of the correlation, difference in the variances, and the sample size

rhoR	$\sigma_T^2 - \sigma_R^2$ V1T	Ignoring correlation in θ, $P\left(C_p < \theta_0\right)$			Accounting for correlation in θ, $P\left(C_p < \theta\right)$		
		N=25	n=50	n=100	n=25	n=50	n=100
-0.2000	0.0000	0.6300	0.9867	1.0000	0.7700	1.0000	1.0000
	0.0100	0.4183	0.9017	1.0000	0.5917	0.9633	1.0000
	0.0200	0.2100	0.6700	1.0000	0.3033	0.8567	1.0000
	0.0300	0.1133	0.3967	0.8133	0.1967	0.6333	0.9467
	0.0400	0.0400	0.1800	0.5433	0.1000	0.3333	0.8067
	0.0600	0.0100	0.0333	0.0400	0.0200	0.0533	0.2300
0.0000	0.0000	0.6733	0.9933	1.0000	0.6733	0.9933	1.0000
	0.0100	0.4333	0.9200	0.9983	0.4333	0.9200	0.9983
	0.0200	0.1867	0.7200	0.9735	0.1867	0.7200	0.9735
	0.0300	0.0867	0.3967	0.8167	0.0867	0.3967	0.8167
	0.0400	0.0567	0.1667	0.4933	0.0567	0.1667	0.4933
	0.0600	0.0033	0.0200	0.0833	0.0033	0.0200	0.0833
0.2000	0.0000	0.7633	0.9933	1.0000	0.6667	0.9800	1.0000
	0.0100	0.4367	0.9133	1.0000	0.3383	0.8367	0.9917
	0.0200	0.1933	0.6167	0.9700	0.1133	0.5000	0.9133
	0.0300	0.1067	0.4533	0.8033	0.0700	0.2933	0.6333
	0.0400	0.0367	0.2300	0.5100	0.0233	0.1433	0.3000
	0.0600	0.0033	0.0133	0.0733	0.0000	0.0067	0.0233
0.4000	0.0000	0.6633	0.9900	1.0000	0.4700	0.9567	1.0000
	0.0100	0.4350	0.9283	1.0000	0.2467	0.7500	0.9800
	0.0200	0.2000	0.7167	0.9767	0.1067	0.4500	0.8633

	0.0300	0.0700	0.4200	0.8233	0.0267	0.1900	0.4900
	0.0400	0.0500	0.1700	0.5533	0.0200	0.0733	0.1667
	0.0600	0.0133	0.0333	0.0733	0.0000	0.0100	0.0033
0.8000	0.0000	0.6967	0.9933	1.0000	0.4133	0.8900	1.0000
	0.0100	0.4517	0.9133	1.0000	0.1650	0.6250	0.9300
	0.0200	0.2467	0.7067	0.9800	0.0733	0.2933	0.6633
	0.0300	0.0833	0.3967	0.8333	0.0200	0.0867	0.2800
	0.0400	0.0567	0.2467	0.4533	0.0033	0.0400	0.0633
	0.0600	0.0033	0.0433	0.0600	0.0000	0.0000	0.0000

Table 6. Percentage of the simulated cases leading to correct decision regarding BE.

		MV BE criterion accounting for the correlation					
		BE			Not BE		
		Incorrect	correct	Total	Incorrect	correct	Total
Intersection of 2 $(1-\alpha/2)CI$ Univariate tests (Bonferoni correction)	Incorrect	3.5	15.56	19.07	11.87	0.91	12.78
	Correct	0.37	80.56	80.93	12.29	74.93	87.22
	Total	3.88	96.12	100	24.16	75.84	100
Intersection of 2 $(1-\alpha)CI$ Univariate tests	Incorrect	3.23	7.81	11.04	15.84	1.72	17.56
	Correct	0.65	88.31	88.96	8.33	74.11	82.44
	Total	3.88	96.12	100	24.16	75.84	100
MV ignoring correlation in rule	Incorrect	0.9	0.52	1.43	22.66	11.04	33.71
	Correct	2.97	95.6	98.57	1.5	64.79	66.29
	Total	3.88	96.12	100	24.16	75.84	100

Figure 4. Power as a function of mean of the test variable

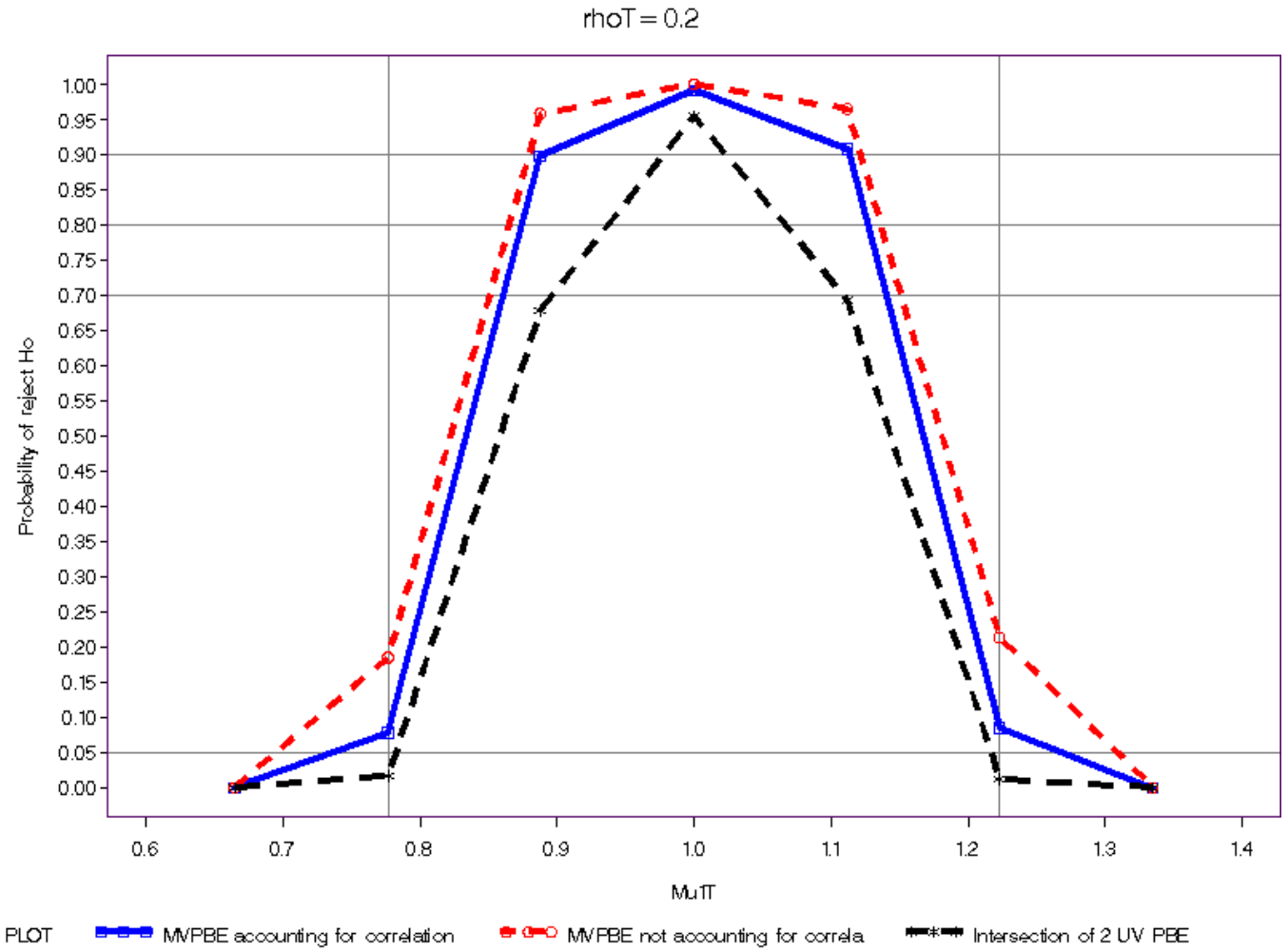


Figure 5. Effect of the sample size and the variance of the test variables on Power

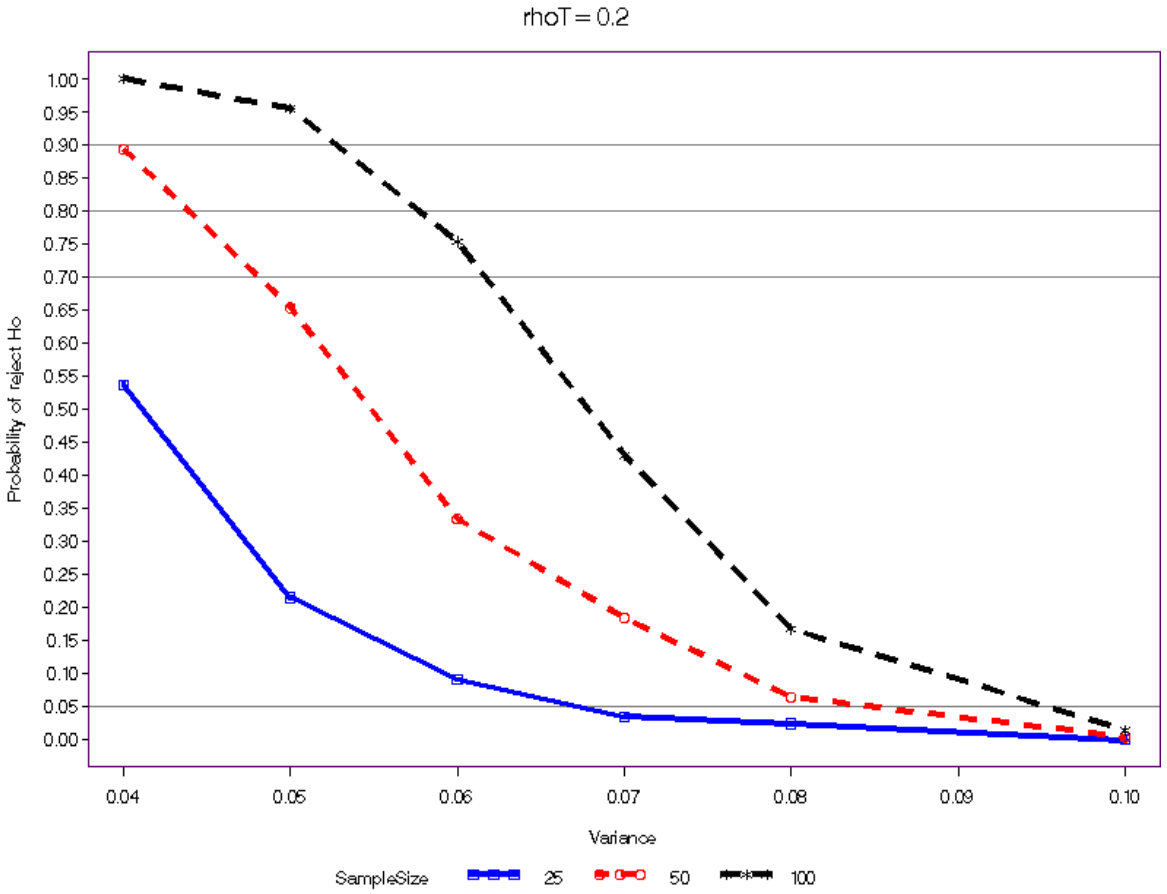


Table 7. Effect of the reference and the test correlations on multivariate PBE criterion (Cp) as a function of θ_0

ρ_R	ρ_T											
	-0.2	-0.1	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
-0.2	+	+	+	+	+	+	+	+	+	+	+	+
-0.1	+	+	+	+	+	+	+	+	+	+	+	+
0	=	=	=	=	=	=	=	=	=	=	=	=
0.1	-	-	-	-	-	-	-	-	-	-	-	-
0.2	-	-	-	-	-	-	-	-	-	-	-	-
0.3	-	-	-	-	-	-	-	-	-	-	-	-
0.4	+	+	-	-	-	-	-	-	-	-	-	-
0.5	+	+	+	-	-	-	-	-	-	-	-	-
0.6	+	+	+	+	+	-	-	-	-	-	-	-
0.7	+	+	+	+	+	+	+	-	-	-	-	-
0.8	+	+	+	+	+	+	+	+	+	-	-	-
0.9	+	+	+	+	+	+	+	+	+	+	+	-

$+: \theta > \theta_0,\ \ =: \theta = \theta_0,\ \ -: \theta < \theta_0,$

θ_0 is θ under no correlation in both reference and test, $\theta_0 = 3.4897$

4.5.3 Asymmetry of test

The upper bound of the multivariate BE region, θ, were calculated according to the correlations between the test, columns in Table 7, and between the reference, rows in the table. The calculated θ's were compared to θ_0, which was calculated assuming independence between the test variables, and independence between the reference variables. The difference between these two θ's represents the difference between the BE regions under correlation and under independence. The '+' represents the condition where the BE region is larger under correlation, while the '-' represents the condition when the BE region was larger under independence. The equal sign represents equal BE regions under both conditions. The diagonal of Table 7 represents the conditions where the reference correlations are equal to the test correlations. The table is clearly asymmetric, because its entries are not mirror images across the diagonal. This means that the BE regions are not equal for the same combinations of correlations depending on which drug is considered the reference. This is due to the scaling of the MV PBE criterion by the reference variance. This results in the possibility of considering a drug A as bioequivalent to a drug B, while Drug B is not bioequivalent to drug A.

4.6. Graphing the BE regions

To graphically illustrate the nature of the regions of equivalence consider the following. As before, let $\mathbf{X}_T$ and $\mathbf{X}_R$ be vectors of random variables representing the metrics used. Let $\mathbf{X}_T$ be distributed as multivariate normal with mean vector $\boldsymbol{\mu}_T$ and variance covariance matrix $\boldsymbol{\Sigma}_T$. Let $\mathbf{X}_R$ be distributed as multivariate normal with mean

vector $\boldsymbol{\mu}_R$ and variance covariance matrix $\boldsymbol{\Sigma}_R$. Then, as described in chapter, section 2.3 bioequivalence acceptance region is defined as

$$C_p = trace\left(\boldsymbol{\Sigma}_T \boldsymbol{\Sigma}_R^{-1}\right) + \left(\boldsymbol{\mu}_T - \boldsymbol{\mu}_R\right)' \boldsymbol{\Sigma}_R^{-1} \left(\boldsymbol{\mu}_T - \boldsymbol{\mu}_R\right) - p \le \theta. \tag{37}$$

Note that this could be rewritten,

$$\left(\boldsymbol{\mu}_T - \boldsymbol{\mu}_R\right)' \boldsymbol{\Sigma}_R^{-1} \left(\boldsymbol{\mu}_T - \boldsymbol{\mu}_R\right) \le \theta + p - trace\left(\boldsymbol{\Sigma}_T \boldsymbol{\Sigma}_R^{-1}\right). \tag{38}$$

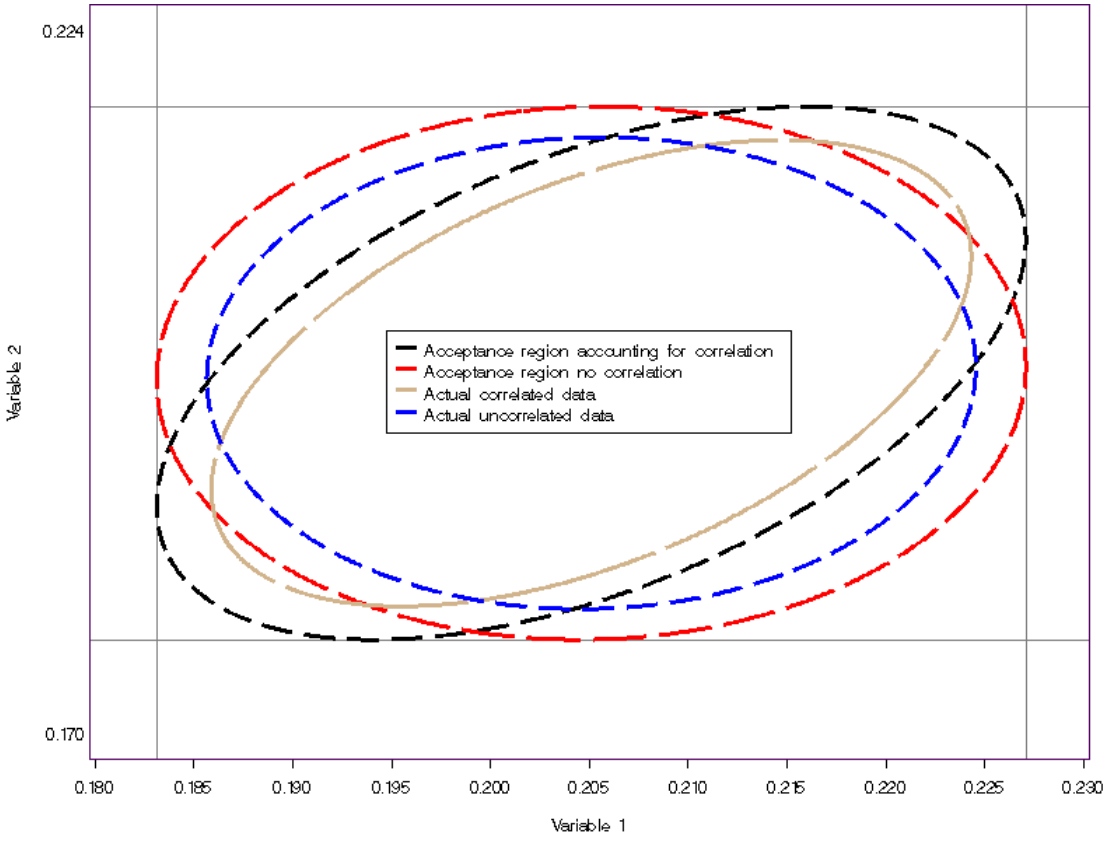
The right hand side of this inequality, which is a scalar, does not depend on the means. Therefore, the left hand side is a quadratic form in terms of the difference of the mean vectors and thus the inequality represents an ellipsoid whose shape is controlled by the variance covariance matrices (Johnson and Wichern 2002). Similarly we could construct an ellipsoid based on the estimates of the parameters on both LHS and RHS of the equation above, say, using $\hat{C}_p \le C_{p95}$. This ellipsoid is formed by the inequality,

$$\left(\hat{\boldsymbol{\mu}}_T - \hat{\boldsymbol{\mu}}_R\right)' \hat{\boldsymbol{\Sigma}}_R^{-1} \left(\hat{\boldsymbol{\mu}}_T - \hat{\boldsymbol{\mu}}_R\right) \le C_{p95} + p - trace\left(\hat{\boldsymbol{\Sigma}}_T \hat{\boldsymbol{\Sigma}}_R^{-1}\right). \tag{39}$$

Then, graphically one could conclude bioequivalence if the ellipsoid based on the 95% region of the data (equation 39) is totally contained within the ellipsoid defined by the θ. In Figure 6 examples of these ellipsoids are presented. The red ellipse represents the BE acceptance region constructed under the FDA defined conditions for two independent variables. The blue ellipse represents the area of the data bounded by the 95th percentiles of the two independent variables. The black ellipse represents the BE acceptance region under correlated variables. Finally, the gold ellipse represents the area of the data bounded by the 95th percentiles of the two correlated variables. The overlap of these ellipses demonstrates how accounting for the correlation between the variables

actually changes the acceptance region. This result could be vastly different depending on how the correlation is incorporated. For example, area in the figure that is outside the BE region that ignores correlation (red ellipse) is actually within the bioequivalence region that accounts for the correlation (the black ellipse). On the other side there are areas within the region that ignores correlation but they are outside the BE regions defined under correlation.

Figure 6 Acceptance bioequivalence regions



5 Multivariate Extensions of Population Bioequivalence: A Comparison Between three Measures

5.1. Abstract

In this article an extension of the univariate methods for evaluating the population bioequivalence (PBE) of a generic drug to a pre-licensed drug, or the bioequivalence of two formulations of a single drug is extended to simultaneously test for multiple correlated pharmacokinetic metrics. Specifically the univariate criterion recommend by the food and drug administration (FDA) is extended. One of the extensions proposed in the literature (Chervoneva, 2007), attempts to extend the univariate PBE through the use of the trace of the matrix of variances covariances of the pharmacokinetic measures. However, the trace, being the sum of the diagonal elements, does not incorporate the covariance. Dragalin et al. (2003) proposed a multivariate criterion using the Kullback–Leibler divergence (KLD) as a measure of discrepancy between the distributions of the two formulations. This criterion does not reduce to the univariate criterion proposed by the FDA, because it is not scaled by the reference variance.

The extension proposed here, similar to the univariate PBE, uses an inequality in quadratic forms. A parametric bootstrap method is used to determine the $(1-\alpha)100\%$ critical point of the distribution of the quadratic form. The performance of the proposed multivariate criterion is evaluated through a simulation study. The results from a simulation study and an application of this method are presented. The three criteria are compared by a simulation and applications.

5.2. Key words

multivariate bioequivalence; population bioequivalence; *AUC*; *Cmax*

5.3. Introduction

Bioequivalence studies are used in the development of generic drugs and the development of new formulations of drugs that were previously approved. Developing a new drug and obtaining approval from the Food and Drug Administration (FDA) requires multiple clinical trials to document the toxicity and the efficacy of the pharmacologically active ingredients of the new drug. A generic formulation of an approved compound is not subject to the multiple clinical trial requirement of a new compound because it is assumed that the active ingredients of the generic drug have the same toxicity and therapeutic efficacy as the approved drug. Thus, a generic must only demonstrate bioequivalence to the approved drug; once bioequivalence is demonstrated, it is also assumed that the therapeutic efficacy is similar between the approved and generic drugs. Thus the bioequivalence studies are designed to establish this expected similarity of the approved drug to the generic drugs having the same active ingredients.

Several pharmacokinetic metrics are used to quantify bioavailability. Experiments are designed to measure the concentration of the active ingredient in the biological active site, like blood, at appropriate time intervals. A profile of the concentration of the drug over time is then generated. The pharmacokinetic metrics, specifically, AUC, the maximum absorbed, $T_{\max}$ the time at which $C_{\max}$ occurs, rates of absorption, rate of elimination, and blood or plasma half lives are calculated.

For any given metric two drugs are defined to be bioequivalent if

$$\delta^2 \leq \theta, \tag{40}$$

where θ is a predefined constant, and for any given metric, δ is a critical value obtained from the distribution of a distance function of the new and the reference drugs. The upper limit θ is often prescribed by regulatory agencies. For example, the US FDA's guideline suggests θ should be 20% of the reference mean. The δ is estimated from the data and is, in general, the $(1 - 2\alpha)$*th* percentile of the distribution of the distance function for a given confidence level α. If δ^2 satisfies the inequality above the two drugs are considered bioequivalent. For instance, if the 90% confidence interval of the distance function falls completely within the interval $[-\theta, \theta]$, bioequivalence is concluded. This procedure is equivalent to testing two one-sided hypotheses each at level α using an analogous test (Schuirmann, 1987).

The average bioequivalence (ABE) test focuses on the differences in the means of the pharmacokinetic parameters. This method of evaluating bioequivalence, does not account for differences in the variability between the reference and test drugs. Population bioequivalence (PBE) was proposed to evaluate prescribability of the drug. Prescribability of a drug is defined as the ability to get the same effect by prescribing the brand-name drug or its generic drug to a new patient (Chow and Liu, 1992). In contrast to average BE, the PBE includes comparisons of the means and the total variability of the pharmacokinetic measures between the reference and test drugs (Hauk and Anderson, 1992).

The PBE was introduced by FDA in 1997 as an alternative method of testing BE. The PBE is a measure of the distance between the test and reference distributions with respect to the first two moments while the ABE is simply the difference between the first

moments only. The PBE may be thought of as the ratio of two expected squared distances where the numerator is the expected squared distance between the reference and the test and the denominator is the expected squared distance between two reference observations. Bioequivalence, then is determined by the ratio of the two expected squared differences is within a predefined distance, θ, from unity. That is,

$$\frac{E\left[\left(y_T - y_R\right)^2\right]}{E\left[\left(y_R - y_{R'}\right)^2\right]} - 1 \le \theta \tag{41}$$

where y_T is a random variable denoting the test PK metrics, y_R and $y_{R'}$ are two realizations of the reference random variable and *E* represents the expectation.

The univariate PBE criterion in (41), by substituting the unit ratio of the denominator term for the 1, could be redefined as (Sheiner 1992, Schall and Luus 1993),

$$\frac{E\left[\left(y_T - y_R\right)^2\right] - E\left[\left(y_R - y_{R'}\right)^2\right]}{E\left[\left(y_R - y_{R'}\right)^2\right] \Big/ 2} \le \theta . \tag{42}$$

Rewriting Eqn (42) in terms of the population mean and variance, it reduces to,

$$C = \frac{\left(\mu_T - \mu_R\right)^2 + \sigma_T^2 - \sigma_R^2}{\sigma_R^2} \le \theta . \tag{43}$$

where μ_R and μ_T are the means of the reference and the test random variables respectively, and σ_R^2 and σ_T^2 are the population variances of the reference and test pharmacokinetics respectively. Thus, the δ^2 from the original inequality in (40) is a function both of a distance metric of the means as well as the variances. The hypothesis test form of PBE uses the hypotheses $H_0 : C > \theta$ vs $H_a : C \le \theta$. Bioequivalence is

concluded with $(1-\alpha)$x100% confidence if $\hat{C}_{(1-\alpha)} \leq \theta$, where $\hat{C}_{(1-\alpha)}$ is the estimate of the upper limit of the one-sided $(1-\alpha)100th$ confidence interval of the PBE criterion defined in (43) using the maximum likelihood estimates (mle's) of the means and variances.

Extending this to more than one metric requires accommodation of the correlation. For example, suppose that the blood absorption coefficient (K_a), and the time (T_{max}) until the maximum concentration (C_{max}) of the blood concentration is reached, and the area under the blood concentration curve (AUC), are all calculated from the same blood concentration-time profile. In this case, the assumption of independence in testing bioequivalence using multiple tests for each of the four parameters is not justifiable. Clearly, the correlations among these variables should be incorporated in the multivariate tests of bioequivalence.

Multiple multivariate extensions for the average BE (Brown, 1995; Berger and Hsu, 1996; Brown, 1997; Munk and Pflujer, 1999; Wang, 1999; and Tamhane and Logan, 2004) have been proposed in the literature. However, there are few that deal with the multivariate PBE. . The first notable exception is Dragalin et al. (2003), in which the Kullback–Leibler divergence (KLD) is used as a measure of discrepancy between the distributions of the two formulations. They propose a generalization of average and PBE measures, and generalized it to the multivariate situation. Their multivariate method could be summarizes as follows. Consider a multivariate random variable $\mathbf{Y}$ representing a set of PK metrics. Suppose $\mathbf{Y}$ is distributed as normal with mean vector $\boldsymbol{\mu}$ and variance covariance matrix $\boldsymbol{\Sigma}$. Let T and R represent treatment and reference groups, respectively. Dragalin et al. (2003) propose a criterion based on the following inequality

$$D_p = \frac{1}{2} trace\left\{\left[\left(\boldsymbol{\mu}_T - \boldsymbol{\mu}_R\right)\left(\boldsymbol{\mu}_T - \boldsymbol{\mu}_R\right)' + \boldsymbol{\Sigma}_T + \boldsymbol{\Sigma}_R\right]\left[\boldsymbol{\Sigma}_T^{-1} + \boldsymbol{\Sigma}_R^{-1}\right]\right\} - 2p < \theta \qquad (44)$$

Here, the left hand side (LHS) of the equation is the KLD. Two formulations are declared bioequivalent if the upper bound of a level-α confidence interval for the KLD is less than a given specific value, θ. This criterion does not reduce to the univariate criterion proposed by the FDA in equation (43); instead it reduces to

$$D_p = \frac{1}{2}\left[\frac{\left(\mu_T - \mu_R\right)^2 + \sigma_T^2 - \sigma_R^2}{\sigma_R^2} + \frac{\left(\mu_T - \mu_R\right)^2 + \sigma_R^2 - \sigma_T^2}{\sigma_T^2}\right] \qquad (45)$$

Thus, this criterion may be seen as the average of two terms where the first term is the same measure of distance scaled by the reference variance proposed by FDA. The second term is similar except that it is scaled by the variance of the test. This criterion is equivalent to the FDA proposed criterion only if the reference and test variances are equal, and then it is only a measure of the squared mean distances and not the differences of the variances. Dragalin et al. (2003) only proposed the multivariate criterion, but never tested it.

The second notable exception is Chervoneva et al. (2007) in which they propose a criterion for the *p*-variate multivariate case using the trace of the variance-covariance matrices. Although this criterion reduces to the univariate PBE when $p = 1$, it is not an appropriate extension to the multivariate case as the method fails to incorporate the correlations. The trace of the matrices, (the sum of the diagonal elements of the covariance matrix), ignores the correlations.

The bioequivalence rule proposed by Chervenova el al. (2007) is,

$$B_p = \frac{(\boldsymbol{\mu}_T - \boldsymbol{\mu}_R)'(\boldsymbol{\mu}_T - \boldsymbol{\mu}_R) + tr(\boldsymbol{\Sigma}_T) - tr(\boldsymbol{\Sigma}_R)}{tr(\boldsymbol{\Sigma}_R)} \leq \theta \tag{46}$$

In applications they propose constructing a 95% confidence interval for B_p, and deem the test and reference distributions bioequivalent when the upper limit of this interval is less than the predefined θ. For p = 1 this rule reduces to the univariate rule in equation (43).

The objective of this study was to develop a multivariate PBE criterion that is equivalent to the univariate criterion approved by the FDA, then comparing the three criteria in testing PBE in a simulation and using examples. The following sections the development of the new criterion are discussed and a standardized method to compare between them is outlined. More detailed discussion of the properties of the newly developed criterion will be presented.

5.4. Methods

5.4.1 Development of the multivariate bioequivalence criterion C_p

The proposed multivariate extension is based on the fundamental definition of the PBE. That is, the difference in the means and the difference in the variances are scaled by the reference covariance matrix and summed. The criterion is given as follows:

$$C_p = trace\left(\boldsymbol{\Sigma}_T \boldsymbol{\Sigma}_R^{-1}\right) + (\boldsymbol{\mu}_T - \boldsymbol{\mu}_R)' \boldsymbol{\Sigma}_R^{-1} (\boldsymbol{\mu}_T - \boldsymbol{\mu}_R) - p < \theta \tag{47}$$

To justify this let $\mathbf{Y}_T$ and $\mathbf{Y}_R$ be p-variate random variables denoting the test and reference PK metrics. Assume, $\mathbf{Y}_T$ is distributed as a p-variate normal with mean vector $\boldsymbol{\mu}_T$ and variance covariance matrix $\boldsymbol{\Sigma}_T$. And let $\mathbf{Y}_R$ and $\mathbf{Y}_{R'}$ be two realizations of the

p-variate normally distributed random variable with mean $\boldsymbol{\mu}_R$ and variance covariance matrix $\boldsymbol{\Sigma}_R$.

The multivariate extension proposed here basically extends the definition (42) for the univariate case. The squares in (41) are replaced by quadratic forms and the denominator is replaced by its multivariate equivalent, namely the inverse of the corresponding matrix

$$\boldsymbol{\Sigma}_R = \frac{1}{2} E\left[\left(\mathbf{Y}_R - \mathbf{Y}_{R'} \right) \left(\mathbf{Y}_R - \mathbf{Y}_{R'} \right)' \right]. \tag{48}$$

Then the multivariate expression similar to (42) is as follows:

$$C_p = E\left[\left(\mathbf{Y}_T - \mathbf{Y}_R \right)' \boldsymbol{\Sigma}_R^{-1} \left(\mathbf{Y}_T - \mathbf{Y}_R \right) \right] - E\left[\left(\mathbf{Y}_R - \mathbf{Y}_{R'} \right)' \boldsymbol{\Sigma}_R^{-1} \left(\mathbf{Y}_R - \mathbf{Y}_{R'} \right) \right] \tag{49}$$

Substituting $\mathbf{Z} = \boldsymbol{\Sigma}_R^{-1/2} \left(\mathbf{Y}_T - \mathbf{Y}_R \right)$; and $\mathbf{K} = \boldsymbol{\Sigma}_R^{-1/2} \left(\mathbf{Y}_R - \mathbf{Y}_{R'} \right)$ in (49), and using the matrix and trace properties, the multivariate criterion in (49) could be expressed as

$$C_p = trace\left(\boldsymbol{\Sigma}_\mathbf{Z} \right) + E\left[\mathbf{Z} \right]' E\left[\mathbf{Z} \right] - trace\left(\boldsymbol{\Sigma}_\mathbf{K} \right) - E\left[\mathbf{K} \right]' E\left[\mathbf{K} \right] \tag{50}$$

It can be shown that the expectation of $\mathbf{Z}$ is equal to $\boldsymbol{\Sigma}_R^{-1/2} \left(\boldsymbol{\mu}_T - \boldsymbol{\mu}_R \right)$, and its variance covariance matrix is $\boldsymbol{\Sigma}_R^{-1/2} \boldsymbol{\Sigma}_T \boldsymbol{\Sigma}_R^{-1/2} + \underline{\mathbf{I}}$, the expectation of $\mathbf{K}$ is $\underline{0}$, and its variance covariance matrix $\boldsymbol{\Sigma}_\mathbf{K}$ is equal to $2\underline{\mathbf{I}}$, where $\underline{\mathbf{I}}$ is a $p \times p$ identity matrix (see Appendix B). Substituting these expectations and variance covariance matrices in (50) reduces to the multivariate criterion in (47) (Dahman 2009 Ch3 for details).

It is simple to show that the resultant multivariate criterion reduces to the univariate criterion (43) when p=1. It also accounts for the total variability and the correlations among the PK metrics used in evaluating bioequivalence.

Using the invariance property, the maximum likelihood estimator of the multivariate PBE criterion could be estimated from the data as:

$$\hat{C}_p = trace\left(\hat{\boldsymbol{\Sigma}}_T \hat{\boldsymbol{\Sigma}}_R^{-1}\right) + \left(\hat{\boldsymbol{\mu}}_T - \hat{\boldsymbol{\mu}}_R\right)' \hat{\boldsymbol{\Sigma}}_R^{-1} \left(\hat{\boldsymbol{\mu}}_T - \hat{\boldsymbol{\mu}}_R\right) - p, \qquad (51)$$

where $\hat{\boldsymbol{\mu}}_T$ and $\hat{\boldsymbol{\mu}}_R$ are the maximum likelihood estimates (mle's) of the population means, and $\hat{\boldsymbol{\Sigma}}_T$ and $\hat{\boldsymbol{\Sigma}}_R$ are mle's of the variance covariance matrices of the test and reference variables.

5.4.2 Constructing the $100(1-\alpha)th$ confidence interval of the MV criteria

The exact distribution of the MV criteria introduced earlier, B_p, D_p, and C_p are not tractable. Therefore, a parametric bootstrap method (Efron & Tibshirani, 1993), as recommended by the FDA, is proposed. This method requires obtaining the mle's of the population parameters $\boldsymbol{\mu}_T$ and $\boldsymbol{\Sigma}_T$ of the test and $\boldsymbol{\mu}_R$ and $\boldsymbol{\Sigma}_R$ of the reference metrics. Random samples of the same size as the original experiment would be generated from multivariate normal distribution with means and variances equal to the mle's of each of the drugs in the original experiment. The MV population criteria would be calculated for each of the bootstrap samples. Then the $100(1-\alpha)th$ percentile of the distribution of criteria based on the *B* bootstrap samples would be determined. PBE is concluded if the $100(1-\alpha)th$ percentile is less than the predefined limit, θ.

5.4.3 Specifying the upper limit, θ, of BE

In the univariate case, θ is defined according to predetermined limits determined by the FDA. The maximum difference between the variances of the test and the reference ($\sigma_T^2 - \sigma_R^2$) allowed by FDA (1997) is 0.02, and the minimum allowed variance of the reference (σ_R^2) is 0.04.

These limits are extended to the multivariate criterion, by setting the maximum difference between the means of the test and reference pharmacokinetic measures as the natural logarithm of 1.25; the maximum difference between the test and reference variances as 0.02, and the lowest variances as 0.04. Since there is no analogous guideline for incorporating the correlations, different combinations of correlations among the test and reference variables and will be used. The value of θ depends on the PBE criterion used. This value would not depend on the correlation for Chervenova's, so it is always constant.

Using the proposed multivariable limits, values of θ were calculate for the case where p = 2 with the BE limits of means and variance differences as defined by FDA. Dahman (2009) have shown that the upper limit of the acceptance region of BE, θ, is affected by the number of parameters, the correlations of test and by the correlations of reference. When these correlations are identical, the bigger the correlation the smaller is θ (data not shown). When the correlations are negative, the value of θ is always greater than θ_0. On the other hand, when the correlations are positive, as expected among PK, θ is always smaller than θ_0. Dahman (2009) demonstrated that this variability in the values of θ affects the results of testing the hypothesis of bioequivalence. This phenomenon

affects both criteria that account for the correlation, i.e. D_p and C_p. This will require the regulatory bodies to study the conditions and effect of the values of the correlations to set the values of the predefined upper limit of the population criterion.

5.5. Properties of the multivariate PBE criterion C_p

Properties of the multivariate PBE criterion C_p proposed in the previous sections are studied using Monte-Carlo simulation methods. The simulation study was designed to evaluate the distribution of the proposed multivariate criterion C_p under different combinations of sample size (number of subjects in the trial), differences in the averages and variances of the pharmacokinetic (PK) parameters between the reference and test drugs, and under different correlations between the PK parameters within each treatment group. This study was mainly designed to guide in the selection of a method to construct the confidence interval for the proposed criterion. Another simulation study was designed to study the size and power of the hypothesis tests, and to compare the multivariate criterion versus the multiple testing using the univariate criteria.

These studies were limited to equal size samples of reference and test drugs. The effect of different sample sizes, missing values and dependence between the treatments drugs should be tested in future studies.

It has been shown in many studies that the log-transformed pharmacokinetic parameters have a normal distribution, and that data extracted from the same concentration time profiles for each subject are correlated. To create samples that preserve these properties we used random sampling from multivariate normal

distribution. A sample $\underline{Y}$ of N sets of two pairs of variables was generated by random selection from a 4-variate normal distribution.

The first pair of variable y_{R1i} & y_{R2i} represent the log-transformed pharmacokinetic parameters of the reference group, namely $\log(C_{\max})$ and $\log(AUC)$. The second pair y_{T1i} & y_{T2i} represent the log-transformed data of the test group, where i represents the i^{th} subject, $i = 1,...,N$. Each pair has a bi-variate normal distribution, which we will denote by $\underline{Y_k} \sim N_p(\underline{\mu}_k, \underline{\Sigma}_k)$, where k is the treatment, and the population means vector for each treatment is $\underline{\mu}_k = [\mu_{k1}, \mu_{k2}]$ and $\underline{\Sigma}_k$ is the variance covariance matrix for each treatment.

SAS IML code calling the function (VNORMAL) was used in a macro to generate the samples of *p*-variate normal data. All simulations and analyses were done using macros in SAS 9.2, SAS Institute, Cary, NC.

The distribution of C_p is not known, and cannot be easily determined. The FDA recommended using the parametric bootstrap method to construct the confidence interval for the univariate PBE criterion rather than applying a normal approximation. It was concluded that similar to the univariate case, using the parametric confidence intervals which assume normality for the multivariate PBE criterion C_p would not be appropriate. In order not to assume any distribution the parametric bootstrap method to evaluate the percentile confidence interval was used as suggested by the FDA. The parametric bootstrap method was used because there is enough evidence that the log-transformed pharmacokinetic measures like AUC, $C_{\max}$ and $T_{\max}$ follow a normal distribution. For

that reason it is convenient to resample (for the bootstrap) from a *p*-variate normal distribution in which means and variance covariance matrix are equal to those estimated from the sample.

The non-parametric confidence interval was constructed using the percentiles from the parametric bootstrap method. From each replication of the experimental settings, the means and the variance covariance matrices were estimated. Then 2000 bootstrap samples were generated by randomly selecting from multivariate normal distribution with mean and variance equal to those estimated from the replication. Then estimates of the multivariate PBE criterion $\hat{C}_p$ were calculated for each of these 2000 bootstrap samples. The 95th bootstrap percentile of $\hat{C}_p$ was determined. This is the upper limit of the one sided 95% confidence interval for $\hat{C}_p$ that would be used in the hypothesis testing.

5.5.1 Simulation Configurations

The simulation study was run with 500 replicates. Each replicate represented a bioequivalence trial with N subjects in each group (treatment or reference). For each subject two measures representing the log-transformed AUC and $C_{\max}$, were selected from the *p*-variate normal distribution as described earlier. To estimate the nonparametric confidence interval in each simulation, 2000 bootstrap samples were randomly selected from the p-variate distributions. Efron (1982) demonstrated that in general 2000 bootstrap samples would be large enough to obtain unbiased confidence interval.

Following FDA guidelines for the univariate case of PBE, and based on the information from published bioequivalence trials several factors were considered at specific levels. It is important to evaluate the performance of any new statistic or test under a variety of sample sizes. The usual number of subjects in most of the bioequivalence drug trials varies between 20 and 100 per group. Three sample sizes were chosen (25, 50 and 100). As discussed earlier the maximum allowable difference between the (log transformed) means of the test and the reference parameters is ln(1.25). the difference in the means was test at 5 values (-2ln(1.25), -ln(1.25)/2, 0, ln(1.25), 2ln(1.25)). As recommended by the FDA, the minimum value for the reference standard deviation of 0.2 was set as the value for the variance of the reference pharmacokinetics. The difference in the variance between the variances of the reference and the test pharmacokinetic were selected from 0.04, 0.06 and 0.1. The correlations between variables were not used in the definition of bioequivalence by the FDA or the previous methods of testing multivariate PBE. The performance of the proposed criterion was evaluated under a wide range of correlations that allow for all possible values including rare cases of negative correlations. The effects of the correlations between -0.2 and 0.8 in increments of 0.2 on the power and the size of the tests for bioequivalence were evaluated.

The size of testing the hypotheses of BE was evaluated under conditions that represent the maximum allowable values of variances and differences between the means of the reference and test drugs. The power of the test was evaluated under conditions that fulfill the BE with respect to the variances and the means. The size and power were

evaluated under different sample sizes and variable correlations between the reference and between the test variables.

5.5.2 Description of the simulation steps

The simulation performed the same steps for each of the combinations of sample size, difference in means, difference in variance and correlations between the reference and the correlations between the test PK measures. Five hundred samples (replicates) of 4 random variables of size N were generated from a multivariate normal distribution with reference means μ_{R1} and μ_{R2} and test means μ_{T1} and μ_{T2}, and correlations ρ_R and ρ_T and variances σ^2_{R1}, σ^2_{R2}, σ^2_{T1}, and σ^2_{T2} that define each configuration or setting. The mean and variance were estimated for each replication. Each replication was bootstrapped by randomly selecting 2000 from multivariate normal distribution with the means and variances equal to those estimated for each replication. The MV criterion $\hat{C}_p$ was calculated fro each bootstrap sample. The 95th percentile for the 2000 calculated bootstrap estimates of $\hat{C}_p$ was determined for each replication. For each replication, BE was determined if the upper limit of the one-sided 95% confidence interval for $\hat{C}_p$ was less than the predefined θ. The number and percent of times the result was correct in agreement with the setting were calculated using the multivariate criterion and using the univariate criterions for each of the variables. The size and power of the tests were calculated by calculating the percentage of times the null hypothesis was rejected among the 500 replicates, if the null hypothesis was true or was not true respectively.

5.5.3 Simulation Results

The size of the test was evaluated using random selection from distributions that have the maximum accepted values of the differences between the means and between the variances. These values define the boundaries of the multivariate PBE acceptance region in all dimensions. These are the values that define the true value of the multivariate PBE criterion as equal to the predefined θ. The table compares between two tests, the first accounts for the correlation in the predefined θ, the second one ignores the correlations when predefined θ. The type I error are very conservative in the case of smallest tested sample sizes of 25, these errors increase by increasing the sample size. These errors never exceed the 0.05 level except in higher correlations than 0.2 between the reference measures, when using the theta that ignores the correlations. The largest type I errors are observed when the correlations of either the reference or the test are very high (0.8), and when the differences between the correlations of the reference and the correlations among the test are large.

Probability of the type I errors are also sensitive to the magnitude of the correlation used in calculating the θ. Testing the bioequivalence hypothesis using negative correlations, on samples drawn from positively correlated random variables, causes the highest increase in type I errors. There is a minimal increase in probability of Type I error when using a positive correlation that is not equal to the true correlation except for testing with lower correlations when the true correlation is high 0.8. This would result in rejecting the null hypothesis, and concluding PBE in more non-BE cases.

As expected the power power of the proposed multivariate PBE test was highest when the two true means of the reference and test variables were equal. The power drops

gradually as the difference between the means increases in either direction. Accounting for the correlation in the upper limit theta, θ, of PBE result in a test with less power than ignoring the correlation. The power increases as the absolute difference between the test and reference means increases. The test that ignores the correlation achieved the highest power, while the test that utilized the intersection of two separate univariate tests was with the smallest power. The power increases as the sample size increases and as the difference between the variances decreases,

Each simulation sample was classified as bioequivalent or not according to the proposed multivariate PBE test accounting for correlations, ignoring the correlations, and using the intersection of two univariate tests of BE with and without Bonferoni correction. Then according to each scenario the simulation samples were classified into correctly classified and incorrectly classified. Table 8 compares the correct classification of all BE and non-BE scenarios using those 4 methods. Ignoring the correlation is superior to all other tests in classifying the scenarios correctly when the truth is BE. However this superiority is at the cost of having the highest proportion of misclassification (Incorrect Classification) when the truth is not BE.

The power function was evaluated as a function of the mean of test drug under fixed values of the reference means and variance covariance matrix and under equal correlations between the test variables and between the reference variables. The power of rejecting the null hypothesis of "no BE" is higher than 0.9 when the reference means are equal to the test means. This power decreases as the difference increases. The power of testing the null hypothesis of 'no BE', under fixed variance of the reference, drops as the difference between the variances of the test and reference increases. This power increases

by the sample size as expected. The power is above 0.7 with sample size of 100 when the difference is smaller that 0.02.

Table 8 Percentage of cases classified correctly/incorrectly

		MV BE criterion accounting for the correlation					
		BE			Not BE		
		Incorrect	correct	Total	Incorrect	correct	Total
Intersection of 2 $(1-\alpha/2)CI$ Univariate tests (Bonferoni correction)	Incorrect	3.5	15.56	19.07	11.87	0.91	12.78
	Correct	0.37	80.56	80.93	12.29	74.93	87.22
	Total	3.88	96.12	100	24.16	75.84	100
Intersection of 2 $(1-\alpha)CI$ Univariate tests	Incorrect	3.23	7.81	11.04	15.84	1.72	17.56
	Correct	0.65	88.31	88.96	8.33	74.11	82.44
	Total	3.88	96.12	100	24.16	75.84	100
MV ignoring correlation in rule	Incorrect	0.9	0.52	1.43	22.66	11.04	33.71
	Correct	2.97	95.6	98.57	1.5	64.79	66.29
	Total	3.88	96.12	100	24.16	75.84	100

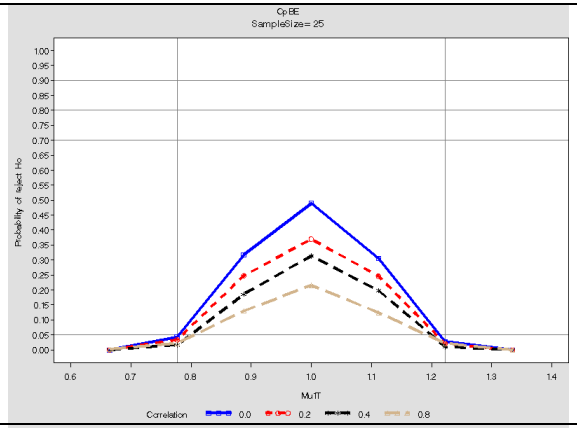
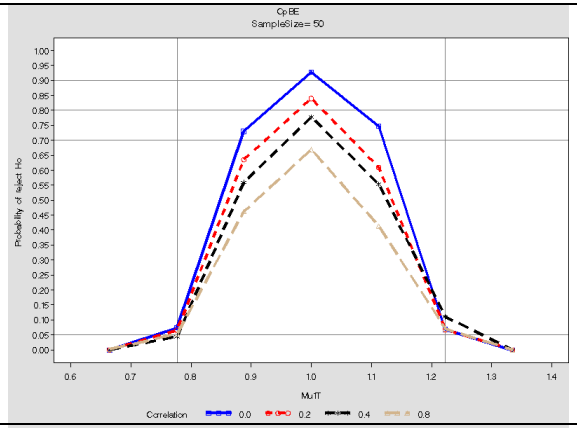
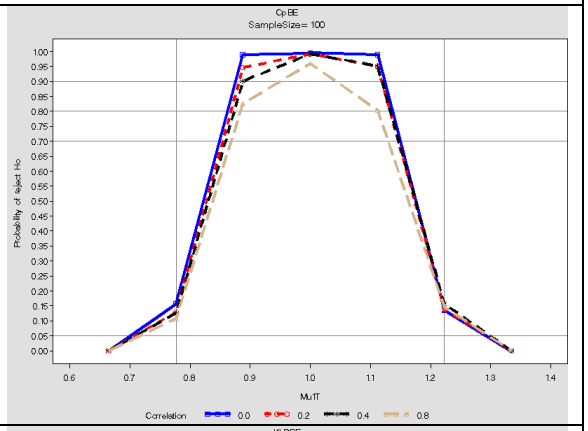
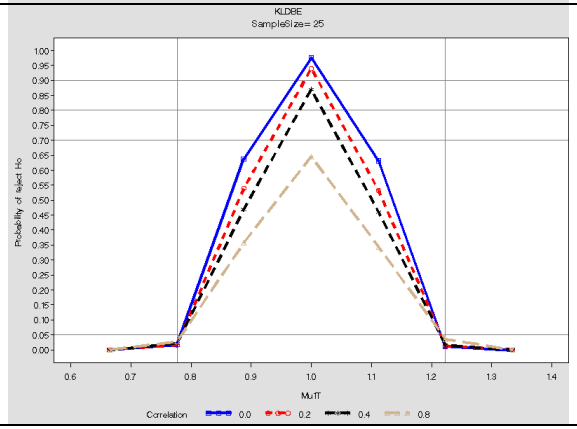
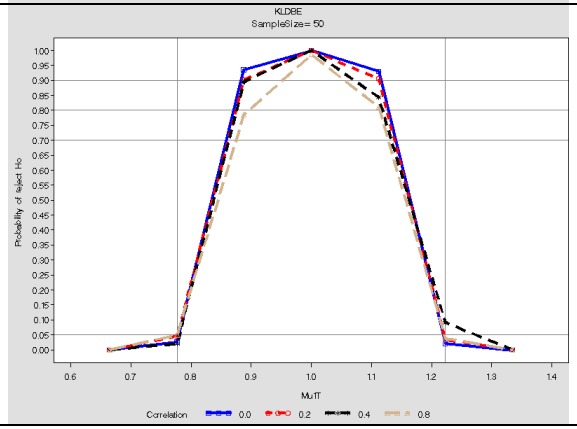
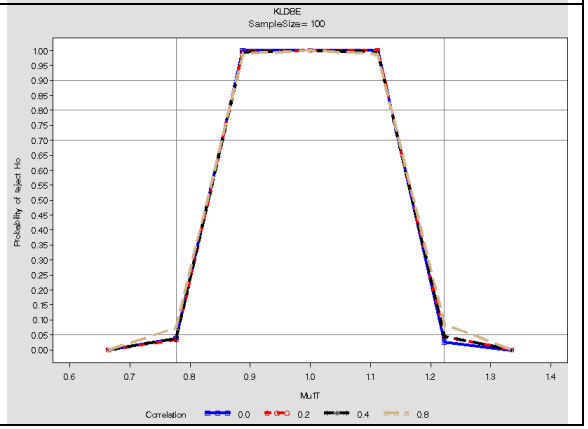
5.6. Comparison Between the Three MV PBE criteria

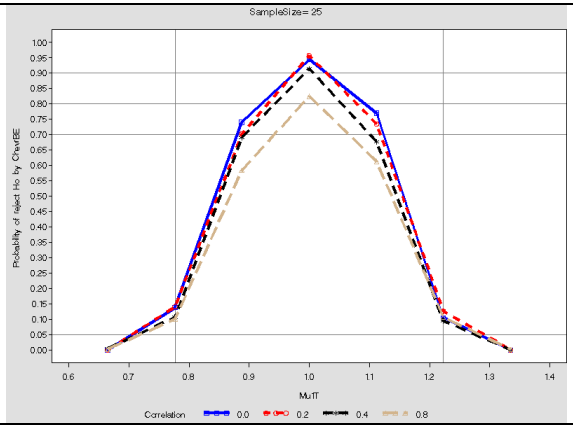
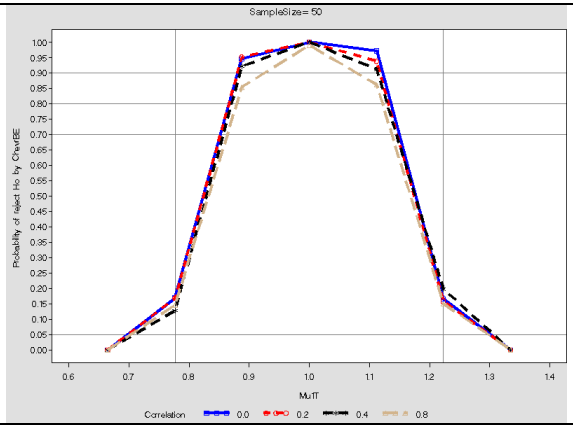
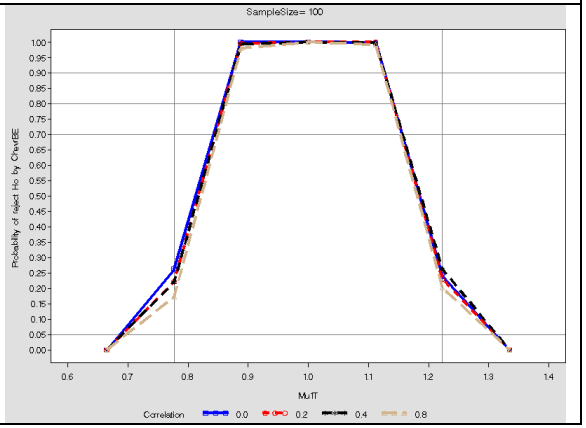
The power functions of testing for bioequivalence using the three multivariate PBE criteria were compared.. Chervenova (2007) linearized the condition of bioequivalence, and constructed confidence intervals using properties of the trace. Their method cannot be applied to the criteria proposed by Dragalin(2003) and Dahman(2009). Dragalin(2003) did not evaluate D_p. To be able to compare the

results properly, one standard method was applied to all three criteria. This method was to apply the simulation conditions demonstrated earlier to generate random samples. For each of those samples the mle's of the mean and variance of the test and reference were estimated. Each of the PBE criteria was calculated for each sample and using bootstrap, the $100(1-\alpha)th$ confidence intervals were constructed for each criterion. Then probabilities of BE were calculated for each scenario.

Figure 7 presents the comparison of the power functions of the three criteria at different sample sizes, and correlations. The plots demonstrate that the power of all tests is greater than 90% when the sample size is large (100). The power of *Cp* is lower than 90% with sample size 50. Although *Bp* does not account for correlation, the power function varied by the correlation when the sample size was small. This might be just random variations, but more tests are needed to verify that. The higher power of *Bp* is compromised by its higher size which exceeds the 0.05 level even when the sample size is only 25.

Figure 7 The Power Function by mean of test and sample size (under fixed variance and reference mean)

	N=25	N=50	N=100
Cp			
Dp			

	N=25	N=50	N=100
Chevr *Bp*			

5.7. Applications

5.7.1 Testing Population Bioequivalence in a parallel design

Data used was from a study performed by Clayton and Leslie (1981) to study the bioequivalence of an enteric-coated erythromycin base (test drug) to the previously available reference formulation of erythromycin stearate. After administering the reference or test drug, venous blood samples were collected at 0.0, 0.5, 1.0, 2.0, 4.0, 6.0, and 8.0 hours. The data are presented in the dissertation appendix F.

Although this study was intended to follow a two period crossover design, due to non randomization and due to the fact that all patients were given the reference drug in the first period, and the test drug in the second period it will be treated as a parallel design (Chinchilli & Elswick 1997). **(**this data set was selected because it provides the blood concentrations for all the collection times, which will be useful in illustrating the nonlinear model method proposed in the next paper.) To illustrate the multivariate method proposed here we first calculated the pharmacokinetic parameters.

Since each subject received both the reference and the treatment drugs, to account for the correlations from these repeated measures a mixed effect linear model was selected to fit this data. Treatment effect will be treated as a fixed effect. The two multiviarate outcome of interest are the Cmax and AUC which are products of the same blood concentration profile for each subject. To account for the repeated measure as well as the correlation of the outcomes, a multivariate mixed effect model was used and the two outcomes were fit simultaneously. Since there were only two periods we modeled the correlation between the two repeated measures as a compound symmetry. In studies with more period other structures like autoregressive(1) AR(1) or Toepliz are more appropriate and should be considered. To account for the correlation

between the outcomes we used an unstructured covariance structure. We also allowed for estimating different variance covariance matrices for the two treatment drugs. One of the limitations of this experiment was the result of administering the same drug to all patients in each period. This results in the redundancy of the period effect and the treatment effect. The period effect in properly designed cross-over studies should be modeled as a fixed effect.

From the blood concentration data we calculated the $C_{\max}$ as the highest concentration of the blood concentration time profile, the AUC using the trapezoidal method (Berger RL, 1996, Gibaldi, 1982). Since these measures are known to be log-normal, they were log-transformed. Table 9 presents the calculated and transformed data for the 20 subjects under the reference and the test drugs.

Table 9 AUC **and** $C_{\max}$ **from Clayton and Leslie**

Reference Treatment					Test Treatment			
SUBJECT	AUC	logAUC	Cmax	logCmax	AUC	logAUC	Cmax	logCmax
1	13.978	2.637	5.350	1.677	10.788	2.378	2.590	0.952
2	13.810	2.625	4.140	1.421	3.150	1.147	0.920	-0.083
3	8.865	2.182	3.460	1.241	5.710	1.742	2.600	0.956
4	9.923	2.295	4.510	1.506	14.033	2.641	3.430	1.233
5	10.545	2.356	3.700	1.308	5.908	1.776	1.430	0.358
6	9.855	2.288	4.100	1.411	7.935	2.071	1.990	0.688
7	23.680	3.165	6.180	1.821	16.295	2.791	3.680	1.303
8	8.320	2.119	4.070	1.404	6.553	1.880	3.580	1.275
9	6.353	1.849	2.500	0.916	7.643	2.034	2.120	0.751
10	12.925	2.559	5.230	1.654	7.658	2.036	1.360	0.307
11	7.918	2.069	4.240	1.445	2.888	1.060	1.010	0.010
12	6.700	1.902	2.520	0.924	9.068	2.205	3.110	1.135
13	10.768	2.377	3.490	1.250	10.743	2.374	2.870	1.054
14	5.970	1.787	2.100	0.742	4.375	1.476	0.820	-0.198
15	10.788	2.378	5.180	1.645	9.290	2.229	3.390	1.221
16	9.345	2.235	3.800	1.335	4.635	1.534	0.920	-0.083
17	5.850	1.766	1.960	0.673	3.050	1.115	0.850	-0.163
18	6.158	1.818	2.360	0.859	2.330	0.846	0.490	-0.713
19	1.650	0.501	0.560	-0.580	5.775	1.754	1.560	0.445

20	5.945	1.783	2.060	0.723	3.933	1.369	1.320	0.278

Using the mixed effects model the mle's of the means and covariances of both the reference and the treatment drugs were estimated. These estimates were

$$\hat{\boldsymbol{\mu}}_T = \begin{bmatrix} \hat{\mu}_{T C\max} \\ \hat{\mu}_{T AUC} \end{bmatrix} = \begin{bmatrix} 0.536 \\ 1.823 \end{bmatrix} \text{and } \hat{\boldsymbol{\mu}}_R = \begin{bmatrix} \hat{\mu}_{R C\max} \\ \hat{\mu}_{R AUC} \end{bmatrix} = \begin{bmatrix} 1.169 \\ 2.134 \end{bmatrix}$$

$$\hat{\boldsymbol{\Sigma}}_T = \begin{bmatrix} 0.2758 & 0.2556 \\ 0.2556 & 0.26325 \end{bmatrix} \text{ and } \hat{\boldsymbol{\Sigma}}_R = \begin{bmatrix} 0.28002 & 0.28035 \\ 0.28035 & 0.359535 \end{bmatrix}$$

The log transformed $C_{\max}$ and AUC were found to be highly correlated, as expected, both in the test group $\hat{\rho}_T$ = 0.88 and in the reference group where $\hat{\rho}_R$ was 0.95.

The multivariate criterion proposed ($\hat{C}_p$) was calculated from these estimates using equation (51). The univariate criterion of PBE for each of AUC and CMax were also calculated according to equation (43). Then the bootstrap method was used to determine the upper limits of the 90% confidence intervals for the MV criterion and the two univariate criteria. The multivariate PBE rule theta which defines the upper boundary were calculated in two ways: i) by ignoring the correlations between the PK (θ_p =3.49), and ii) by accounting for these correlations (θ_p =4.32).

Parametric bootstrap was set to generate 2000 samples and calculate the multivariate criterion for each sample and then generating the 90% confidence interval of original $\hat{C}_p$ and to determine the upper limits of the 90% confidence intervals for the MV criterion and the two univariate criteria, these were found as 16.87, 2.32 and 5.30. The qualitative results of these tests as bioequivalent or not are performed by comparing the upper limit of the one-sided 95% CI to the predetermined acceptance boundary.

To compare between the multivariate test and the univariate tests, we used the bootstrap samples to evaluate the probabilities of rejecting bioequivalence using the multivariate test, and the univariate tests. Those probabilities are actually the p-values of those tests. We also tested bioequivalence using the two univariate test simultaneously under the assumption of independence. This is a rectangular test which means that the p-value represents the probability of rejecting bioequivalence with respect to *AUC* and/or *Cmax*.

The estimates of the upper limit of the one-sided 95% CI were greater than the upper boundaries of the bioequivalence acceptance limits in the multivariate and the two univariate cases. So these results show that for this study the two formulations are not bioequivalent using either the multivariate or the univariate population bioequivalence testing. It also shows that for this specific example accounting for the correlations did not make a difference in the qualitative result.

Table 10 presents the results of the comparison of the tests p-values. Although all tests rejected bioequivalence in this study, there are noticeable differences between the p-values of the bivariate test p=0.8425 and the intersection of the two univariate tests p=0.5885. the value of the multivariate test is larger so it is more conservative in accepting BE the using the univariate test separately or simultaneously.

Table 10 result of PBE testing of Clayton data

		p
Bivariate test with rule theta no corr	P(Cp>3.49)	0.9185
Univariate AUC	P(C1>1.74)	0.5430
Univariate Cmax	P(C2>1.74)	0.0995
Bivariate test with rule theta with corr	P(Cp>4.32)	0.8425
Assuming independence Intersection of 2 univariate	1-[P(C1<1.74)* P(C2<1.74)]	0.5885

5.7.2 Testing multivariate PBE in a crossover design:

Twenty eight subjects were recruited in the study, and were given 3 doses of each of the two drugs, with a washout period between each of the two sessions. The order in which the two drugs were given was selected at random using a block randomization method to ensure the balance within each subject. Although there are 20 possible ways to assign three A's and three B's, the random process did not cover all possibilities.

At the beginning of each session, baseline data were collected 30 seconds prior to the administration of the drug. The pharmacological baseline measurement was supposed to be zero if the wash out period was appropriately long enough. Then the drug was administered, and the specified pharmacological measurement (level of specific active material in the blood) was obtained over several minutes at very short intervals.

Each subject was administered one of the two drugs in all six sessions, however due to mechanical failure, or lower than qualified measures, some of the sessions were completely missing all blood measures. No information about those sessions was provided other than the drug given.

From the blood concentration by time data presented in the APPENDIX E, the pharmacokinetic PK metrics were calculated. The maximum absorbed blood level of the active material (C_{max}) was defined as the highest observed concentration. The time associated with that concentration (C_{max}) was considered the T_{max}. The total absorbed drug was calculated as the AUC using the trapezoidal method. Table 11 presents the natural logs of C_{max} and AUC for each subject in each session of the study. The sequence in the table determines which drug was given in each period.

This study was powered to test the bioequivalence using one outcome at a time. Since we are interested in accounting for the correlations between the evaluated outcomes, the multivariate mixed model could be applied. We used a bivariate mixed model where the two outcomes log(*Cmax*) and log(*AUC*) were estimated simultaneously.

It also assumes that the correlation is decaying as the periods are further apart.

Using the estimates of log*Cmax* and log*AUC* for each of the treatment and the variance covariance structures displayed in Table 12, the univariate and multivariate criteria for PBE we calculated according to (43) and (51). Using bootstrap method, the upper limit of the 90% confidence interval of these criteria were calculated.

The correlations between the log*Cmax* and log*AUC* were found to be .9 for both the treatment and the the test drugs. The PBE criteria were estimated as $\hat{C}_p$ = -0.03132, univariate C1 = 0.057199, univariate C2 = -0.06109. The upper limits of the 90% CI were 1.75622, 1.17345, 0.95773 estimated by the bootstrap were compared to the acceptance boundaries 2.3103, 1.74483, 1.74483 for the multivariate, the univariate for *Cmax* and the univariate for the *AUC* respectively.

The comparison of the p-values for these tests are presented in Table 13. In this example again, all the tests resulted in the same qualitative result that those too drugs are bioequivalent. The differences in the p-values suggest the high possibility of getting different qualitative results in other samples or examples. The results in Table 13 display that the tests that accounts for the correlation between the outcomes has higher probability of rejecting bioequivalence than the test that ignore the correlation.

Table 11 PM data

sequence	logCmax1	logAUC1	logCmax2	logAUC2	logCmax3	logAUC3	logCmax4	logAUC4	logCmax5	logAUC5	logCmax6	logAUC6
ABABAB	1.668	5.112	.	.	.	.	.	.	.	.	.	.
BABABA	2.588	6.073	2.313	5.733	2.152	5.745	2.079	5.671	2.416	5.819	2.163	5.585
BABBAA	2.632	6.246	2.313	5.896	2.342	5.853	2.603	6.035	2.116	5.701	2.282	5.850
AABABB	3.025	5.930	2.407	6.004	2.067	5.667	2.001	5.563	2.434	5.839	2.241	5.813
ABABAB	.	.	.	.	2.001	5.608	.	.	.	.	.	.
ABBBAA	.	.	1.131	2.442	.	.	1.361	4.210	1.526	4.398	1.386	4.467
ABBAAB	2.976	6.276	3.077	6.460	2.912	6.455	2.981	6.471	2.815	6.276	2.912	6.449
BABBAA	1.482	4.417	1.668	4.488	1.411	4.377	2.907	6.159	2.791	5.972	2.653	6.157
ABAABB	.	.	2.851	6.125	2.695	5.677	2.398	5.670	2.510	5.849	2.754	5.906
ABAABB	2.241	5.628	2.262	5.798	2.028	5.592	2.092	5.550	2.175	5.621	2.241	5.800
BBBAAA	1.932	5.474	1.825	5.433	1.808	5.432	1.792	5.376	2.092	5.730	1.988	5.678
AABBAB	.	.	.	.	1.792	5.103	1.361	4.684	1.335	4.545	1.435	4.623
ABAABB	1.504	4.465	.	.	1.932	4.579	1.411	3.411	1.932	4.719	.	.
BABAAB	1.723	5.394	1.917	5.429	1.902	5.459	2.054	5.484	2.282	5.716	2.079	5.725
ABBAAB	1.335	4.624	.	.	1.686	5.371	2.152	5.594	1.974	5.147	1.361	5.017
BABBAA	1.740	5.348	1.163	3.984	1.723	4.433	2.425	5.595	1.361	3.671	1.386	4.520
BABABA	2.493	5.974	2.833	6.279	2.960	6.457	2.741	6.246	3.699	6.395	2.653	6.247
BAABAB	3.307	6.406	3.343	6.205	2.976	6.211	3.250	6.535	3.195	6.447	3.339	6.093
BBAAAB	2.251	5.760	2.116	5.697	2.175	5.864	2.028	5.460	1.932	5.642	.	.
ABBAAB	2.734	6.181	2.701	6.198	2.001	5.633	2.588	6.084	2.833	6.444	2.741	6.353
BBABAA	3.219	6.428	3.091	6.271	3.364	6.479	.	.	3.100	6.329	.	.
BABABA	2.610	6.007	2.760	6.154	3.030	6.524	2.728	6.197	.	.	3.049	6.660
BBABAA	2.896	5.933	3.367	6.204	3.144	6.097	2.944	6.055	3.250	6.398	3.127	6.296
AABABB	2.851	6.393	2.803	6.256	2.827	6.393	.	.	2.518	6.145	2.603	6.138
AABABB	2.342	5.791	2.688	6.351	2.407	6.093	2.868	6.359	3.118	6.838	2.797	6.339
AABABB	3.025	6.403	3.780	6.887	3.630	7.202	3.170	6.366	3.140	6.440	3.481	6.793
BBAABA	2.603	5.605	2.617	5.782	2.493	5.711	2.380	5.501	2.493	5.504	2.632	5.781
ABBABA	.	.	1.932	5.592	2.067	5.600	2.001	5.615	1.856	5.534	1.723	5.466

Table 12 Mean and covariance estimates of PM data using multivariate mixed model

	$\hat{\mu}_A$	$\hat{\mu}_B$	$\hat{\Sigma}_A =$	$\hat{\Sigma}_B$
logCmax	2.2018	2.2891	$\begin{bmatrix} 0.135903 & 0.175565 \\ 0.175565 & 0.283044 \end{bmatrix}$	$\begin{bmatrix} 0.152862 & 0.187008 \\ 0.187008 & 0.284740 \end{bmatrix}$
logAUC	5.5115	5.6456		

Table 13 Results of test comparison

		p
Bivariate test with rule theta no corr	P(Cp>3.49)	0.0005
Univariate *AUC*	P(C1>1.74)	0.0140
Univariate *Cmax*	P(C2>1.74)	0.0070
Bivariate test with rule theta with corr	P(Cp>2.31)	0.023
Assuming independence Intersection of 2 univariate	1-[P(C1<1.74)* P(C2<1.74]	0. 0209

5.8. Conclusion

In this article the univariate definition of population bioequivalence was extended to account for simultaneous multivariate testing of bioequivalence. A new multivariate criterion that is an extension to the FDA approved univariate criterion was developed. The statistical properties of this criterion in testing for bioequivalence were investigated using a simulation study. It was presented that this newly developed and proposed MV criterion could be a reliable aggregate measure for bioequivalence with good size and power properties. This study presented the importance for accounting for the correlation in defining the acceptable bioequivalence region. Further work is required to provide a guideline on how to use the correlation in defining the acceptable PBE regions.

The problem of asymmetry could result in conflicting determinations of BE depending on the drug used as a reference. Solutions to this problem need to be investigated. This article did not study the multivariate testing using more than two outcomes. This could be a subject for a separate study. There are computational difficulties when attempting to use more complex covariance structures to account for differences in variability between treatment drugs and to account for different covariance structures other than compound symmetry and autoregressive(1) when using multivariate mixed effect models. The effect of these limitations warrants further investigation. Finally, the patterns and effects of the missing observations in these studies on the multivariate evaluation of population bioequivalence were not evaluated.

The methodology developed in this article does not utilize the complete blood concentration data collected in bioequivalence studies to generate the blood concentration profiles. This method only uses the point estimates of the PKs estimated from the blood concentration profiles

and assumes they are measured without any error. This assumption is not reliable, so other analysis models should be considered to account for the variability of these profiles.

6 Testing Population Bioequivalence Using Non Linear Mixed Effects Models

6.1. Introduction

Bioequivalence testing is an important part of the drug approval process. Companies that make generic drugs are required, by the Food and Drug Administration (FDA), to show enough evidence that the generic drug is bioequivalent to the originally approved and patented drug. It is also required from the original manufacturers when they introduce a new formulation of their previously approved drugs. It is assumed that any drugs with the same chemical composition or active ingredient will have the same biological effect if they have similar absorption rates, blood/serum concentration levels and similar excretion rates. With this assumption the approval of generic drugs or new formulations of previously approved drugs does not require the extensive testing of toxicity and efficacy that is required for the approval of new drugs.

The FDA publishes and updates, daily, the "*Approved Drug Products with Therapeutic Equivalence Evaluations*", which is commonly known as the Orange Book. The designation of 'therapeutic equivalence' indicates that the generic formulation is (among other things) bioequivalent to the original (or innovator) formulation and signifies the FDA's expectation that the formulations are likely to have equivalent clinical effect and no difference in their potential for adverse effects. The main criterion for the inclusion of any product is that the product is the subject of an application with an effective approval that has not been withdrawn for safety or efficacy reasons. In addition, the Orange Book contains therapeutic equivalence evaluations for approved multisource prescription drug products (FDA, 1998).

Pharmaceutical equivalence means that two drugs contain the same amounts of the active pharmaceutical ingredients in the same dosage and route of administration. Therapeutic equivalence requires, in addition to the pharmaceutical equivalence, the two drugs to be bioequivalent. Bioequivalence is defined by FDA in section 505(j)(8)(B) of the Federal food, Drug, and Cosmetic Act as: "the rate and extent of absorption of the drug do not show a significant difference from the rate and extent of absorption of the listed drug when administered at the same molar dose of the therapeutic ingredient under similar experimental conditions in either a single dose or multiple doses" or "the extent of absorption of the drug does not show a significant difference from the extent of absorption of the listed drug when administered at the same molar dose of the therapeutic ingredient under similar experimental conditions in either a single dose or multiple doses and the difference from the listed drug in the rate of absorption of the drug is intentional, is reflected in its proposed labeling, is not essential to the attainment of effective body drug concentrations on chronic use, and is considered medically insignificant for the drug."(Congress, 2008)

Bioequivalence studies attempt to gain insight on formulation "switchability" (*i.e.*, the ability to substitute one formulation or another without concern of the potential for reduced effectiveness or increased probability of adverse effects). A key assumption is that switchability may be inferred from plasma concentration vs. time data and metrics reflecting the rate and extent of drug absorption. The area under the plasma concentration vs. time curve (*AUC*) is commonly employed as the metric describing the extent of drug absorption, while the maximal concentration observed following drug administration (*Cmax*) is the metric recommended by the FDA to evaluate the rate of drug absorption (FDA, 1992).

In the July 1992 guidance on *Statistical Procedures for Bioequivalence Studies Using a Standard Two-Treatment Crossover Design*, it was recommended that a standard in vivo BE study design be based on the administration of either single or multiple doses of the T and R products to healthy subjects on separate occasions, with random assignment to the two possible sequences of drug product administration (FDA, 1992).

In addition to the constants of absorption and elimination, the maximum concentration of the drug *Cmax*, and the total amount of the drug available in the compartment of effect *AUC* are usually used to compare the bioavailability of drugs, and hence to compare their effectiveness.

6.2. Estimation of PK metrics

It is widely known that drugs or other chemicals are absorbed, metabolized, and eliminated from the body according to specific mechanisms that are peculiar to each drug or groups of drugs and its physio-chemical metabolic pathways, and to the anatomical compartments it is distributed through. These pharmacological compartments are instantaneously well mixed and kinetically homogeneous. These specific mechanisms result in specific pharmacological compartmental model. These compartmental models are postulations of how the pharmacological system is believed to function. They are composed of finite number of components, like one-, two- or higher order- compartmental models. These compartments are specifically connected with each other, and each of them has specific input and output routes. The absorption of each drug into each of these compartments follows specific models like zero- or first-order models. Similarly the excretion of any drug from any of these compartments has its own models and constants.

6.2.1 Non-model based estimation

The metrics used to evaluate the bioequivalence of drugs are usually estimated from the blood-concentration-time profiles using non-model or non-compartmental methods. These methods do not assume any knowledge of the pharmaceutical models the drug could follow. They use algebraic equations to describe the blood-concentration profiles. For example, the *AUC* resultant from a single dose of a drug formulation is commonly assessed with the linear trapezoidal method (Berger RL, 1996, Gibaldi, 1982). The area of the trapezoids connecting the consecutive time points and their measured blood concentrations are calculated and summed to produce the *AUC*. The trapezoidal formula used for *AUC* is an approximation of the total area under the curve. The further the distance between the time points when the blood concentrations are measured, the larger the inaccuracy of the calculated *AUC*. Depending on the original profile of blood concentration curve, this could be an underestimation in some cases and an overestimation in other cases.

$C_{\max}$ is measured as the highest observed concentration. Although this measure rarely coincides with the true $C_{\max}$ (the estimate is biased downward), this measure is widely used in bioequivalence determinations. It is not unusual for plasma concentration profiles that reach a peak then the concentration drops, only for the concentration to peak again. The second peak may be higher or lower than the first peak. In these situations, the $C_{\max}$ is usually estimated as the concentration of the highest peak in profile. However, the first peak may be used as the estimate of $C_{\max}$ when used as a measure of absorption. The $T_{\max}$ is defined as the time when $C_{\max}$ is observed, and similar to $C_{\max}$, it is rarely accurate.

The rate of absorption could be measured in two ways. The first method is based on the linear fit of the first few points (at least three points) from beginning of the concentration profile

to the first peak. This absorption constant, usually noted as k_0, is calculated as the slope of that linear fit. The number of points chosen for this fit is based on the R-squares of the fits. Other methods uses nonlinear models to estimate the absorption rate constant denoted as k_a. These estimates of the bioequivalence parameters are non-model based calculations and they cannot account for the uncertainty in measuring the drug concentration. Alternatively, these parameters could also be estimated by fitting mechanistically meaningful non-linear models.

These non-model based estimates assume that the drug concentration measurements are made without errors, hence they ignore the uncertainty of these measurements.

6.2.2 Model-based estimation

Pharmacokinetic curves of approved drugs are always studied extensively during and after the approval of the drug. In the early phases of drug development these models might not be known. However, after the approval of the drug, and before the development of new formulations or generic drugs all model characteristics would be studied from which they could be well specified and all the characteristics of the plasma-concentration curves could be determined and the compartmental models could be defined. These compartmental models can handle non-linearities in the functions of blood concentration with time. They are often used in the experimental designs and to estimate the dosing regiments for phase I clinical trials. Acquiring this knowledge and the availability of advance analytic software makes it logical to use compartmental or model-based methods for estimating the metrics used in evaluating bioequivalence.

Non-linear models have been used in pharmacology to study the pharmacokinetics of drugs for a long time. These models could be based on the theoretical compartmental models describing the underlying mechanism that produces the data. As a consequence, the non-linear

model parameters have a more physical interpretation (Adams, 2002). However, these models are not generally used in drug testing, except for very limited tasks, like estimating the constants of absorption and elimination. Even in situations where non-linear models are used to estimate the other PK metrics such as AUC and $C_{\max}$ only point estimates are used in the bioequivalence testing. The uncertainties in the estimation are ignored (FDA 1992-2001, Chow SC, Liu JP 2000).

We demonstrate in this article that the two steps, namely the estimation of the pharmacokinetic measures from the concentration time curves and the subsequent multivariate bioequivalence testing with respect to these measures could be combined into one complete analysis. The proposed method is easy to implement in part due to the emergence of non-linear mixed model methods (Davidian 2003, Galecki 2004) and the subsequent software developments (Wolfinger 1999, SAS 2008, R-manual 1996, Pinheiro and Bates 2000).

Consider the one-compartment pharmacological model that determines the drug concentration in the plasma or blood at any time point according to this function:

$$C = \frac{k_a k_e D}{Cl\left(k_a - k_e\right)}\left[e^{-k_e t} - e^{-k_a t}\right], \qquad (52)$$

where C is the plasma concentration, D is the dose, Cl is the clearance, t is the time of the measurement, k_a is the constant of absorption, and k_e is the constant of elimination. These are known as the primary parameters of the blood concentration function. Note that the clearance rate of the drug is $Cl = k_e V$, where V is the volume of the active compartment. The area under the curve AUC, could be estimated by integrating the plasma concentration function with respect to time of the concentration function in (52). That yields a closed form to estimate the AUC as the ratio of the dose (D) to the clearance (Cl). If one is interested in AUC alone,

the function in (52) could be re-parameterized in terms of AUC by substituting AUC for that ratio. That is, the model could be rewritten

$$C = \frac{AUC * k_a k_e}{(k_a - k_e)}\left[e^{-k_e t} - e^{-k_a t}\right]. \qquad (53)$$

Similarly $C_{\max}$ could be calculated by differentiating (52) with respect to t and equating it to zero then solving for $T_{\max}$ and $C_{\max}$. This yields $C_{\max} = \frac{k_a k_e D}{Cl(k_a - k_e)}\left[e^{-k_e T_{\max}} - e^{-k_a T_{\max}}\right]$, where $T_{\max}$ is calculated as $\frac{\ln(k_e) - \ln(k_a)}{(k_e - k_a)}$. The calculations of the PK parameters that are usually used as metrics in bioequivalence studies are presented here. However, all other PK parameters could be estimated as functions of the primary parameters of the concentration function in (52) and its derivatives. The first order absorption k_a which is the rate of absorption, and the rate of drug elimination, k_e, are both estimated directly from these models.

Theory and algorithms for fitting nonlinear models (NLM) have been extensively developed and have been in use for decades (Wolfinger 1999, Davidian 2003, Galecki 2004, Adams 2002). However, applications of these in the bioequivalence literature seem limited. When the blood concentration data is collected a smoothed profile is drawn to represent the estimated function. Many PK measures are actually measured from these non linear fits, which are usually done on individual subject basis. However these nonlinear models are not used in statistical testing of bioequivalence. Only few attempts to fit nonlinear models and use them in testing hypotheses in drug studies could be found in the literature.

Pinheiro and Bates (1995) applied a nonlinear mixed effect model on data collected to study the drug theophylline. The serum concentrations of the drug were measured in 12 subjects

over a 25-hour period after oral administration. They applied the nonlinear modeling on a one compartment pharmacological model. They used this methodology for studying the pharmacodynamics of theophylline. Pinheiro and Bates (2000) also demonstrated that non-linear models, unlike linear models, provide more reliable predictions for responses outside the observed range of the data.

Panhardt (2007) recently used simulation studies on the theophylline data to show that nonlinear mixed effect models (NLMEM) could be used in testing pharmacological interactions and bioequivalence. They considered a cross-over PK study of a drug that follows a first order model. They implemented the same model used by Pinheiro and Bates (1995) to compare between two drugs using simulated data based on the theophylline data. They used this simulation to test for average bioequivalence. They studied only one outcome. They did not apply their methods on any multivariate cases.

The literature is really scarce in examples of using the nonlinear models in drug bioequivalence testing. Even finding a data set that records all the plasma concentrations and time for two groups of drugs is very hard.

Multivariate testing of multiple outcomes was never tested using the nonlinear models. Multivariate bioequivalence testing using these models was also never done.

6.3. Multivariate Analysis of PBE:

To establish bioequivalence in PK the FDA recommends testing the *AUC* and *Cmax*. It has been demonstrated that these two PK metrics are highly correlated. However, most of the tests of bioequivalence ignore this correlation and test the two variables separately, often without multiple comparisons adjustment. Further, even when multivariate methods are used they are

essentially an intersection test of the two individual univariate tests. That is, these tests do not account for the correlations.

The acceptance region of bioequivalence in these tests are rectangular (in the case of bivariate) or hyper-rectangular (in the case of multivariate). Munk, and Pflujer (1999) argue that the convex alternatives of this rectangular region (e.g., ellipsoid) is more appropriate in the case of multivariate testing if one wishes to appropriately account for the correlated nature of these measure. They also show that the $1-\alpha$ confidence rules for convex alternatives are actually $\alpha/2$ level tests—with applications to the multivariate assessment of bioequivalence. They suggest a Hotellings' T^2 statistic for multivariate version of the average bioequivalence testing.

We developed a multivariate test for population bioequivalence (in the first paper in Ch5), that is based on comparing the upper limit of a $100(1-\alpha)$ confidence interval of the PBE multivariate criterion C_p to a fixed bound θ_p that is defined by the regulators (FDA 1999).

6.4. Statistical Method:

Consider a bioequivalence study comparing a new test drug to a reference drug. In such studies, which are usually designed as cross-over studies, each subject receives both treatments, and he/she might receive each treatment multiple times. So these correlated repeated measures need to be accounted for when estimating the fixed effects. Suppose, the blood-concentration by time profile of the reference drug could be represented by a known non-linear function *f* linking concentrations to sampling times of all the subjects with subject specific PK parameters, such as

absorption (k_a), elimination (k_e)rate constants, and clearance half-life (Cl). An example of this function is the one compartment model function in (52).

Suppose several subjects are observed over a time interval at different occasions (of periods) on different treatments. At time point, t_{ijpk}, let C_{ijpk} represent the blood-concentration of the k^{th} treatment given to the i^{th} subject at the j^{th} time point, in the p^{th} period. Here i = 1, 2, …, n, k = T or R, p= 1, 2, …, P, and j = 1, 2, …, t_{ipk}, for, n subjects, 2 treatments, P periods and t_{ipk} time points. Assume that the sampling times are fixed and identical, for each treatment, period, and for all subjects, as often is the case in cross-over trials. Then for all i, j, p, and k, the time t_{ijpk} could be simplified to t_j. Let $\boldsymbol{\lambda}_{ipk}$ be the vector of the PK parameters of the subject i for treatment k in period p. Then the nonlinear model for the concentration profile is,

$$C_{ijpk} = f\left(t_j, \boldsymbol{\lambda}_{ipk}\right) + \varepsilon_{ijpk}, \tag{54}$$

where ε_{ijpk} is the measurement error. It is also assumed that ε_{ijpk} are independent of $\boldsymbol{\lambda}_{ipk}$, and they are normally distributed with mean zero and variance σ^2_{ijpk}.

Assume that the parameters $\boldsymbol{\lambda}_{ipk}$ are random vectors that could be decomposed for each period and treatment as

$$\boldsymbol{\lambda}_{ipk} = \boldsymbol{\mu}_{\cdot} + \boldsymbol{\beta}_k + \boldsymbol{\gamma}_p + \mathbf{u}_{ik} \tag{55}$$

where $\boldsymbol{\mu}_{\cdot}$ is the overall mean, $\boldsymbol{\beta}_k$ is the fixed effect of the treatment, $\boldsymbol{\gamma}_p$ is the fixed effect of the period, and $\mathbf{u}_{ik}$ is the random effect of subject i for treatment k, it is also assumed that $\mathbf{u}_{ik}$ is distributed as a multivariate normal with mean zero vector and a variance covariance matrix

$\boldsymbol{\Psi}_k$. To ensure that the estimates are always positive, $\boldsymbol{\lambda}_{ipk}$ elements are the natural logarithms of the original PK parameters in the function *f*. The elements of $\boldsymbol{\lambda}_{ipk}$ are $\log\left(Cl_{ipk}\right)$, $\log\left(k_{a_{ipk}}\right)$, and $\log\left(k_{e_{ipk}}\right)$.

The mle's of the original PK parameters k_a , k_e, and Cl could be estimated using this nonlinear mixed effects model. Since $C_{\max}$ and AUC or other metrics are functions of these PK parameters, then the mle's of these metrics for each treatment group could be estimated as functions of the mle's of the PK. The asymptotic approximation of the variance covariance matrices of these metrics could be estimated using the Taylor series expansion theorem. Where the second partial derivatives are derived and the mle's are estimated. This method is known as the delta method. When closed form derivatives are available, the estimation of the information matrix is easy. Otherwise other iterative methods are utilized.

Using these estimates of the means and variance covariance matrices of the test and the reference drugs, the multivariate PBE criterion (C_p) that was proposed in (the first paper in Ch5) would be estimated. Then the 90% confidence interval could be constructed using the parametric bootstrap method as suggested by the FDA and as shown in (our first paper). Two thousand samples are randomly generated from a multivariate distribution with means and variances equal to the mle's obtained from the NLMEM. The upper limit of the resultant confidence interval is compared to the predefined limit of bioequivalence θ described earlier in the first paper in Ch5. Bioequivalence is concluded if the upper limit of the 90% confidence interval is smaller than the predefined limit θ.

6.5. Examples:

6.5.1 Cross over design

Twenty eight subjects were recruited in the study, and were given 3 doses of each of the two drugs, with a washout period between each of two sessions. The order in which the two drugs were given was selected at random using a block randomization method to ensure the balance within subjects. Although there are 20 possible ways to assign three A's and three B's, the random process did not cover all possibilities.

At the beginning of each session, baseline data were collected 30 seconds prior to the administration of the drug. The pharmacological baseline measurement is supposed to be zero if the wash out period is long enough. After the drug was administered, the specified pharmacological measurement (level of specific active material in the blood) was obtained over several minutes at very short intervals.

Each subject was administered one of the two drugs in all six sessions, however due to mechanical failure, or lower than qualified measures, some of the sessions were completely missing all blood measures. The data collected from the subjects are presented in (APPENDIX E). Only the first two periods were used in this analysis due to computational issues in proc nlmixed in SAS9.2

The goal was to determine if the two drugs are bioequivalent. The NLMEM presented above in equation (55) is a suitable model for this data because the administration was repeated in nature on subjects, and due to the missing data in one of the periods for few patients. However due to computational limitations the random effects were at the subject level only, and they were the same for both treatments.

The logarithmic transformations of $C_{\max}$ and AUC and their variance covariance matrices were estimated for each of the treatment by fitting the model in equation (55) . These mle estimates are displayed in **Error! Reference source not found.**. The univariate and multivariate criteria for PBE were calculated according to (43) and (51). Then using the bootstrap method, the upper limit of the 90% confidence interval of these criteria were calculated. The 95th percentile of the MV criterion was 2.33274, and the univariate 95th percentile for Cmax was 0.88843, while the 95th percentile for the AUC was 1.39511.

The multivariate rule θ was calculated accounting for and ignoring the correlations between the variables in the test ($\hat{\rho}_T = 0.6$) and reference ($\hat{\rho}_R = 0.7$) groups. The rule theta accounting for these correlations was 2.67, while the upper bound of bioequivalence ignoring the correlation was 3.49.

The p-values for testing for bioequivalence were calculated as the probability of rejecting bioequivalence among the bootstrap samples. This probability was calculated as the proportion of bootstrap samples that had C_p greater than the rule θ.

The upper limits of the confidence intervals are all smaller than the predetermined bounds. It is concluded that these two drugs (A and B) are bioequivalent. This conclusion did not differ by the method used to test in this example.

Table 14 shows that the probabilities of rejecting bioequivalence. The difference in the p-values between the test that used the intersection of the two univariate tests and the MV test that accounts for correlation in the rule theta was small. On the other hand, the rectangular test differed with the MV test that ignored the correlation in the rule theta by 10-fold difference.

Table 14 Mean and covariance estimates of PM data using multivariate mixed model

		$\hat{\mu}_A$	$\hat{\mu}_B$	$\hat{\Sigma}_A =$	$\hat{\Sigma}_B$
Cross over example	logCmax	1.9738	1.9904	$\begin{bmatrix} 0.803000 & 0.427108 \\ 0.427108 & 0.56892 \end{bmatrix}$	$\begin{bmatrix} 0.69168 & 0.365640 \\ 0.36564 & 0.419672 \end{bmatrix}$
	logAUC	6.6503	6.4239		
Parallel example	logCmax	0.1639	1.0525	$\begin{bmatrix} 0.915564 & 0.81042 \\ 0.81042 & 2.39928 \end{bmatrix}$	$\begin{bmatrix} 0.292656 & 0.209664 \\ 0.209664 & 0.770328 \end{bmatrix}$
	logAUC	2.259	2.3062		

Table 15 result of PBE testing of PM data

		p
Bivariate test with rule theta no corr	P(Cp>3.49)	0.0035
Univariate AUC	P(C1>1.74)	0.0310
Univariate Cmax	P(C1>1.74)	0.0045
Bivariate test with rule theta with corr	P(Cp>2.67)	0.0245
Assuming independence Intersection of 2 univariate	1-P(C1<1.74)* P(C2<1.74	0.0354

6.5.2 Parallel design

We used the data from a study performed by Clayton and Leslie (1981) to study the bioequivalence of an enteric-coated erythromycin base (test drug) to the previously available reference formulation of erythromycin stearate. Twenty subject received the reference drug, and after a washout period each subject received the test drug. After administering the reference or test drug, venous blood samples were collected at 0.0, 0.5, 1.0, 2.0, 4.0, 6.0, and 8.0 hours. The data is presented in the dissertation APPENDIX F.

This study was described as a parallel design by (Chinchilli and Elswick 1997) because the treatments were not randomly assigned to each period. However, we have to account for the correlated nature of the repeated measures within each subject. The lack of the randomization, makes the period effect redundant to the treatment effect, so only the treatment fixed effect should be included in the model. The same nonlinear mixed effects model (NLMEM) used in the cross over design above (without the period effect) was used to model the one-compartment pharmacological model that fits this data.

The mle estimates of means and covariance were obtained by fitting this models using proc nlmixed in SAS 9.2. These estimates are presented in Table 14.

$C_{\max}$ and AUC were found to be moderately highly correlated, as expected, both in the test group $\hat{\rho}_T$ was 0.55, and in the reference group where $\hat{\rho}_R$ was 0.44. The upper limit of bioequivalence, θ was defined accordingly as 3.48 not accounting for the correlation and as 2.87 when accounting for the correlations. The multivariate criterion was calculated from these estimates and found to be 7.14. The univariate criterion of PBE for each of AUC and $C_{\max}$ were found 4.83 and 2.12 respectively. The upper limits of the 90% confidence intervals constructed

by the bootstrap for the MV criterion and the two univariate criteria, these were found as 7.14, 4.83 and 2.12.

Qualitatively the conclusion in this example did not differ according to which method was used. All the tests rejected bioequivalence by a very high p-value. Table 16 shows that the differences in the values of p-values are so subtle. This might be a reflection to the huge differences between the estimates of the means of the reference and they are so big that with any test bioequivalence is rejected easily.

Table 16 result of PBE testing of Clayton data using NLMEM

		p
Bivariate test with rule theta no corr	P(Cp>3.49)	0.9730
Univariate AUC	P(C1>1.74)	0.9635
Univariate Cmax	P(C1>1.74)	0.6740
Bivariate test with rule theta with corr	P(Cp>2. 55)	0.9910
Assuming independence Intersection of 2 univariate	1-[P(C1<1.74)* P(C2<1.74)]	0.9881

6.6. Conclusion

We have discussed in this article the methods and processes of estimating the pharmacokinetic metrics used in testing bioequivalence. Although the nonlinear models have seen a considerable amount of applications and software development , its use in the field of bioequivalence testing is in its infancy. We demonstrated how to apply the nonlinear models in estimating and testing for bioequivalence. We presented the importance of multivariate testing

for bioequivalence, and we implemented the use of the nonlinear mixed effects models to account for correlated data in the bioequivalence studies repeated measure.

We faced several computational limitations in the implementation of these mixed models. These are important issues that need to be investigated and studied. As an example of these limitations, we were unable to model more than one random effects in the nonlinear models. Another example is the convergence issues with higher number of periods in the cross-over designs. That forced us to limit the analysis to two periods only. Increasing the number of variables fixed or random effects greatly reduces the speed of the computations, and many time the estimations fail due to the large number of parameters to estimate. The starting values that are required to get good estimates are many times hard to determine, which results in local convergences, that are far from the real estimates.

We did not study the effect of choosing various models of the variance covariance structures for the repeated measures. This might affect the estimation of the fixed effects of interest. This could be the subject of a future study.

7 Summary and Recommendations for Future Work

In this study we extended the definition of the population bioequivalence into the multivariate dimension. We derived a criterion that is an actual multivariate extension of the univariate PBE approved by the FDA. Using a simulation study we have shown that this multivariate criterion has satisfactory properties (size and power) in testing for population bioequivalence.

The simulation study has also shown that this criterion suffers from two problems. The first is being asymmetric with respect to the Reference and the Test, where the exchanging the reference for the test metrics will result in un-identical results. This problem makes it possible to consider drug A bioequivalent to the reference drug B, however drug B is not bioequivalent to drug A.

This problem is not in the multivariate test only. It could be displayed in the univariate tests. It is due to the scaling of this criterion by the reference variance, or variance covariance matrix in the multivariate case. One may argue that in bioequivalence studies, the reference and test formulation cannot be treated equally, and that the test should be proved to be equivalent to the reference, and not the other way around. In other words the FDA puts extra burden on the generic drug producers to prove bioequivalence of their products to the novel drug.

For patients, bioequivalent drugs are drugs that could be used interchangeably to treatment certain ailments or to alleviate specific symptoms. From this respect, bioequivalence should not be different between two drugs depending on which was considered the reference

drug. Actually by definition the bioequivalence studies are surrogates for clinical trials in evaluating the therapeutic equivalence of the generic drugs to the novel drug.

The asymmetric property might be especially inconvenient if the equivalence of two generic drugs, or two new drugs, is to be evaluated. The KLD method might be a good solution for this issue. Testing the merits of this method in studying multivariate population bioequivalence might convince the FDA to change its stance of rejecting it.

.

Another problem is due to the aggregate nature of this criterion which is a sum of two entities. The upper limit of this criterion could be reached by increasing either one of the entities while keeping the other as small as possible, or by increasing both entities simultaneously. The impact of this is that multivariate bioequivalence could be met even if it is violated at one of the univariate tests when the other metric is close to zero. This is a limitation of all aggregate tools. That is why the FDA requires that to show univariate population bioequivalence, univariate average bioequivalence should be presented. Similarly we might need to show multivariate BE after showing univariate population BE. This is an important research question that needs to be investigated to try to find measures or methods that have the hierarchical nature of average bioequivalence and population bioequivalence.

The distance approach in evaluating bioequivalence is very intuitive. That's why it is easy to extend from the average bioequivalence into the population bioequivalence by adding the distance between the variances. However, when we looked into the multivariate case of PBE, another parameter appears, that is the correlation. It is hard to understand the correlations between the reference and between the test drug metrics as part of the distance between the two distributions of the reference and the test. That's why it is hard to select a meaningful difference in the correlations that could affect the bioequivalence. We tested the acceptance regions with

and without correlations, but our selection of the correlation to be equal to the estimated correlation of the data was arbitrary. Future research might be able to create some guidelines on how to define the acceptable differences in correlations.

We compared between the three PBE criteria. We examined Dragalin's criterion of PBE that was based on the KLD. This criterion was not accepted by the FDA although it was never tested before. The reason for the rejection is related to put the burden on the producer of the test drug to prove that it is equivalent to the reference drug. Although this is reasonable in most of the cases, there are cases where more than one generic drugs could be compared, and such cases require that both drugs be proven equivalent to each other regardless of which one is considered the reference. More testing of Dragalins criterion as a true distance measure between any two distributions might prove useful in other applications of equivalence testing where no preference is given to any distribution.

We also discussed and demonstrated how nonlinear models are good tools in evaluating bioequivalence by showing how to use it in population bioequivalence studies. We demonstrated the use of nonlinear mixed effects models in bioequivalence studies. We have shown that the development in statistical software had made it much easier to use these complex models in estimating the metrics and their variances. However there are still many limitations. For example, there is no direct way to implement a specific covariance pattern for repeated correlated measure. There is no specific way to have more than one random effect, although in some situations it is required to have random effects for different units of analysis.

We demonstrated the use of the NLMEM on the one-compartment pharmacological model. There are much more complex models in pharmacological studies. However, it is well known that the nonlinear models are very mechanistic. We could use these models to model

almost all intrinsic mechanisms for any pharmaceutical processes, like absorption, elimination, effective range, onset of activity, lag of action, etc. Nonlinear models could be used in modeling more complex models than the one-compartment model that we demonstrated. It is important to test bioequivalence applications using more complex models.

Bioequivalence problems could be handled by a frequentist or a Bayesian approach, but frequentist solutions are more common in the literature and much more often used in practice, despite the nice properties of many Bayesian solutions. This may be due in part to the weight of regulatory agencies in bioequivalence testing, the most significant application area. There might be a regulatory bias towards the frequentist approach, but it is also likely that the level of difficulty and the weak understanding of the Bayesian approach might reducing the use of such methods in this field. It would be interesting to approach PBE and it multivariate testing from a Bayesian approach.

The studies in this dissertation were limited to equal size samples of reference and test drugs. This might not be realistic in a lot of BE studies. The simulation studies also did not have any missing values. The effect of different sample sizes, missing values and dependence between the treatments drugs should be tested in future studies.

.

Bibliography

Bibliography

1. Adams E., Coomans D.,Smeeyers-Verbeke J. Massart DL (**2002**). Non-linear mixed effects models for the evaluation of dissolution profiles. International Journal of Pharmaceutics 240 37–53
2. Berger RL, Hsu JC. (**1996**). Bioequivalence trials, intersection union tests and equivalence confidence sets. Statistical Science **1996**; **11**:**283**–319.
3. Brown LD, Casella G, Hwang JT. (**1995**). Optimal confidence sets, bioequivalence, and the Limacon of Pascal. Journal of the American Statistical Association **1995**; **90**:**880**–**889.**
4. Brown LD, Hwang JT, Munk A. (**1997**). An unbiased test for the bioequivalence problem. Annals of Statistics **1997**; **25**:**2345**–2367.
5. Chervoneva I, Hyslop T, Hauck WW (**2007**). A multivariate test for population bioequivalence, Statistics in Medicine, 26(6):1208-23, 2007
6. Chinchilli VM, Elswick Jr RK (**1997**)**.** The multivariate assessment of bioequivalence. Journal of Biopharmaceutical Statistics **1997**; **7**:**113**–123.
7. Chow, S.C., and Liu, J.P. (**1995**). Statistical Design and Analysis in Pharmaceutical Science. *Marcel Dekker, New York*.
8. Chow SC, Liu JP. (**2000**) Design and Analysis of Bioavailability and Bioequivalence Studies - Revised and Expanded, Second Edition, Marcel Dekker, Inc., New York, New York. **14-19**

9. Chow SC, Liu JP. (**2009**) Design and Analysis of Bioavailability and Bioequivalence Studies, Third Edition, Chapman & Hall, New York, New York.
10. Chow, S**.C.**, and Liu, J**.P.** (**1995**) "Current issues in bioequivalence trials", Drug Information Journal, Vol. **29**(**3**), **795**-**804.**
11. Clayton D, Leslie A (**1981**): The bioavailability of erythromycin stearate vesus enteric-coated erythromycin base when taken immediately before and after food. Journal of International Medical Research **9:4770-4777**.
12. Dahman, Bassam (2009) "Nonlinear Models in Multivariate Population Bioequivalence testing". Ph.D. Dissertation. Virginia Commonwealth University.
13. Davidian, M. and Giltinan, D.M. (**2003**), "Nonlinear Models for Repeated Measurement Data: An Overview and Update," *JABES* 8, 387-419
14. Dragalin V, Fedorov V, Patterson S, Jones B. (**2003**) Kullback–Leibler divergence for evaluating bioequivalence. Statistics in Medicine **2003**; **22**:**913**–930.
15. Efron B. (**1982**) The Jackknife, the Bootstrap and other Resampling Plans. SIAM: Philadelphia, **1982.**
16. Efron, B., Tibshirani, R**.J.** (**1993**) An Introduction to the Bootstrap. Chapman and Hall, NY.
17. Erickson, J., Seaman, J., Stamey, J. (**2005**). Bayesian Methods for Bioequivalence Studies. *Eli Lilly and Company, Indianapolis, IN, Department of Statistical Science, Baylor University, Waco, TX, Stephen F. Austin University, Nacogdoches, TX*
18. FDA (**1997**) - Center for Drug Evaluation and Research, In vivo bioequivalence studies based on population and individual bioequivalence approaches - Draft guidance [may be accessed from: http://www.fda.gov/cder/guidance/index.htm], (**1997**)**.**

19. FDA (**1992).** Guidance on Statistical Procedures for Bioequivalence Studies Using a Standard Two-Treatment Crossover Design. Office of Generic Drugs, Center for Drug Evaluation and Research, Food and Drug Administration, Rockville, Maryland.
20. FDA (**1999).** In Vivo Bioequivalence Studies Based on Population and Individual Bioequivalence Approaches. Food and Drug Administration, Rockville, Maryland, August, **1999.**
21. FDA (**2000).** Guidance for Industry: Bioavailability and Bioequivalence Studies for Orally Administered Drug Products — General Considerations. Center for Drug Evaluation and Research, Food and Drug Administration, Rockville, Maryland, **2000**
22. FDA (**2001).** Guidance for Industry: Statistical Approaches to Establishing Bioequivalence. 1Food and Drug Administration, Rockville, Maryland, January, **2001.**
23. FDA (**2000**) CITE: 21CFR320, Code of Federal Regulations, Title **21**, Volume **5**, Parts **300** to **499**, [accessed on 2009 from http://www1.va.gov/oro/apps/compendium/Files/21CFR320.htm]
24. FDA (**1998**) Center for Drug Evaluation and Research, Approved drug products with therapeutic equivalence evaluations, [accessed on October 2009 from: http://www.accessdata.fda.gov/scripts/cder/ob/default.cfm]
25. FDA (**1992**) - Office of Generic Drugs - Division of Bioequivalence, Guidance on statistical procedures for bioequivalence studies using a standard two-treatment crossover design [accessed from: http://www.fda.gov/cder/guidance/index.htm]
26. Galecki AT, Wolfinger RD, Linares OA, Smith MJ, Halter JB. (**2004**). "*Ordinary differential equation PK/PD models using the SAS macro NLINMIX*". J Biopharm Stat. 2004 May;14(2):483-503..

27. Gibaldi, M. and Perrier, D.(**1982**), *Pharmacokinetics*, *2nd ed*., Marcel Dekker, New York NY (1982)

28. Hauck, W**.W.**, and S. Anderson, (**1992)**, "Types of Bioequivalence and Related Statistical Considerations," Int. J. Clin. Pharmacol. Therap., **30**:**181-7.**

29. http://www1.va.gov/oro/apps/compendium/Files/21CFR320.html (last checked on **5**/**3**/**2009**)]

30. Lindstrom MJ, Bates DM. (**1990**) Nonlinear mixed-effects models for repeated measures data. Biometrics; 46:673–687.

31. Mehrotra N, M Gupta M, A Kovar A, Meibohm B. (**2007**). The role of pharmacokinetics and pharmacodynamics in phosphodiesterase-5 inhibitor therapy. *International Journal of Impotence Research* **19,** 253–264

32. Munk A, Pflujer R**. (1999)** $1-\alpha$ equivariant confidence rules for convex alternatives are $\alpha/2$-level tests—with applications to the multivariate assessment of bioequivalence. *Journal of the American Statistical Association* 1999; **94**: 1311–1319.

33. Pinheiro, J.C. and Bates, D.M. (**1995**), "Approximations to the Log-likelihood Function in the Nonlinear Mixed-effects Model," *Journal of Computational and Graphical Statistics*, 4, 12-35.

34. Pinheiro, J.C., Bates, D.M. (**2000**). 'Mixed-Effects Models in S and S-plus, Statistics and Computing'. New York: Springer–Verlag.

35. R-manual (**1996**). "*Documentation for package 'nlme' version 3.1-96*". *[http://stat.ethz.ch/R-manual/R-devel/library/nlme/html/00Index.html]*.

36. SAS Institute Inc. (**2008**). *SAS/IML® 9.2User's Guide*. Cary, NC: SAS Institute Inc.

37. Schuirmann DJ (**1987**). A comparison of the two one-sided tests procedure and the power approach for assessing the equivalence of average bioavailability. Journal of Pharmacokinetics and Biopharmaceutics **1987**; **15**:**657**–680.

38. Sheiner LB. (**1992**) Bioequivalence revisited. Statistics in Medicine **1992**; 11:1777–1788.

39. Schall R, Luus HG. (**1993**). On population and individual bioequivalence. Statistics in Medicine; **12**:**1109**–1124.

40. Tamhane AC, Logan BR. (**2004**). A superiority-equivalence approach to one-sided tests on multiple endpoints in clinical trials. Biometrika **2004**; **91**:**715**–727.

41. US Congress, Federal food, Drug, and Cosmetic Act (revision *February **2008**) [accessed from* http://www.fda.gov/RegulatoryInformation/Legislation/FederalFoodDrugandCosmeticActFDCAct/FDCActChapterVDrugsandDevices/ucm108125.htm on 10/2009]

42. Wang W, Hwang JT, DasGupta A. (**1999**). Statistical tests for multivariate bioequivalence. Biometrika **1999**; **86**:**395**–402.

43. Wolfinger, RD. (**1999**). "*Fitting Nonlinear Mixed Models with the New NLMIXED Procedure*" SUGI 24, SAS Institute Inc., Cary, NC. [http://*support.sas.com/rnd/app/papers/nlmixedsugi.pdf*, accessed on Octoner 2009].

44. Zarifa NM-D, Patterson SD, Boyle D, Hyneck H. (**2000**). Case studies, practical issues, and observations on Population and Individual Bioequivalence. Statistics in Medicine **2000**; **19**:**2811**–2820.

45. Zarifa NM-D, Patterson SD. (**2001**). Population and individual bioequivalence: lessons from real data and simulation studies. Journal of Clinical Pharmacology **2001**; **41**:**811**–**822**.

APPENDIX A: Population Bio-Equivalence a distance measure

The population bioequivalence was proposed by [Ref] as a measure of the distance between the distribution of the T treatment and the reference R treatment. This distance is scaled by the variance of the reference treatment.

$$\frac{E\left[\left(Y_T - Y_R\right)^2\right] - E\left[\left(Y_R - Y_{R'}\right)^2\right]}{E\left[\left(Y_R - Y_{R'}\right)^2\right]\Big/2} \tag{56}$$

The first term of denominator, and second term of the numerator:

$$\begin{aligned} E\left[\left(Y_R - Y_{R'}\right)^2\right] &= EY_R^2 - 2E\left[Y_R Y_{R'}\right] + EY_{R'}^2 \\ &= \mu_R^2 + \sigma_R^2 + \mu_R^2 + \sigma_R^2 - 2\mu_R^2 = 2\sigma_R^2 \end{aligned} \tag{57}$$

Note the division by two in the denominator allows for scaling by the reference variance σ_R^2.

The first term of the numerator:

$$\begin{aligned} E\left[\left(Y_T - Y_R\right)^2\right] &= EY_T^2 - 2E\left[Y_T Y_R\right] + EY_R^2 \\ &= \mu_T^2 + \sigma_T^2 + \mu_R^2 + \sigma_R^2 - 2E\left[Y_T Y_R\right] \\ &= \mu_T^2 + \sigma_T^2 + \mu_R^2 + \sigma_R^2 - 2\left(\rho\sigma_T\sigma_R + \mu_T\mu_R\right) \end{aligned} \tag{58}$$

Under the assumption of independence between the reference R and Test distributions ($\rho = 0$)

$$E\left[\left(Y_T - Y_R\right)^2\right] = \mu_T^2 + \sigma_T^2 + \mu_R^2 + \sigma_R^2 \tag{59}$$

Then the numerator in (56) will reduce to

$$\begin{aligned} E\left[(Y_T - Y_R)^2\right] - E\left[(Y_R - Y_{R'})^2\right] &= \mu_T^2 + \sigma_T^2 + \mu_R^2 + \sigma_R^2 - 2\mu_T\mu_R - 2\sigma_R^2 \\ &= \left(\mu_T^2 - 2\mu_T\mu_R + \mu_R^2\right) + \sigma_T^2 - \sigma_R^2 \\ &= (\mu_T - \mu_R)^2 + \sigma_T^2 - \sigma_R^2 \end{aligned} \tag{60}$$

Then the univariate population bioequivalence is expressed as:

$$\frac{(\mu_T - \mu_R)^2 + \sigma_T^2 - \sigma_R^2}{\sigma_R^2} \tag{61}$$

This was used by Zarifa in testing bioequivalence, by comparing this criterion to a predetermined upper limit θ

$$\begin{aligned} &\frac{(\mu_T - \mu_R)^2 + \sigma_T^2 - \sigma_R^2}{\sigma_R^2} \le \theta \\ &(\mu_T - \mu_R)^2 + \sigma_T^2 - \sigma_R^2 \le \theta\sigma_R^2 \\ &(\mu_T - \mu_R)^2 + \sigma_T^2 - \sigma_R^2 - \theta\sigma_R^2 \le 0 \\ &(\mu_T - \mu_R)^2 + \sigma_T^2 - \sigma_R^2(1+\theta) \le 0 \end{aligned} \tag{62}$$

The upper limit of the 90% confidence interval for the linearized criterion $(\mu_T - \mu_R)^2 + \sigma_T^2 - \sigma_R^2(1+\theta)$ should be negative to accept bioequivalence.

APPENDIX B: Derivation of MV PBE

Derivation of MV PBE

The univariate population bioequivalence (PBE) criterion is expressed as

$$\frac{E\left[\left(Y_T - Y_R\right)^2\right] - E\left[\left(Y_R - Y_{R'}\right)^2\right]}{E\left[\left(Y_R - Y_{R'}\right)^2\right]/2} \tag{63}$$

This expression could be rewritten as

$$\frac{\left(\mu_T - \mu_R\right)^2 + \sigma_T^2 - \sigma_R^2}{\sigma_R^2} \tag{64}$$

To develop the equivalent multivariant equivalent for the criterion in 1.1, let $\mathbf{Y}_T$ and $\mathbf{Y}_R$ be *p*-variate random variables denoting the test and reference PK metrics. Let's assume that $\mathbf{Y}_T$ is distributed as a *p*-variate normal with a mean $\boldsymbol{\mu}_T$ and a variance covariance matrix $\boldsymbol{\Sigma}_T$. And let $\mathbf{Y}_R$ and $\mathbf{Y}_{R'}$ be two realizations of the *p*-variate normally distributed random variable with mean $\boldsymbol{\mu}_R$ and a variance covariance matrix $\boldsymbol{\Sigma}_R$.

Then the multivariate equivalent of the denominator in (63) is

$$E\left[\left(\mathbf{Y}_R - \mathbf{Y}_{R'}\right)\left(\mathbf{Y}_R - \mathbf{Y}_{R'}\right)'\right]/2 = \boldsymbol{\Sigma}_R \tag{65}$$

Then the multivariate criterion that is equivalent to (63) could be written as:

$$C_p = E\left[\left(\mathbf{Y}_T - \mathbf{Y}_R\right)' \boldsymbol{\Sigma}_R^{-1}\left(\mathbf{Y}_T - \mathbf{Y}_R\right)\right] - E\left[\left(\mathbf{Y}_R - \mathbf{Y}_{R'}\right)' \boldsymbol{\Sigma}_R^{-1}\left(\mathbf{Y}_R - \mathbf{Y}_{R'}\right)\right] \tag{66}$$

Let $\mathbf{Z} = \boldsymbol{\Sigma}_R^{-1/2}(\mathbf{Y}_T - \mathbf{Y}_R)$; and let $\mathbf{K} = \boldsymbol{\Sigma}_R^{-1/2}(\mathbf{Y}_R - \mathbf{Y}_{R'})$, then by substitution with $\mathbf{Z}$ and $\mathbf{K}$ in (66), the multivariate criterion could be expressed as

$$C_p = E[\mathbf{Z}'\mathbf{Z}] - E[\mathbf{K}'\mathbf{K}] \tag{67}$$

Note that

$$\begin{aligned} E[\mathbf{Z}'\mathbf{Z}] &= E\left[\sum_{i=1}^{p} z_i^2\right] = \sum_{i=1}^{p} E\left[z_i^2\right] \\ &= \sum_{i=1}^{p}\left(\sigma_{z_i}^2 + E[z_i]^2\right) \\ &= \sum_{i=1}^{p}\left(\sigma_{z_i}^2 + E[z_i]^2\right) \\ &= \sum_{i=1}^{p}\sigma_{z_i}^2 + \sum_{i=1}^{p} E[z_i]^2 \\ &= trace(\boldsymbol{\Sigma}_{\mathbf{Z}}) + E[\mathbf{Z}]' E[\mathbf{Z}] \end{aligned} \tag{68}$$

Then the multivariate population bioequivalence criterion could be written as;

$$C_p = trace(\boldsymbol{\Sigma}_{\mathbf{Z}}) + E[\mathbf{Z}]' E[\mathbf{Z}] - trace(\boldsymbol{\Sigma}_{\mathbf{K}}) - E[\mathbf{K}]' E[\mathbf{K}] \tag{69}$$

The expectation of $\mathbf{Z}$ is equal to

$$\begin{aligned} E[\mathbf{Z}] &= E\left[\boldsymbol{\Sigma}_R^{-1/2}(\mathbf{Y}_T - \mathbf{Y}_R)\right] \\ &= \boldsymbol{\Sigma}_R^{-1/2} E\left[(\mathbf{Y}_T - \mathbf{Y}_R)\right] \\ &= \boldsymbol{\Sigma}_R^{-1/2}(\boldsymbol{\mu}_T - \boldsymbol{\mu}_R) \end{aligned} \tag{70}$$

So the second term of the right hand side of (69) is

$$\begin{aligned} E[\mathbf{Z}]' E[\mathbf{Z}] &= (\boldsymbol{\mu}_T - \boldsymbol{\mu}_R)' \boldsymbol{\Sigma}_R^{-1/2} \boldsymbol{\Sigma}_R^{-1/2} (\boldsymbol{\mu}_T - \boldsymbol{\mu}_R) \\ &= (\boldsymbol{\mu}_T - \boldsymbol{\mu}_R)' \boldsymbol{\Sigma}_R^{-1} (\boldsymbol{\mu}_T - \boldsymbol{\mu}_R) \end{aligned} \tag{71}$$

The last term of (69) is zero because, the expectation of $\mathbf{K}$ is

$$
\begin{aligned}
E[\mathbf{K}] &= E\left[\boldsymbol{\Sigma}_R^{-1/2}\left(\mathbf{Y}_R - \mathbf{Y}_{R'}\right)\right] \\
&= \boldsymbol{\Sigma}_R^{-1/2} E\left[\left(\mathbf{Y}_R - \mathbf{Y}_{R'}\right)\right] \\
&= \boldsymbol{\Sigma}_R^{-1/2}\left(\boldsymbol{\mu}_R - \boldsymbol{\mu}_R\right) \\
&= \underline{\mathbf{0}}
\end{aligned}
\tag{72}
$$

Note that the covariance of $\mathbf{Z}$ is:

$$
\begin{aligned}
\boldsymbol{\Sigma}_{\mathbf{Z}} &= Cov(\mathbf{Z}) \\
&= Cov\left(\boldsymbol{\Sigma}_R^{-1/2}\left(\mathbf{Y}_T - \mathbf{Y}_R\right)\right) \\
&= \boldsymbol{\Sigma}_R^{-1/2} Cov\left(\mathbf{Y}_T - \mathbf{Y}_R\right)\boldsymbol{\Sigma}_R^{-1/2} \\
&= \boldsymbol{\Sigma}_R^{-1/2}\left(Cov\left(\mathbf{Y}_T\right) + Cov\left(\mathbf{Y}_R\right)\right)\boldsymbol{\Sigma}_R^{-1/2} \\
&= \boldsymbol{\Sigma}_R^{-1/2}\left(\boldsymbol{\Sigma}_T + \boldsymbol{\Sigma}_R\right)\boldsymbol{\Sigma}_R^{-1/2} \\
&= \boldsymbol{\Sigma}_R^{-1/2}\boldsymbol{\Sigma}_T\boldsymbol{\Sigma}_R^{-1/2} + \boldsymbol{\Sigma}_R^{-1/2}\boldsymbol{\Sigma}_R\boldsymbol{\Sigma}_R^{-1/2}
\end{aligned}
\tag{73}
$$

Since the last term of the above equation reduces to a *pxp* identity matrix, then the variance-covariance matrix of $\mathbf{Z}$ reduces to $\boldsymbol{\Sigma}_R^{-1/2}\boldsymbol{\Sigma}_T\boldsymbol{\Sigma}_R^{-1/2} + \mathbf{I}$

And the covariance of $\mathbf{K}$

$$
\begin{aligned}
\boldsymbol{\Sigma}_K &= Cov(K) \\
&= Cov\left(\boldsymbol{\Sigma}_R^{-1/2}\left(\mathbf{Y}_R - \mathbf{Y}_{R'}\right)\right) \\
&= \boldsymbol{\Sigma}_R^{-1/2} Cov\left(\mathbf{Y}_R - \mathbf{Y}_{R'}\right)\boldsymbol{\Sigma}_R^{-1/2} \\
&= \boldsymbol{\Sigma}_R^{-1/2}\left(Cov\left(\mathbf{Y}_R\right) + Cov\left(\mathbf{Y}_{R'}\right)\right)\boldsymbol{\Sigma}_R^{-1/2} \\
&= \boldsymbol{\Sigma}_R^{-1/2}\left(\boldsymbol{\Sigma}_R + \boldsymbol{\Sigma}_R\right)\boldsymbol{\Sigma}_R^{-1/2} \\
&= 2\mathbf{I}
\end{aligned}
\tag{74}
$$

Substituting into (69) and using the cyclical properties of the trace, the multivariate criterion could be expressed as

$$
\begin{aligned}
C_p &= trace(\boldsymbol{\Sigma}_Z) + E[\mathbf{Z}]' E[\mathbf{Z}] - trace(\boldsymbol{\Sigma}_K) - E[\mathbf{K}]' E[\mathbf{K}] \\
&= trace\left(\boldsymbol{\Sigma}_R^{-1/2}\boldsymbol{\Sigma}_T\boldsymbol{\Sigma}_R^{-1/2} + \mathbf{I}\right) + (\boldsymbol{\mu}_T - \boldsymbol{\mu}_R)' \boldsymbol{\Sigma}_R^{-1}(\boldsymbol{\mu}_T - \boldsymbol{\mu}_R) - trace(2\mathbf{I}) \\
&= trace\left(\boldsymbol{\Sigma}_T\boldsymbol{\Sigma}_R^{-1}\right) + (\boldsymbol{\mu}_T - \boldsymbol{\mu}_R)' \boldsymbol{\Sigma}_R^{-1}(\boldsymbol{\mu}_T - \boldsymbol{\mu}_R) - p
\end{aligned}
$$

So the Multivariate criterion is

$$
C_p = trace\left(\boldsymbol{\Sigma}_T\boldsymbol{\Sigma}_R^{-1}\right) + (\boldsymbol{\mu}_T - \boldsymbol{\mu}_R)' \boldsymbol{\Sigma}_R^{-1}(\boldsymbol{\mu}_T - \boldsymbol{\mu}_R) - p \qquad (75)
$$

.

APPENDIX C: Distribution of Multivariate PBE criterion

Using a simulation with sample sizes 25, 50, and 100 per treatment group, and with a difference between the treatment and reference groups μ_d, as: 0.7, 1, and 1.3. and differences between the variances of the treatment and reference groups $\sigma_T^2 - \sigma_R^2$ as: 0.0, 0.02, and 0.04. The multivariate criterion was calculated, and its histograms representing its distribution were plotted.

The following two figures show that the distribution of C_p get more toward the normal as the sample size increases when the variances the variances are between the two treatment groups are equal, however this trend is not seen when the variances are between the two treatment groups are not equal.

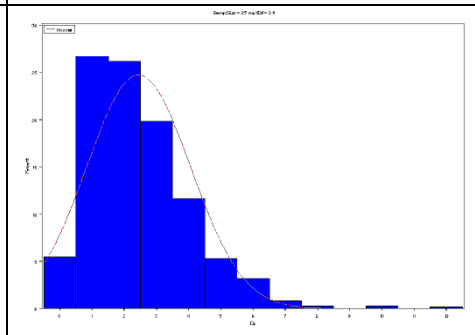
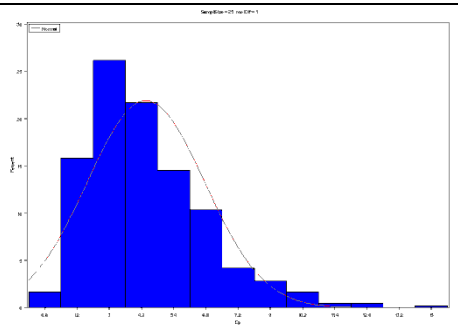
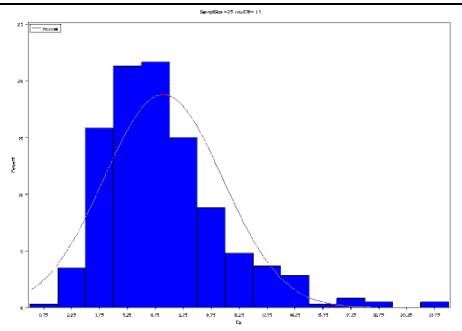
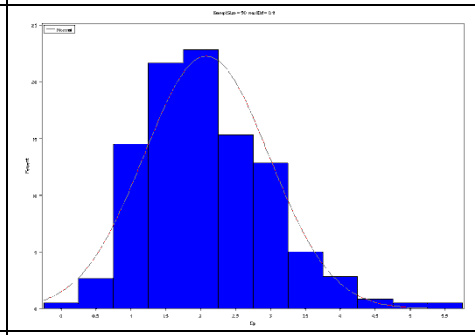
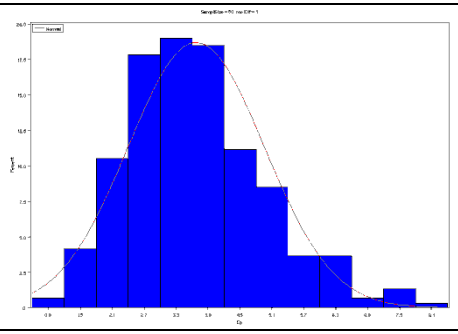
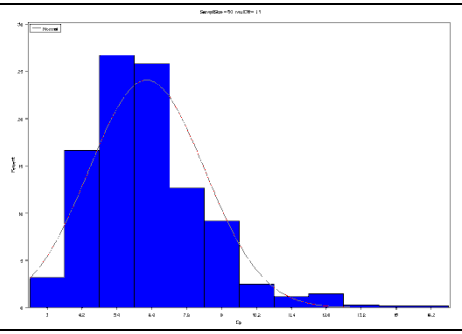
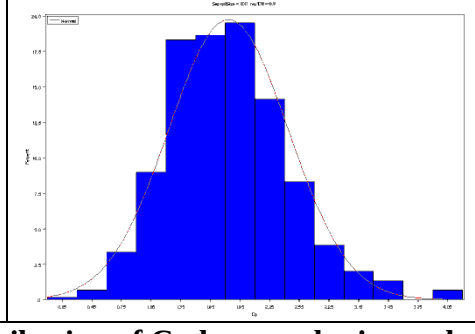
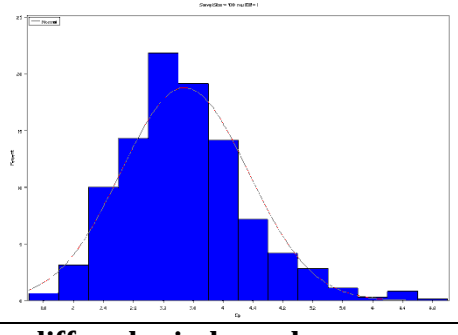
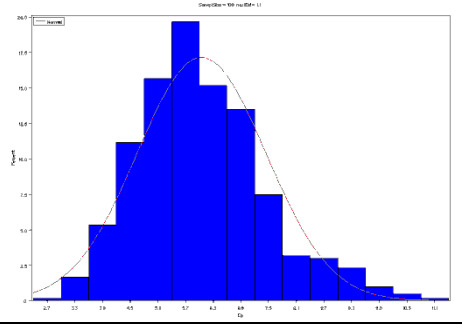
N \ μ_d	**.7**	**1**	**1.3**
25			
50			
100			

Figure 8 Distribution of Cp by sample size and mean diff under independence

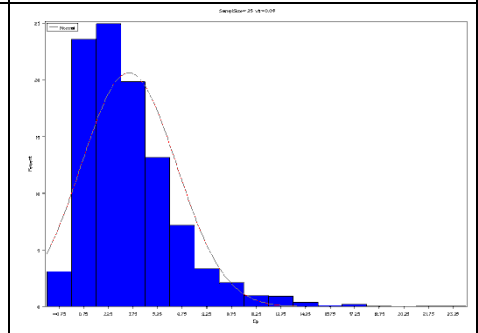
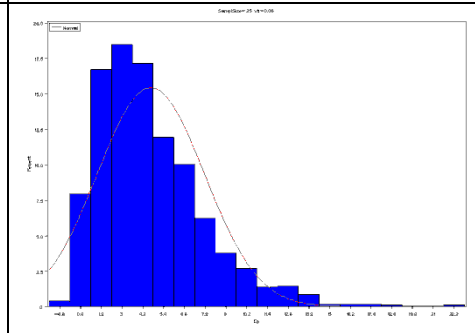
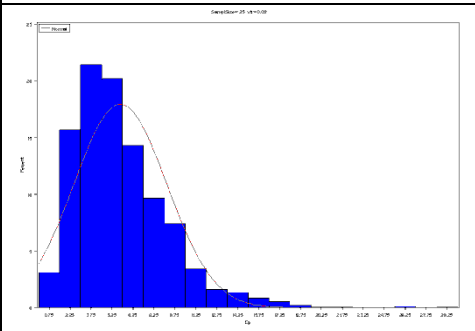
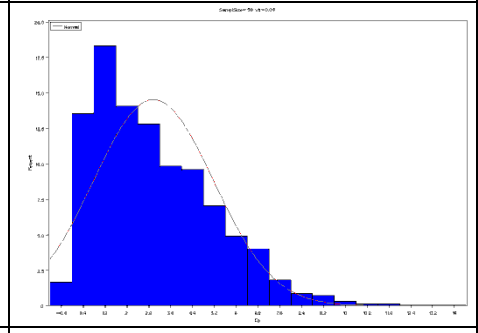
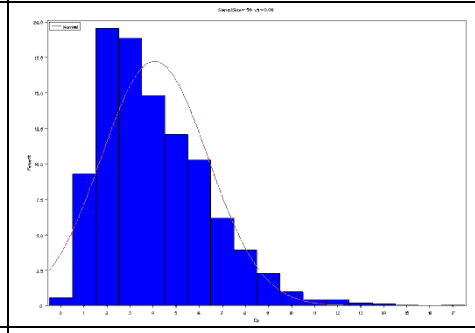
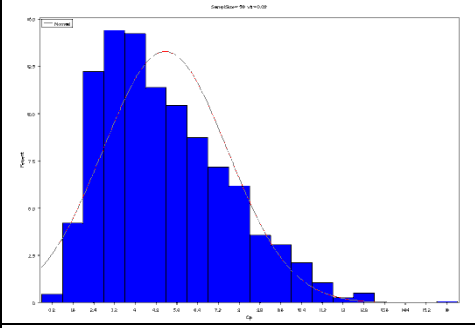
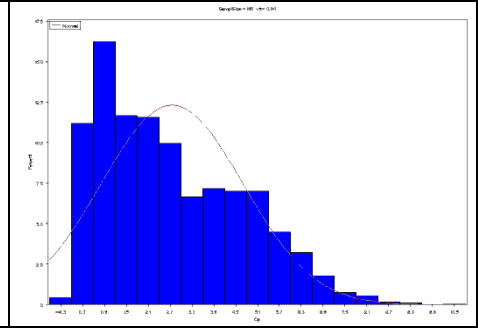
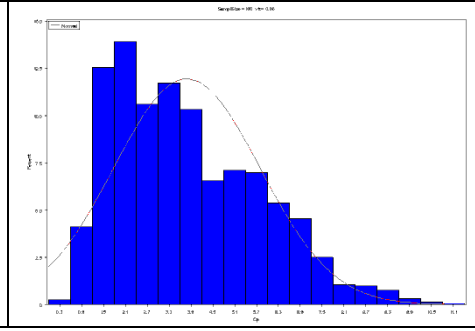
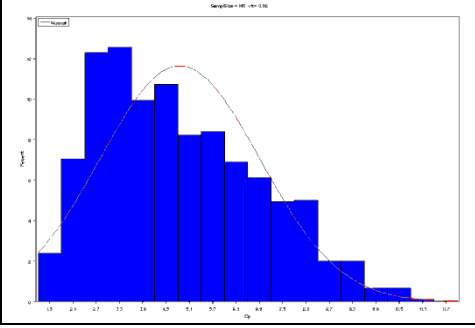
N \ $\sigma_T^2 - \sigma_R^2$	**0**	**0.02**	**0.04**
25			
50			
100			

Figure 9 Distribution of Cp by sample size and difference of variances under independence

APPENDIX D: Graphical presentation of PBE Acceptance regions

To graphically illustrate the ellipsoidal nature of the regions of equivalence developed by the test consider the following. Let $\mathbf{X}_T$ and $\mathbf{X}_R$ be vectors of random variables representing the metrics used in bioequivalence testing. Let $\mathbf{X}_T$ be distributed as multivariate normal with mean vector $\boldsymbol{\mu}_T$ and variance covariance matrix $\boldsymbol{\Sigma}_T$. Let $\mathbf{X}_R$ be distributed as multivariate normal with mean vector $\boldsymbol{\mu}_R$ and variance covariance matrix $\boldsymbol{\Sigma}_R$. Then bioequivalence acceptance region is defined as

$$C_p = trace\left(\boldsymbol{\Sigma}_T \boldsymbol{\Sigma}_R^{-1}\right) + \left(\boldsymbol{\mu}_T - \boldsymbol{\mu}_R\right)' \boldsymbol{\Sigma}_R^{-1} \left(\boldsymbol{\mu}_T - \boldsymbol{\mu}_R\right) - p \le \theta_p \qquad (76)$$

Note that this could be reordered as

$$\left(\boldsymbol{\mu}_T - \boldsymbol{\mu}_R\right)' \boldsymbol{\Sigma}_R^{-1} \left(\boldsymbol{\mu}_T - \boldsymbol{\mu}_R\right) \le \theta_p + p - trace\left(\boldsymbol{\Sigma}_T \boldsymbol{\Sigma}_R^{-1}\right) \qquad (77)$$

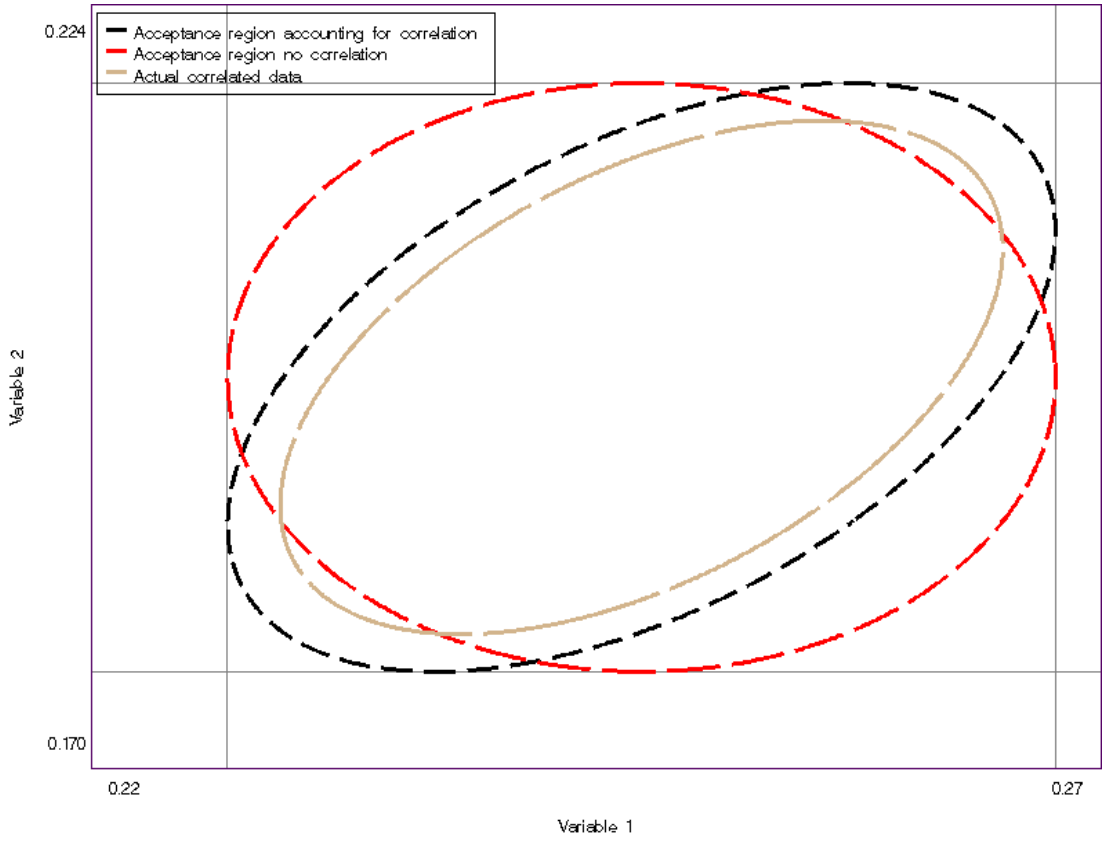
The right hand side of this inequality, which is a scalar, does not depend on the means. Therefore, the left hand side is a quadratic form in terms of the difference of the mean vectors and thus the inequality represents an ellipsoid whose shape is controlled by the variance covariance matrices (Johnson and Wichern.xxx). Similary we could construct an ellipsoid based on the estimates of the parameters on both LHS and RHS of the equation above such that $\hat{C}_p \le C_{p95}$. That is, the ellipsoid formed by the inequality,

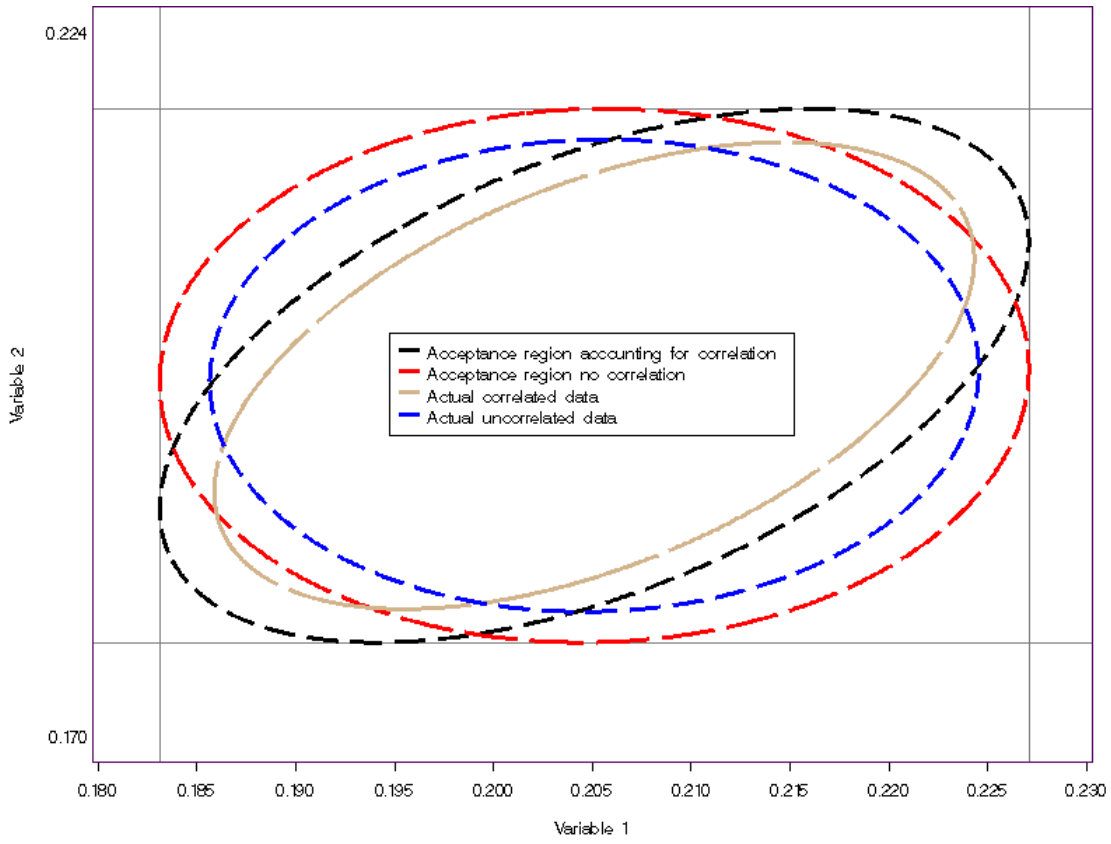
$$\left(\hat{\boldsymbol{\mu}}_T - \hat{\boldsymbol{\mu}}_R\right)' \hat{\boldsymbol{\Sigma}}_R^{-1} \left(\hat{\boldsymbol{\mu}}_T - \hat{\boldsymbol{\mu}}_R\right) \le C_{p95} + p - trace\left(\hat{\boldsymbol{\Sigma}}_T \hat{\boldsymbol{\Sigma}}_R^{-1}\right). \qquad (78)$$

Bioequivalence is then pictorially presented if the second ellipsoid (actual data) is totally contained within the first ellipsoid defined by the regulators' limits. In Figure 10 examples of these ellipsoids are presented. The values were chosen to demonstrate, accounting for the

correlation between the variables actually changes the acceptance region. This result could be vastly different depending on how the correlation is incorporated. For example, the area in the figure that is outside the red ellipse (that ignores correlation) is outside the bioequivalence acceptance region if we ignore correlation, however it will be within the bioequivalence region if we account for the correlation.

Figure 10 Acceptance bioequivalence regions





Create two data sets

Test and reference with the following variance and means:

sigmaT		sigmaR		mubarT	mubarR	Mubar d
0.0407217	-0.000756	0.0412375	-0.000168	1.1275743	1.0130981	0.1144762
-0.000756	0.0448253	-0.000168	0.0410481	1.1182117	0.9863747	0.131837

Theta for the bivariate is 3.5, and theta for the univariate is 1.75

The point estimates of the PBE criteria

Bivariate Cp_Data	Univariate C1_data	Univariate C2_data
0.8236143	0.3052789	0.5154513

The upper limit of the 95% CI for these found by 2000 bootstraps as

theta	Cp_95	C1_95	C2_95
3.489652	1.313103	0.604775	0.881751

In the same bootstrap, also calculated the upper limit of the quadratic form

quad_C	quad_theta
1.053592	3.661915

The ellipses of the quadratic term with the following limits were plotted:

$$(\hat{\mu}_T - \hat{\mu}_R)' \hat{\Sigma}_R^{-1} (\hat{\mu}_T - \hat{\mu}_R) \le \theta_p - trace \hat{\Sigma}_T \hat{\Sigma}_R^{-1} + p$$
$$(\hat{\mu}_T - \hat{\mu}_R)' \hat{\Sigma}_R^{-1} (\hat{\mu}_T - \hat{\mu}_R) \le \hat{C}_{p95} - trace \hat{\Sigma}_T \hat{\Sigma}_R^{-1} + p$$

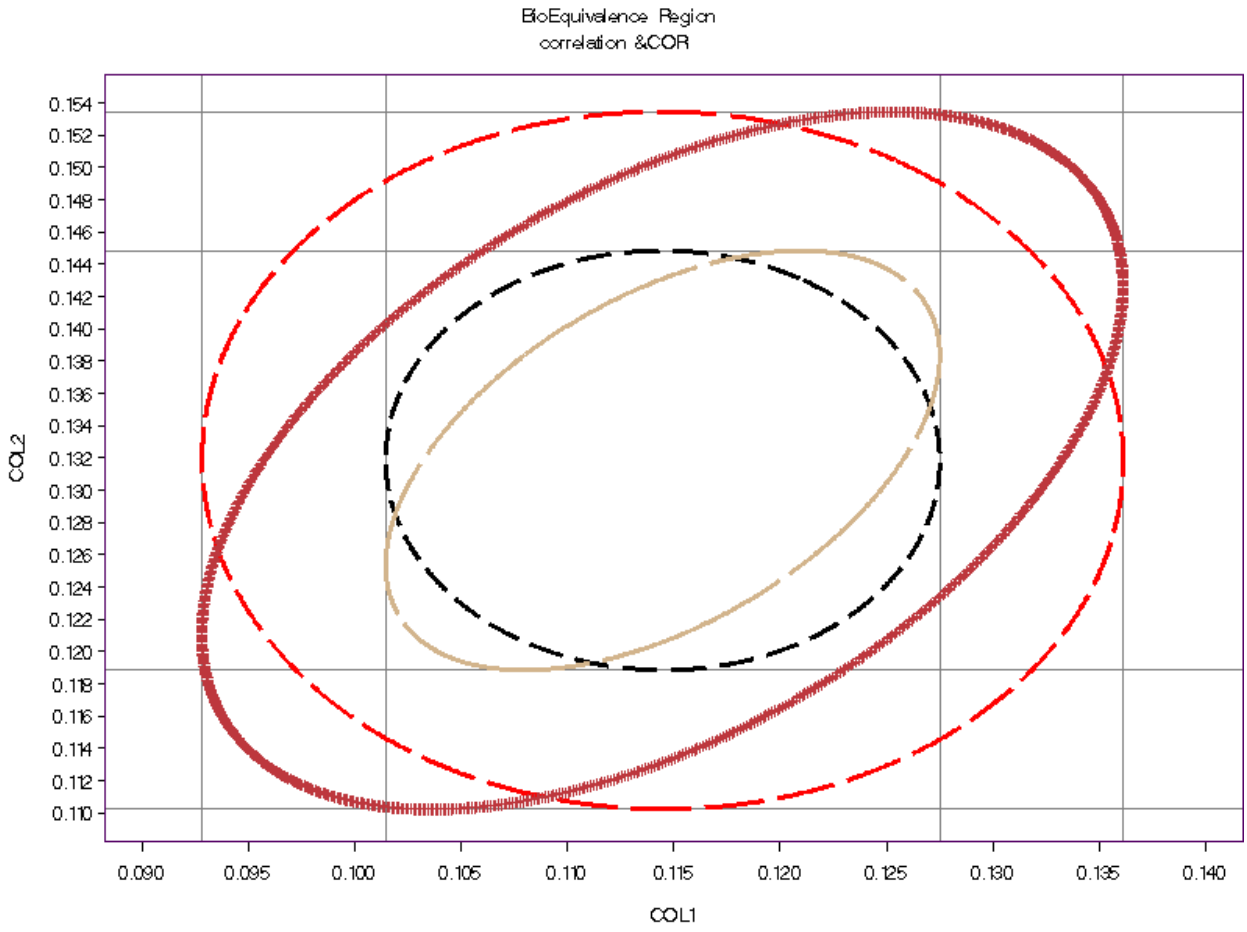
Then $\hat{\Sigma}_R$ & $\hat{\Sigma}_T$ where modified to have a correlation=0.5 and the same variances.

The quadratic forms were plotted again;

The plot shows that the although the angle changes from 0 to 45, the upper and lower limits do not change

The outer ellipses are from $(\hat{\mu}_T - \hat{\mu}_R)' \hat{\Sigma}_R^{-1} (\hat{\mu}_T - \hat{\mu}_R) \le \theta_p - trace \hat{\Sigma}_T \hat{\Sigma}_R^{-1} + p$

And the inner ellipses are from $(\hat{\mu}_T - \hat{\mu}_R)' \hat{\Sigma}_R^{-1} (\hat{\mu}_T - \hat{\mu}_R) \le \hat{C}_{p95} - trace \hat{\Sigma}_T \hat{\Sigma}_R^{-1} + p$



APPENDIX E: PM Data Blood Concentration Curves

PM Data Blood Concentration Curves

PM Data
All data from 1-7 seconds were zeros.

Subject	period	Treat	_0	_8	_9	_10	_11	_12	_13	_14	_15	_16	_17	_18	_19	_20	_22	_24	_26	_28	_30	_34	_38	_42	_46	_50	_54	_58	_62
7	1	A	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.1	4.1	3.9	5.3	4.7	4.9	5.2	4.8	5.2	4.4	4.2	3.6	3.2	3.8	3.1	3.1	0.0	2.6
8	1	B	0	0.0	0.0	0.0	4.0	7.2	10.3	11.0	12.9	11.5	11.8	13.3	12.1	10.6	12.7	11.0	9.6	10.3	8.7	9.3	9.5	6.5	6.5	7.4	5.7	5.5	4.6
8	2	A	0	0.0	0.0	0.0	2.9	6.3	8.7	9.2	9.7	8.8	9.3	10.1	8.7	8.3	7.3	7.5	6.7	4.2	6.4	6.3	6.2	5.1	7.0	4.4	4.1	3.1	3.6
8	3	B	0	0.0	0.0	0.0	2.5	4.3	6.4	6.5	7.1	7.7	7.9	8.0	8.6	8.0	7.6	6.8	7.6	7.0	6.1	6.6	6.6	5.6	5.5	5.3	4.8	4.7	4.4
8	4	A	0	0.0	0.0	0.0	5.4	6.1	6.7	7.4	8.0	7.6	7.4	7.6	7.8	6.2	6.3	6.4	6.5	6.2	5.9	5.9	5.9	4.7	4.7	4.6	4.7	4.5	4.2
8	5	B	0	0.0	0.0	0.0	3.8	6.6	7.2	6.8	8.4	10.0	9.2	8.0	7.8	8.3	11.2	8.2	7.3	6.9	7.3	6.4	6.1	6.1	6.0	5.8	5.0	4.2	4.5
8	6	A	0	0.0	0.0	0.0	2.9	4.5	4.9	8.7	7.1	7.5	7.9	7.4	6.7	8.3	6.6	6.6	6.0	5.7	5.8	5.4	5.6	4.7	5.2	4.7	3.7	3.0	0.0
9	1	B	0	0.0	0.0	0.0	0.0	4.8	7.1	9.9	12.2	13.0	12.9	13.9	13.7	13.3	13.9	13.3	12.2	12.5	12.0	10.9	9.4	9.6	9.7	8.7	8.3	7.4	6.4
9	2	A	0	0.0	0.0	0.0	0.0	2.7	4.2	5.7	7.1	7.7	8.1	6.6	7.7	8.9	9.2	9.2	9.7	10.1	10.0	10.0	6.7	6.7	6.6	5.1	7.0	4.6	3.7
9	3	B	0	0.0	0.0	0.0	0.0	0.0	4.0	7.4	7.5	10.4	10.1	9.6	9.2	9.8	9.1	8.4	8.5	8.1	9.1	7.7	7.4	8.3	5.7	5.0	4.5	4.3	4.2

Subject	period	Treat	_0	_8	_9	_10	_11	_12	_13	_14	_15	_16	_17	_18	_19	_20	_22	_24	_26	_28	_30	_34	_38	_42	_46	_50	_54	_58	_62
9	4	B	0	0.0	0.0	0.0	2.9	4.3	5.7	8.7	10.6	10.4	11.4	13.5	11.9	11.2	9.8	10.5	9.9	9.6	9.3	9.0	7.9	7.8	7.3	7.3	6.1	5.3	5.7
9	5	A	0	0.0	0.0	0.0	0.0	0.0	3.1	4.9	6.2	7.0	7.3	7.9	7.6	6.8	8.3	7.9	7.9	6.9	6.9	6.0	5.9	5.7	5.8	5.6	5.2	4.2	4.4
9	6	A	0	0.0	0.0	0.0	0.0	0.0	2.8	4.3	6.6	8.0	8.5	8.0	8.5	8.7	8.9	9.8	9.0	8.9	8.7	8.0	7.1	6.7	6.6	5.7	5.2	5.1	5.3
11	1	A	0	0.0	0.0	0.0	0.0	0.0	0.0	3.8	5.5	7.2	4.7	20.6	9.8	9.0	9.2	11.3	9.4	7.1	6.6	9.8	0.0	8.8	8.2	7.7	6.2	8.4	10.5
11	2	A	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	9.3	10.4	10.8	10.7	11.1	11.1	8.4	9.4	8.6	7.8	7.3	8.0	9.0	10.0	8.3	6.8	9.3	10.0	0.0
11	3	B	0	0.0	0.0	0.0	0.0	3.4	4.8	6.9	7.3	7.0	7.2	7.9	7.0	7.5	6.9	7.2	7.9	6.3	6.3	6.2	4.8	5.2	5.1	4.8	4.7	4.9	4.1
11	4	A	0	0.0	0.0	0.0	0.0	0.0	2.9	4.3	5.9	6.0	6.7	7.3	7.4	6.7	7.3	7.0	5.7	5.9	6.3	5.8	6.6	4.8	4.4	4.3	3.6	3.9	2.9
11	5	B	0	0.0	0.0	0.0	0.0	2.9	4.2	4.6	8.0	8.4	10.1	9.3	11.4	9.0	10.4	10.0	9.4	8.7	7.3	6.5	7.5	6.2	5.0	4.9	4.8	5.4	5.4
11	6	B	0	0.0	0.0	0.0	0.0	4.4	5.8	7.2	8.8	9.2	8.5	7.4	8.3	8.4	7.7	9.4	8.0	8.4	7.1	7.1	6.4	6.6	6.4	4.9	4.7	5.0	4.8
12	3	A	0	0.0	0.0	0.0	2.8	3.4	5.2	6.3	5.0	5.7	5.4	7.4	7.2	6.8	7.1	6.1	6.3	5.4	5.2	4.7	6.8	5.1	5.1	4.7	4.6	4.0	4.3
13	2	B	0	0.0	0.0	0.0	2.7	3.1	2.9	2.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
13	4	B	0	0.0	0.0	0.0	0.0	0.0	3.9	3.3	3.8	3.1	3.0	2.7	2.9	2.9	2.9	2.7	2.6	2.5	2.7	2.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0
13	5	A	0	0.	0.	0.0	0.0	0.0	0.0	0.0	2.6	3.9	3.3	2.7	2.6	2.9	3.2	2.8	2.7	2.7	4.6	0.0	0.0	3.8	2.5	0.0	0.0	0.0	0.0

Subject	period	Treat	_0	_8	_9	_10	_11	_12	_13	_14	_15	_16	_17	_18	_19	_20	_22	_24	_26	_28	_30	_34	_38	_42	_46	_50	_54	_58	_62
				0	0																								
13	6	A	0	0.0	0.0	0.0	0.0	0.0	3.4	4.0	3.4	2.7	2.8	2.9	2.9	3.8	3.4	2.9	2.7	2.8	2.7	2.7	0.0	0.0	0.0	0.0	2.8	0.0	2.8
14	1	A	0	0.0	0.0	0.0	5.1	10.6	7.8	11.3	18.0	17.3	19.6	16.0	14.8	14.6	18.3	13.9	15.3	14.7	11.2	9.2	9.1	8.7	7.9	7.4	5.6	6.8	6.3
14	2	B	0	0.0	0.0	3.2	7.0	10.7	16.1	18.7	19.8	19.6	21.7	19.0	18.3	17.4	17.1	19.9	16.8	13.7	14.0	12.1	11.0	11.4	9.7	7.4	7.4	7.3	8.2
14	3	B	0	0.0	0.0	0.0	2.9	8.2	11.4	14.4	16.1	16.5	17.3	18.4	17.7	16.4	17.1	16.3	16.1	16.1	13.6	13.2	13.5	11.9	10.4	8.6	8.8	9.0	7.5
14	4	A	0	0.0	0.0	0.0	0.0	4.0	7.9	12.4	15.1	18.3	18.1	17.5	17.4	18.4	17.7	16.4	19.7	14.3	14.2	14.9	12.2	12.0	11.9	9.6	9.2	9.2	6.4
14	5	A	0	0.0	0.0	0.0	2.7	5.0	8.3	11.7	12.4	13.1	13.7	13.8	15.6	15.1	16.7	14.4	12.9	11.5	11.2	11.0	10.3	9.7	10.8	7.8	7.6	6.7	6.2
14	6	B	0	0.0	0.0	0.0	0.0	0.0	4.6	9.6	10.2	15.3	15.8	17.2	18.0	16.5	16.5	15.2	15.4	11.6	18.4	13.8	13.7	12.7	10.0	12.0	10.4	9.4	8.0
17	1	B	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.4	2.8	4.4	3.9	3.9	4.3	3.7	3.6	3.4	3.1	2.9	2.9	0.0	0.0	0.0	0.0	0.0	0.0
17	2	A	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.6	5.1	4.9	4.7	4.4	4.1	5.3	3.9	0.0	3.0	3.5	2.8	0.0	3.5	0.0	0.0	0.0	0.0	0.0
17	3	B	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.3	3.7	3.8	3.2	3.6	4.1	4.0	3.7	3.8	0.0	2.7	0.0	0.0	0.0	0.0	0.0
17	4	B	0	0.0	0.0	0.0	0.0	0.0	3.3	6.2	11.6	12.1	16.8	13.8	14.8	18.3	12.5	12.6	12.3	11.1	10.6	10.6	9.7	9.0	7.0	7.2	6.3	6.8	5.8
17	5	A	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.6	3.9	10.1	13.8	11.8	16.3	11.4	14.5	13.1	11.7	11.0	9.0	7.5	5.9	6.5	6.5	4.2	5.6	5.2

Subject	period	Treat	_0	_8	_9	_10	_11	_12	_13	_14	_15	_16	_17	_18	_19	_20	_22	_24	_26	_28	_30	_34	_38	_42	_46	_50	_54	_58	_62
17	6	A	0	0.0	0.0	0.0	0.0	0.0	0.0	2.6	7.9	11.1	11.7	14.1	13.5	13.2	14.2	14.0	12.7	12.8	13.6	11.4	8.0	9.1	8.0	6.8	7.4	7.3	5.6
18	2	B	0	0.0	0.0	2.9	6.1	10.2	13.2	14.9	15.7	17.2	17.3	15.1	12.9	13.9	11.6	11.6	11.4	9.9	10.1	9.2	7.9	7.7	6.0	5.7	5.3	4.3	3.6
18	3	A	0	0.0	4.6	7.6	11.0	14.8	13.5	12.5	11.8	9.9	8.3	8.6	7.1	6.8	6.4	6.7	5.9	5.1	4.8	4.2	4.5	4.1	4.8	2.9	4.1	2.8	0.0
18	4	A	0	0.0	2.6	5.8	8.0	11.0	10.5	9.3	10.2	8.3	8.0	9.8	7.2	7.0	7.5	7.4	7.2	6.5	5.6	6.1	4.6	5.3	5.2	4.4	3.1	0.0	0.0
18	5	B	0	2.6	7.6	8.3	9.3	10.2	11.0	12.3	10.3	9.8	10.5	9.9	7.9	8.1	8.9	8.1	7.8	5.7	5.9	6.1	6.0	8.9	4.4	4.0	3.6	3.6	0.0
18	6	B	0	0.0	3.5	8.2	11.7	13.0	13.7	15.7	14.2	12.5	11.3	9.8	9.0	10.5	8.2	7.7	7.9	7.2	7.2	7.8	5.0	5.4	5.1	4.9	3.4	3.2	3.0
19	1	A	0	0.0	0.0	0.0	0.0	0.0	0.0	3.1	4.8	6.9	8.7	7.3	7.0	7.2	7.3	9.4	6.7	6.6	6.2	5.7	5.2	5.3	7.6	4.7	3.8	4.0	2.9
19	2	B	0	0.0	0.0	0.0	0.0	0.0	0.0	3.1	5.5	6.9	7.9	9.6	9.2	9.0	9.1	9.0	9.2	8.9	7.8	7.4	7.1	7.0	6.1	5.4	5.0	4.6	3.9
19	3	A	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.9	4.6	5.7	6.8	6.2	6.8	6.5	7.1	7.1	7.6	7.0	7.0	5.6	5.2	6.3	4.6	4.2	3.8	3.7
19	4	A	0	0.0	0.0	0.0	0.0	0.0	0.0	3.4	5.6	5.8	7.2	7.4	7.4	8.1	6.9	7.0	6.4	8.1	6.4	5.6	5.3	5.0	4.3	3.9	3.6	3.9	3.0
19	5	B	0	0.0	0.0	0.0	0.0	0.0	0.0	2.5	3.8	5.9	6.6	6.7	8.4	7.3	8.8	6.7	7.3	8.0	8.6	5.4	6.6	6.2	5.4	4.0	3.6	3.1	3.3
19	6	B	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	4.6	5.6	7.8	8.3	9.0	8.2	8.6	9.0	9.3	9.3	9.4	7.4	7.9	6.2	8.0	5.4	4.1	5.0	3.1
20	1	B	0	0.	0.	0.0	0.0	0.0	0.0	0.0	2.9	5.2	5.2	5.5	5.5	6.5	5.8	6.2	6.0	6.9	6.2	5.9	5.6	5.7	6.0	5.1	5.7	0.0	0.0

Subject	period	Treat	_0	_8	_9	_10	_11	_12	_13	_14	_15	_16	_17	_18	_19	_20	_22	_24	_26	_28	_30	_34	_38	_42	_46	_50	_54	_58	_62
				0	0																								
20	2	B	0	0.0	0.0	0.0	0.0	0.0	2.6	3.0	2.6	3.3	4.7	4.3	4.4	5.6	5.8	5.8	5.8	5.8	6.2	5.3	5.2	0.0	5.2	4.4	4.1	5.7	5.5
20	3	B	0	0.0	0.0	0.0	0.0	0.0	0.0	3.4	2.6	3.3	4.0	4.9	4.7	3.2	5.2	4.2	3.1	4.0	4.8	4.8	5.9	3.9	4.9	6.1	5.5	4.8	4.9
20	4	A	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.5	2.8	3.2	3.7	3.7	4.5	4.7	5.4	6.0	5.5	4.9	4.9	4.5	4.7	4.4	4.9	4.6	4.6
20	5	A	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.8	4.2	5.7	6.1	6.9	6.8	5.8	5.7	6.3	6.7	6.2	6.2	6.4	6.6	6.0	8.1	7.7	7.2	5.9
20	6	A	0	0.0	0.0	0.0	2.5	2.7	2.9	3.1	4.1	4.0	4.9	6.0	5.9	6.4	6.9	7.2	6.6	7.3	6.4	6.7	6.1	6.1	6.0	4.4	6.0	5.2	4.8
21	3	B	0	0.0	0.0	0.0	0.0	2.8	4.8	4.1	5.1	6.0	4.7	4.7	5.3	4.7	5.0	4.6	4.1	4.0	3.9	3.6	3.5	2.8	2.5	3.3	2.5	0.0	0.0
21	4	B	0	0.0	0.0	0.0	0.0	0.0	2.5	2.8	3.6	3.4	3.4	3.0	3.2	2.9	3.8	3.9	3.0	3.1	2.9	0.0	2.6	3.4	2.6	0.0	2.8	0.0	0.0
21	5	A	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.7	2.7	2.9	3.4	3.2	3.3	3.1	3.8	0.0	2.7	3.0	0.0	0.0	3.6	2.8	0.0	2.8
21	6	B	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.9	3.8	3.1	3.7	4.2	3.1	3.0	2.9	2.9	3.3	3.6	2.5	2.6	3.5	0.0	0.0	0.0	2.5	0.0
23	1	A	0	0.0	0.0	0.0	0.0	0.0	0.0	3.4	3.6	4.5	3.9	3.4	3.1	3.5	3.3	3.0	3.8	3.4	3.2	2.8	3.0	0.0	0.0	0.0	0.0	0.0	0.0
23	3	A	0	0.0	0.0	0.0	0.0	2.7	3.5	5.7	3.4	3.7	5.0	6.9	6.9	4.9	3.3	3.7	3.3	3.8	2.7	0.0	4.0	0.0	0.0	0.0	0.0	0.0	0.0
23	4	A	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.4	4.1	3.5	0.0	2.8	3.0	2.7	0.0	3.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Subject	period	Treat	_0	_8	_9	_10	_11	_12	_13	_14	_15	_16	_17	_18	_19	_20	_22	_24	_26	_28	_30	_34	_38	_42	_46	_50	_54	_58	_62
23	5	B	0	0.0	0.0	0.0	0.0	3.6	3.3	6.9	3.1	2.6	3.0	3.2	3.3	2.9	2.9	3.2	4.3	0.0	5.3	0.0	4.0	2.6	3.9	0.0	0.0	0.0	0.0
24	1	B	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.9	2.6	3.7	4.1	4.0	4.6	5.4	4.9	5.2	5.2	5.6	4.9	4.8	4.6	4.6	4.9	4.2	4.3	4.2
24	2	A	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.6	3.9	4.5	4.1	4.6	5.1	5.4	6.8	5.6	5.7	5.8	5.9	4.5	5.1	4.8	3.7	4.1
24	3	B	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.5	3.6	4.7	5.0	4.9	4.7	5.3	5.6	6.1	6.7	5.6	5.4	4.4	4.7	4.9	4.8	5.1
24	4	A	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	4.1	4.1	4.8	6.2	5.9	5.9	5.9	6.1	7.8	5.9	5.0	6.4	5.3	4.7	4.8	4.8	0.0
24	5	A	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.8	5.6	7.2	7.6	8.2	9.8	8.9	8.0	8.2	7.4	7.6	5.3	5.3	6.1	6.2	5.3	4.7
24	6	B	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.0	6.0	5.2	5.7	6.3	6.8	7.3	7.8	7.9	8.0	7.2	8.0	7.2	6.3	6.0	5.7	5.4	5.1
26	1	A	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.2	2.7	3.1	3.5	2.8	3.1	3.0	2.9	3.5	3.2	2.8	3.8	3.2	2.6	0.0	0.0	2.5	0.0	0.0
26	3	B	0	0.0	0.0	0.0	0.0	0.0	0.0	3.1	3.9	4.7	4.5	5.4	4.8	4.7	5.0	5.0	4.9	5.3	4.9	4.5	4.2	4.2	4.0	4.7	4.2	4.0	3.7
26	4	A	0	0.0	0.0	0.0	0.0	0.0	0.0	2.5	5.7	5.3	3.8	5.4	4.9	4.4	3.7	5.6	7.3	4.7	4.2	6.8	5.6	8.6	5.4	6.3	5.7	5.0	2.9
26	5	A	0	0.0	0.0	0.0	3.3	2.9	2.8	0.0	2.9	3.2	0.0	4.3	3.5	3.9	0.0	0.0	6.4	5.0	3.6	7.2	5.2	3.2	3.0	2.7	3.5	2.6	0.0
26	6	B	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.9	3.3	3.5	3.7	3.7	3.4	3.9	3.6	3.3	3.3	3.5	3.5	3.3	3.2	3.2	2.5	3.3
27	1	B	0	0.	0.	0.0	0.0	0.0	0.0	3.2	4.0	5.7	5.6	4.9	4.9	4.7	5.3	5.1	4.2	4.7	4.9	4.9	5.2	5.0	3.9	4.1	3.8	3.5	0.0

Subject	period	Treat	_0	_8	_9	_10	_11	_12	_13	_14	_15	_16	_17	_18	_19	_20	_22	_24	_26	_28	_30	_34	_38	_42	_46	_50	_54	_58	_62
				0	0																								
27	2	A	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.5	2.7	2.9	2.9	2.8	2.7	2.9	3.2	2.7	0.0	2.5	0.0	0.0	0.0	0.0	0.0	0.0
27	3	B	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.5	2.5	2.5	2.5	3.1	3.9	2.9	2.9	0.0	0.0	2.8	0.0	5.6	0.0	0.0
27	4	B	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.5	6.3	5.4	4.5	4.1	4.4	6.2	11.3	4.1	5.1	5.1	7.2	4.7	5.0	5.7	6.2	6.6	5.3	4.1
27	5	A	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.9	3.3	3.9	0.0	3.8	0.0	0.0	0.0	0.0	0.0	0.0
27	6	A	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.8	2.8	2.8	2.8	3.0	3.5	3.9	2.9	3.8	3.5	2.8	0.0	0.0	4.0	0.0	0.0	0.0
28	1	B	0	0.0	2.5	5.2	7.2	8.3	12.0	10.1	12.1	11.1	10.0	10.5	9.7	10.3	9.5	9.6	7.4	7.3	6.6	7.1	6.9	8.2	6.2	6.1	5.9	5.3	4.3
28	2	A	0	0.0	0.0	0.0	3.6	6.1	8.3	8.8	16.6	15.1	13.6	15.0	17.0	13.9	11.1	11.5	12.9	12.6	11.4	11.0	10.5	8.3	9.5	10.5	6.8	8.6	8.5
28	3	B	0	0.0	3.2	4.8	11.3	8.3	19.3	14.1	13.2	12.6	13.3	13.9	13.3	14.6	17.1	11.8	11.0	12.5	13.9	11.5	9.0	9.5	13.4	12.0	12.3	11.9	11.5
28	4	A	0	0.0	0.0	2.5	5.2	9.2	10.4	10.9	11.9	10.1	11.8	11.0	15.5	11.3	10.1	11.5	12.6	10.3	10.8	10.9	12.7	9.6	8.9	8.5	7.7	7.4	8.2
28	5	B	0	0.0	0.0	0.0	6.4	6.5	7.5	7.7	8.8	10.0	8.6	9.3	8.5	11.9	9.7	9.5	9.2	10.0	9.7	7.9	12.1	10.6	40.4	10.2	6.9	7.9	8.9
28	6	A	0	0.0	0.0	2.6	5.2	8.1	9.2	10.3	10.9	11.3	11.5	10.5	10.6	10.3	9.5	12.7	9.1	9.8	8.9	9.7	10.0	10.3	14.2	11.1	8.4	8.3	6.9
29	1	B	0	0.0	0.0	5.9	13.2	19.9	23.7	27.0	27.3	23.8	23.3	21.8	19.2	18.5	15.6	13.9	13.2	12.1	11.1	10.3	9.4	8.5	7.5	8.7	6.0	5.0	4.2

Subject	period	Treat	_0	_8	_9	_10	_11	_12	_13	_14	_15	_16	_17	_18	_19	_20	_22	_24	_26	_28	_30	_34	_38	_42	_46	_50	_54	_58	_62
29	2	A	0	0.0	0.0	0.0	5.1	9.1	12.3	14.0	28.3	19.4	19.6	18.5	17.4	15.7	14.8	13.5	12.6	11.4	10.2	9.7	8.5	7.4	7.0	5.7	4.4	3.9	3.2
29	3	A	0	0.0	0.0	0.0	0.0	7.4	11.0	14.6	16.0	18.2	18.4	19.6	17.0	16.8	15.8	13.9	14.8	11.6	10.2	9.5	8.5	7.8	6.9	6.4	5.1	5.1	5.4
29	4	B	0	0.0	0.0	2.7	7.5	12.3	17.8	22.6	24.2	25.2	25.8	24.1	23.4	20.5	20.3	16.9	17.8	15.0	13.5	12.4	11.0	10.5	9.9	8.8	8.0	8.8	7.2
29	5	A	0	0.0	0.0	0.0	2.6	7.3	12.4	17.3	20.5	24.4	23.4	23.6	21.7	24.3	22.0	17.8	16.4	16.6	14.6	12.3	12.0	9.4	7.4	8.4	5.8	5.4	4.6
29	6	B	0	0.0	0.0	0.0	0.0	5.4	5.9	9.5	12.5	12.2	14.4	15.7	28.2	13.5	13.6	13.7	11.9	9.5	9.2	11.1	7.0	7.7	5.7	5.0	4.2	5.5	4.4
30	1	B	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	8.8	6.9	8.5	9.5	8.4	7.3	6.9	7.0	6.6	8.3	6.1	6.9	7.2	7.3	5.8	5.5	5.7
30	2	B	0	0.0	0.0	0.0	0.0	2.7	3.8	4.8	5.6	5.2	6.9	6.5	6.1	6.3	6.9	6.2	6.6	6.6	6.5	6.3	6.3	8.3	4.9	5.0	6.3	4.4	4.4
30	3	A	0	0.0	0.0	0.0	0.0	0.0	3.1	4.2	4.6	5.6	6.5	6.8	7.4	7.9	7.4	8.4	7.5	7.9	7.3	6.9	8.8	8.6	7.2	7.5	6.5	6.1	5.7
30	4	A	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	3.5	5.0	5.4	5.9	4.4	7.6	4.9	6.3	5.2	5.5	5.8	5.3	6.2	4.8	4.1	4.2	4.2	3.8	4.0
30	5	A	0	0.0	0.0	0.0	0.0	0.0	2.5	3.1	4.9	5.4	6.0	6.1	6.9	6.8	6.8	6.6	6.2	6.1	6.0	6.1	5.1	6.1	5.4	5.7	5.6	5.4	4.9
31	1	A	0	0.0	0.0	0.0	0.0	0.0	0.0	5.7	9.5	12.6	13.5	15.4	13.9	12.4	12.5	12.5	12.5	13.1	12.4	12.1	11.3	10.4	9.1	6.8	5.4	6.3	5.1
31	2	B	0	0.0	0.0	0.0	0.0	0.0	3.0	8.1	13.6	14.9	14.0	13.8	12.2	12.7	12.6	12.3	13.7	11.5	11.5	10.8	9.7	10.6	8.1	7.8	7.3	7.3	6.0
31	3	B	0	0.	0.	0.0	0.0	0.0	0.0	0.0	0.0	4.2	4.8	5.5	7.3	5.9	7.4	7.2	6.7	6.4	7.0	6.8	6.2	5.8	7.1	5.1	4.9	4.6	5.2

Subject	period	Treat	_0	_8	_9	_10	_11	_12	_13	_14	_15	_16	_17	_18	_19	_20	_22	_24	_26	_28	_30	_34	_38	_42	_46	_50	_54	_58	_62
				0	0																								
31	4	A	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	4.3	6.5	8.8	11.1	10.1	10.5	11.3	13.3	9.4	11.2	13.0	9.8	8.6	8.9	9.3	8.8	8.0	5.8	8.0
31	5	A	0	0.0	0.0	0.0	0.0	0.0	2.9	6.5	11.3	12.5	15.9	16.0	13.3	15.4	17.0	16.0	16.1	14.4	13.5	15.8	15.9	11.6	11.6	11.4	9.3	9.6	9.4
31	6	B	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	5.3	8.2	15.1	13.4	14.5	14.6	14.4	15.5	15.1	15.4	14.0	12.7	11.3	11.6	13.0	11.0	12.0	8.2	6.9
32	1	B	0	0.0	0.0	3.8	7.3	15.6	15.7	18.2	23.2	25.0	21.5	18.5	24.4	17.4	18.4	17.0	15.2	15.3	13.9	10.5	8.8	9.3	8.5	10.0	5.9	5.7	5.6
32	2	B	0	0.0	3.4	7.7	18.1	21.5	20.7	22.0	19.3	21.6	19.8	16.0	18.5	16.5	18.1	12.2	12.5	10.3	9.8	8.0	8.2	7.4	5.5	4.4	5.0	4.6	3.9
32	3	A	0	0.0	4.7	14.7	18.1	23.4	26.2	27.1	28.9	27.8	26.2	23.4	20.2	18.7	16.0	16.5	15.5	14.2	11.8	11.5	9.3	8.7	8.1	6.1	5.2	4.9	3.7
32	5	A	0	0.0	0.0	2.6	7.1	11.0	15.1	20.3	19.6	22.2	20.9	19.1	19.1	17.2	16.5	16.9	13.5	10.7	11.4	10.0	8.9	8.6	7.3	7.3	6.2	6.5	4.6
33	1	B	0	0.0	0.0	3.1	7.4	10.3	11.9	13.2	13.4	13.6	13.0	11.7	9.3	10.5	10.1	10.4	9.5	8.9	8.2	7.3	6.4	6.6	6.1	6.3	5.3	5.1	4.4
33	2	A	0	0.0	0.0	3.2	3.6	7.3	11.0	12.2	14.1	12.8	15.8	15.1	12.9	13.7	13.2	11.3	12.1	11.2	9.6	8.5	9.5	8.2	7.1	6.7	6.7	4.9	5.7
33	3	B	0	0.0	0.0	0.0	5.3	9.7	14.8	18.6	19.2	20.7	19.4	18.6	15.7	18.5	19.6	16.4	14.1	14.8	14.2	14.5	12.0	11.5	10.5	13.3	9.3	9.8	7.8
33	4	A	0	0.0	0.0	0.0	4.1	5.8	10.5	13.1	14.1	14.6	13.6	15.3	13.3	12.5	11.8	11.2	13.1	11.4	10.9	9.6	8.8	8.1	7.6	5.3	11.8	5.8	6.2
33	6	A	0	2.6	5.2	11.2	16.4	17.0	20.1	17.3	18.6	19.5	19.4	21.1	17.8	16.8	18.9	15.4	16.0	15.4	14.5	14.6	14.4	13.6	14.1	12.8	11.1	13.3	9.3

Subject	period	Treat	_0	_8	_9	_10	_11	_12	_13	_14	_15	_16	_17	_18	_19	_20	_22	_24	_26	_28	_30	_34	_38	_42	_46	_50	_54	_58	_62
35	1	B	0	0.0	0.0	0.0	0.0	3.4	7.8	11.9	16.1	18.1	16.3	16.8	15.0	14.7	14.2	11.1	10.1	9.0	8.1	7.4	6.1	4.3	4.3	3.9	3.4	3.2	3.1
35	2	B	0	0.0	0.0	3.0	7.7	13.1	16.1	19.0	29.0	23.6	17.9	18.1	13.2	25.7	13.0	9.7	10.7	11.2	8.8	6.8	7.1	6.0	5.3	5.7	0.0	12.1	4.0
35	3	A	0	0.0	0.0	0.0	10.2	14.5	18.5	19.9	23.2	18.4	20.3	19.1	15.1	12.0	12.0	12.8	9.7	9.1	9.1	10.9	8.2	5.5	6.5	3.7	3.4	0.0	0.0
35	4	B	0	0.0	0.0	2.8	6.2	9.4	15.1	19.0	17.6	17.1	17.9	12.7	14.0	12.6	11.5	10.0	12.9	8.5	8.1	7.2	6.5	5.8	5.7	3.9	5.4	5.0	3.7
35	5	A	0	0.0	0.0	0.0	0.0	7.0	13.4	16.4	25.8	22.9	24.1	22.5	17.8	17.9	17.5	16.6	16.4	12.4	12.7	11.9	10.5	9.5	9.0	8.1	7.6	5.4	6.0
35	6	A	0	0.0	0.0	0.0	3.7	6.0	10.8	13.5	20.0	22.8	16.2	18.4	15.3	17.0	12.3	11.0	14.8	13.7	14.0	11.1	9.6	8.4	8.4	7.5	8.6	5.4	4.2
36	1	A	0	0.0	0.0	2.6	5.0	14.8	12.2	15.0	16.6	17.1	17.3	16.3	15.7	16.2	15.2	15.5	13.9	13.7	12.7	11.3	11.0	10.0	9.0	8.9	8.5	9.4	6.7
36	2	A	0	0.0	0.0	0.0	3.1	6.1	5.3	8.3	9.9	12.0	14.8	13.4	14.7	14.0	16.5	14.9	13.5	13.3	10.5	9.9	10.1	11.5	7.3	8.0	7.8	7.9	7.4
36	3	B	0	0.0	0.0	0.0	0.0	4.7	7.2	9.7	12.1	16.9	14.5	16.3	14.9	15.3	14.5	15.5	13.9	11.4	16.6	12.3	12.1	10.9	11.3	10.6	9.5	8.9	7.7
36	5	B	0	0.0	0.0	0.0	0.0	5.2	10.3	9.1	7.8	10.1	12.2	12.3	12.4	12.1	12.3	12.1	11.6	11.4	9.4	10.1	10.3	9.2	8.5	7.2	6.8	6.5	5.7
36	6	B	0	0.0	0.0	0.0	3.1	4.2	7.3	9.5	10.9	12.1	12.2	13.5	13.4	13.0	11.8	11.9	11.5	10.6	10.5	9.6	9.7	7.9	7.4	6.5	7.1	6.5	7.8
37	1	A	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	2.7	8.0	0.0	0.0	7.0	10.4	9.6	10.3	8.8	9.2	8.9	7.4	7.7	7.3
37	2	A	0	0.	0.	7.5	3.8	9.5	10.	12.	11.	12.	11.	12.	11.	11.	13.	12.	12.	14.	12.	11.	11.	12.	10.	11.	8.8	7.9	6.3

Subject	period	Treat	_0	_8	_9	_10	_11	_12	_13	_14	_15	_16	_17	_18	_19	_20	_22	_24	_26	_28	_30	_34	_38	_42	_46	_50	_54	_58	_62
				0	0				8	1	1	5	9	5	5	3	1	8	5	7	6	5	7	1	4	7			
37	3	B	0	0.0	0.0	0.0	3.5	6.1	7.6	7.3	10.4	9.3	10.6	10.5	10.1	9.1	9.7	10.2	10.2	10.2	11.1	10.7	8.3	8.8	8.3	8.4	6.2	6.3	5.8
37	4	A	0	0.0	0.0	0.0	0.0	2.5	7.0	8.2	9.9	9.5	13.6	11.6	10.6	13.0	13.1	12.1	12.8	9.9	11.5	11.9	12.7	17.6	10.7	10.1	11.3	8.2	12.5
37	5	B	0	0.0	3.7	5.1	5.3	9.0	10.0	15.2	16.4	18.7	17.0	19.2	19.2	19.4	19.5	19.9	22.4	19.0	21.9	22.6	19.4	19.2	17.1	19.4	15.7	15.3	11.4
37	6	B	0	0.0	0.0	0.0	0.0	3.9	5.8	7.4	9.1	9.6	10.1	12.9	13.2	10.5	13.3	14.0	8.6	13.3	13.6	13.2	12.8	14.3	16.4	11.6	7.9	4.2	8.9
38	1	A	0	0.0	0.0	0.0	4.6	10.1	15.4	14.9	19.5	20.6	19.0	19.8	18.4	16.4	16.6	16.3	15.7	14.5	14.0	11.1	11.4	9.9	9.9	8.2	7.6	6.1	5.8
38	2	A	0	0.0	0.0	0.0	7.2	12.1	17.1	21.7	43.8	34.7	32.4	26.7	29.6	23.9	26.4	20.9	26.1	28.5	20.7	17.1	17.6	18.0	15.3	13.2	11.4	16.7	7.6
38	3	B	0	0.0	0.0	0.0	11.1	18.8	19.6	35.8	36.4	30.8	37.7	36.5	33.5	34.0	35.8	33.4	23.3	32.7	37.5	22.1	19.4	15.1	32.8	27.6	22.5	17.3	20.4
38	4	A	0	0.0	0.0	0.0	5.3	14.9	11.8	18.8	19.8	19.6	19.1	17.7	16.4	16.1	23.8	13.4	12.5	11.4	12.5	11.0	9.6	9.6	8.4	7.0	7.8	7.6	5.1
38	5	B	0	0.0	0.0	0.0	6.5	10.8	15.2	19.4	23.0	21.7	23.1	22.8	18.0	18.7	17.2	15.2	15.5	13.0	15.7	11.3	12.1	8.6	9.5	8.8	7.9	6.3	5.4
38	6	B	0	0.0	0.0	0.0	6.3	12.7	20.6	29.2	25.8	32.5	28.5	25.4	24.8	27.8	23.6	22.9	21.3	22.6	18.2	18.7	16.9	15.8	12.5	12.3	9.9	11.3	9.7
39	1	B	0	0.0	0.0	0.0	0.0	3.1	6.2	10.1	11.3	13.5	13.4	10.8	11.2	10.8	9.7	8.4	7.6	7.1	6.2	4.0	4.5	4.3	3.2	2.7	2.7	0.0	3.1
39	2	B	0	0.0	0.0	0.0	0.0	6.6	10.5	13.7	12.4	13.4	13.2	12.4	10.0	9.2	8.2	7.6	8.6	6.2	7.0	5.4	6.0	6.0	5.4	3.6	3.5	2.9	2.5

Subject	period	Treat	_0	_8	_9	_10	_11	_12	_13	_14	_15	_16	_17	_18	_19	_20	_22	_24	_26	_28	_30	_34	_38	_42	_46	_50	_54	_58	_62
39	3	A	0	0.0	0.0	0.0	2.9	6.6	8.4	11.6	12.0	12.1	11.0	9.6	9.5	8.7	7.7	7.2	7.0	7.1	6.5	5.3	5.8	5.4	3.4	4.2	3.3	2.8	3.6
39	4	A	0	0.0	0.0	0.0	2.8	5.1	7.2	10.8	10.1	10.2	9.3	8.6	10.0	8.5	7.4	6.5	4.5	5.4	4.6	3.9	3.5	3.0	2.8	2.8	4.9	3.3	0.0
39	5	B	0	0.0	0.0	0.0	0.0	4.0	6.6	9.4	12.1	11.3	9.1	7.9	11.3	7.7	6.6	6.0	5.1	4.8	4.9	5.2	3.4	4.0	2.8	3.0	3.1	2.6	3.2
39	6	A	0	0.0	0.0	0.0	4.7	6.1	9.0	11.9	11.7	11.9	11.9	13.9	10.1	10.0	9.1	9.2	7.3	8.0	5.6	6.8	4.9	5.2	4.7	3.3	3.1	3.4	4.1
40	2	B	0	0.0	0.0	0.0	6.9	6.6	5.1	4.1	5.4	4.7	4.6	5.0	6.1	6.1	5.5	4.8	6.7	5.1	5.5	5.9	6.3	6.8	4.0	4.3	3.8	4.1	4.5
40	3	B	0	0.0	0.0	0.0	0.0	2.9	3.4	4.0	4.4	5.4	7.9	5.8	6.3	6.7	6.2	6.1	5.9	5.9	6.2	5.7	5.6	5.3	4.7	5.0	4.6	5.4	4.2
40	4	A	0	0.0	0.0	0.0	0.0	0.0	0.0	3.5	4.0	4.5	5.8	5.1	5.2	6.3	7.4	5.9	5.4	6.8	5.1	5.8	6.1	5.0	7.3	4.9	5.0	5.2	6.7
40	5	B	0	0.0	0.0	0.0	2.6	4.1	4.4	3.7	5.5	6.2	6.3	5.6	5.7	5.5	5.9	4.8	5.0	5.0	5.6	4.9	6.4	5.8	4.3	2.8	4.9	4.5	4.1
40	6	A	0	0.0	0.0	0.0	0.0	2.6	3.6	3.6	3.6	4.4	4.2	4.6	4.5	4.4	5.2	5.1	5.2	5.6	5.1	5.6	5.2	4.9	4.5	4.5	4.0	4.5	4.3

Figure 11 Blood conc.by time curves averaged over periods

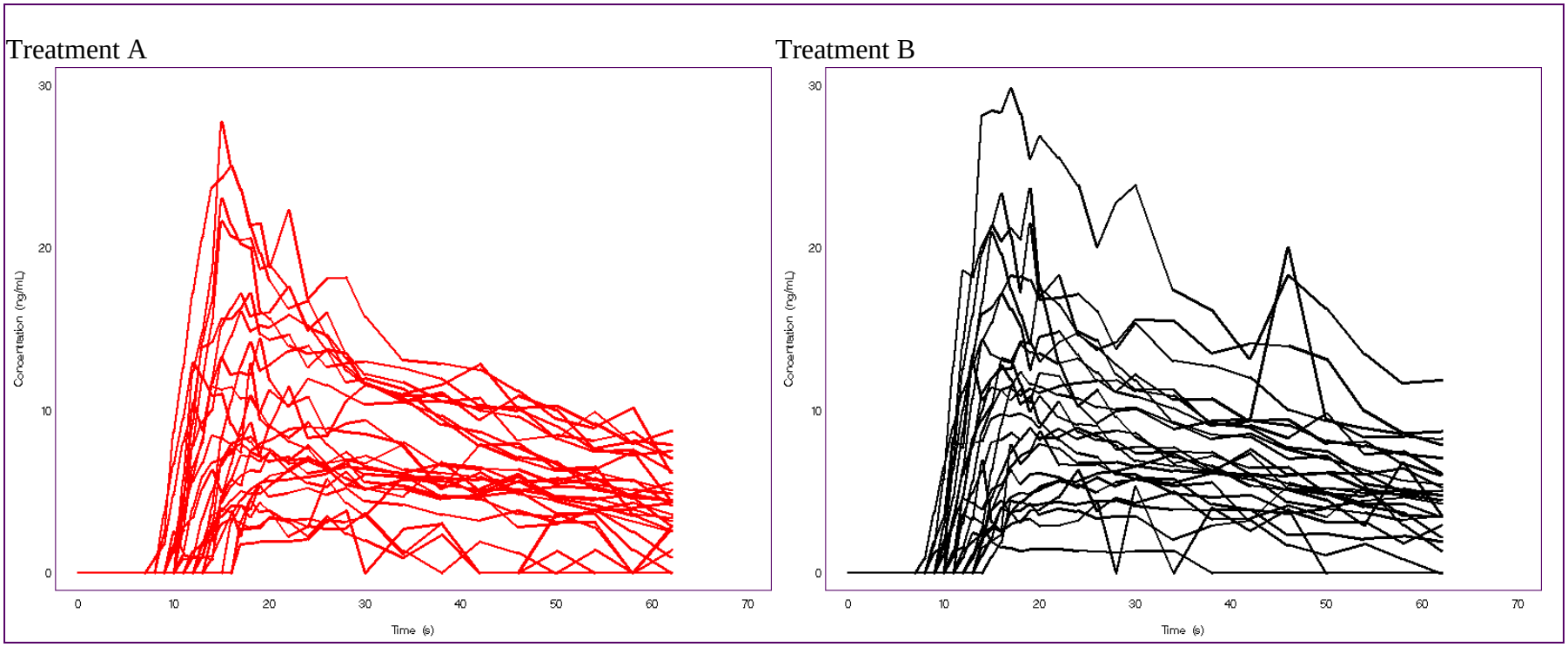
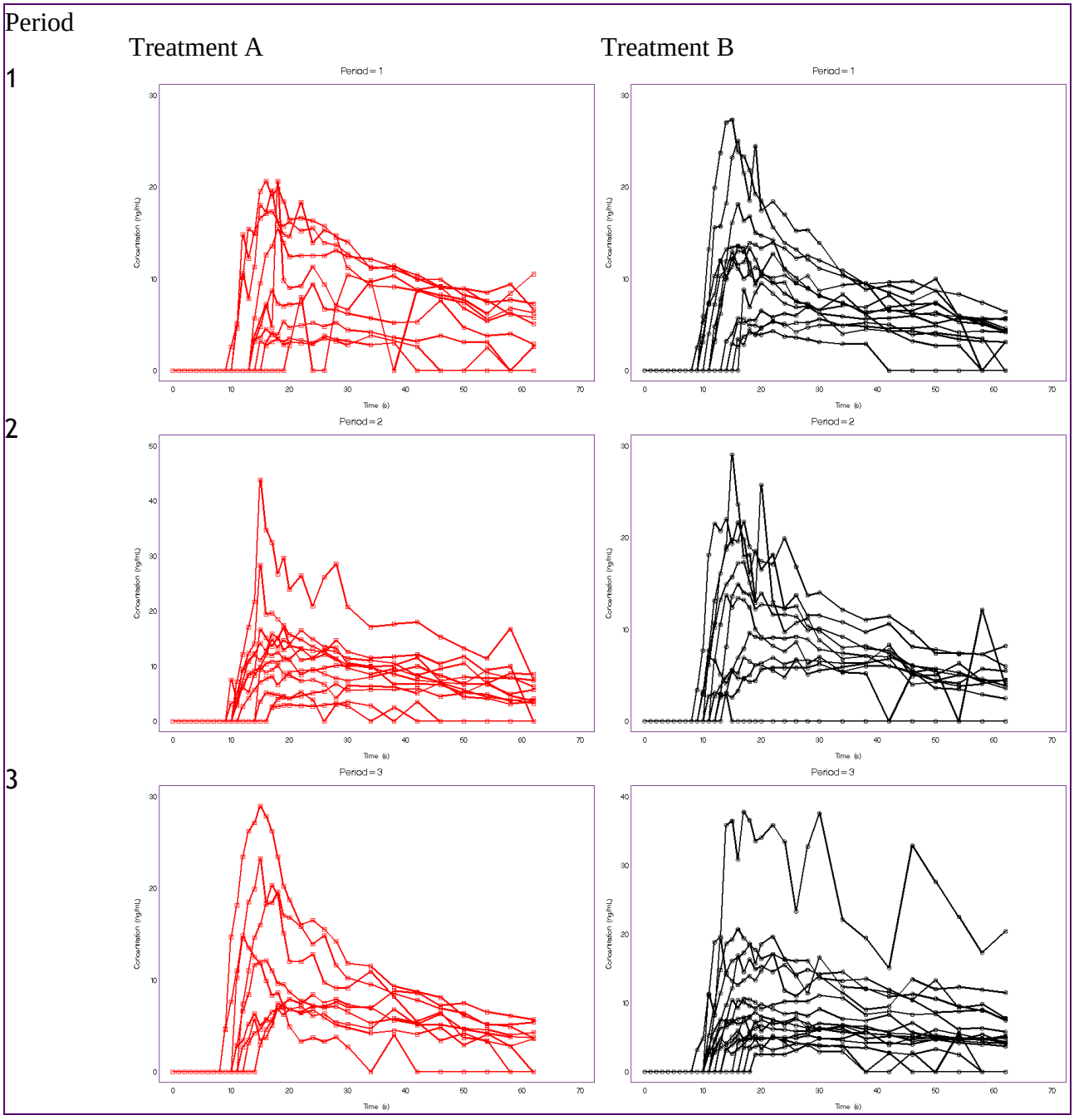
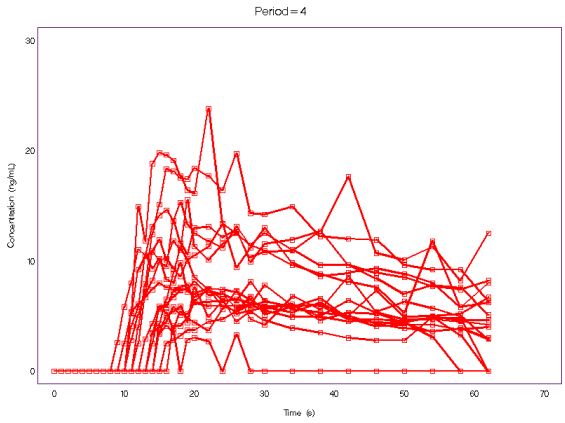
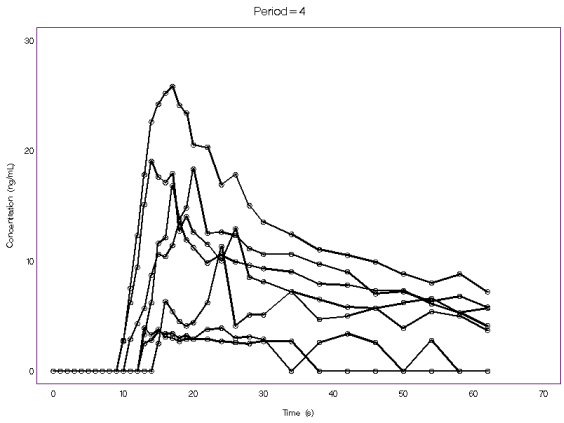
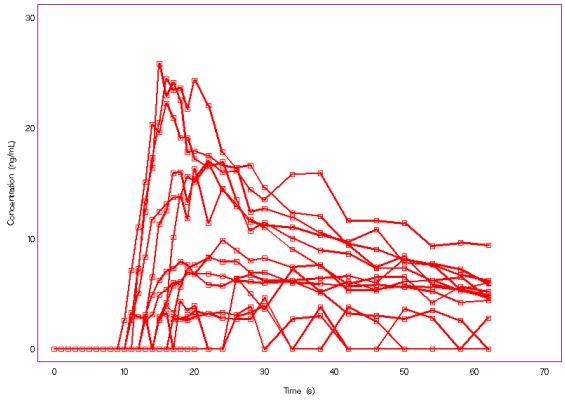
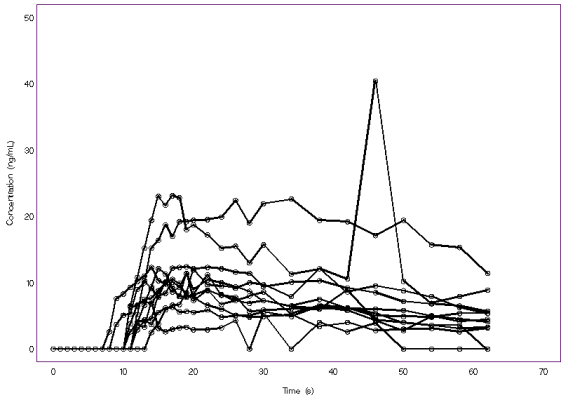
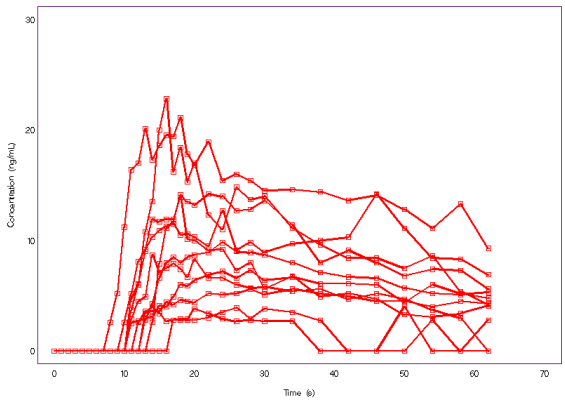
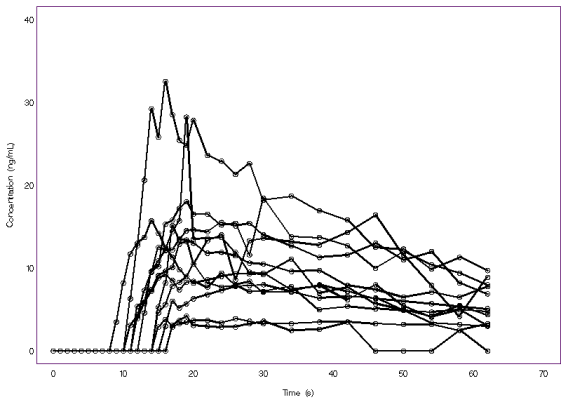


Figure 12 Blood conc curves by treatment and period



Period	Treatment A	Treatment B
4		
5		
6		

.

APPENDIX F: Parallel Design Data

This data from:

"Clayton D, Leslie A: The bioavailability of erythromycin stearate vesus enteric-coated erythromycin base when taken immediately before and after food. Journal of International Medical Research 9:4770-4777, 1981"

Variable name	t0	t05	t10	t15	t20	t40	t60	t80
Concentration @ time	0	30min	60min	90min	2hours	4hours	6hours	8hours

Table 17 Reference Concentrations by subject and time

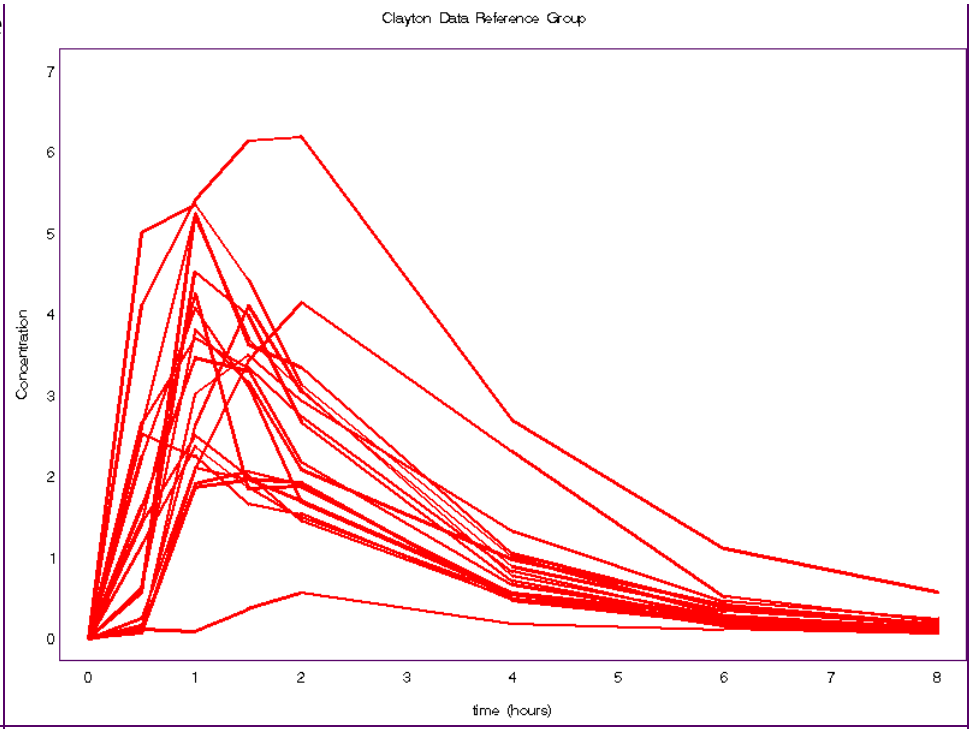
SUBJECT	Treatment	t0	t05	t10	t15	t20	t40	t60	t80
1	Reference	0	5.00	5.35	4.42	3.13	1.00	0.28	0.12
2	Reference	0	0.07	2.08	3.42	4.14	2.30	0.52	0.21
3	Reference	0	1.63	3.46	3.31	2.18	0.66	0.25	0.12
4	Reference	0	0.62	4.51	3.98	2.67	0.71	0.23	0.15
5	Reference	0	2.66	3.70	3.34	2.74	0.83	0.23	0.15
6	Reference	0	0.17	2.62	4.10	3.04	0.88	0.34	0.17
7	Reference	0	4.12	5.40	6.13	6.18	2.68	1.10	0.57
8	Reference	0	2.24	4.07	3.11	1.68	0.55	0.16	0.09
9	Reference	0	1.43	2.50	2.01	1.45	0.49	0.22	0.15
10	Reference	0	2.63	5.23	3.62	3.34	1.04	0.38	0.17
11	Reference	0	1.44	4.24	1.84	1.87	0.56	0.28	0.14
12	Reference	0	2.52	2.25	1.66	1.54	0.48	0.23	0.14
13	Reference	0	0.09	3.01	3.49	2.93	1.32	0.46	0.25
14	Reference	0	0.26	2.10	1.96	1.92	0.47	0.19	0.09
15	Reference	0	0.57	5.18	3.69	3.07	0.77	0.28	0.13
16	Reference	0	0.65	3.80	3.16	2.08	0.97	0.41	0.18
17	Reference	0	0.09	1.86	1.96	1.70	0.56	0.26	0.13
18	Reference	0	1.16	2.36	1.87	1.49	0.54	0.20	0.12
19	Reference	0	0.11	0.09	0.36	0.56	0.18	0.12	0.07
20	Reference	0	0.12	1.90	2.06	1.90	0.49	0.21	0.13

Table 18 Test concentration by subject and time

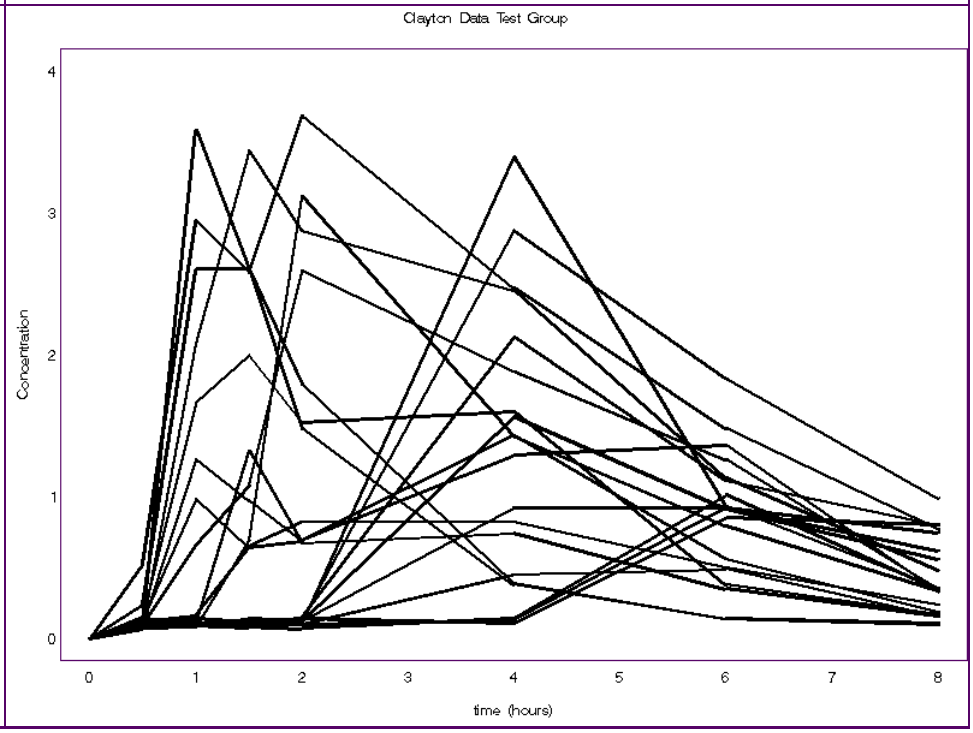
SUBJECT	Treatment	t0	t05	t10	t15	t20	t40	t60	t80
21	Test	0	0.10	0.66	1.08	2.59	1.88	1.26	0.35
22	Test	0	0.12	0.10	0.13	0.10	0.13	0.92	0.75
23	Test	0	0.24	2.60	2.60	1.48	0.38	0.14	0.10
24	Test	0	0.10	2.08	3.43	2.87	2.45	1.13	0.48
25	Test	0	0.13	0.98	0.64	0.69	1.43	0.56	0.19
26	Test	0	0.16	1.66	1.99	1.52	1.60	0.38	0.17
27	Test	0	0.53	2.94	2.58	3.68	2.47	1.48	0.77
28	Test	0	0.11	3.58	2.60	1.79	0.39	0.14	0.11
29	Test	0	0.13	0.14	0.12	0.15	2.12	1.11	0.80
30	Test	0	0.10	1.26	0.97	0.69	1.29	1.36	0.33
31	Test	0	0.10	0.12	0.08	0.07	0.14	1.01	0.35
32	Test	0	0.07	0.08	0.69	3.11	1.42	0.79	0.34
33	Test	0	0.13	0.11	0.14	0.13	2.87	1.83	0.99
34	Test	0	0.12	0.17	0.65	0.82	0.82	0.50	0.24
35	Test	0	0.14	0.09	0.10	0.10	3.39	0.93	0.36
36	Test	0	0.13	0.10	0.14	0.12	0.92	0.92	0.62
37	Test	0	0.10	0.09	0.12	0.14	0.11	0.85	0.80
38	Test	0	0.13	0.10	0.10	0.10	0.45	0.49	0.16
39	Test	0	0.08	0.09	0.08	0.12	1.56	0.91	0.56
40	Test	0	0.08	0.11	1.32	0.67	0.74	0.35	0.16

Figure 13 Reference and Test serum concentration profiles

reference



Test



AUC was calculated using the trapezoidal method. CMax was determined as the highest observed concentration.

Table 19 AUC and Cmax from non-compartmental methods

Reference Treatment					Test Treatment				
SUBJECT	AUC	logAUC	Cmax	logCmax	SUBJECT	AUC	logAUC	Cmax	logCmax
1	13.978	2.637	5.350	1.677	21	10.788	2.378	2.590	0.952
2	13.810	2.625	4.140	1.421	22	3.150	1.147	0.920	-0.083
3	8.865	2.182	3.460	1.241	23	5.710	1.742	2.600	0.956
4	9.923	2.295	4.510	1.506	24	14.033	2.641	3.430	1.233
5	10.545	2.356	3.700	1.308	25	5.908	1.776	1.430	0.358
6	9.855	2.288	4.100	1.411	26	7.935	2.071	1.990	0.688
7	23.680	3.165	6.180	1.821	27	16.295	2.791	3.680	1.303
8	8.320	2.119	4.070	1.404	28	6.553	1.880	3.580	1.275
9	6.353	1.849	2.500	0.916	29	7.643	2.034	2.120	0.751
10	12.925	2.559	5.230	1.654	30	7.658	2.036	1.360	0.307
11	7.918	2.069	4.240	1.445	31	2.888	1.060	1.010	0.010
12	6.700	1.902	2.520	0.924	32	9.068	2.205	3.110	1.135
13	10.768	2.377	3.490	1.250	33	10.743	2.374	2.870	1.054
14	5.970	1.787	2.100	0.742	34	4.375	1.476	0.820	-0.198
15	10.788	2.378	5.180	1.645	35	9.290	2.229	3.390	1.221
16	9.345	2.235	3.800	1.335	36	4.635	1.534	0.920	-0.083
17	5.850	1.766	1.960	0.673	37	3.050	1.115	0.850	-0.163
18	6.158	1.818	2.360	0.859	38	2.330	0.846	0.490	-0.713
19	1.650	0.501	0.560	-0.580	39	5.775	1.754	1.560	0.445
20	5.945	1.783	2.060	0.723	40	3.933	1.369	1.320	0.278

.

Figure 14 Raw and Predicted Blood Concentration Profiles of Reference drug

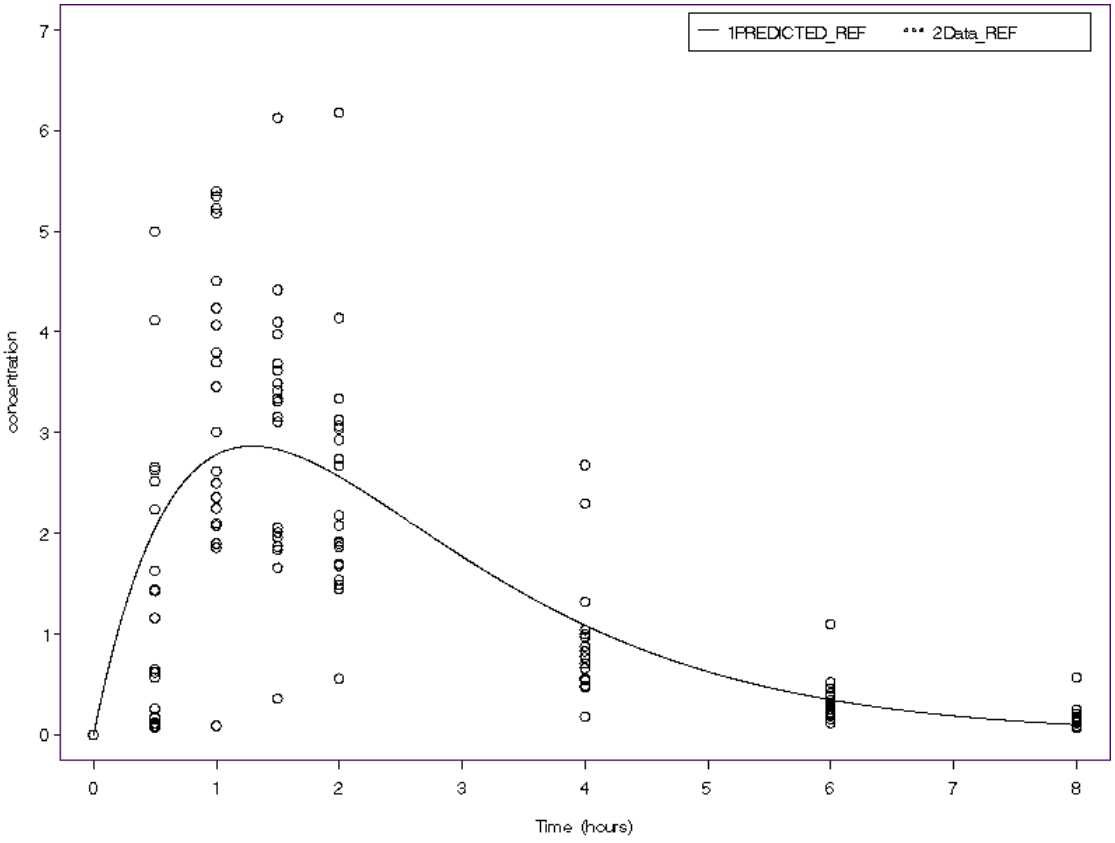


Figure 15 Raw and Predicted Blood Concentration Profiles of Test drug

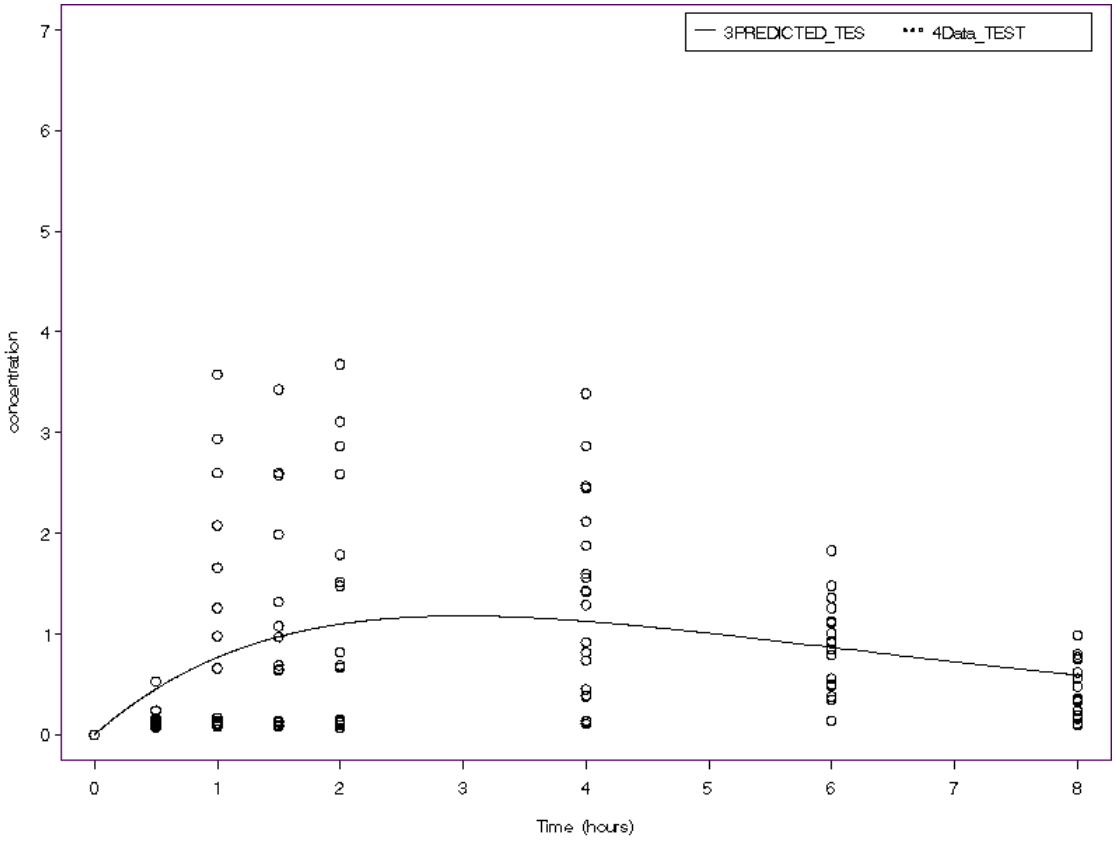
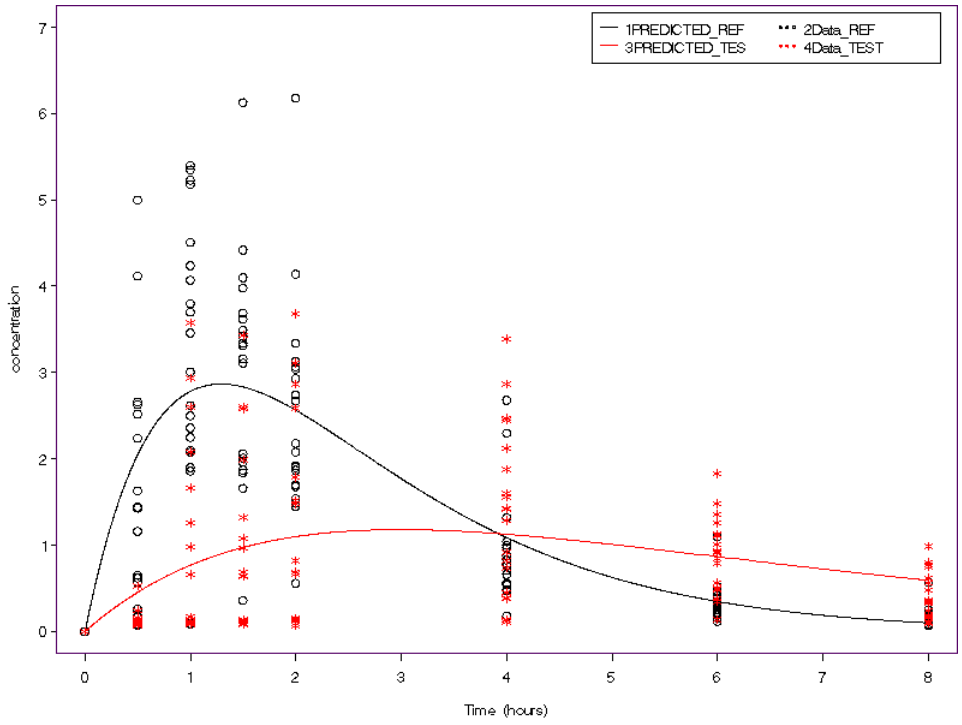


Figure 16 Raw and Predicted Blood Concentration Profiles of Reference and Test Drugs



.

APPENDIX G: Confidence intervals of the estimated size in table 1

Table 20 95% CI of estimates sizes in table 1

		Cp95<theta			Cp95<theta-c		
rhoR	rhoT	n=25	n=50	n=100	n=25	n=50	n=100
-0.2	-0.2	-0.029 0.049	’0 , 0	0 , 0	-0.032 0.058	,-0.013 0.019	,-0.009 , 0.036
	0	0 , 0	-0.016 0.029	’0 , 0	0 , 0	-0.016 0.029	,-0.005 , 0.058
	0.2	-0.025 0.039	’0 , 0	0 , 0	-0.025 0.039	,-0.018 0.045	’0.002 , 0.078
	0.4	-0.019 0.026	’0 , 0	-0.008 0.015	,-0.034 0.067	,-0.018 0.045	,-0.006 , 0.053
	0.8	0 , 0	-0.016 0.029	’-0.01 , 0.03	-0.035 0.075	,-0.017 0.077	,-0.005 , 0.058
0	-0.2	-0.019 0.026	,-0.018 0.045	,-0.009 0.036	,-0.019 0.026	,-0.018 0.045	,-0.009 , 0.036
	0	-0.019 0.026	,-0.019 0.059	’-0.01 , 0.03	-0.019 0.026	,-0.019 0.059	’-0.01 , 0.03
	0.2	-0.019 0.026	,-0.018 0.045	,-0.002 0.068	,-0.019 0.026	,-0.018 0.045	,-0.002 , 0.068
	0.4	0 , 0	-0.018 0.045	’-0.01 , 0.03	0 , 0	-0.018 0.045	’-0.01 , 0.03
	0.8	-0.029 0.049	,-0.019 0.052	,-0.003 0.063	,-0.029 0.049	,-0.019 0.052	,-0.003 , 0.063
0.2	-0.2	-0.029 0.049	’0 , 0	-0.009 0.036	,-0.025 0.039	’0 , 0	-0.01 , 0.03
	0	0 , 0	-0.012 0.105	,0.011 , 0.102	0 , 0	-0.019 0.052	,-0.006 , 0.053
	0.2	-0.032 0.058	,-0.018 0.071	,0.005 , 0.088	-0.029 0.049	,-0.018 0.045	,-0.009 , 0.023
	0.4	-0.032 0.058	,-0.012 0.105	,0.025 , 0.129	-0.025 0.039	,-0.018 0.071	,-0.006 , 0.053
	0.8	-0.035 0.135	,-0.002 0.136	,***0.075, 0.212***	-0.036 0.082	,-0.018 0.045	’0.003 , 0.083
0.4	-0.2	-0.019 0.026	’0 , 0	-0.01 , 0.03	-0.019 0.026	,-0.013 0.019	,-0.009 , 0.036

		Cp95<theta			Cp95<theta-c		
rhoR	rhoT	n=25	n=50	n=100	n=25	n=50	n=100
	0	0 , 0	-0.016 0.029	, -0.002 0.068	’ 0 , 0	0 , 0	-0.007 0.047 ,
	0.2	-0.025 0.039	, -0.016 0.083	’ 0.018 , 0.116	-0.019 0.026	, -0.019 0.059	, -0.008 0.042 ,
	0.4	-0.029 0.049	, -0.012 0.105	, ***0.064 , 0.196***	-0.025 0.039	, -0.018 0.045	, -0.009 0.023 ,
	0.8	-0.035 0.141	, ***0.066, 0.274***	***0.275 , 0 .465***	-0.029 0.049	, -0.019 0.065	, -0.008 0.042 ,
0.8	-0.2	0 , 0	0 , 0	0 , 0	-0.025 0.039	, -0.016 0.083	, -0.002 0.068 ,
	0	0 , 0	0 , 0	0 , 0	-0.025 0.039	, -0.018 0.038	, -0.007 0.047 ,
	0.2	0 , 0	0 , 0	0 , 0	-0.025 0.039	, -0.018 0.045	, -0.008 0.015 ,
	0.4	0 , 0	0 , 0	0 , 0	-0.019 0.026	, -0.016 0.029	, -0.002 0.068 ,
	0.8	-0.032 0.058	, ***0.023, 0.197***	***0.198, 0.375***	0 , 0	0 , 0	-0.007 0.047 ,

.

APPENDIX H: Results of Crossover Design mixed model

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
outcome	1	38.2	5284.2	<.0001
sequence	14	13.4	2.26	0.0726
period	5	104	0.91	0.4807
Treatment	1	105	5.2	0.0247
Treatment*outcome	1	92.2	1.6	0.2097

Solution for Fixed Effects

Effect		Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept		1.9832	0.4428	13	4.48	0.0006
outcome	logAUC	3.3565	0.04935	51.9	68.01	<.0001
outcome	logCmax	0	.	.	.	.
sequence	AABABB	0.8618	0.4899	12.5	1.76	0.103
sequence	AABBAB	-0.4223	0.6364	13.8	-0.66	0.5179
sequence	ABAABB	0.263	0.508	12.7	0.52	0.6136
sequence	ABABAB	0.001884	0.5766	16.4	0	0.9974
sequence	ABBAAB	0.4606	0.506	12.5	0.91	0.3799
sequence	ABBABA	-0.07047	0.627	13.1	-0.11	0.9122
sequence	ABBBAA	-0.6698	0.6278	13.1	-1.07	0.3052
sequence	BAABAB	1.3201	0.6196	12.5	2.13	0.0537
sequence	BABAAB	0.01746	0.6196	12.5	0.03	0.978
sequence	BABABA	0.7231	0.506	12.5	1.43	0.1775
sequence	BABBAA	0.1171	0.5059	12.5	0.23	0.8207
sequence	BBAAAB	0.161	0.627	13.1	0.26	0.8013
sequence	BBAABA	0.6355	0.6196	12.5	1.03	0.3245
sequence	BBABAA	1.1973	0.5389	12.7	2.22	0.0452
sequence	BBBAAA	0	.	.	.	.
period	1	-0.06025	0.1114	70	-0.54	0.5904
period	2	0.007936	0.1062	78.4	0.07	0.9406
period	3	-0.06229	0.09954	90.4	-0.63	0.533
period	4	0.03087	0.09082	104	0.34	0.7346
period	5	0.08034	0.07215	111	1.11	0.2679
period	6	0	.	.	.	.

Treatment	A		-0.08724	0.04502	110	-1.94	0.0552
Treatment	B		0	.	.	.	.
Treatment*outcome	A	logAUC	-0.04687	0.0371	92.2	-1.26	0.2097
Treatment*outcome	A	logCmax	0	.	.	.	.
Treatment*outcome	B	logAUC	0	.	.	.	.
Treatment*outcome	B	logCmax	0	.	.	.	.

Least Squares Means

Effect	trt		Estimate	Standard Error	Lower	Upper
Treatment*outcome	A	logAUC	5.5115	0.1112	5.2814	5.7416
Treatment*outcome	A	logCmax	2.2018	0.09974	1.9895	2.4142
Treatment*outcome	B	logAUC	5.6456	0.1113	5.4154	5.8758
Treatment*outcome	B	logCmax	2.2891	0.09996	2.0764	2.5017

.

Table 21 Example of R covariance Matrix for subject 20 Example of the R covariance matrix for subject # 20

0.1531	0.152	0.09697	0.09628	0.06143	0.06099	0.03891	0.03864	0.02465	0.02448	0.01562	0.0
0.152	0.2687	0.09628	0.1702	0.06099	0.1078	0.03864	0.06832	0.02448	0.04328	0.0155	0.0
0.09697	0.09628	0.1531	0.152	0.09697	0.09628	0.06143	0.06099	0.03891	0.03864	0.02465	0.0
0.09628	0.1702	0.152	0.2687	0.09628	0.1702	0.06099	0.1078	0.03864	0.06832	0.02448	0.0
0.06143	0.06099	0.09697	0.09628	0.1531	0.152	0.09697	0.09628	0.06143	0.06099	0.03891	0.0
0.06099	0.1078	0.09628	0.1702	0.152	0.2687	0.09628	0.1702	0.06099	0.1078	0.03864	0.0
0.03891	0.03864	0.06143	0.06099	0.09697	0.09628	0.1531	0.152	0.09697	0.09628	0.06143	0.0
0.03864	0.06832	0.06099	0.1078	0.09628	0.1702	0.152	0.2687	0.09628	0.1702	0.06099	0.1
0.02465	0.02448	0.03891	0.03864	0.06143	0.06099	0.09697	0.09628	0.1531	0.152	0.09697	0.0
0.02448	0.04328	0.03864	0.06832	0.06099	0.1078	0.09628	0.1702	0.152	0.2687	0.09628	0.1
0.01562	0.0155	0.02465	0.02448	0.03891	0.03864	0.06143	0.06099	0.09697	0.09628	0.1531	0.1
0.0155	0.02742	0.02448	0.04328	0.03864	0.06832	0.06099	0.1078	0.09628	0.1702	0.152	0.2

.

APPENDIX I: Results of Crossover Design NLMEM model

```
proc nlmixed data=BE.concentrations2 ecov ;
where period<3;
a= treatment='A';
per=period=2;

dose=1;

      parms beta1=-5 beta2=-2 beta3=0  b1=-1 b2=0 b3=1  s2=57
            s2b1 =0.03  cb12 =0    s2b2 =0.4  ;
      cl   = exp(beta1 + b1*a+ u1);
      ka   = exp(beta2 + b2*a+ u2);
      ke   = exp(beta3+b3*a+p1*per);
      pred = dose*ke*ka*(exp(-ke*time)-exp(-ka*time))/cl/(ka-ke);
      model concentration ~ normal(pred,s2);
      random u1 u2 ~ normal([0,0],[s2b1,cb12,s2b2]) subject=subject;

            cla   = exp(beta1+ b1);
            kaa   = exp(beta2+ b2);
            kea   = exp(beta3+b3);

            kab= exp(beta2);
            keb= exp(beta3);
            clb= exp(beta1);
            estimate        'logCmaxModel        a'        log(kaa*kea*(exp(-
kea*log(kea/kaa)/(kea-kaa))-exp(-kaa*log(kea/kaa)/(kea-kaa)))/(cla*(kaa-
kea))) ;
            estimate 'logAUCModel a' log(1/cla)  ;
            estimate        'logCmaxModel        b'        log(kab*keb*(exp(-
keb*log(keb/kab)/(keb-kab))-exp(-kab*log(keb/kab)/(keb-kab)))/(clb*(kab-
keb))) ;
            estimate 'logAUCModel b' log(1/clb)  ;

      predict pred out=predvals;

   run;
```

The NLMIXED Procedure

Specifications	
Data Set	BE.CONCENTRATIONS2
Dependent Variable	concentration
Distribution for Dependent Variable	Normal
Random Effects	u1 u2
Distribution for Random Effects	Normal
Subject Variable	Subject
Optimization Technique	Dual Quasi-Newton
Integration Method	Adaptive Gaussian Quadrature

Parameters											
beta1	beta2	beta3	b1	b2	b3	s2	s2b1	cb12	s2b2	p1	NegLogLike
-5	-2	0	-1	0	1	57	0.03	0	0.4	1	5407.07691

Iteration History						
Iter		Calls	NegLogLike	Diff	MaxGrad	Slope
1		5	5307.82215	99.25476	156.5468	-4082.82
2		7	5174.03882	133.7833	68.23715	-83.7382
80		897	4347.94491	0.000617	0.248288	-0.0006
81		898	4347.94407	0.000845	0.093543	-0.00168
82		900	4347.94402	0.000047	0.072078	-0.00004

NOTE: GCONV convergence criterion satisfied.

Fit Statistics	
-2 Log Likelihood	8695.9
AIC (smaller is better)	8717.9
AICC (smaller is better)	8718.1
BIC (smaller is better)	8738.0

Parameter Estimates									
Parameter	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper	Gradient
beta1	-6.4239	0.09766	44	-65.78	<.0001	0.05	-6.6207	-6.2271	0.023142
beta2	-3.6162	0.1890	44	-19.13	<.0001	0.05	-3.9971	-3.2353	0.019694
beta3	-3.2253	0.1315	44	-24.53	<.0001	0.05	-3.4903	-2.9603	-0.0112
b1	-0.2264	0.1125	44	-2.01	0.0503	0.05	-0.4531	0.000261	0.048394
b2	-0.3056	0.2409	44	-1.27	0.2113	0.05	-0.7910	0.1799	0.042683
b3	-0.1579	0.1653	44	-0.96	0.3445	0.05	-0.4910	0.1752	0.048333
s2	13.6338	0.4990	44	27.32	<.0001	0.05	12.6280	14.6396	0.025545
s2b1	0.08768	0.04568	44	1.92	0.0614	0.05	-0.00438	0.1797	0.044931
cb12	-0.1306	0.04242	44	-3.08	0.0036	0.05	-0.2161	-0.04515	0.00585
s2b2	0.3163	0.1401	44	2.26	0.0289	0.05	0.03404	0.5986	-0.07208
p1	0.05966	0.1311	44	0.45	0.6514	0.05	-0.2047	0.3240	0.038647

Additional Estimates								
Label	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper
logCmaxModel a	1.9738	0.1351	44	14.61	<.0001	0.05	1.7015	2.2460
logAUCModel a	6.6503	0.1137	44	58.48	<.0001	0.05	6.4211	6.8795
logCmaxModel b	1.9904	0.1254	44	15.87	<.0001	0.05	1.7377	2.2432
logAUCModel b	6.4239	0.09766	44	65.78	<.0001	0.05	6.2271	6.6207

Covariance Matrix of Additional Estimates					
Row	Label	Cov1	Cov2	Cov3	Cov4
1	logCmaxModel a	0.01825	0.009707	0.000999	0.001117
2	logAUCModel a	0.009707	0.01293	0.001121	0.004910
3	logCmaxModel b	0.000999	0.001121	0.01572	0.008310
4	logAUCModel b	0.001117	0.004910	0.008310	0.009538

.

APPENDIX J: Asymmetry of PBE

Table 22 Asymmetric effect of the reference and test correlations on true multivariate PBE criterion Bp's relation to θ_0

ρR	ρT											
	-0.2	-0.1	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
-0.2	+	+	+	+	+	+	+	+	+	+	+	+
-0.1	+	+	+	+	+	+	+	+	+	+	+	+
0	=	=	=	=	=	=	=	=	=	=	=	=
0.1	-	-	-	-	-	-	-	-	-	-	-	-
0.2	-	-	-	-	-	-	-	-	-	-	-	-
0.3	-	-	-	-	-	-	-	-	-	-	-	-
0.4	+	+	-	-	-	-	-	-	-	-	-	-
0.5	+	+	+	-	-	-	-	-	-	-	-	-
0.6	+	+	+	+	+	-	-	-	-	-	-	-
0.7	+	+	+	+	+	+	+	-	-	-	-	-
0.8	+	+	+	+	+	+	+	+	+	-	-	-
0.9	+	+	+	+	+	+	+	+	+	+	+	-

$+: \theta_p > \theta_0,\ =: \theta_p = \theta_0,\ -: \theta_p < \theta_0,$

θ_0 is θ under no correlation in both reference and test, $\theta_0 = 3.4897$

This table was generated by comparing θ_p which is calculated according to the row and column correlation to θ_0

It shows that asymmetry with respect to the correlation between the reference and between the tests.for example 0.8-0.9 combination is on the other side of than 0.9-0.8 combination.

Table 22 was generated by comparing θ which is calculated according to the row and column correlation to θ_0. If θ was greater than θ_0 a '+' was entered in the cell. If it was smaller a '-' is entered, and if they were equal an '=' sign was entered. The table was color coded according to the signs. This displays the asymmetry of θ with respect to the correlations between the reference and between the tests. For example note the signs of 0.8-0.9 combination, although the correlations are both large, and they are close in value, the sign comparing an θ to θ_0 is the reverse of the 0.9-0.8 combination. These results agree with Figure 17 and Figure 18 and emphasize the importance of investigating how to define BE in terms of the correlation, and the need for regulatory guidelines.

.

APPENDIX K: Effect of Reference and Test Correlations on θ

Figure 17: Effect of ρ_R and ρ_T on the rule θ

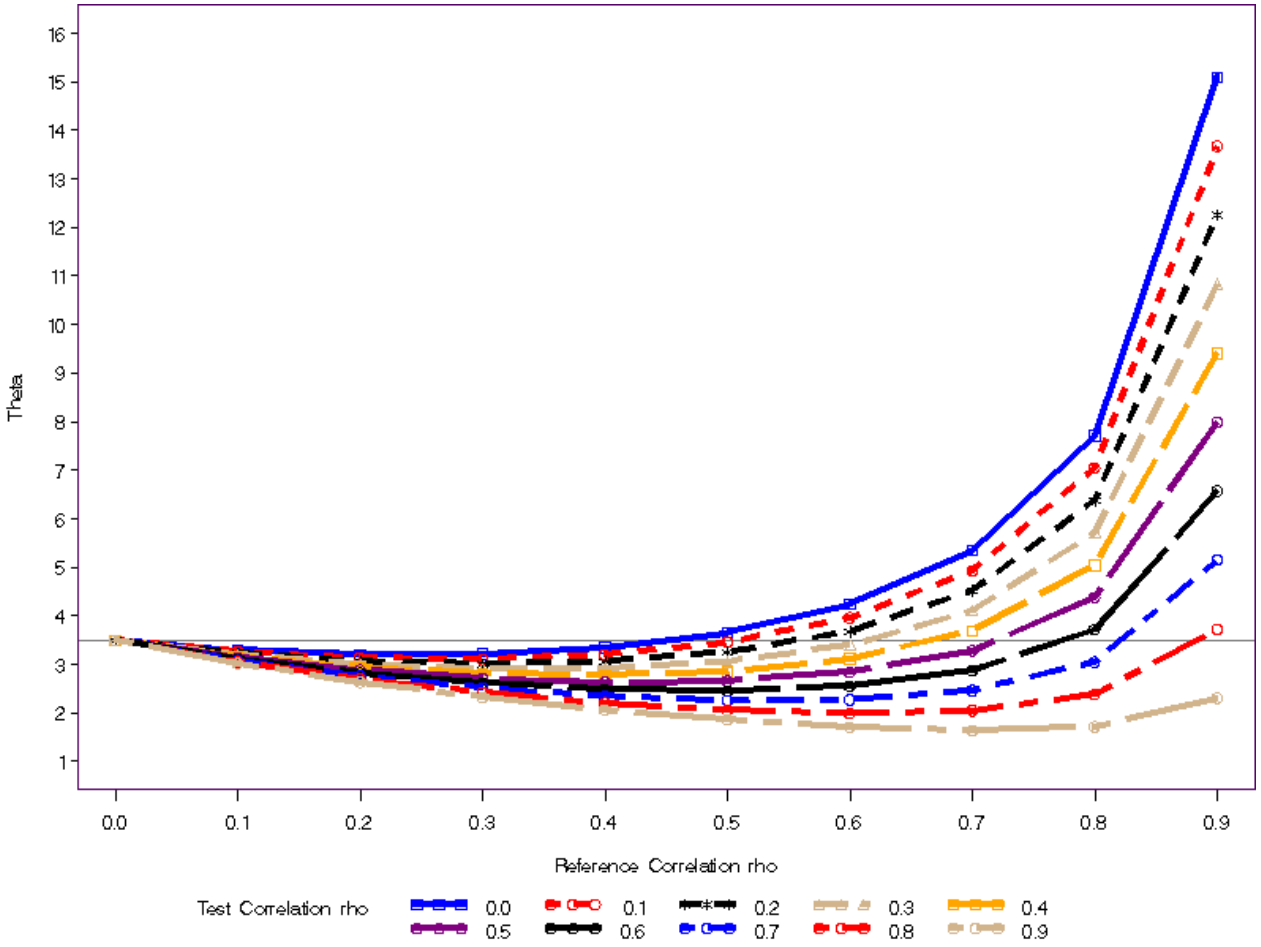
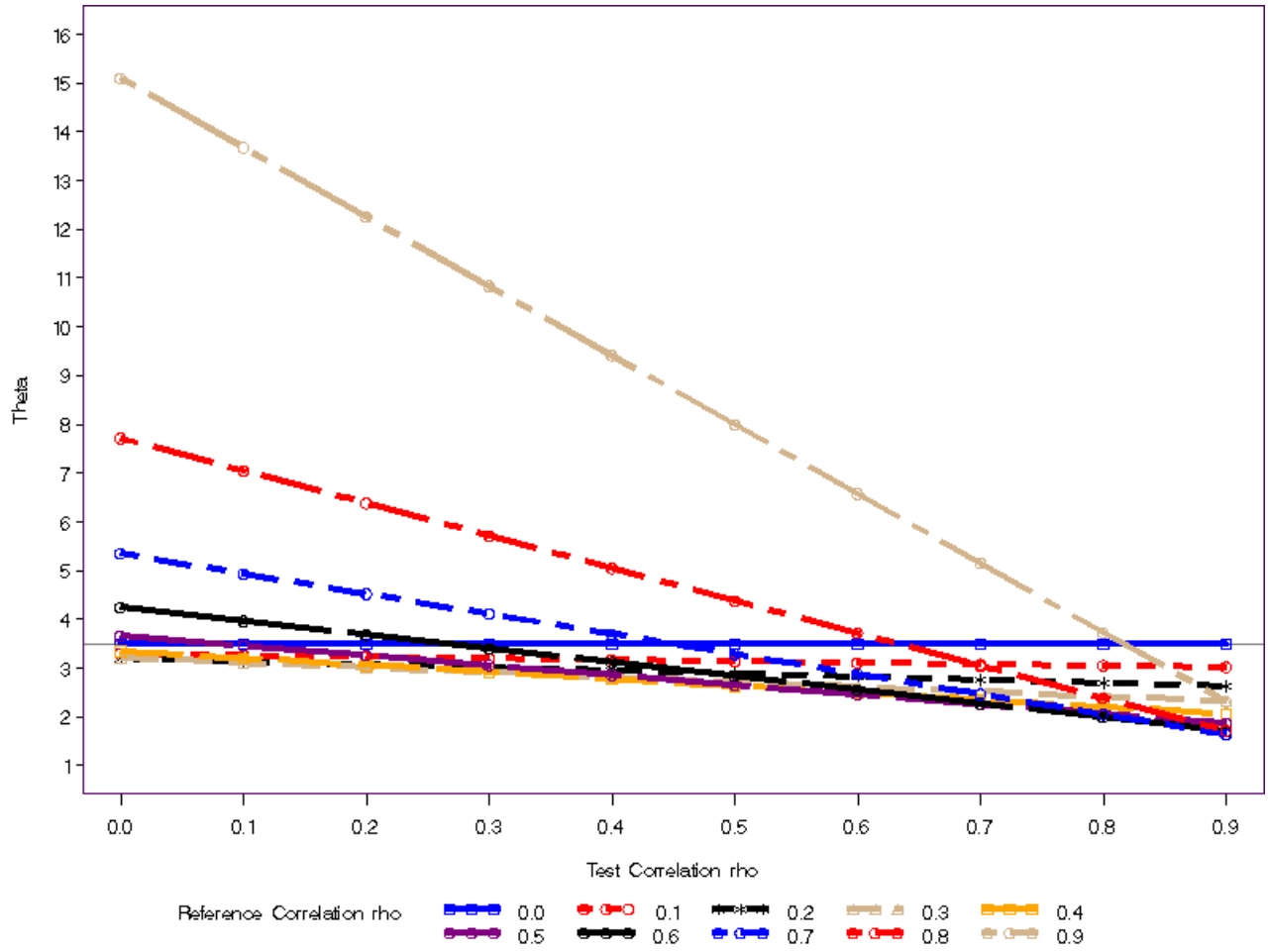


Figure 18: Effect of ρ_R and ρ_T on the rule θ



.

APPENDIX L: Direct Product AR(1) Covariance Structure

. Due to the limited software ability to choose the covariance structure of Bivariate Mixed model in SAS, we selected the Direct product AR(1) structure, which is:

Direct Product AR(1): UN@AR(1)

$$\begin{pmatrix} \sigma_1^2 & \sigma_{21} \\ \sigma_{21} & \sigma_2^2 \end{pmatrix} \otimes \begin{pmatrix} 1 & \rho & \rho^2 & \rho^3 & \rho^4 & \rho^5 \\ & 1 & \rho & \rho^2 & \rho^3 & \rho^4 \\ & & 1 & \rho & \rho^2 & \rho^3 \\ & & & 1 & \rho & \rho^2 \\ & & & & 1 & \rho \\ & & & & & 1 \end{pmatrix}$$

$$= \begin{pmatrix} \sigma_1^2 & \cdots & \rho^5\sigma_1^2 & \sigma_{21} & \cdots & \rho^5\sigma_{21} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \rho^5\sigma_1^2 & & \sigma_1^2 & \rho^5\sigma_{21} & \cdots & \sigma_{21} \\ \sigma_{21} & \cdots & \rho^5\sigma_{21} & \sigma_2^2 & \rho & \rho^5\sigma_{21} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \rho^5\sigma_{21} & \cdots & \sigma_{21} & \rho^5\sigma_2^2 & \cdots & \sigma_2^2 \end{pmatrix}$$

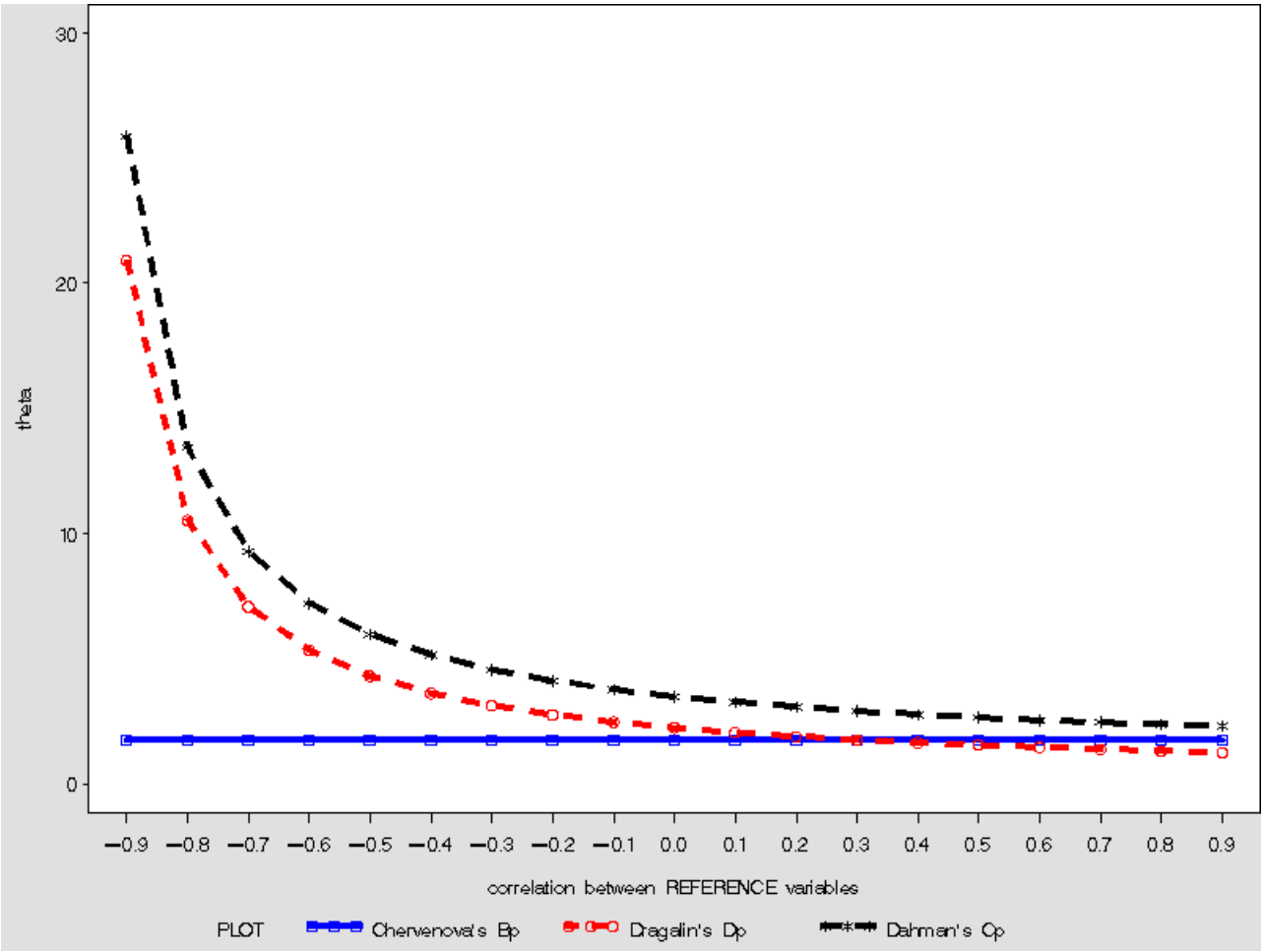
This structure assumes a constant correlation between any two measures taken at consecutive periods.

APPENDIX M: calculated θ 's for the three MV PBE criteria

rhoR	rhoT	theta	Chervo_theta	_theta1	_theta2	KLD_theta
-0.2	-0.2	4.11207	1.74483	1.74483	1.74483	2.76005
0.0	-0.2	3.48965	1.74483	1.74483	1.74483	2.47663
0.2	-0.2	3.32471	1.74483	1.74483	1.74483	2.42193
0.4	-0.2	3.63547	1.74483	1.74483	1.74483	2.60509
0.6	-0.2	4.80603	1.74483	1.74483	1.74483	3.21815
0.8	-0.2	9.04981	1.74483	1.74483	1.74483	5.36781
-0.2	0.0	4.23707	1.74483	1.74483	1.74483	2.61508
0.0	0.0	3.48965	1.74483	1.74483	1.74483	2.24138
0.2	0.0	3.19971	1.74483	1.74483	1.74483	2.09641
0.4	0.0	3.34975	1.74483	1.74483	1.74483	2.17143
0.6	0.0	4.24353	1.74483	1.74483	1.74483	2.61832
0.8	0.0	7.71647	1.74483	1.74483	1.74483	4.35479
-0.2	0.2	4.36207	1.74483	1.74483	1.74483	2.59482
0.0	0.2	3.48965	1.74483	1.74483	1.74483	2.13084
0.2	0.2	3.07471	1.74483	1.74483	1.74483	1.89559
0.4	0.2	3.06404	1.74483	1.74483	1.74483	1.86248
0.6	0.2	3.68103	1.74483	1.74483	1.74483	2.14320
0.8	0.2	6.38314	1.74483	1.74483	1.74483	3.46647
-0.2	0.4	4.48707	1.74483	1.74483	1.74483	2.69345
0.0	0.4	3.48965	1.74483	1.74483	1.74483	2.13125
0.2	0.4	2.94971	1.74483	1.74483	1.74483	1.79779
0.4	0.4	2.77832	1.74483	1.74483	1.74483	1.64860
0.6	0.4	3.11853	1.74483	1.74483	1.74483	1.75522
0.8	0.4	5.04981	1.74483	1.74483	1.74483	2.65736
-0.2	0.6	4.61207	1.74483	1.74483	1.74483	2.99138
0.0	0.6	3.48965	1.74483	1.74483	1.74483	2.30517
0.2	0.6	2.82471	1.74483	1.74483	1.74483	1.84770
0.4	0.6	2.49261	1.74483	1.74483	1.74483	1.55665

rhoR	rhoT	theta	Chervo_theta	_theta1	_theta2	KLD_theta
0.6	0.6	2.55603	1.74483	1.74483	1.74483	1.46336
0.8	0.6	3.71647	1.74483	1.74483	1.74483	1.91858
-0.2	0.8	4.73707	1.74483	1.74483	1.74483	3.97773
0.0	0.8	3.48965	1.74483	1.74483	1.74483	3.05772
0.2	0.8	2.69971	1.74483	1.74483	1.74483	2.36646
0.4	0.8	2.20689	1.74483	1.74483	1.74483	1.82375
0.6	0.8	1.99353	1.74483	1.74483	1.74483	1.42078
0.8	0.8	2.38314	1.74483	1.74483	1.74483	1.31928

Figure 19 The upper limit of each of PBE criteria as a function of the correlation



VITA

Bassam Dahman is a US citizen who was born in Gaza/Palestine on 8/5/1959. He graduated for Kuwait University with BSc in Medical Sciences in 1982.

He finished his Masters degree in biostatistics at Department of Biostatistics/School of Medicine/VCU on 2007.

He has been working with the Division of Quality Health Care in the Department on Internal Medicine at VCU since 2003 as an analyst. He collaborated with several departments in the School of Medicine and the Department of Health Administration in School of Applied Health.

He is now an Instructor at the Department of Public Health Management and policy at the School of Medicine at Virginia Commonwealth University.