Bioequivalence Tests Based on Individual Estimates using Non-compartmental or Model-Based Analysis

School of Statistics and Actuarial Science

By

Mzamo Makulube

Student Number: 853536

Supervisor

Renette Krommenhoek

Submitted in partial fulfilment of Mathematical Statistics Masters by Coursework and Research Report.

04-11-2019

DECLARATION:

I, Mzamo Makulube, declare that this Research Report is my own unaided work. It is being submitted for the Degree of Master of Science at the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at any other University.

Signature: _____

Date: _____

ABSTRACT

The growing demand for generic drugs has led to an increase in the generic drug industry. As a result, there has been a growing demand for bioequivalence studies. The challenges with the bioequivalence studies arose with the method used to quantify bioavailability. Bioavailability is commonly estimated by the area under the concentration-time curve (AUC), which is traditionally estimated by Non-Compartmental Analysis (NCA) such as interpolation in aid of the trapezoidal rule. However, when the number of samples per subject is insufficient, the NCA estimates may be biased and this can result in incorrect conclusions about bioequivalence. Alternatively, AUC can be estimated by the Non-Linear Mixed Effect Model (NLMEM).

The objective of this study is to evaluate bioequivalence on lnAUC estimated by using a NCA approach to those based on the lnAUC estimated by the NLMEM approach.

The NCA and NLMEM approaches are compared on the resulting bias when the linear mixed effect model is used to analyse the lnAUC data estimated by each method. The methods are evaluated on simulated and real data. The 2x2 crossover designs of different sample sizes and sampling time intensities are simulated using two null hypotheses. In each crossover design, concentration profiles are simulated with different levels of between-subject variability, within-subject variability and residual error variance.

A higher bias is obtained with the lnAUC estimated by the NCA approach for trials with a limited number of samples per subject. The NCA estimates provide satisfactory global Type-I-error results. The NLMEM fails to distinguish between the existing formulation differences when the residual variability is high.

KEYWORDS

Average bioequivalence, bioequivalence, bioequivalent, generic (test) drug, intermediate design, non-compartmental analysis, non-linear mixed effects model, pharmacokinetic parameters, reference drug, rich design, sample size, simulation, sparse design.

ACKNOWLEDGEMENTS

I would like to thank my supervisor Renette Krommenhoek for all her guidance and assistance rendered to me.

I thank my colleagues Mr Heyns Kotze and Mr Morries Chauke for their guidance while compiling this report and for helping me to understand some of the statistical techniques.

I thank Dr Josef Makanda for his guidance and opinions while compiling this report.

I thank my siblings and my mother Mrs Elizabeth Makulube for their support throughout my studies.

Thanks are also due to my partner Gugulethu Ndabeni and my friends for encouraging me to keep going.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

ABE Average Bioequivalent

Act 21CFR Code of Federal Regulations

AIC Akaike Information Criterion

ANOVA Analysis of Variance

AUC Area Under the Concentration-time Curve

AUC_{0-t} Area Under the Concentration-time Curve from 0 to t

$AUC_{0-\infty}$ Area Under the Concentration-time Curve from 0 to infinity

BIC Bayesian Information Criterion

BSV Between-Subject variability

cAIC Conditional AIC

CI Confidence Interval

C_{max} Maximum Concentration

EBE Empirical Bayes Estimates

EM Expectation Maximisation

EMEA European Medicines Evaluation Agency

IBE Individual Bioequivalence

IWRES Individually Weighted Residuals

FDA Food and Drug Administration

LMEM Linear Mixed Effects Model

LS means Least squares means

MCCSA Medicines Control Council of South Africa

NCA Non-Compartmental Analysis

NLMEM Non-Linear Mixed Effects Model

NPDE Normalised Prediction Distribution Errors

OFV Objective Function Value

PBE Population Bioequivalence

pcVPC Prediction Corrected Visual Predictive Check

PK Pharmacokinetic

RMSE Root Mean Square error

PWRES Population-Weighted Residuals

SAEM Stochastic Expectation Maximisation

VPC Visual Predictive Check

TIER Test of Individual Equivalence Ratios

T_{max} Time taken to reach Maximum Concentration

TOST Two One-Sided Test

WSV Within-Subject variability

CHAPTER 1 INTRODUCTION

Chow and Liu (2004) argue that the growing demand for generic drugs is leading to an increase in the generic drug industry. As a result, there is a growing demand for bioequivalence studies. Bioequivalence studies compare the bioavailability of two drugs or two formulations of the same drug. The term bioavailability refers to biological availability (Metzler and Huang, 2009). According to Jones and Kenward (2003), the United States of America (USA) was the first country to carry out bioequivalence studies. The definition of bioequivalence is derived from *Act 21CFR (Part 320.1, 2009)*. The United States Food and Drug Administration (USFDA) adopted this definition. According to *Act 21 CFR (Part 320.1, 2009),* "bioavailability is the rate and extent to which the active drug ingredient from a drug product is absorbed and becomes available at the site of drug action". When two drugs or formulations are deemed bioequivalent they are expected to be therapeutically equivalent or to have a similar therapeutic effect and can be used interchangeably (Chow and Liu, 2004). Alternatively, if two drugs are pharmaceutically equivalent or alternatives, then they are considered bioequivalent. Pharmaceutical equivalence suggests that the drugs have a similar amount of the same active ingredient.

According to Birkett (2003), bioequivalence studies are mostly conducted to parallel treatments derived from the same drug: These are the reference (R) treatment (the already approved drug that is on the market) and the test treatment (T). The T treatment is a different formulation of the R treatment, which is not yet approved and is usually a cheaper version of the R treatment. The T treatment is commonly referred to as a generic drug (Birkett, 2003).

Bioequivalence studies have become more attractive in drug industries because they are cheaper to conduct and can be done in a short space of time. As a result, they make medicine more affordable. This is because under the Medicines Control Council of South Africa (MCCSA), for an applicant to register a generic drug or new formulation of the reference drug in their New Drug Application (NDA) submission, they must have proof of quality, efficacy and safety of the new formulation (Department of Health, 2006). The NDA must also include the comparison of the risk benefits, efficacy, and safety of the new and reference drug. However, the common obstacles with the NDA submission are the cost and time required to obtain comparative results. These are often acquired through expensive long-term experiment such as Phase I to Phase III clinical trials. However, with the introduction of bioequivalence

studies, manufacturers are able to obtain these results at a lower cost and in a short space of time. That is, if the manufacturer can present the proof of bioequivalence of the new drug to the reference drug, then the new drug can be accepted as an alternative to the reference.

In South Africa, the MCCSA regulates the generic drug industry. One of the functions of the MCCSA is to ensure that all medicines used by the state are safe, therapeutically effective and adhere to the regulatory standards (MCCSA, 2003). The MCCSA (2003) has put into place the policies, guidelines and methods to be followed when conducting bioequivalence studies. However, there are some issues with the methods used to estimate the measures of bioavailability. There are three pharmacokinetic (PK) parameters used to measure bioavailability: the area under the concentration-time curve (AUC), maximum concentration (C_{max}) of the drug and time taken to reach maximum concentration T_{max} (Chow, Shao and Wang, 2003). C_{max} is observed from the data and T_{max} is the time where C_{max} occurred. The PK parameter AUC is traditionally estimated by non-compartmental analysis (NCA) such as interpolation in aid of the trapezoidal rule (Yen and Kwan, 1978). Though, if the information at the individual level is sparse NCA approach yields biased estimates (Korkmaz and Orman, 2012). On some occasions, the approach may result in missing values. In the sampling phase of the experiment, it is possible that some values are recorded incorrectly or unexpected values are obtained which results in missing values (Chow and Liu, 2004). When the number of missing values is large, the bias in the estimate of AUC could be large, which consequently could affect the result of the study or lead to incorrect conclusions (Chow and Liu, 2004).

Alternatively, PK parameters can be estimated by the use of a non-linear mixed effect model (NLMEM). In this method, the NLMEM is fitted on the concentration data, and its parameter estimates are then used to calculate the PK parameters. The NLMEM method is more flexible in explaining the repeated measured data (Gabrielsson and Weiner, 2001). The NLMEM have proven to be powerful in sparsely sampled data. These are cases where there are difficulties in attaining measurements for the target population such as children, elderly people or for those whom the stage of the disease is acute (Hu, Moore, Kim and Sale, 2004; Mentré, Dubruc and Thénot, 2001).

This study compares bioequivalence results obtained using the NCA and the NLMEM methods. This study is conducted on a well-known antibiotic used in pigs. To compare the two methods, bioequivalence analyses are conducted on the AUC estimated by the NCA method

and on *AUC* estimated by the NLMEM method. The magnitude of the resulting errors in bioequivalence analysis is then used to compare the results. To evaluate the two methods beyond available data in this study, data sets of different sample sizes and sample sizes per subject are simulated. In addition, the bioequivalence test is conducted on simulated data sets to compare the two methods.

The report is structured as follows:

- Chapter 1: Introductory Chapter of the study.
- Chapter 2: Description of The NCA and NLMEM methods.
- Chapter 3: Description and details of the data set used for this study as well as the simulation procedure.
- Chapter 4: Results.
- Chapter 5: Conclusion and Recommendation.

1.1 Problem Statement

There are several studies which used the *AUC*, estimated by the NLMEM method, to determine bioequivalence and compared the results to bioequivalence based on the NCA estimates of *AUC* (Panhard and Mentre, 2005; Panhard, Taburet, Piketti and Mentré, 2007; Dubois, Gsteiger, Pigeole and Mentré, 2010; Mentré, Dubruc and Thénot, 2001). These studies have indicated the NLMEM to be more advantageous than the NCA methods especially in the case of sparse data. However, these studies have all used the same anti-asthmatic Theophylline drug data set, which is administrated orally. The Theophylline drug data set is usually used for software testing (Panhard and Mentre, 2005). With the availability of the new data set for a different drug, administrated intramuscular, it is of great importance to test the two methods on a different data set and compare the results with previous studies. Furthermore, the use of the NLMEM estimates in bioequivalence studies with reduced sample size is still rare; most often, it is used when the sample size is large especially in phase III trials. The phase III trial is the third step of the clinical trial. The phase III trial is designed to compare the clinical usefulness and the safety of the new treatment with standard treatment or to compare different dosing regimens of the standard treatment (Friedman, Furberg, DeMets, Reboussin and Granger, 2015). The phase III trial usually involves a bigger number of subjects to be able to detect differences in treatments; they may involve as many as hundreds of subjects.

1.2 Aims and Objectives

1.2.1 Aims

The aims of the research were to:

- Determine which approach is more precise when variability in the data is high or low;
- Determine which method between the NCA and NLMEM is more appropriate to use when the sample per subject is small; and
- Pinpoint when the NCA method should be avoided.

1.2.2 Objectives

The objective of this study was to compare bioequivalence tests using the NCA approach with the NLMEM approach, using simulated concentration data and real data.

1.3 The Specific Research Questions Addressed in this Research Report

The following research questions are addressed in this research report:

- How does the variability level affect NCA and NLMEM estimation precision?
- How does sample per subject affect estimation precision?
- When is the NCA favourable to the NLMEM approach?
- How small should the sample per subject be for NCA method to fail?

CHAPTER 2 LITERATURE REVIEW

2.1 Introduction

In establishing bioequivalence, the method suggested in FDA (2013) guideline includes information on pharmacokinetics, pharmacodynamics, clinical, and in vitro studies. Researchers must select the most accurate, sensitive, suitable and reproducible studies. For drugs intended for system absorption and circulation, the FDA (2013) recommends the use of PK bioequivalence based studies to compare different drug formulations. The PK parameters are sensitive to formulation differences, making them ideal for bioequivalence analysis. In bioequivalence studies, C_{max} , $AUC,$ and T_{max} are considered as a valid surrogate for clinical efficacy, safety, and they are mostly used for bioequivalence analysis (Morais and Lobato, 2010). This section covers bioequivalence and the methods used to test bioequivalence and introduces the NCA and NLMEM methods.

2.2 Bioavailability

Bioequivalence studies are designed to measure the rate and extent to which the active ingredient or active moiety is absorbed from a drug product. According to the FDA (2003) for drugs that are not intended to be absorbed into the blood stream, 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. The bioavailability studies are commonly carried out on healthy subjects. As a result, the chosen design and the statistical method must be well suited for a particular study. The relationship between the two aspects is of great importance, since the statistical method to be used is detected by the design of choice. Furthermore, to obtain valid conclusions, data must be collected using a design that is scientifically sound, and uses statistical methods that are fitting for the design.

Chow and Liu (2004, 31) pointed out three questions that need to be answered before a bioavailability study design can be carried out:

- “What is to be studied, i.e. what are the study objectives”?
- “How are the data to be collected, i.e. what design is to be employed”?

- "How are the data to be analysed, i.e. what statistical methods are to be used"?

2.3 Pharmacokinetics and Pharmacodynamics Studies

Pharmacokinetics involves understanding how the body affects the active ingredient from the point of administration up to the point of elimination (Gabrielsson and Weiner, 2001), by mathematically modelling the time course of drug concentrations in body fluids. Pharmacodynamics focuses on modelling the effect of the drug to the body, whether human or experimental animal. Greenblatt, Moltke, Harmatz, and Shader (1998) argues that the pharmacokinetics and pharmacodynamics methods have been applied in the course of new drug development as well as for enhancing understanding of the clinical activities of drugs that are at present in the market. In pharmacokinetics studies, the main objective is to study the effect of the body to the active ingredient; most often distinguish by absorption, distribution, metabolism, and elimination of the drug after administration. The serum or plasma concentration of the drug is evaluated over time to characterise these PK parameters. Whereas, pharmacodynamics, study the link between the concentration of the drug at the site of action and its resultant effects on the body (Schwinghammer and Kroboth, 1988). The two techniques pharmacokinetics and pharmacodynamics have been improved on over time and become more sensitive and accurate in measuring treatment effects and concentration. The improvement of these techniques goes hand in hand with the improvement of the kinetic-dynamic modelling, where time is integrated in the effect to concentration model. Figure 2.1 illustrates the relation between pharmacokinetics, pharmacodynamics, and kinetic-dynamic modelling. The concentration effect relationship illuminates the clinical rationality of reading concentration from the plasma or serum. In clinical psychopharmacology, kinetic-dynamic studies involves quantitation of the drug concentration and clinical effect at different sampling times after dosing, commonly under placebo-controlled (double blind laboratory conditions) (Greenblatt et al., 1998).

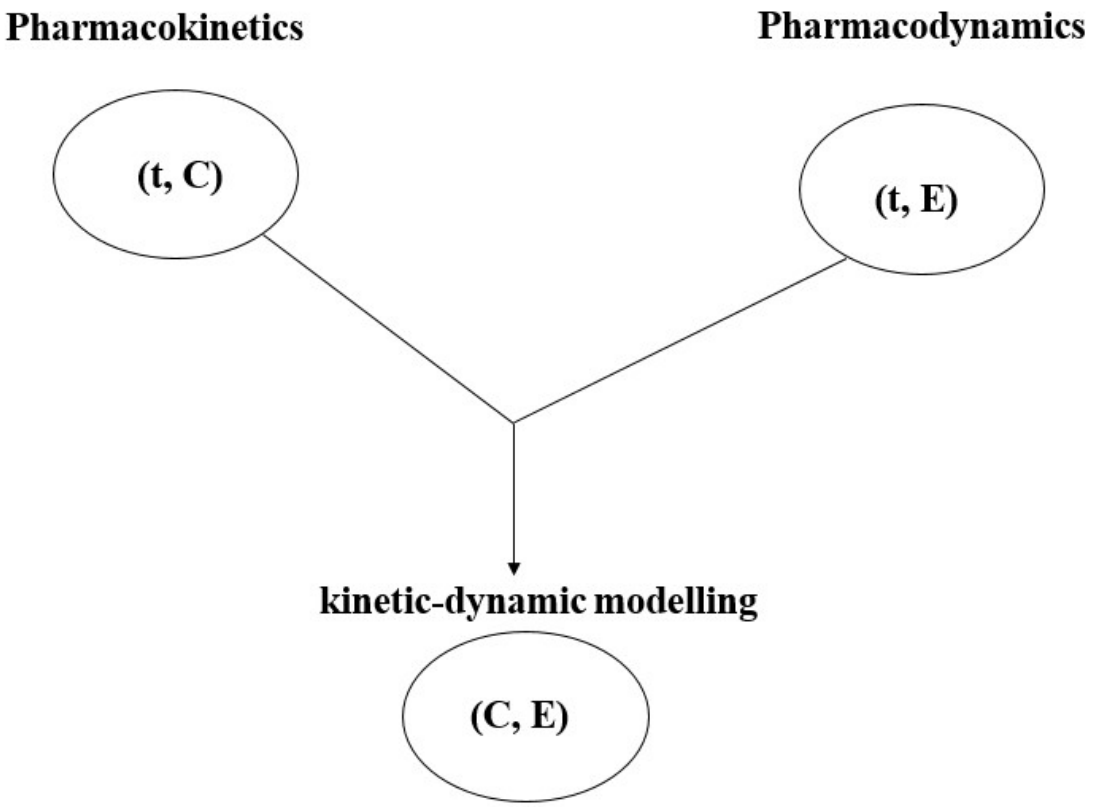


Figure 2.1: Relationship between kinetic-dynamic modelling, pharmacokinetics and pharmacodynamics.

Figure 2.1 illustrates the association between pharmacokinetics, pharmacodynamics and kinetic-dynamic modelling that was adapted from Greenblatt et al. (1998). In the diagram, C, E and t respectively represent concentration, effect and time. The kinetic-dynamic modelling involves pharmacokinetics and pharmacodynamics, and time integrated into the relationship of effect and concentration.

The sigmoid, exponential and linear model are common models used to explain the link between pharmacodynamics effects and concentration. These three models are as follows (Greenblatt et al., 1998):

Sigmoid $E_{\max}$:

$$E = \frac{E_{\max} C^A}{C^A + EC_{50}^A};$$

Exponential: $E = mC^A$;

Linear: $E = mc$.

Where $E_{\max}$ is maximum pharmacodynamics effect, EC_{50} is the concentration producing E that is 50% of $E_{\max}$, A is the exponent, and m is a slope factor. The strength of the concentration response relationship reflects in the exponent A. Greenblatt et al. (1998, 510) say, *"* $E_{\max}$ and EC_{50} values allow inferences about questions such as the relative potency or efficacy of drugs producing the same clinical effect, individual differences in drug sensitivity, the mechanism of action of pharmacologic potentiators or antagonists, and the possible clinical role of new medications*"*. However, the sigmoid $E_{\max}$ may not always explain the concentration effect relationship in every study. Simpler models, such as exponential or linear equations might be suitable for some data.

2.4 PK Parameters

The collection of plasma or biological fluid samples from subjects must be timed in a manner that will fully characterise the concentration-time profiles of the R treatment. For fed treatments, eating may have some influence in the course of plasma drug concentrations, whereas in the fasted treatment the food does not alter with the time course. Hence, the scheduling of the sampling times should be planned accordingly (Niazi, 2014).

The PK parameters are estimated from the resulting plasma concentration-time curve and are classified according to full exposure, peak exposure, and time-to-peak exposure (Niazi, 2014). The PK parameters are:

- The total exposure measured (AUC_{0-k}) the AUC from time 0 (pre-dose) to time *k* (the last quantifiable drug concentration sample),
- The peak exposure measured or maximum concentration (C_{max}),
- The time taken to reach maximum concentration or the time-to-peak exposure $T_{max.}$

Prior to the analysis, the FDA (2013) recommends that the exposure measure AUC_{0-k} or C_{max} to be log-transformed. That is because the distribution of the PK parameters (C_{max} and AUC_{0-k})

are both skewed. By applying the log-transformation to these parameters, the resulting log-transformed C_{max} and $AUC_{0\text{-}k}$ are normally distributed (Patterson and Jones, 2005).

2.4.1 Estimation of PK Parameters

The PK parameter AUC is traditionally estimated by NCA method or indirectly by fitting a NLMEM on the concentration data.

2.4.2 Non-Compartmental Approach (NCA)

Amongst available methods, the interpolation using the trapezoidal method is customarily used in estimating AUC. This method is most favoured due to the computational simplicity and the concept behind it. The trapezoidal method is ideal when the data to be compared share a common linear shape and sampling scheme. The linear trapezoidal is used to estimate AUC from time to time t_k, denoted by AUC_{0-k}. To describe the linear trapezoidal, let $c_0, c_1, \ldots, c_k$ represent the oberved concentration values at time $t_0, t_1, \ldots, t_k$, respectively. The linear trapezoidal rule estimate AUC_{0-k} as follows:

$$AUC_{0-k} = \sum_{i=2}^{k} \frac{c_{i-1} + c_i}{2}(t_i - t_{i-1}) . \tag{2.1}$$

After determining time t_k, a significant fraction of the drug concentration may be present and therefore, to capture the full exposure, AUC must be estimated from time zero to time infinity (Martinez and Jackson, 1991). The AUC from time zero to time infinity can be estimated as follows (Rowland and Tozer, 2005):

$$AUC_{0-\infty} = AUC_{0-k} + \frac{c_k}{\lambda}, \tag{2.2}$$

where λ is the rate of decrease in concentration per unit time, which is obtained by regressing on log-linear terminal part of the concentration-time curve, then λ is the regression slope multiplying by -2.303 (Chow and Liu, 2004, Jones and Kenward, 2003). The maximum

concentration obtained as follows, $C_{\max} = \max\{_0, c_1, \ldots, c_k\}$. Similarly, T_{max} is the time where $C_{\max}$ is observed.

The drawback with the traditional NCA approach is that to estimate *AUC* with accuracy, the information at the individual level must be dense. Most often in sparsely sampled individuals, it is difficult to obtain such information (Hu et al., 2004). The NCA method uses linear regression to estimate the terminal slope, assuming that the terminal part of the concentration-time curve is linear. However, for drugs with non-linear elimination (particularly at the terminal phase) the NCA approach may poorly estimate $AUC_{0-\infty}$ (Panhard and Mentre, 2005; Gabrielsson and Weiner, 2001). Hence, the NLMEM have been proposed as a powerful alternative in modelling sparse repeated measured data or for estimating PK parameters (Wu and Wu, 2002).

2.4.3 Non-linear Mixed Effects Model (NLMEM) Approach

For the NLMEM approach the concentration y_{ijk} for the *ith* subject $i(i = 1.2, \ldots, N)$ at sampling time t_{ijk}, $j(j = 1, \ldots, n_i)$ on treatment $k(k = T, R)$ can be modelled by the following structure model (Lavielle and Mentre, 2007):

$$y_{ijk} = f(t_{ijk}, \phi_{ik}) + g(t_{ijk}, \phi_{ik})\varepsilon_{ijk};\ \varepsilon_{ijk} \sim N(0, \sigma^2) \tag{2.3}$$

Where f and g are parametric functions of the structure model and the residual error model respectively, $\phi_{ik} = (\phi_{ikl}; l = 1, \ldots p)'$ represent the vector of PK parameters for a subject in treatment k and ε_{ijk} is an error in measurements of the *ith* observation for *ith* individual in treatment k. The parameters are assumed to be random vectors, in such a way that for the *kth* treatment and in the absence of covariates they can be broken down as follows:

$$\phi_{ik} = \mu_k e^{b_{ik}},\ b_{ik} \sim N(0, \Omega). \tag{2.4}$$

Where μ_k is the average response for treatment k (fixed effect), b_{ik} denotes a vector of random effects for the ith individual on treatment k. To avoid any negative parameter estimates the parameters ϕ are modelled in a log scale (Panhard and Mentre, 2005). Then the model is:

$$\log \phi_{ik} = \log \mu_k + b_{ik}. \tag{2.5}$$

A global fit is obtained by including a treatment effect β on a component/s on a vector of fixed effects. When the global analysis is conducted, another component c_{ik} is incorporated to ϕ_{ik}, this is done to model the intra-subject variability. c_{ik} is assumed to be independently and normally distributed with a mean zero and variance ψ. In this case, a variance of ϕ_{ik} is thus a combination of Ω and ψ. Therefore the global model is:

$$\log \phi_{ik} = \log \mu_k + \beta_k + b_i + c_{ik},\ b_{ik} \sim N(0,\Omega), c_{ik} \sim N(0,\psi). \tag{2.6}$$

Where $\log \mu_k + \beta_k$ is the average response for treatment k (fixed effect), b_i is $p-vector$ of random effects for the ith individual, c_{ik} represents a random effects $p-vector$ for the ith subject on the kth treatment. For treatment R, $\beta_R = 0$, as a result, μ_k is the mean effect of R formulation, hence, $\log \mu_T = \log \mu_k + \beta_k$ is the mean effect of the T formulation. In addition, β_k represents the difference between treatment R and T. The distribution of the individual PK parameters $\phi_i = (\phi_{iR}, \phi_{iT})$ of the subject i is assumed Gaussian with a mean $(\mu, \mu + \beta_T)$ and the variance covariance matrix Γ (Panhard and Samson, 2009).
Where:

$$\Gamma = \begin{pmatrix} \Omega+\Psi & \Omega & \dots & \Omega \\ \Omega & \Omega+\Psi & \ddots & \vdots \\ \vdots & \ddots & \ddots & \Omega \\ \Omega & \dots & \Omega & \Omega+\Psi \end{pmatrix}.$$

Here $\Omega+\Psi$ is the sum of the two matrices Ω and Ψ. When introducing the categorical covariates to the fixed effect component in Equation (2.5) the lth parameter can be rewritten as:

$$\log \mu_{ikl} = \log \mu_{Rl} + \beta_{T,l} T_{ik} + \beta_{p,l} P_k + \beta_{s,l} S_i, \tag{2.7}$$

where μ_{ikl} is the fixed component, μ_{Rl} is R formulation mean for the *lth* parameter, T_{ik} is the formulation, P_k is the period and S_i is the sequence, $\beta_{T,l}, \beta_{p,l}$, and $\beta_{s,l}$ are the effects of the covariates T_{ik}, P and S_i respectively. The covariate takes a value of zero for reference and one for the test treatment.

The residual error model g can take three different forms, such as the constant, proportional and the combined residual error model (Buonaccorsi, 2010).

- In the constant residual error model $g = p_i$ where p_i is a constant,
- For proportional residual error model, g is proportional to f, such that for some scalar $g = q_i f$,
- The combined error model can be defined as $g\left(t_{ij}, \phi_i\right) = \left(p_i + q_i f\left(t_{ij}, \phi_{ij}\right)\right)^2$.

where ϕ_i is the vector of the individual's parameters and ϕ_{ij} is the vector of PK parameters for the *ith* individual at the *jth* sampling time $\left(t_{ij}\right)$. The combined residual error model incorporates both constant and proportional forms. Which makes it more advantageous to both the constant and proportional residual error model (Comets and Mentre, 2001).

For sparse data, the estimation procedures for NLMEM are based on a linearisation of the likelihood, because of the small sample size and larger inter-individual variation (Davidian and Giltinan, 1995). The Lindstrom and Bates (1990) algorithm is the most commonly used linearisation procedure in pharmacokinetics studies. However, the likelihood approximates based on the linearisation procedures may significantly deviate from the true likelihood. As such, inferences based on the linearisation procedures may not be reliable, and the parameter estimator is not a true maximum likelihood estimate (Vonesh and Chinchilli, 1996). Furthermore, the linearisation procedures have been shown to have poor convergence, poor

quartiles of parameter estimates and often fail to estimate some regression parameters, when the data is sparse.

Kuhn and Lavielle (2005) proposed a Stochastic Expectation Maximisation (SAEM) algorithm to estimate the parameters of the NLMEM. The SAEM is the stochastic version of the expectation-maximisation (EM) algorithm, such that the expectation step in the EM algorithm is replaced with a single iteration of a stochastic approximation procedure (Delyon, Lavielle and Moulines, 1999). Historically the EM algorithm was developed for modelling data with missing values. The EM consists of two steps: (a) The E step for computing the conditional expectation for the complete data, and (b) the M step that evaluates the maximum likelihood of the complete data (Dempster, Laird and Rubin, 1977). When applied to the NLMEM the random effects are treated as missing data, and the E step is replaced with the stochastic approximation. In contrast to the linearisation approach, the SAEM algorithm uses the exact likelihood instead of the likelihood approximation, which improves the convergence of the parameters and typical maximum likelihood inferences are easily obtained directly from the likelihood. Therefore, this study uses the SAEM algorithm to estimate the parameters of the NLMEM, more details of this algorithm are in Appendix A.

2.4.4 The Evaluation of the NLMEM Model

Intuitively the term itself indicates that the model must be assessed for its ability to predict the observed data, the data it was trained on. However, model evaluation is a complex task, because some models are only developed to understand the underlying phenomena, and some are developed for predicting the future outcome or under new conditions. A clear understanding of the purpose of the model being developed helps in setting up a clear path on what tools to use for model evaluation (Comets, Brendel and Mentre, 2010). The steps to model evaluation may begin with model diagnostics, where candidate models are selected from the pool of possible models. Most techniques of model diagnostics are usually visual; such as the individual fits assessment, the plot of observed data against the model, and the plot of the residual against the predicted values. From the remaining candidates, the model selection tools are then used to pick the ideal model.

Unlike with the model diagnostics, the model selection tools are analytical. A calculated criterion is used to compare the candidate models. A common technique in model evalution is

to divide the data into three subsets; the training subset, where the model is fitted, the subset for validation, and test subset, for assessing superiority of the selected model. Alternatively, model selection can be done internally, where the similar data used to fit the model is used for evaluation (Lavielle, 2014).

2.4.5 Model Diagnostics

Individual Fits

The individual fit are used to assess the inter-subject variability, because the SAEM algorithm allows the estimation of the individual parameters ϕ_i and population parameters ϕ_{pop}. As such, the average profile can be predicted using population parameters, and individual profiles from individual parameters. For each individual, the predicted population profile is plotted against the predicted individual profiles. If the model represents the data well, the predicted individual profiles are expected to behave like the population profile or vary around the predicted population profile (Lavielle, 2014).

The Plot of the Model against the Observed

The plot of the predicted against the observed values is the simplest and most common method; it offers a view of the model capabilities in representing the observed data. The slope or shape and the intercept of the predicted and observed can be compared to assess the model correctness. A good model should represent the majority of the data, and predict a similar shape to the observed data. No consensus exists on which variable (predicted or observed) should be plotted on each axis. In addition, the choice of the axis produces different evaluation results. However, placing the predicted values in the *y*-axis and observed on the *x*-axis, may lead to an incorrect conclusion about model correctness, because the slope and intercept differ depending on the choice of the axis (Pineiro, Perelman, Guerschman and Paruelo, 2008). Hence, the observed values should be placed on the *y*-axis and the predicted on *x*-axis.

The Residual Plots

Residual plots are frequently created to assess the residual distribution assumptions, where the predicted values (*x*-axis) are plotted against the residuals (*y*-axis). Using this plot, non-linearity, heteroscedasticity, independence, outliers and normality can be assessed (Fox, 1997). In pharmacokinetics modelling the Normalised Prediction Distribution Errors (NPDE) and Individually Weighted Residuals (IWRES) are the common residuals for assessing model

misspecification. The IWRES are based on individual prediction. The IWRES are defined as follows:

$$\mathrm{IWRES}_{ij} = \frac{y_{ij} - f\left(t_{ij}, \widehat{\phi_{ij}}\right)}{g\left(t_{ij}, \widehat{\phi_{ij}}\right)}. \tag{2.8}$$

In case of the correlated residuals, which is common in repeated measured data, the relation can be removed by multiplying each individual vector of residuals by the inverse of the estimated correlation matrix. The NPDE is a non-parametric form of the Population-Weighted Residuals (PWRES) based on the rank statistics (Comets, Brendel and Mentre, 2008), where the PWRES is defined as the normalised residuals of the population prediction. That is if $\mathrm{E}\left(y_i\right) = \mathrm{E}\left(f\left(t_{ij}, \widehat{\phi_{ij}}\right)\right)$ is the population prediction and $y_{\ i} = \left(y_{ij,} 1 < j < n_i\right)$ represent observed values vector. The PWRES for each individual can be defined as follows:

$$\mathrm{PWRES}_{ij} = \frac{y_i - E(y_i)}{\sqrt{\Gamma}}. \tag{2.9}$$

Where $\sqrt{\Gamma}$ is the $n_i \times n_i$ variance covariance matrix for each subject. Let F_{ij} be a cumulative distribution of PWRES_{ij}. By definition F_{ij} is the uniform distribution in the interval [0, 1]. Let $\Phi^{-1}\left(F_{ij}\right)$ be the standard normal distribution of F_{ij}, then NPDE for each subject is defined as $\mathrm{NPDE}_{ij} = \Phi^{-1}\left(\widehat{F_{ij}}\right)$. Where $\Phi^{-1}\left(\widehat{F_{ij}}\right)$ are empirical estimations of $\Phi^{-1}\left(F_{ij}\right)$ by Monte Carlo simulation (Comets et al., 2008). Residuals with inconsistency in amplitude indicate the need for proportional error model, to explain the data properly (Lavielle, 2014). The common and more informative approach is to summarise the distribution of weighted residuals by their empirical percentiles, usually of order 10%, 50% and 90%. To assess the model misspecification the quartiles are plotted with the 90% prediction interval produced by the model for each empirical quartiles. If the model poorly represents the data, the empirical percentiles and their prediction intervals will show discrepancies.

The Empirical Bayes Shrinkage Estimator

The SAEM algorithm in the Monolix software estimates individual parameters, subject predicted values, and residuals during a second "Bayes" estimation stage, by the means of a different objective function (also known as the post hoc, empirical Bayes or conditional estimation step). According to Mould and Upton (2013), this step can use the weighted least squares Bayes objective function, where for subject $i\left(i=1,2,\ldots,N\right)$ with $j\left(j=1,2,\ldots,N\right)$ samples (y_j) and a model with $l\left(l=1,2,\ldots,p\right)$ parameters (ϕ_l), the Bayes Objective Function Value (OFV) can be described as follows (Mould and Upton, 2013):

$$\text{OFV}=\sum_{j=1}^{n_i}\frac{(y_j-\widehat{y_j})^2}{\sigma^2}+\sum_{l=1}^{p}\frac{(\phi_l-\phi_{pop})^2}{\varphi_{l,pop}^2}, \tag{2.10}$$

where $\varphi_{l,pop}^2$ is the variance of the l^{th} population parameter and σ^2 is the residual variance. When using the OFV the best parameters for each subject are estimated by reducing the deviation of each subject's parameter estimate values (ϕ_l) from the population parameter value (ϕ_{pop}) and the subject's predicted values ($\widehat{y_j}$) from the actual values (y_j). In the case of sparsely sampled subjects, Equation 2.10 illustrates that, the population data carry more weight than the individual parameter estimate. Consequently, ϕ_l will shrink towards ϕ_{pop}. The shrinkage on the l^{th} Empirical Bayes Estimates (EBE) (η_l) is:

$$\text{shrinkage}(\eta_l)=1-\frac{\text{var}(\eta_{il})}{\omega_l^2}, \tag{2.11}$$

where ω_l^2 represents the value of the estimated population variance and $\text{var}(\eta_{il})$ represents the empirical variance of the l^{th} individual random effects. Mould and Upton (2013) observed that when the shrinkage is high, the PK parameters are poorly estimated. The poorly estimated parameters may lead to poor conclusions about the relationship between response and exposure rate. Furthermore, model diagnostic using plots of the individual predicted and residual values maybe misleading when the shrinkage is high. Hence, when shrinkage is high model diagnostic should be at the population level, since the population parameters are less affected by

shrinkage. If the shrinkage is higher than 20–30%, model diagnostic at the individual level may be misleading (Xu, Yuan, Karlsson, Dunne, Nandy and Vermeulen, 2012). That is if the shrinkage is higher than 20–30% the individual level diagnostic reflect the population behaviour. Furthermore, more weight is provided by the population mean values when the shrinkage is greater than 50% (Xu et al., 2012).

The Visual Predictive Check (VPC)

The VPC is popular in the NLMEM model development because it allows the diagnostics of both structural and statistical modelling. In many cases, this involves the grouping of the data into bins over time (the independent variable) and computing the empirical quartiles to compare them with the simulated quartiles within bins. The diagnostic works well for single dose experiments. However, when the dose variability is high and in the presence of dominant covariates, VPC is disadvantaged when binning across time. In this case, the Prediction Corrected VPC (pcVPC) is mostly used. In the pcVPC, the observed and simulated values are normalised to eliminate the dose-induced variability in the bins (Bergstrand, Hooker, Wallin and Karlsson, 2011).

Random Effect Diagnostics

For a given individual the random effects are assumed to be independent and normally distributed (Bergstrand et al., 2011). If the data are analysed using a model that comprises of both random and fixed effects, it is necessary to check the assumption compliance of the estimated random effects.

2.4.6 Model Selection

The model selection is based on some calculated criterion, such as the Bayesian Information Criterion (BIC) and the Akaike Information Criterion (AIC). For hierarchal models, the selection can also be based on a Wald test and likelihood ratio test. The BIC and AIC are defined as follows:

$$\mathrm{BIC} = -2\log(L) + P\log(n_{tot}) \tag{2.12}$$

$$\mathrm{AIC} = -2\log(L) + 2P\ . \tag{2.13}$$

Where L denotes the likelihood of the data, P is the number of parameters in the model and n_{tot} is the number of observations used. In the context of NLMEM, it is not clear whether to use N the total number of subjects or to use n_{tot} in the penalty term of the BIC, this is because there is no clear defined effective sample size. Hence the computation of the BIC differs from software to software, i.e. the R package nlme and the SPSS mixed procedure use n_{tot}, whereas the SAS proc NLMIXED procedure and the Monolix software use N (Lavielle, 2014). For studies with small sample sizes, the traditional AIC is inconsistent and may result in overfitting. Hence, the conditional AIC is more appropriate. The conditional AIC (cAIC) can be computed as follows (Vaida and Blanchard, 2005):

$$\text{cAIC} = \text{AIC} + \frac{2P(P+1)}{N-P+1}. \tag{2.14}$$

This method should be used for fitting NLMEM using the data with a small sample size, instead of the traditional AIC (Lavielle and Mentré, 2007).

2.5 Crossover Design

In this study, a two period two sequence crossover (2x2 crossover) design was used for data collection. The 2x2 crossover is the most popular statistical design for average bioequivalence and recommended by the FDA (2013) and the Department of Health (2006). The 2x2 crossover design has been used in many cases for establishing new concepts in bioequivalence studies (Walker, Kanfer and Skinner, 2006). If R and T represent the reference and test formulations respectively the 2x2 crossover design can be described as follows:

- RT represent a sequence from drug R to drug T, and TR represent sequence from drug T to drug R,
- Let PI and PII be the first and second dose periods,
- In a 2x2 crossover each individual is randomly allocated to either RT or TR. For example, a particular subject is assigned to RT, at the first dosing period (PI), the subject will receive drug R, after a suitably chosen time, known as a washout period the subject will crossover to drug T in period PII.

The 2x2 crossover is designed to eliminate any component of variation that is due to between-subject variations, by allowing subjects to crossover from one treatment to another (Jones and Kenward, 2003, Chow and Liu, 2004 and Fleiss, 1989). However, the disadvantages of the 2x2 crossover are the possibility of the presence of the residual effect of the treatment from the first dose period on the second dose period. Such a confounding factor may lead to an ambiguous conclusion about the individual treatment (Jones and Kenward, 2003). Fleiss (1989) advocated that the 2x2 crossover design must be implemented whenever prior information about the disease progression is available, in addition, while the study is in progress on average the disease will not advance dramatically, and when equal or close carry-over effect is expected. Schall (1995) motivated for a 2x3 crossover design for IBE and PBE. However, a 2x3 crossover design requires more samples per subject because of additional periods making it less ethical as compared to the 2x2 crossover, and it runs for a longer period due to prolonged washout period leading to more subject drop out (Oh, Ko and Oh, 2003).

The 2x2 crossover design can be described by the model Chow and Liu (2004):

$$Y_{ijk} = \mu + S_{ik} + P_j + T_{(j,k)} + C_{(j-1,k)} + e_{ijk,} \tag{2.15}$$

where:

Y_{ijk} : is the observed AUC or C_{max} of the i^{th} subject in the k^{th} sequence at the j^{th} period,

μ : is the overall mean of the model,

S_{ik} : is the random sequence effect,

P_j : is the fixed effect of the j^{th} period, where $j=1,...,p$ and $\sum_j P_j = 0$,

S_{ik} : is the direct fixed effect of the formulation effect at j^{th} period in k^{th} sequence.

In the 2x2 crossover design of two formulations, $T_{(j,k)}$ is of the form:

$$T_{(j,k)} = \begin{cases} T_R \; if \; k = j \\ T_T \; if \; k \neq j \end{cases} \quad for \; k = 1,2; j = 1,2.$$

$C_{(j-1,k)}$: is the residual effect carried from the first period (*j-1*) to the second period (*j)* in the k^{th} sequence. The carry-over effects is only present on the second period in the 2x2 crossover design, thus,

$$C_{(j,k)} = \begin{cases} C_R \; if \; k = 1 \quad j = 2 \\ C_T \; if \; k = 2 \quad j = 2 \end{cases},$$

where C_R and C_T denote the residuals of the reference and test formulation respectively.

$e_{ijk,}$: is the random error in observation Y_{ijk} .

2.5.1 Model Assumptions

In the bioequivalence analysis, the interest is to test the direct formulation effects independent of the period and carry-over effects. Thus, in practice, the period and carry-over effects are assumed to be zero. This is because a long enough washout period should eliminate the residuals of the formulation in the first period. A properly designed study eliminates the period effects (Jones and Kenward, 2003). Usually the washout period lasts at least five times the half-life time of the formulation.

2.6 Sample Size Calculation

In practice, the sample size should be large enough to provide the power of the test to show group differences. Bioequivalence studies with adequate sample size have the potential to produce results that are accurate, valid and reliable. The minimum required sample for animal trials and human trials are different. According to the MCCSA (2003), animal studies usually begin with eight animals, apart from the modified release oral dosage forms, which begins with twelve animals. If eight animals fail to produce 80% power, additional animals are added until the power of 80% is achieved. Studies with human subjects begin with twelve subjects, for a power of 80%, extra subjects are added if the power is less than 80%. The modified release oral dosage form studies usually start with twenty subjects. The sample size for the bioequivalence study can be calculated as follows (Rani and Pargal, 2004):

$$n_e \geq 2\left(t_{\frac{\alpha}{2};2n-2} + t_{(\beta;2n-2)} \right)^2 \left(\frac{\widehat{\sigma_d}}{\Delta} \right)^2 \tag{2.16}$$

where n_e represents the total subjects in the sequence, with both sequences allocated equal subjects, α is the Type-I-error, β represents the power, Δ represents the assumed detectable minimum difference between the formulations and $\widehat{\sigma_d}$ represents the pooled sample standard deviation for the differences of the period from the two sequences, commonly obtained from past studies. The degrees of freedom $2n-n$ are obtained using a numeric iterative search, which alternates the values of n, until a stable solution is found (Chow, Wang and Shao, 2007). The sample size is $N = 2n$, a small value of α leads to a large critical value $t_{\frac{\alpha}{2};2n-2}$, which then increases the sample size (Rani and Pargal, 2004).

Prior to the planned study, a pilot study is recommended to define optimal sample collection time intervals, drug variability and to gain more insight about the drug (FDA, 2013). The FDA (2013) recommends that, the sampling schedule be arrange in a manner to capture the changes in absorption, distribution and elimination. Depending on the pharmaceutical product about twelve to eighteen samples must be taken per subject per dose, including the pre-dose sample. The samples must be taken at three or more terminal elimination half-lives of the drug, and spaced appropriately to estimate $C_{\max}$ and elimination rate with accuracy (FDA, 2013). However, in some studies, it is difficult to obtain an adequate sampling internsity per individual, due to subject's nature or the phase of the illness, for example, children and late phase of the illness (Sheiner and Steimer, 2000).

2.7 Types of Bioequivalence Studies

Bioequivalence studies can be categorised into three groups: the Average Bioequivalence (ABE), Individual Bioequivalence (IBE) and Population Bioequivalence (PBE). If the average or the first moment of the marginal distribution of the PK parameter for the new drug is sufficiently close to the first moment or the mean of the marginal distribution of the R formulation, ABE is declared (Anderson and Hauck, 1990; Chow and Liu, 2004). For two formulations to be IBE, their bioavailability must be equivalent for all individuals (Anderson and Hauck, 1990). IBE is an extension of average bioequivalence, by including the comparison

of within-individual variation and individual by formulation interaction in the analysis (Chen, Patnaik, Hauck, Schuirmann, Hyslop and Williams, 2000). PBE is based on comparing the similarities of the overall distribution of the bioavailability for the two formulations (Anderson and Hauck, 1990). The IBE addresses whether a patient can be safely and effectively switch from their current formulation to another. PBE addresses whether the new formulation can be safely and effectively prescribed to a patient.

2.7.1 Average Bioequivalence Analysis

For a test based on ABE, Niazi (2014) recommended the analysis to be based on the quotient of T and R formulation geometric means, either using AUC_{0-k} , $AUC_{0-\infty}$ or $C_{\max}$ as the bioavailability measure. When submitting bioequivalence results to the regulating bodies, they require amongst other things the report about the summary statistics that includes the group averages, standard deviations and coefficients of variation (FDA, 2013; MCCSA, 2003).

The common practice is to construct a 90% confidence interval (CI) of the ratio of the means. If the 90% CI falls within specified limits known as the acceptance range then ABE is concluded. ABE can also be concluded based on hypothesis testing such as the Schuirmann's two one-sided test (TOST) (Schuirmann, 1987). Both methods are acceptable under the regulator's guidelines (FDA, 2013; MCCSA, 2003).

Interval Approach

To demonstrate ABE, the FDA (2013) recommends the use of $80/125$ rule. In the $80/125$ rule, ABE is accepted when the two formulation's bioavailability quotient is between 80% and 125% with 90% confidence (Jones and Kenward, 2003). That is if μ_T and μ_R are the average bioavailabilities (either AUC or $C_{\max}$) of the T formulation and the R formulation respectively, the two formulations are declared ABE if

$$80\% < \frac{\mu_T}{\mu_R} < 125\%. \tag{2.17}$$

That is if the CI falls within (80%-125%) then it can be concluded that the two drugs are ABE. The PK data are usually skewed. As a result, the FDA (2013) recommends log-transformed data to construct the CI. The CI on the original scale can be obtained by back transforming the

CI on the log scale. Of course, ABE can also be tested on the log scale by transforming the ABE acceptance range into log scale (Chow and Liu, 2004). Hence, the ABE is concluded if for a 90% CI the $\ln \mu_T - \ln \mu_R$ is contained within the interval (-0.2231, 0.2231). The problem with the CI interval approach is that the constructed CI is itself a random interval, thus there are no guarantees that ABE conclusion will hold if a new sample is selected. As a result it is recommended to simulate several trials and test ABE on those simulated trials to determine if ABE conclusion holds (Chow and Liu, 2004).

Hypothesis Approach

The purpose of hypothesis testing is to show that the two groups or treatments are different. The test is done by comparing the two statements:

$$\begin{gathered} \text{The null hypothesis is: } H_0 : \mu_1 = \mu_2, \\ \text{versus} \\ \text{The alternative: } H_1 : \mu_1 \neq \mu_2. \end{gathered} \tag{2.18}$$

In this test, if the alternative hypothesis is not accepted, it does not imply that the two means are equivalent. However, it is an indication of a lack of evidence to reject the null hypothesis. In ABE testing, the goal is to demonstrate that the two treatments are equivalent. Hence, instead of the traditional hypothesis testing the TOST method is used (Schuirmann (1987). The TOST method demonstrate similarity between the two treatments. This is done by formulating two null hypotheses. When both null hypotheses are rejected, then the two treatments are deemed equivalent. The two TOST hypotheses are as follows:

$$\begin{gathered} \text{The null hypothesis is: } H_{01} : \mu_T - \mu_R \leq \Delta_1, \\ \text{versus} \\ \text{The alternative } H_{11} : \mu_T - \mu_R > \Delta_1. \end{gathered} \tag{2.19}$$

And

$$\begin{gathered} \text{The null hypothesis is: } H_{02} : \mu_T - \mu_R \geq \Delta_2, \\ \text{versus} \\ \text{The alternative } H_{12} : \mu_T - \mu_R < \Delta_2. \end{gathered} \tag{2.20}$$

Where μ_T and μ_R are the average bioavailabilities of the T and the R formulations, respectively, Δ_1 and Δ_2 are standards specified by the authorities that decide how close the formulations must be to be confirmed bioequivalent.

The ANOVA for the 2x2 Crossover Design.

According to the FDA (2013) the 90% CI approach, is the best available method for testing bioequivalence using AUC and C_{max}. The log–transformation must be applied to the data before the analysis. Let $\overline{X}_{T1}$ and $\overline{X}_{T2}$ be the means for the T formulation in the first period and second period, respectively. Let $\overline{X}_{R1}$ and $\overline{X}_{R2}$ be the respective means for the R formulation. The means of the T and R formulations over the two periods can be estimated as $\overline{X}_T = \frac{\overline{X}_{T1} + \overline{X}_{T2}}{2}$ and $\overline{X}_R = \frac{\overline{X}_{R1} + \overline{X}_{R2}}{2}$, respectively. At the beginning of the study, both sequences have an equal number of subjects (n_1 and n_2), however, at end of the study some subjects may drop out. Hence, the above means $\overline{X}_T$ and $\overline{X}_R$ are equivalent to the Least Squares means (LS means) of μ_T and μ_R, the target population means of the T and R formulations (Niazi, 2014). In each sequence, the LS means are not calculated using sample size. Thus, the Analysis Of Variance (ANOVA) is used to find the root mean square error (RMSE), to be used to estimate $\hat{\sigma}_d$ (the pooled sample standard deviation of period differences from both sequences). This pooled sample standard deviation of period differences from both sequences is used to compute the 90% CI (Niazi, 2014). The ANOVA (Table 2.1) is used to compute the RMSE.

Table 2.1: ANOVA table for a 2x2 crossover design.

Source	Degrees of freedom (*df*)	Sum of squares (SS)	Mean square (MS)	F statistic
Formulation	1	SSF	MSF	MSF/MSE
Subject (sequence)	$n_1 + n_2 - 2$	SSS	MSS	MSS/MSE
Period	1	SSP	MSP	MSP/MSE
Error	$n_1 + n_2 - 2$	SSE	MSE	
Total	$2n_1 + 2n_2 - 2$	SST	MST	

The Classic (Shortest) Confidence Interval for the Difference of the Means

Let $\overline{Y_R}$ and $\overline{Y_T}$ be the sequence by period LS means of the R and T formulations, respectively. If there is no period and carry-over effects, the classic (shortest) $(1-2\alpha)\times 100$ CI for $\mu_T - \mu_R$ can be derived from the following two samples *t* statistic under normality assumption with $n_1 + n_2 - 2$ degrees of freedom:

$$T = \frac{\left(\overline{Y_R} - \overline{Y_T}\right) - \left(\mu_T - \mu_R\right)}{\sqrt{0.5\hat{\sigma}_d^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}, \tag{2.21}$$

where n_1 and n_2 are the number of subjects in sequences 1 and 2, respectively and $\widehat{\sigma_d}$ is the pooled sample standard deviation of period differences from both sequences, $\widehat{\sigma_d}$ can be obtained from the ANOVA Table 2.1 as half the square root of MSE,

If L_1 and U_1 are the lower and the upper limits, the classic or shortest $(1-2\alpha)\times 100$ CI for $\mu_T - \mu_R$ is as follows:

$$\begin{aligned} L_1 &= \left(\overline{Y_T} - \overline{Y_R}\right) - t_{(\alpha, n_1+n_2-2)}\sqrt{0.5\hat{\sigma}_d^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} \\ U_1 &= \left(\overline{Y_T} - \overline{Y_R}\right) + t_{(\alpha, n_1+n_2-2)}\sqrt{0.5\hat{\sigma}_d^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}. \end{aligned} \tag{2.22}$$

Under the $\mu_R \pm 20$ rule, the ABE is accepted if $(L_1; U_1) \in (-0.20\mu_R; 0.20\mu_R)$.

The Classic (Shortest) Confidence for the Ratio of the Means

The classical $(1-2\alpha)\times 100$ CI for the ratio of the means (μ_T / μ_R) can be derived from the above classical or shortest CI of the difference of the means, by dividing by $\overline{Y_R}$ to obtain the following CI:

$$L_2 = \left(\frac{L_1}{\overline{Y_R} + 1} \right) \times 100\%,$$

$$U_2 = \left(\frac{U_1}{\overline{Y_R} + 1} \right) \times 100\%. \tag{2.23}$$

Under the $\mu_R \pm 20$ rule, the ABE is accepted if $(L_2; U_2) \in (80\%; 120\%)$.

For both transformed and untransformed data, an identical ANOVA model is used to analyse the 2x2 crossover design. Hence, the method presented here is for the untransformed data. Note, the means obtained based on the original scale differs from the means obtained when back-transforming the mean from log-transformation. To illustrate this, let $\overline{\ln AUC_T}$ be the mean of the log-transformed AUC ($\ln AUC$) for the subject receiving the T formulation, then,

$$\overline{\ln AUC_T} = \frac{1}{n_2} \sum_{i}^{n_2} \ln AUC_i .$$

When $\overline{\ln AUC_T}$ is back-transformed to the original scale, the resulting mean ($e^{\overline{\ln AUC_T}}$) is the geometric mean (Chow and Liu, 2004). If the data is log-transformed, the procedure to compute the 90% CI for the difference of the means is followed. Hence, if L_1^* is the lower bound of the 90% CI for the log-transformed data and U_1^* is the respective upper bound, the 90% CI on the original scale can be obtained as follows;

$$\text{90\% lower bound} = (e^{L_1^*}) \text{ and}$$

$$\text{90\% upper bound} = (e^{U_1^*}). \tag{2.24}$$

The resulting CI in (2.24) is for geometric means ratio, and their interpretation is identical to the CI in (2.23) (Niazi, 2014). Using the accepted confidence interval of 80% to 125% the ABE is accepted if the geometric means quotient is within 0.80 and 1.25.

The Two One-Sided Test (TOST)

The FDA (2013) recommends the TOST procedure for bioequivalence analysis. The TOST for a balanced design is based on the two statistics:

$$T_1 = \frac{\left(\mu_T - \mu_R\right) - \Delta_1}{S\sqrt{\frac{2}{n}}} \text{ and } T_2 = \frac{\Delta_2 - \left(\mu_T - \mu_R\right)}{S\sqrt{\frac{2}{n}}}. \quad (2.25)$$

Where S is the standard deviation and n is the number of subjects per treatment. The TOST tests the two null hypotheses specified in (2.19) and (2.20), if both null hypotheses are rejected it is tantamount to concluding that the two drugs are ABE. Under the assumption of normal distribution of the means, the two hypotheses are tested using one-sided *t*-test with *v df*.
That is, if

$$T_1 > t_{(\alpha,\ df)} \text{ and } T_2 < -\ t_{(\alpha,\ df)}, \quad (2.26)$$

where $t_{(\alpha,\ df)}$ represents the upper 100α percentile of the Student's t-distribution with *v df.* The TOST method is operationally identical to the method of declaring bioequivalence if the $1-2\alpha$ CI for $\mu_T - \mu_R$ lies within $[\Delta_1, \Delta_2]$ the equivalence interval (Westlake, 1979).

2.7.2 Individual Bioequivalence and Population Bioequivalence

The measures of PBE and IBE can be grouped into three categories: disaggregate criterion, moment-based aggregate criterion, probability-based aggregate criterion. The moment-based criterion is based on the differences in bioavailability for a subject receiving reference formulation in two periods. This criterion is the function of the ratio of the T formulation within-subject variability to the R formulation within-subject variability, the difference is in variability of subject-by-formulation interaction and average bioavailability (Chen et al., 2000). For PBE or IBE studies based on moment criterion, the conclusions are recommended based on a non-parametric bootstrap 95% upper bound confidence limit. However, the bootstrap is computationally intensive and as such; Hyslop, Hsuan and Holder (2000) proposed an alternative method of using the one-sided 95% upper confidence bound. The method involves the use of the Howe's approximation I to a Cornish-Fisher expansion. Under the probability-based criterion, BE is accepted if for a specified population proportion BE holds. The required proportion of individuals is specified by a regulatory agency. This method was first proposed by Anderson and Hauck (1990), known as the Test of Individual Equivalence Ratios (TIER). In a disaggregate criterion, the concept of ABE is kept, since it partly involves the comparison of means, with the addition of variance comparison. The mean and variance components have different criterion, to avoid the mean and variance trade-offs. However,

regulatory agencies will need to specify multiple acceptance limits when using the criterion, due to a multiplicity of tests this method is consequently less favourable (Chen et al., 2000).

In the FDA (2001) draft, the PBE criterion, θ_{PBE} the aggregated, scaled moment-based criterion is defined as:

$$\theta_{PBE} = \frac{\left[\left(\mu_T - \mu_R\right)^2 + \left(\sigma_{TT}^2 - \sigma_{TR}^2\right)^2\right]}{\max\left\{\sigma_0^2, \sigma_{TR}^2\right\}}, \tag{2.27}$$

where;

$\sigma_{TT}^2 = \sigma_{WT}^2 + \sigma_{BT}^2$, $\sigma_{RT}^2 = \sigma_{WT}^2 + \sigma_{BR}^2$,

μ_T : is the mean for the T treatment,

μ_R : is the mean for the R treatment,

σ_{TT}^2 : is the total variance for the T treatment,

σ_{TR}^2 : is the total variance for the R treatment,

σ_0^2 : is the specified threshold value of the total variance,

σ_{WT}^2 : is the within-subject variance for the T treatment,

σ_{BT}^2 : is the between-subject variance for the T treatment,

σ_{WT}^2 : is the within-subject variance for the R treatment,

σ_{BR}^2 : is the between-subject variance for the R treatment.

The specified threshold value of the total variance is $\sigma_0^2 = 0.04$ (Jones and Kenward, 2003).

The hypothesis tested in PBE is

$$H_0^{PBE} : \theta_{PBE} \geq \theta_P \text{ versus } H_1^{PBE} : \theta_{PBE} < \theta_P . \tag{2.28}$$

Where $\theta_P = 1.7448$, obtained as follows:

$$\theta_P = \frac{\left(\mu_T - \mu_R\right)^2 + \sigma_{TT}^2 - \sigma_{TR}^2}{0.04}. \tag{2.29}$$

Under H_0^{PBE}, $\left(\mu_T - \mu_R\right)^2 = \ln(1.25)^2$ and $\sigma_{TT}^2 - \sigma_{TR}^2$ =0.02, substituting into (2.29) yields to,

$$\theta_P = \frac{\ln(1.25)^2 + 0.02}{0.04} = 1.7448.$$

Then, the PBE is accepted if H_0^{PBE} is rejected.

The IBE, the aggregated scaled moment-based criterion θ_{IBE} is defined as follows:

$$\theta_{IBE} = \frac{\left[\left(\mu_T - \mu_R\right)^2 + \sigma_D^2 + \left(\sigma_{WT}^2 - \sigma_{WR}^2\right)^2\right]}{\max\left\{\sigma_0^2, \sigma_{WR}^2\right\}}. \tag{2.30}$$

Where σ_D^2 is the subject-by-formulation interaction variance, σ_{WT}^2 and σ_{WR}^2 are respectively the within-subject variances for the T and R formulations. The within-subject variances cannot be separately estimated from the between-subject variance estimates using the 2x2 crossover design. Hence, a replicated crossover design is generally required to evaluate IBE (Jones and Kenward, 2003).

2.8 Summary

Authors have suggested simulation studies for comparing the NCA and NLMEM estimation methods. One such study is that of Panhard and Mentré (2005), where they used empirical Bayes estimates of individual estimates, for the analysis of bioequivalence based on results obtained from NLMEM and compared the results to the results obtained from the standard test based on NCA. Panhard and Mentré (2005) found that, when the individuals are densely sampled, tests based on the NCA and NLMEM methods produced similar results, except for the special studies of sparsely sampled individuals. However, Panhard and Mentré (2005) used the linearisation procedure to obtain the parameter estimates of NLMEM. The SAEM algorithm is more appropriate to estimate parameters of the NLMEM. Another study is that of Dubois et al. (2010) where they test bioequivalence based on the NCA method and compared

the results to the NLMEM methods individual EBE estimates. The tests based on the NCA estimates are more appealing for studies with high and small between-subject variability (Dubois et al., 2010). The study by Dubois et al. (2010) also revealed that bias is high when subjects are sparsely sampled. This study used the formulation effect to test ABE. On the other hand, the use of geometric means to test ABE in bioequivalence studies is a more popular approach. Consequently, in this study, the ABE test based on the NCA and the NLMEM methods are compared. The SAEM algorithm is used to estimate the parameters of the NLMEM method. The bioequivalence test is based on the geometric means as recommended by the FDA (2001) guidelines. Trials of different sample sizes, different sampling time intensity and varying variability levels are simulated to be equivalent with the bioequivalence results obtained using the NCA and the NLMEM methods.

CHAPTER 3 METHODOLOGY

3.1 Introduction

This study compare the bioequivalence test based on the NCA and the NLMEM estimates. In this section, the simulation procedure and the comparison methods are discussed. The pig data set from Onderstepoort, University of Pretoria is presented. The structural model that is used for simulation is discussed.

3.2 Ethical Guidelines for Clinical Trials

The ethical guidelines are there to ensure that certain individuals are not exposed to danger or risk in the society's benefit. The guidelines provide accepted standards and safeguard their implementation. In South Africa, the MCCSA (2003) provides the recommendations and guidelines to be followed when conducting animal or human clinical trials. These guidelines include the selection of subject and study conditions, as discussed in Section 3.2.1 and 3.2.2 below.

3.2.1 Selection of Subject

To maintain the focus of distinguishing between the pharmaceutical products the selected animal population must have less variability and only healthy subjects must be considered. The criterion used for excluding or including animals in the study must be vividly stated.

Human trials can only be conducted on healthy human subjects if there is no health risk. The use of healthy subjects allow the escalation of the bioequivalence results to the population on whom the R treatment is used, for in vivo trials. As with the animal trials, the criterion used for including or excluding subjects must be vividly stated. In the MCCSA (2003) guidelines, the required characteristics for human subjects are detailed as follows:

- *"*Sex: Subjects may be selected from either sex. However, the risk to women of childbearing potential should be considered on an individual basis",
- "Age: Subjects should be between 18 and 55 years of age",
- "Mass: Subjects should have a body mass within the normal range according to accepted normal values for the Body Mass Index [BMI = mass in kg divided by height

in meters squared, i.e. kg/m2), or within 15 % of the ideal body mass, or any other recognised reference.]",

- "Informed Consent: All subjects participating in the study should be capable of giving informed consent",
- "Medical Screening: Subjects should be screened for suitability by means of clinical laboratory tests, an extensive review of medical history, and a comprehensive medical examination. Depending on the API's therapeutic class and safety profile, special medical investigations may have to be carried out before, during and after the completion of the study",
- "Smoking/Drug and Alcohol Abuse: Subjects should preferably be non-smokers and without a history of alcohol or drug abuse. If moderate smokers are included they should be identified as such and the possible influences of their inclusion on the study results should be discussed in the protocol".

3.2.2 Study Conditions

Similar study conditions must be used to minimise variability. As a result, factors such as fluid intake, diet, and exercise must be standardised across subjects.

3.3 Data Set

In this study, the concentration data set of a well-recognised antibiotic will be used, the data set was obtained from fourteen healthy pigs after the antibiotic was administrated intramuscular. Each subject received 5 mg/kg dose of the antibiotic at the scheduled time. For each subject, measurements were taken at 0.25, 0.5, 0.75, 1, 1.5, 2, 4, 6, 9, 12, 18, 24 and 36 hours after administration to measure plasma concentration.

3.4 The Non-linear Model for Population Pharmacokinetics Modelling

The one-compartment model with first order absorption and first order elimination is often adequate to link plasma concentration and time profile (Panhard and Mentre, 2005, Dubois et al., 2010). The model can be defined as follows (Vonesh and Chinchilli, 1996):

$$f(t,\phi) = \frac{FDose}{V} e^{\frac{CL}{V}t}. \tag{3.1}$$

Where,

F: is the bioavailability of the drug, for drugs with intramuscular administration F=1,

$Dose$: is the amount of the drug administrated to each subject,

V : is the volume of distribution,

CL : is clearance rate,

t : is the sampling time,

ϕ : is the vector of population PK parameters equivalent to $(\ln CL/F, \ln V/F)$.

In the NLMEM method, $AUC_{0-\infty}$ is estimated as follows:

$$AUC_{0-\infty} = Dose/CL\,. \tag{3.2}$$

The parameters C_{max} and T_{max} are estimated directly from the observed concentrations.

3.5 Settings for the 2x2 Crossover Trial Simulation

The model in (3.1) is fitted to the data using the SAEM algorithm in the Monolix 2.4 software (Samson, Lavielle and Mentre, 2007). To simulate the 2x2 crossover trials the fitted model is used. For simplicity, the sequence effect or period effect is not simulated. To simulate the data for the R formulation, the PK parameters were kept at the average values estimated using SAEM. The vector of the PK parameters is modified by adding the treatment effect $(0, \widehat{\beta}_{T,CL/F}, \widehat{\beta}_{T,V/F})'$, to alter bioavailability for the T treatments.

To mimic a change in variability, both within-subject variability (WSV) and between-subject variability (BSV) are simulated. First simulation is with high-level BSV, and high-level WSV. Second simulation is with low-level BSV and low-level WSV. The level of variability used in the simulation study is summarised in Table 3.1.

Table 3.1: The variability settings used in the simulation study

Variability settings	BSV	WSV	Residual error
S_{ll}	20% for $\ln(CL/F)$ 10% for $\ln(V/F)$	10% for $\ln(CL/F)$ 5% for $\ln(V/F)$	$p=0.1$
S_{hl}	50% for $\ln(CL/F)$ 50% for $\ln(V/F)$	15% for $\ln(CL/F)$ 15% for $\ln(V/F)$	$p=0.1$
S_{hh}	50% for $\ln(CL/F)$ 50% for $\ln(V/F)$	15% for $\ln(CL/F)$ 15% for $\ln(V/F)$	$p=1$

The BSV and WSV are the standard deviations of each parameter of model (3.1). Similar to the study by Dubois et al. (2010), the residual error is changed twice; by changing the value of p in the error model as presented in Section 2.4.3. For the low-level, $p=0.1mg/l$ and for the high-level $p=1mg/l$. Hence, this implies that there are three variability settings, S_{ll} for low BSV, low WSV and low residual error, S_{hl} for low residual error, high BSV and high WSV, and S_{hh} for high BSV, WSV and high residual error.

3.5.1 Simulation Process

For each trial m for $m=1,\ldots,M$, a vector of between-subject random effect b_j is simulated together with two vectors of the within-subject random effects c_{ik} for a period and for each individual. To obtain θ_{ikl}, the random effects b_j and c_{ik} is added to μ_{Rl} and to the treatment effect β_{Tl}. Then the estimated individual parameters is used to find the concentration $f(t_j,\theta_{ik})$ at time t_j for each treatment group k. The simulated concentration y_{ijk} defined in Equation (2.15) is obtained by adding the residual error, which is produced by a $N(0,p^2)$ to each predicted concentration.

3.5.2 Simulation Designs

The M trials presented in the simulation process (Section 3.5.1) are simulated with four types of designs. The first design is based on the original data, the second is the intermediate design, the third one is the sparse design and the fourth one is the rich design, to mirror the real situations in bioequivalence trials, where studies can be rich and sparse depending on the drug and the nature of the subjects. Firstly, simulating data that has an identical design to the observed data, where $N = 14$ and $n = 12$, with measurements obtained at the following times 0.25, 0.5, 0.75, 1, 1.5, 2, 4, 6, 9, 12, 18 and 24 hours after dosing. Secondly, data with $N = 30$ subjects and $n = 5$ samples per subject per period measured at 0.25, 1, 4, 12 and 24 hours is simulated for the intermediate design. Thirdly, data with $N = 50$ and $n = 3$, with measurements obtained at 0.25, 3.25 and 24 hours after dosing is simulated for a sparse design. Fourthly, a rich design is simulated with $N = 50$ and $n = 12$ measurements taken at 0.25, 0.5, 0.75, 1, 1.5, 2, 4, 6, 9, 12, 18 and 24 hours after dosing. The variability settings S_{ll}, S_{hl} and S_{hh} is used in each simulation for each design. For the intermediate design, S_{hh} is added. For each combination of variability setting and design, a trial is simulated under two bioequivalence hypotheses recommended by FDA (2013): $H_{0;80\%}$ and $H_{0;125\%}$, where $\beta_T = (0, \ln(0.8))'$ and $\beta_T = (0, \ln(1.25))'$, correspondingly. The R treatment is not affected by the change in bioavailability due to $H_{0;80\%}$ or $H_{0;125\%}$, hence the concentration values are identical for the design and variability setting under $H_{0;80\%}$ and $H_{0;125\%}$.

3.6 Estimation of the Individual PK Parameters

The bioequivalence is evaluated on $C_{\max}$, AUC_{0-k} and $AUC_{0-\infty}$ for the original data set. For simulated trials bioequivalence is only evaluated on $AUC_{0-\infty}$, since $C_{\max}$ is directly observed from the data (i.e. is not estimated by the trapezodial method) and $AUC_{0-\infty}$ captures the full exposure. With the NCA approach AUC_{0-k} and $AUC_{0-\infty}$ are estimated using Equation (2.1) and (2.2), respectively. To estimate the parameters of the NLMEM, the SAEM algorithm is used, and $AUC_{0-\infty}$ is estimated by Equation (3.2) where the CL is the estimated parameter of the NLMEM. In all, there are $2N$ AUC values and since each subject i for $i = 1, \ldots, N$ is measured twice, once for the R formulation and once for the T formulation. Hence, $\widehat{AUC_i^R}$ is the AUC

estimate for subject i on R formulation and $\widehat{AUC_i^T}$ is the corresponding AUC estimate for the T formulation. $\widehat{C_{\max i}^R}$ is the estimated $C_{\max}$ of the i^{th} individual in the R formulation and $\widehat{C_{\max i}^T}$ is the corresponding estimated $C_{\max}$ of the T treatment. Note that $C_{\max}$ is observed directly from the data, hence the ABE test based on C_{max} is not used for comparison.

3.7 Analysis of the 2x2 Crossover Design

In practice, the assumption is that a well-designed study should be able to eliminate the carry-over effects in (2.15), hence, the assumption is that there is no carry-over effect (Jones and Kenward, 2003). Thus, the 2x2 crossover design is analysed using the following model derived from (2.15).

$$Y_{ijk} = \mu + T_{ik} + P_k + S_{ik} + e_{ijk}, \tag{3.3}$$

where Y_{ijk} represents the log of the individual PK parameter (AUC_{0-k} or $C_{\max}$ and or $AUC_{0-\infty}$), μ is the overall mean, and e_{ijk} is the residual error which is also normally distributed with the mean zero and constant variance, covariates T_{ik} , P_k and S_{ik} are treatment, period and sequence fixed effects, respectively.

3.7.1 Statistical Assumptions for the ANOVA model

The following assumptions have to be adhered to for the ANOVA model:

- The covariate S_{ik} follows an independent normal distribution with a mean zero and constant variance σ_s^2 ,
- e_{ijk} and S_{ik} are mutually independent,
- e_{ijk} follows an independent normal distributed with a mean zero and constant variance σ_e^2.

Under these assumptions, the inference about approximation, CI, the fixed effects and hypothesis testing can be constructed (Jones and Kenward, 2003). As recommended by the MCCSA (2003), Walker et al. (2006) and FDA (2013), the TOST procedure and shortest confidence interval approach described in Section 2.7.1 are used to test average bioequivalence on the geometric means.

3.8 Type-I-Error Evaluation

Bioequivalence tests are conducted for $\ln AUC_{0-\infty}$ obtained using either the NCA or the NLMEM estimates from simulated data. Berger and Hsu (1996) defined the TOST Type-I-error as the largest of the Type-I-errors over the null space.

*Definition 3.1 (**Type-I-error For the TOST procedure**)*

The Type-I-error for the TOST procedure is defined in Berger and Hsu (1996) as the probability of rejecting the null hypothesis (H_{02} or H_{01}) while it is true or the probability of declaring bioequivalence of the drugs while they are not.

The Type-I-error of the TOST can be assessed for each limit of the null hypothesis space (Liu and Weng, 1995). As a result, the trials are simulated under $H_{0;80\%}$ and $H_{0;125\%}$ null hypotheses. For each null hypothesis $H_{0;80\%}$ and $H_{0;125\%}$, the rate of the Type-I-error is determined by the proportion of the simulated trials for which Type-I-error has been committed. If tests are done on the true values of AUC the rate of theType-I-error is identical for the two null hypotheses, since the two null hypotheses are symmetric. In the case of estimates, the rate of the Type-I-error for $H_{0;80\%}$ and $H_{0;125\%}$ might not be identical. Subsequently, the global Type-I-error is used.

Definition 3.2 ***(global Type-I-error)***

The global Type-I-error is defined as the largest between the two resulting Type-I-errors (Panhard and Mentre, 2005).

The method with less global Type-I-error is deemed reliable or better in estimating parameters.

3.9 The Population Bioequivalence

The criterion specified in Section 2.7.2 is used to evaluate PBE as recommended by the FDA (2001). The NCA or NLMEM estimates are used to evaluate PBE for the original data set of fourteen healthy pigs. The FDA recommends the BE measures be log-transformed using log base-10 or the natural log. In this study, the natural log is used. Henceforth, the PBE is evaluated on $\ln AUC_{0-\infty}$, $\ln AUC_{0-k}$ and $\ln C_{\max}$.

3.10 Estimates of the Sample Mean Evaluation

For the actual bioequivalence test, both AUC_i estimated from the simulated trials will be used. The precision of each approach (NCA and EBE) in estimating the sample means of $\ln AUC_{0-\infty}$ is evaluated by comparing estimation errors for each treatment. For each treatment group, the estimation error is computed for each simulated trial. The estimation errors of AUC_i^R and AUC_i^T is calculated as the difference between the predicted values and the estimated values. The estimation error for $\ln AUC_{0-\infty}$ of the T treatment, can be computed as follows (Jones and Kenward, 2003)

$$ee_{AUC}^T = AUC_i^T - \widehat{AUC_i^T}\ , \tag{3.4}$$

where $\widehat{AUC_i^T}$ is the predicted AUC value obtained by the ANOVA model for individual $i = 1, \ldots, \ddot{N}$ and AUC_i^T represents the estimated value for $i = 1, \ldots, N$ by either the NCA or the NLMEM method. When the NCA method is used, missing values may occur, therefore, $\ddot{N} \le N$. For the NLMEM method there are no missing values, that is $\ddot{N} = N$.

The Root Mean Square error (RMSE) and Mean Bias (MB) for the T treatment are calculated as follows:

$$\text{MB} = \sum_i^{2\ddot{N}} \frac{(AUC_i^T - \widehat{AUC_i^T})}{2\ddot{N}} \tag{3.5}$$

and

$$\text{RMSE} = \sqrt{\sum_i^{2\ddot{N}} \frac{\left(AUC_i^T - \widehat{AUC_i^T}\right)^2}{2\ddot{N}}}\ . \tag{3.6}$$

For each design variability combinations, the bias is tested for significance using the 95% CI. The standard error of the mean and the 97.5% normal distribution quantile are used to compute the 95% CI of the bias. If the 95% CI of bias does not include zero, then the bias is significantly different from zero at a 5% significance level (Jones and Kenward, 2003). Between the NCA method and the NLMEM method, the method that leads to a smaller bias and a smaller RMSE is considered more reliable.

3.11 The Bayes Shrinkage Estimator and Tests Based on the NLMEM Method

The EBEs are affected by shrinkage when the data are uninformative at the individual level. This shrinkage leads to the poor estimation of PK parameters; hence, this may lead to lower power to detect the exposure-response relationship. Therefore, the relationship between the Type-I-error of the TOST procedure and shrinkage is assessed for the NLMEM estimates. To estimate shrinkage for $\ln AUC_{0-\infty}$ of the R formulation Equation (2.11) can be adapted as follows:

$$\text{shrinkage}(\ln AUC^{R}_{0-\infty}) = 1 - \frac{\text{var}(\ln AUC^{R}_{0-\infty i})}{\widehat{\omega^2_{\ln AUC^{R}_{0-\infty}}}}, \tag{3.7}$$

where $\text{var}(\ln AUC^{R}_{0-\infty i})$ represents the variance of each estimate of $\ln AUC^{R}_{0-\infty i}$ and $\widehat{\omega^2_{\ln AUC^{R}_{0-\infty}}}$ is the corresponding calculated population variance.

- Since $AUC^{R}_{0-\infty i} = Dose / CL^{R}_{i}$ (where CL^{R}_{i} represents R the formulation clearance for each individual).
- Then $\ln AUC^{R}_{0-\infty i} = \ln Dose + \ln CL^{R}_{i}$.
- It follows that $\text{var}(\ln AUC^{R}_{0-\infty i}) = \text{var}(\ln CL^{R}_{i})$ and $\widehat{\omega^2_{\ln AUC^{R}_{0-\infty}}} = \widehat{\omega^2_{\ln Cl^{R}}}$.

In this study, the median shrinkage of the R formulation from1000 simulated trials of each design and for each variability setting is used to assess the link between shrinkage and global Type-I-error for the NLMEM estimates.

3.12 Summary

This chapter covered the methods and data used in this study. These include:

- The ethical guidelines for clinical trials to ensure that certain individuals are not exposed to danger or risk in the society's benefit,
- The overview and source of the data set that was used in this study,
- The non-linear model for population modelling; the settings for the 2x2 crossover trial which include the initial values of the model parameters that were simulated,

- The simulation process for simulating concentration data; simulation designs for simulating varies 2x2 crossover trials,
- The individual parameters estimation for the bioequivalence measures,
- Analysis of the 2x2 crossover design which includes the details of the model for analysing the 2x2 crossover design data,
- The population bioequivalence,
- The estimation of sample means for evaluating the precision of the NCA and the NLMEN methods,
- Type-I-error evaluation for comparing the bioequivalence results based on the NCA and NLMEM estimates and the shrinkage for validating NLMEM estimates,
- The data are analyzed using SAS and Monolix 2.4 software.

CHAPTER 4 RESULTS

4.1 Introduction

The previous chapter dealt with the research methodology and outlined the research design, sampling technique and explained the instrument used for data collection.

This chapter presents the results and discusses the analysis of the study. The data were analysed using the NCA or NLMEM approaches. A comparison of the estimation methods based on simulated data was conducted using the simulation methods presented in Chapter 3.

The concentration-time profiles of the fourteen healthy pigs used in the study are in Appendix C. As detailed in Chapter 3 only the bioequivalence tests based on $AUC_{0-\infty}$ are used in comparing NCA and NLMEM methods. To assess ABE the classic confidence interval approach and the interval hypothesis testing approach using Schuirmann's TOST procedure are used.

The results will be presented and subsequently, interpretation and analysis thereof will follow, leading to the recommendations and conclusion of the study which form Chapter 5.

The analysis and results thereof are reported under five sections as follows:

- The NCA Approach,
- The NLMEM Approach,
- Population Bioequivalence Analysis,
- Simulation Results and
- Summary of the Results.

4.2 The NCA Approach

The values of the C_{max} and $\ln C_{max}$ are presented in Table 4.1 and the values of the AUC and $\ln AUC$ are presented in Table 4.2 for the NCA approach.

Table 4.1: C_{max} and $\ln C_{max}$ for the R and the T formulation for both periods estimated by the NCA method.

Sequence	Pig	Period		Period	
		I	II	I	II
		C_{max}	C_{max}	$\ln C_{max}$	$\ln C_{max}$
TR	6	1.4410	1.1380	0.3653	0.1293
	10	1.3720	1.4900	0.3163	0.3988
	12	1.3560	1.2200	0.3045	0.1989
	13	1.3940	1.4720	0.3322	0.3866
	14	1.3300	1.3810	0.2852	0.3228
	23	1.2690	1.1820	0.2382	0.1672
	25	1.2600	1.2520	0.2311	0.2247
RT	4	1.3620	1.1560	0.3090	0.1450
	9	1.3710	1.0580	0.3155	0.0564
	11	1.3610	1.3360	0.3082	0.2897
	15	1.2580	1.4440	0.2295	0.3674
	18	1.5230	1.3280	0.4207	0.2837
	19	1.3680	1.3320	0.3133	0.2867
	24	1.3290	1.4410	0.2844	0.3653

Table 4.1 above shows the original and log-transformed values of the PK parameter C_{max} of the fourteen healthy pigs used in this study. Each pig was randomly allocated to either sequence TR or sequence RT. For all the pigs allocated in sequence TR, in the first dosing period, each pig initially received the T formulation and after a sufficient washout period, the subject crossed over to the R formulation in the second period. For sequence RT, the procedure is reversed. Soon after, the dosing samples were collected at 0.25, 0.5, 0.75, 1, 1.5, 2, 4, 6, 9, 12, 18, 24 and 36 hours.

Table 4.2: AUC_{0-k} and $AUC_{0-\infty}$ estimated by the NCA method.

Sequence	Pig	Period		Period	
		I	II	I	II
		AUC_{0-k}	$\ln AUC_{0-k}$	$AUC_{0-\infty}$	$\ln AUC_{0-\infty}$
TR	6	12.1904	7.6844	12.5544	7.7014
	10	15.0854	15.4897	15.1153	15.5489
	12	14.0283	9.5135	14.0819	9.5413
	13	13.4765	9.1711	12.4400	9.1891
	14	14.9046	11.0338	14.9657	11.0652
	23	15.6595	12.223	15.7567	12.2847
	25	15.4633	11.9016	15.4930	11.9906
RT	4	13.8473	8.3514	14.2341	8.4112
	9	14.7908	14.1151	15.0575	14.1463
	11	15.2143	11.9056	15.2417	11.9494
	15	14.9325	11.804	15.0128	11.8166
	18	15.8373	12.3218	15.8904	12.3291
	19	15.4373	11.9164	15.4717	11.9359
	24	16.4227	12.4213	16.5370	12.4977

The AUC_{0-k} was estimated using trapezoidal rule defined by Equation (2.1), and $AUC_{0-\infty}$ was computed using Equation (2.2). The values of AUC_{0-k} and $AUC_{0-\infty}$ for the R formulation and the T formulation are in Table 4.2 above. The values $AUC_{0-\infty}$ are higher than that of AUC_{0-k}. This is due to the significant fraction that remains after time t_k as noted by Martinez and Jackson (1991).

4.2.1 Classic (Shortest) Confidence Interval Approach

The model specified in (3.3) was used to analyse the 2x2 crossover design, the ANOVA table of the model is in Table 4.3. The ANOVA is used to find the pooled sample standard deviation

($\widehat{\sigma_d}$) needed when calculating 90% CI. It is generally assumed that there are no carry-over effects because the chosen length of the washout period is adequate to eliminate the carry-over effects. Hence, this justifies the use of the model specified in (3.3).

Table 4.3: The ANOVA for AUC_{0-k} using SAS.

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	15	0.9699	0.06466	6.01	0.0017
Error	12	0.1290	0.01075		
Corrected Total	27	1.09886			
R-Square	Coeff Var	Root MSE	ln AUC_{0-k} Mean		
0.8826	4.05766	0.1037	2.5552		
Source	DF	Type I SS	Mean Square	F Value	Pr > F
Sequence	1	0.03520	0.03520	3.27	0.0955
period	1	0.5267	0.5267	49.00	<.0001
formulation	1	0.001381	0.00138	0.13	0.7262
pig	12	0.4065	0.03388	3.15	0.0288

The model is significant (*P*-value=0.0017) at 5% level of significance. The model fit R^2 is 0.8826, which means that 88.26% of the variation is explained by the model. The CV is 4.06%, which is an indication that the data are not highly variable. At 5% level of significance, the hypothesis of no period effect is rejected (*P*-value <0001). However, the significant period effect do not invalidate the 2x2 crossover design. There is no sequence effect (*P*-value =0.0955) and no formulation effect (*P*-value=0.7262). This suggests that there are no carry-over effect and no formulation effect in this study at 5% significance level. The pooled sample variance ($\widehat{\sigma_d}^2$) is the mean square error, that is $\widehat{\sigma_d}^2 = 0.01075$. The pooled sample standard deviation ($\widehat{\sigma_d}$) is the square root of the pooled sample variance, which is 0.1037, is used to calculate the confidence intervals in Table 4.5.

Table 4.4: The means for $\ln AUC_{0-k}$ and the geometric means using SAS.

Formulation	$\ln AUC_{0-k}$ lsmeans	Standard error	Geometric means for AUC_{0-k}
R	2.5482	0.02771	12.7841
T	2.5623	0.02771	12.9649

Table 4.4 above provides the lsmeans for the transformed data and respective geometric means for the two treatments. These means are used in constructing the shortest CI for bioequivalence testing as provided in Table 4.5.

Table 4.5: The classic (shortest) CI for $\ln AUC_{0-k}$ and for the geometric mean.

AUC	$\widehat{\sigma}_d^2$	lower bound	90% CI lower limit	90% CI upper limit	upper bound
CI for difference of the means (Log-transformed data)	0.01075	-0.2231	-0.02329	0.05138	0.2231
CI for the ratio of the geometric means		0.8	0.977	1.05272	1.25

Table 4.5 above portrays the 90% classic (shortest) CI of difference between formulations on the log-transformed scale and the 90% CI of the geometric means ratio (back-transformed from the former 90% CI).

Based on 90% classic (shortest) CI for the difference between formulations and for the corresponding ratio of geometric means, ABE can be claimed at 5% level of significance. Since the CI of the difference between formulations is (-0.02329; 0.05138) which is contained inside the accepted limits of - 0.2231 and 0.2231, and the 90% CI of geometric means ratio (0.9770; 1.05272) is contained within the accepted limits of 0.8 and 1.25; the T and the R formulations are thus average bioequivalent based on AUC_{0-k} using the classic (shortest) CI approach.

Table 4.6: The ANOVA for $\ln AUC_{0-\infty}$ using SAS.

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	15	0.9735	0.0649	5.9100	0.0018
Error	12	0.1319	0.0110		
Corrected Total	27	1.1054			
R-Square	Coeff Var	Root MSE	$\ln AUC_{0-\infty}$ Mean		
0.8807	4.0980	0.1048	2.5581		
Source	DF	Type I SS	Mean Square	F Value	Pr > F
Sequence	1	0.0422	0.0422	3.84	0.0737
period	1	0.5223	0.5223	47.53	<.0001
formulation	1	0.0003	0.0003	0.03	0.8667
pig	12	0.4087	0.0341	3.1	0.0306

The fitted model is significant (*P*-value =0.0018) at 5% level of significance. The model R^2 is 0.8807, which means that the model explains 88.07% of variation in the data. The CV is 4.1%, which is an indication that the data are not highly variable. The period effect is significant with the (*P*-value<.0001). The sequence effect is not significant (*P*-value=0.0737). The formulation effect is not significant (*P*-value =0.8667). Thus, there are no carry-over effect and no formulation effect at 5% significance level. From Table 4.6 $\widehat{\sigma_d}^2 = 0.0110$ and $\widehat{\sigma_d}$ =0.1049, the predicted value for the variance is calculated to be used in constructing the confidence intervals in Table 4.8.

Table 4.7: The lsmeans for $\ln AUC_{0-\infty}$ and the geometric means.

Formulation	$\ln AUC_{0-\infty}$ lsmeans	Standard error	Geometric means for $AUC_{0-\infty}$
R	2.5547	0.0280	12.8677
T	2.5615	0.0280	12.9555

The lsmeans of $\ln AUC_{0-\infty}$ and the geometric means are given in Table 4.7 and are being used to construct the classic CI for $\ln AUC_{0-\infty}$ and the geometric means in Table 4.8.

Table 4.8: The classic CI for $\ln AUC_{0-\infty}$ and the geometric means.

$\ln AUC_{0-\infty}$	$\widehat{\sigma}_d^2$	lower bound	90% CI lower limit	90% CI upper limit	upper bound
CI for difference of the means(Log-transformed data)	0.0110	-0.2231	-0.0310	0.0445	0.2231
CI for the ratio of the geometric means		0.80	0.9695	1.0456	1.25

Table 4.8 above depicts the classic CI for $\ln AUC_{0-\infty}$ and the geometric means. The 90% CI of the difference between treatments on the log-transformed scale (-0.0310; 0.0445) and the 90% CI of the ratio of the geometric means (0.9695; 1.0456) are both contained within the accepted range for each. Thus, the ABE is accepted at 5% significance level. Therefore, using $AUC_{0-\infty}$, the R and the T treatments are average bioequivalent, based on the classic CI.

Table 4.9: The ANOVA for $\ln C_{\max}$ from SAS using SAS.

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	15	0.1018	0.006783	0.84	0.6269
Error	12	0.09636	0.008030		
Corrected Total	27	0.1981			
R-Square	Coeff Var	Root MSE	$\ln C_{\max}$ Mean		
0.513604	31.85764	0.08961	0.2813		
Source	DF	Type I SS	Mean Square	F Value	Pr > F
Sequence	1	0.0001940	0.0001940	0.02	0.8791
period	1	0.01423	0.01423	1.77	0.2079
formulation	1	0.0007201	0.0007201	0.09	0.7697
pig	12	0.08661	0.0072178	0.90	0.5718

Table 4.9 above illustrates the ANOVA table obtained by fitting the model specified in (3.3) on the $\ln C_{\max}$ data presented in Table 4.1. The model is not significant (*P*-value = 0.6269), this can occur due to the small sample size and the variability of the data. There is no sequence effect (*P*-value= 0.8791), no period effect (*P*-value=0.2079) and no formulation effect (*P*-value=0.7697) at a 5% level of significance. The pooled sample variance ($\widehat{\sigma}_d^2$) is 0.008030 and $\widehat{\sigma}_d$ is 0.08961. The $\widehat{\sigma}_d$ value is used to calculate the confidence intervals in Table 4.11.

The CV is 31.86%, which is higher than 30% and as such not acceptable. In addition, when the CV is high the CI approach may not have the required level of assurance to assess ABE (Chow and Liu, 2004).

Table 4.10: The lsmeans for $\ln C_{max}$ and the geometric means.

Formulation	$\ln C_{max}$	Standard error	Geometric means for C_{max}
R	0.2864	0.02395	1.3316
T	0.2762	0.02395	1.3181

The lsmeans of $\ln C_{max}$ and the geometric means are given in Table 4.10, which are used in Table 4.11 to construct the 90% CI for $\ln C_{max}$ and the geometric means for C_{max}.

Table 4.11: The classic (shortest) CI for $\ln C_{max}$ and the geometric means.

C_{max}	$\widehat{\sigma}_d^2$	lower bound	90% CI lower limit	90% CI upper limit	upper bound
CI for difference of the means(Log-transformed data)	0.008030	-0.2231	-0.04241	0.02212	0.2231
CI for the ratio of the geometric means		0.80	0.9585	1.02237	1.25

The 90% CI of the difference between formulations on the log-transformed scale and the 90% CI of the ratio of the geometric means based on C_{max} are given in Table 4.11. The 90% CI for the difference between treatment means and that of the ratio of the geometric means are both contained within the accepted ranges. Thus, the ABE is accepted at a 5% significance level. Therefore, using C_{max}, the two treatments are average bioequivalent, based on the classic (shortest) CI.

4.2.2 The Schuirmann's Two One-Sided Tests Procedure (TOST)

The FDA (2013) advocated for the TOST procedure defined in section 2.7.1 for evaluating ABE on the PK parameters, AUC and C_{max}. Recall that, in the TOST procedure ABE is concluded if the two hypotheses H_{01} and H_{02} are both rejected at 5% level of significance.

Alternatively, if the TOST 90% CI for the geometric means are contained within the acceptance limits, (0.80; 1.25), then the ABE is concluded under the ± 20 rule.

Table 4.12: The TOST for $\ln AUC_{0-k}$

N	Geometric Mean	Coefficient of Variation	Minimum	Maximum	
14	1.0141	0.3258	0.6031	1.5864	
Geometric Mean	95% CL Mean		Coefficient of Variation	95% CL CV	
1.0141	0.8442	1.2183	0.3258	0.2333	0.5471
Geometric Mean	Lower Bound	90% CI			upper bound
1.0141	0.8	0.8726	1.1787		1.25
Test	Null	DF	t Value	P-Value	
Lower	0.8	13	2.79	0.0076	
Upper	1.25	13	-2.46	0.0142	
Overall				0.0142	

The 90 % CI (0.8726; 1.1787) in Table 4.12 above is contained within the acceptance range of (0.80; 1.25), thus the two hypotheses of the TOST procedure in Section 2.7.1 are rejected since both *P*-values are less than 0.05. ABE is consequently claimed at a 5% significance level, using the ±20 rule.

Table 4.13: The TOST for $\ln AUC_{0-\infty}$

N	Geometric Mean	Coefficient of Variation	Minimum	Maximum	
14	1.0068	0.3254	0.5909	1.6301	
Geometric Mean	95% CL Mean		Coefficient of Variation	95% CL CV	
1.0068	0.8383	1.2092	0.3254	0.2331	0.5464
Geometric Mean	Lower Bound	90% CI			upper bound
1.0141	0.8	0.8664	1.1699		1.25
Test	Null	DF	t Value	*P*-Value	
Lower	0.8	13	2.71	0.0089	
Upper	1.25	13	-2.55	0.0121	
Overall				0.0121	

The 90 % CI (0.8664; 1.1699) for the geometric means of $AUC_{0-\infty}$ in Table 4.13 above is contained within the acceptance range of (0.80; 1.25). Thus, the two hypotheses of the TOST procedure are rejected, since both *P*-values are less than 0.05. Hence, ABE is claimed at a 5% significance level, using the $\pm$ 20 rule.

Table 4.14: The TOST for $\ln C_{max}$

N	Geometric Mean	Coefficient of variation	Minimum	Maximum	
14	0.9899	0.131	0.7717	1.2663	
	95% CL Mean		Coefficient of variation	95% CL CV	
0.9899	0.9181	1.0673	0.131	0.0948	0.2125
Geometric Mean	Lower Bound	90% CL Mean		Upper Bound	Assessment
0.9899	0.8	0.9306	1.0529	1.25	Equivalent
Test	Null	DF	t Value	P-Value	
Lower	0.8	13	6.11	<.0001	
Upper	1.25	13	-6.69	<.0001	
Overall				<.0001	

The 90% CI (0.9306; 1.0529) of the geometric means ratio for C_{max} in Table 4.14 above is contained entirely within the acceptable bounds (0.8; 1.25) and both *P*-values are less than 0.05. Hence, both hypotheses are rejected and thus, ABE is concluded using the $\pm$ 20 rule. Both the 90% CI and TOST confirmed that the R and T formulations are ABE for $\ln C_{max}$.

4.3 The NLMEM Approach

4.3.1 Population Modelling

The model specified in (3.1) was used for population modelling and the constant error model was selected to model the error. On visual examination of the residual against the predicted, the visual predictive check (VPC) and the individual fit statistics for the constant error model was justified, and it has the lowest AIC criterion. The estimates of the model specified in (3.1) are in Table 4.15 below. The SAEM algorithm implemented in the Monolix 2.4 software was used for paramater estimation.

Table 4.15: Parameter estimates and fit statistics for model NLMEM.

Parameter	Estimates	standard error	Treatment effect test (*P*-values)
V	3.770	0.13	
β_{V_T}	0.0357	0.047	0.45
Cl	0.334	0.02	
β_{CL_T}	-0.033	0.085	0.75
b_V	0.021	0.1	
b_{CL}	0.00681	1.1	
c_v	0.114	0.026	
c_{cl}	0.192	0.055	
a	0.139	0.0057	

In Table 4.15 above V = volume; β_{V_T} = T treatment's volume effect; Cl=clearance; b_V = between population subject random effect for volume; b_{CL} = between population subject random effect for clearance; c_v = within population subject random effect for volume; c_{cl} = within population subject random effect for clearance and a = constant term of the error model. The *P*-values of the Wald test in Table 4.15 confirmed that the clearance and volume effects of the T formulation and the R formulation are not significantly different with *P*-values of 0.45 and 0.75, respectively. However, such conclusions may be misleading because of a small sample size. Thus, bioequivalence is not concluded at this stage.

4.3.2 Model Evaluation

In this study the profiles of the two treatments are compared, therefore the model is developed to understand the two formulations. It is important that the invented model complies with the assumption and elucidates the underlying phenomenon. The residuals versus the predicted values were used to assess the normality and constant error variance assumptions, and the plots of the individual fits were used to assess the inter-subject variation. Regarding the random effects, their joint distribution was utilized to assess normality and independence of the random effects.

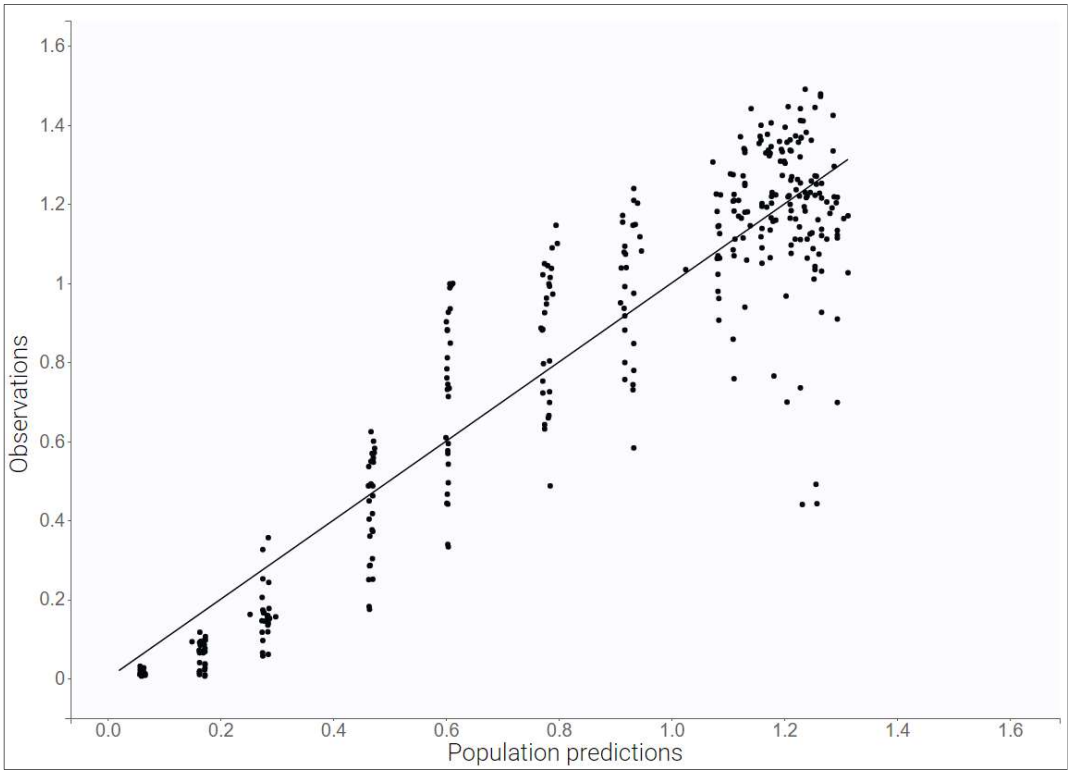


Figure 4.1: The observed against the population model prediction.

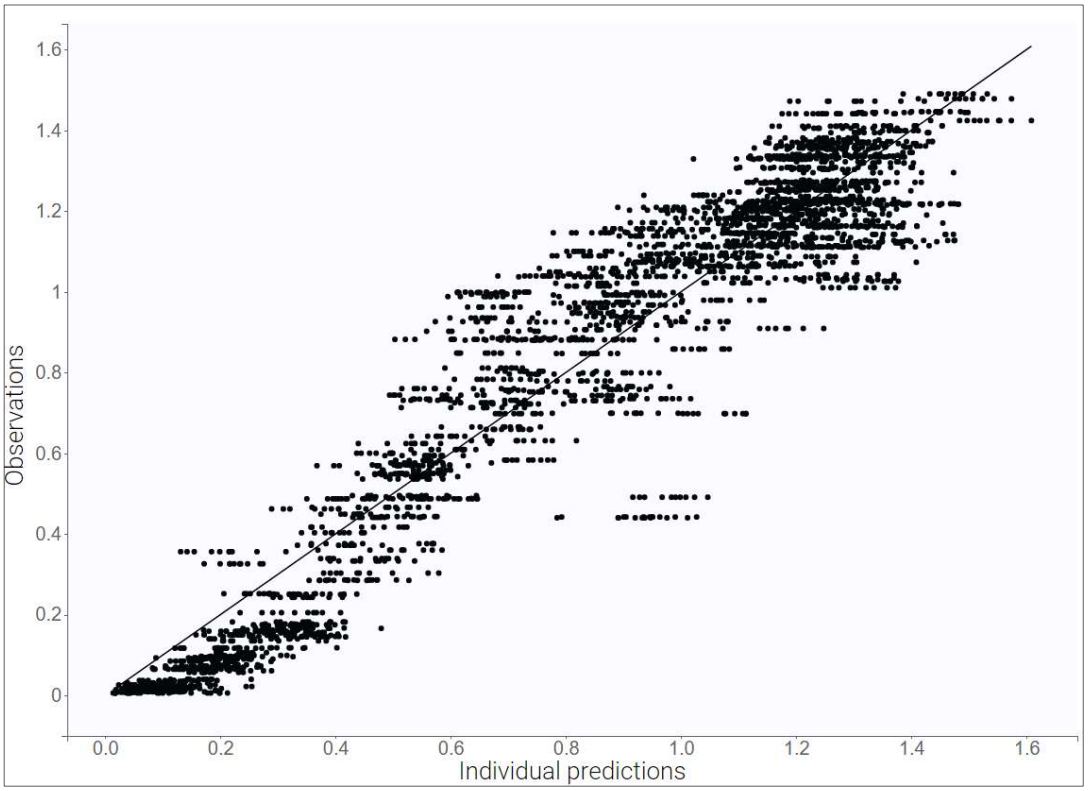


Figure 4.2: The observed against the individual prediction.

Figure 4.1 above illustrates the plot of the observed data plotted against the population model prediction and Figure 4.2 illustrates the plot of the observed data against the individual predictions. The points in the observed vs population predictions plot are scattered around the identity line. The points in the observations vs individual predictions plot are evenly distributed around the identity line. It is evident from both plots in Figure 4.1 and Figure 4.3 that the model adequately represents the data. The distribution of points around the identity line is narrower for the observation vs individual predictions when compared to that of the observations vs population predictions. This affirms that the variability between-subjects is high and the data is better modelled at the individual level.

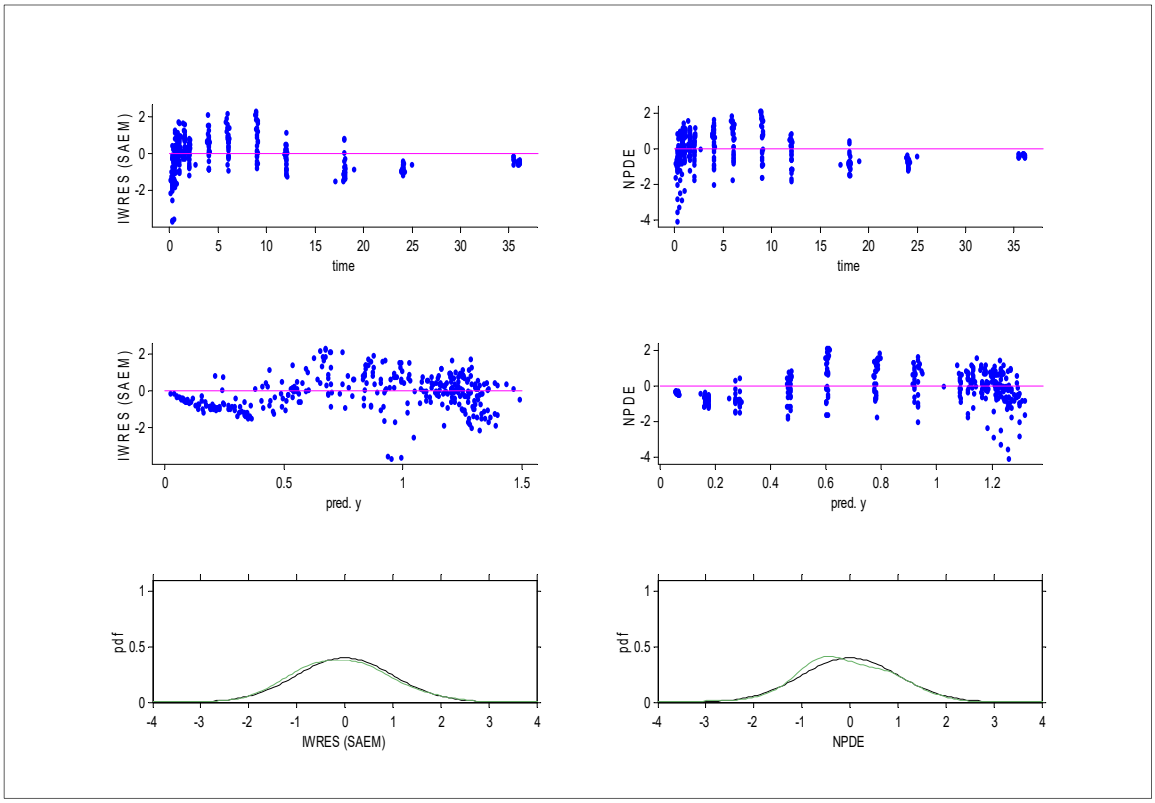


Figure 4.3: The plot of the individually weighted residuals against the sampling time

Figure 4.3 above portrays the plot of the NPDE and IWRES of the model. The first two plots on top are for IWRES and NPDE plotted against sampling time, the second plots in the middle are for IWRES and NPDE plotted against predicted values, and the last plots at the bottom illustrate the distributions of IWRES and NPDE over the theoretical normal distribution (dark line). Both the IWRES and NPDE show a higher amplitude at the beginning, which declines over time, this is common with the concentration data (which declines over time). The plot of the IWRES against the predicted values shows a presence of heteroscedasticity. The plot of the NPDE against the predicted values also shows a presence of heteroscedasticity, noted by a funnel like shape. The IWRES follows a normal distribution. The distribution of the NPDE shows a slight deviation from the theoretical normal distribution. The NPDE distribution is skewed to the left, indicating that on average the model is over-predicting. This is also confirmed by the behavior displayed in NPDE vs predicted plot.

The random effects are assumed to be normally and independently distributed. The box plots were used to summarise the empirical distributions of the standardised simulated random effects. The Shapiro –Wilk test was used to test the normality assumption of each random

effect. Pearson's correlation test was used to test the correlation between the random effects. The results of the Shapiro –Wilk and Pearson's correlation tests are respectively presented in Table 4.16 and Table 4.17 below.

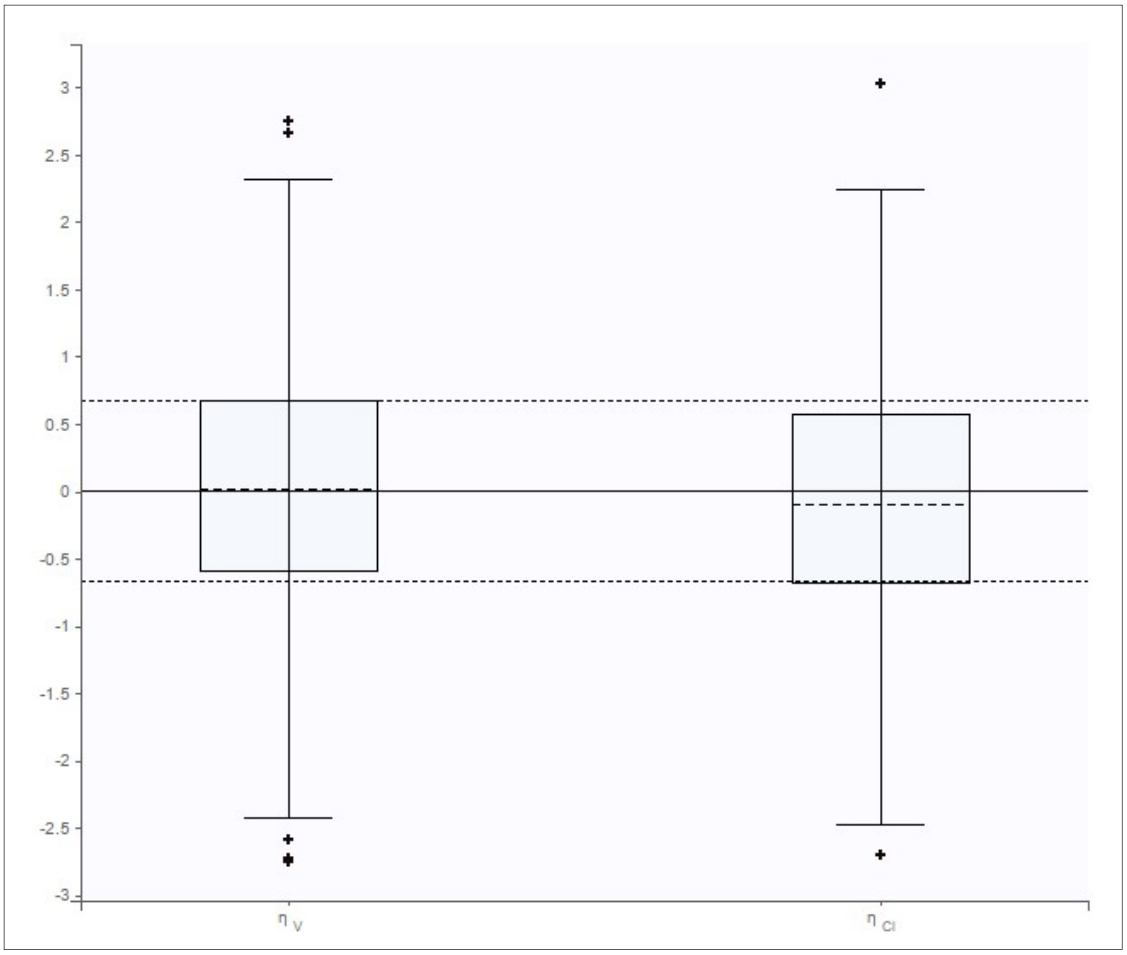


Figure 4.4: The box plots of the empirical distributions of the standardised simulated random effects.

The standardised empirical distribution of the random effect is illustrated in Figure 4.4 above. The horizontal lines show the theoretical quantiles of the standard normal distribution. The bottom dotted line represents the first quantile, the solid line represents the median and the top dotted line represents the third quartile. The boxplot represents the simulated empirical distribution of each random effect. The horizontal line segmenting the boxplot are quartiles of the simulated empirical distribution, the bottom and the top lines represent the first and third quartiles respectively. The median is represented by the inner line and the extremities by the whiskers. The empirical distribution of the volume and clearance random effects are similar to

the theoretical distribution. Consequently, the choice of the log-normal model for the parameters seems to be correct.

Table 4.16 presents the Shapiro –Wilk test for testing the normality assumption of the random effects. The null hypothesis of the test is:

H_0: the random effects are normally distributed.

Vs the alternative:

H_1: the random effects are not normally distributed.

Table 4.16: The Shapiro –Wilk tests of normality.

Parameter	Test-statistic	*P*-value
η_v	0.9604	0.2404
η_{Cl}	0.9638	0.3924

For the volume random effect, the *P*-value is 0.24, and 0.39 for the clearance random effect. Both *P*-values are higher than 0.05, hence at the 5% significance level, the normality assumption of the random effects cannot be rejected.

Table 4.17 below displays the results of Pearson's correlation test for testing the independence of the random effects. The null hypothesis is:

H_0: the correlation between the random effects of the clearance and volume is zero.

Vs the alternative:

H_1: the correlation between the random effects is different from zero.

Table 4.17: The Pearson's correlation test.

Coefficient	Test-statistic	*P*-value
-0.0417	-0.2160	0.6801

At the 5% significance level, the null hypothesis cannot be rejected (*P*-value=0.68>0.05). Therefore, the assumption of independence of random effects cannot be questioned. Which also justifies the use of the diagonal variance covariance matrix.

4.3.3 Average Bioequivalence Based on AUC Estimated by the NLMEM Method

The individual parameter estimates for the model specified in (3.1) for each subject were used to estimate $AUC_{0-\infty}$ using Equation (3.2). The estimated $AUC_{0-\infty}$ and $\ln AUC_{0-\infty}$ values are in Table 4.18 below.

Table 4.18: $AUC_{0-\infty}$ and $\ln AUC_{0-\infty}$ values for each pig estimated by the NLMEM Model.

Sequence	Pig	Period		Period	
		I	II	I	II
		$AUC_{0-\infty}$	$AUC_{0-\infty}$	$\ln AUC_{0-\infty}$	$\ln AUC_{0-\infty}$
TR	6	14.8403	10.4371	2.6973	2.3454
	10	17.4101	17.2135	2.8570	2.8457
	12	16.3532	12.02443	2.7944	2.4869
	13	15.7694	11.2663	2.7581	2.4218
	14	17.4502	13.4485	2.8593	2.5989
	23	18.07599	14.08292	2.8946	2.6450
	25	17.8171	12.8386	2.8802	2.5525
RT	4	16.5240	11.9289	2.8048	2.4790
	9	17.5426	17.3136	2.8646	2.8515
	11	17.1004	14.7037	2.8391	2.6881
	15	16.9365	13.7866	2.8295	2.6237
	18	18.05967	15.1405	2.8937	2.7174
	19	17.8183	13.8068	2.8802	2.6252
	24	18.6937	13.7540	2.9282	2.6213

Table 4.18 presents the results for $AUC_{0-\infty}$ for both periods as well as for both sequences. The results for $\ln AUC_{0-\infty}$ are also presented for both periods and both sequences. These results are used in the analyses in Table 4.19.

Table 4.19: The ANOVA for $\ln AUC_{0-\infty}$ estimated by the NLMEM Model.

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	15	0.6314	0.04209	6.69	0.001
Error	12	0.07550	0.006291		
Corrected Total	27	0.7069			
R-Square	Coeff Var	Root MSE	$\ln AUC_{0-\infty}$ Mean		
0.8932	2.9114	0.07932	2.7244		
Source	DF	Type I SS	Mean Square	F Value	Pr > F
Sequence	1	0.03637	0.03637	5.78	0.0332
period	1	0.3840	0.3840	61.03	<.0001
formulation	1	0.006030	0.006030	0.96	0.3469
Pig	12	0.2051	0.01709	2.72	0.04820

The model is significant (*P*-value=0.001). The fit R^2 is 0.8932, which means that 89.32% of the variation is explained by the model. The CV is 2.91%, which is an indication that the data are not highly variable. There is a sequence effect and a period effect, since both *P*-values are below 0.05. There is however no formulation effect (*P*-value=0.3469). Based on the ANOVA results in the above Table 4.19 acquired from SAS using the Proc GLM, the pooled sample variance ($\widehat{\sigma_d}^2$) is 0.006291 and $\widehat{\sigma_d}$ is 0.07932. The $\widehat{\sigma_d}$ value is calculated to be used in calculating the CI in Table 4.21.

Table 4.20: The lsmeans for $\ln AUC_{0-\infty}$ and the geometric means.

Formulation	$\ln AUC_{0-\infty}$ lsmeans	Standard error	Geometric means for $AUC_{0-\infty}$
R	2.7097	0.02120	15.02522
T	2.7391	0.02120	15.4727

The R and T formulation geometric means and lsmeans for the log-transformed $AUC_{0-\infty}$ are presented in Table 4.20 above, which are being used for the analysis in Table 4.21 below.

Table 4.21: The classic CI for lsmeans of $\ln AUC_{0-\infty}$ and the geometric means.

$\ln AUC_{0-\infty}$	$\widehat{\sigma}_d^2$	lower bound	90% CI lower limit	90% CI upper limit	upper bound
CI for difference of the means (Log-transformed data)	0.006291	-0.2231	0.0007885	0.05791	0.2231
CI for the ratio of the geometric means		0.80	1.00079	1.05962	1.25

Table 4.21 above depicts the classic CI for lsmeans of $\ln AUC_{0-\infty}$ and the geometric means. The 90% CI for the difference of the formulation means of the log-transformed data is (0.0007885; 0.05791) and the resulting 90% CI for the geometric means ratio is (1.00079; 1.05962). Both 90% CI are contained within the respective acceptance range of (-0.2231, 0.2231) and (0.80 and 1.25), hence, ABE is concluded. The R and T formulations are therefore ABE when employing the classic CI approach.

Table 4.22: The TOST for geometric means under the NLMEM Model.

N	Geometric Mean	Coefficient of variation	Minimum	Maximum
14	1.0298	0.0865	0.9163	1.2208

Geometric Mean	95% CL Mean		Coefficient of variation	95% CL CV	
1.0298	0.9797	1.0824	0.0865	0.0626	0.1397

Geometric Mean	Lower Bound	90% CL Mean		Upper Bound	Assessment
1.0298	0.8	0.9886	1.0727	1.25	Equivalent

Test	Null	DF	*t* Value	P-value
Lower	0.8	13	10.94	<.0001
Upper	1.25	13	-8.4	<.0001
Overall				<.0001

Table 4.22 above depicts the TOST for geometric means under the NLMEM Model wherein the 90% CI (0.9886; 1.0727) of the geometric means ratio for AUC lies entirely in the acceptance range, thus ABE is accepted. The R and T formulations are ABE based on the TOST method under ± 20 rule.

Both the 90% CI approach and the TOST confirm that the R and T formulations are ABE for $\ln AUC_{0-\infty}$.

4.4 Population Bioequivalence Analysis

Equation (2.27) was used to find the value of θ_{PBE}, as recommended by FDA (2001), $\theta_P = 1.7448$. The actual PBE test is presented in Table 4.23 below.

Table 4.23: The Population Bioequivalence Test

NCA	θ_{PBE}	θ_P	Conclusion
$\ln AUC_{0-k}$	0.003774	1.7448	PBE
$\ln AUC_{0-\infty}$	0.003491		PBE
NLMEM			PBE
$\ln AUC_{0-\infty}$	0.04070		PBE
$\ln C_{\max}$	0.002604		PBE

Table 4.23 above presents the PBE test results using $\ln AUC_{0-k}$ and $\ln AUC_{0-\infty}$ estimated by the NCA method. The results for $\ln AUC_{0-\infty}$ estimated by the NLMEM and $\ln C_{\max}$ are also available. From Table 4.23 all parameters are less than 1.7448 PBE limit and consequently H_0^{PBE} is rejected. Likewise, PBE of the R and T formulations have been achieved using the moment-based criterion. The value of θ_{PBE} is larger with NLMEM method. That may be due to a higher difference in the formulation means of the NLMEM estimates. Nonetheless, PBE is achieved utilizing either NLMEM or NCA method. The conclusion of PBE implies that the T formulation can safely and effectively be prescribed to a patient.

4.5 Simulation Results

Four trial designs were used to simulate the concentration data as detailed in Section 3.5.2 (the rich design, the original design, the sparse design and the intermediate design). With each design two variability settings were used (S_{ll} and S_{hl}) but for the intermediate design an extra variability setting S_{hh} was added. To introduce a difference in bioavailability between the R and T formulations, the trials for the T formulation were simulated under two hypotheses the $H_{0;80\%}$ and $H_{0;125\%}$, as detailed in Section 3.5.1. The plots of the simulated trials are presented in the Appendix C (from Figure A.29 to Figure A.42). A small value of $0.0001 mg/l$ as recommended by the (FDA, 2013) was assigned to the concentration values that were less than zero after simulating.

With the NCA method it is possible to have missing AUC_i values due to increasing last concentrations. The proportion of these missing AUC_i values may vary for each design, each variability setting and for the two hypotheses (Dubois, et al., 2010). It is worth noting that for this study, there were no missing values for the sparse, original, rich and intermediate design simulated under both hypotheses and all variability settings (S_{ll} , S_{hl} and S_{hh}).

4.5.1 The TOST Type-I-Error Analysis.

The TOST was conducted on $\ln AUC_{0-\infty}$ estimated using the two methods (NCA and NLMEM) for each trial. These methods were compared based on the rate of the Type-I-error acquired from testing ABE on 1000 replicates of each design. The results of this analysis are presented in Table 4.24 below.

Table 4.24 The TOST Type-I-error rate for NCA and NLMEM.

Variability settings	Hypothesis	Sparse design (N=50,n=3)		Original design (N=14,n=12)		Rich design (N=50,n=12)		Intermediate design (N=30,n=5)	
		NCA	NLMEM	NCA	NLMEM	NCA	NLMEM	NCA	NLMEM
S_{ll}	$H_{0;80\%}$	5.6	20	6	10.9	5.0	7.9	4.4	16.8
	$H_{0;125\%}$	6.4	2.2	5.9	12.7	5.0	10.5	5.8	12.4
S_{hl}	$H_{0;80\%}$	6.3	14.2	5.1	8.6	6.1	6.8	5.2	7
	$H_{0;125\%}$	4.8	1.8	5.0	6.8	6.0	6.8	5.8	5.5
S_{hh}	$H_{0;80\%}$							0	73
	$H_{0;125\%}$							0	99

The 95% prediction interval for 5% Type-I-error obtained from 1000 replicates is (3.7%; 6.4%). For the tests performed on the NCA estimates, the Type-I-error is contained within the 95% prediction interval for all designs simulated with S_{ll} and S_{hl} and for both hypotheses. The Type-I-error is at the minimum (0%) for the intermediate design simulated with the S_{hh}. Furthermore, the Type-I-errors are similar for both hypotheses for each design variability combination on NCA estimates.

On the other hand, with NLMEM estimates, the Type-I-errors are distinct for the two hypotheses and mostly lie outside of the 95% prediction interval. For example, it can be observed in Table 4.24 that the Type-I-error is extremely large for the sparse design simulated under $H_{0;80\%}$and small under $H_{0;125\%}$ for the two variability settings. Similarly, the comparative results are seen for S_{hh}, where the Type-I-error is zero when the tests are performed on the NCA estimates and surprisingly huge when the NLMEM estimates are performed.

Figure 4.5 presents the global Type-I-error of the bioequivalence tests, obtained as the biggest Type-I-error between the two Type-I-errors from each hypothesis ($H_{0;80\%}$ and $H_{0;125\%}$).

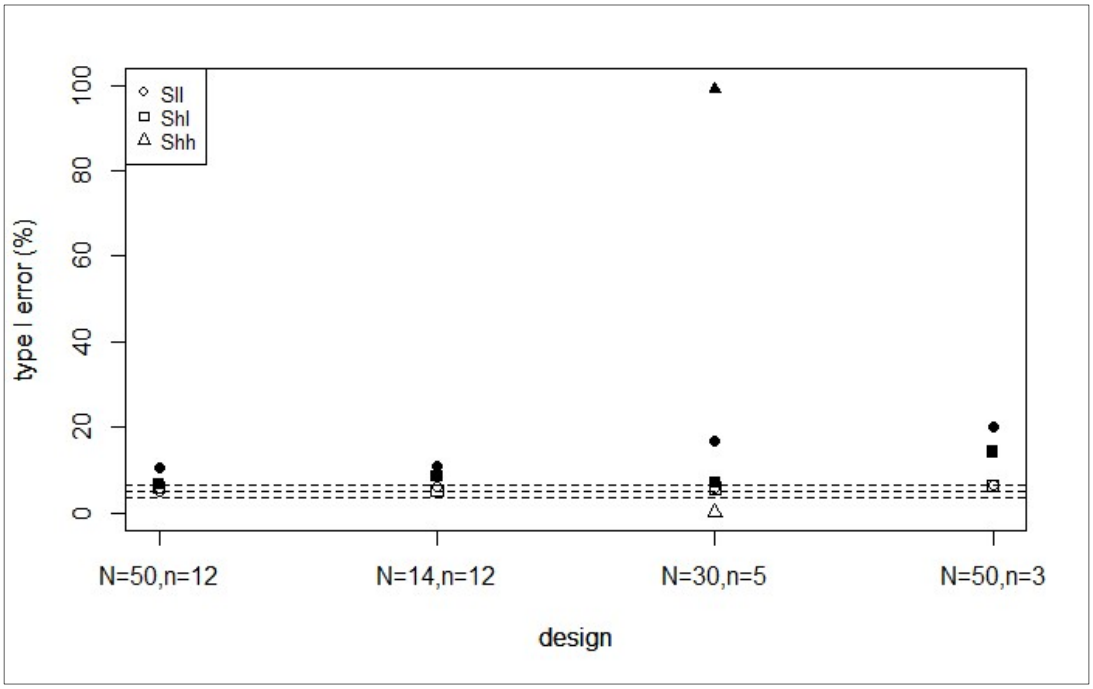


Figure 4.5: The global Type-I-error for all variability settings.

In Figure 4.5, *N* and *n* represent the number of subjects and the number of samples per subject, respectively. The combination of *N* and *n* represents the simulated designs as in Table 4.24. The filled symbols represent the global Type- I- error acquired using the NLMEM estimates and the non-filled symbols utilising the NCA estimates. The dotted lines represent the 5% Type-I-error and its 95% prediction interval.

The S_{hh} is the only variability setting where the global Type-I-error lies below the 95% prediction-interval. For all designs simulated with S_{ll} and S_{hl},, the global Type-I-error lies within the 95% prediction interval for tests performed on NCA estimates. For S_{ll}, the global Type-I-error increases with a dwindled number of samples per subject.

The global Type-I-error of the tests performed on NLMEM, estimates is lower for S_{hl} compared to those of S_{ll} and S_{hh}. The global Type-I-error is high for the sparse design and highly inflated for S_{hh} on the NLMEM estimates. Similar to the result on the NCA estimates, the global Type-I-error increases with the decreasing number of samples per subject for S_{ll}.

4.5.2 The Shrinkage Evaluation for the NLMEM estimates

Figure 4.6 presents global Type-I-error of the tests based on the NLMEM estimates over the median shrinkage.

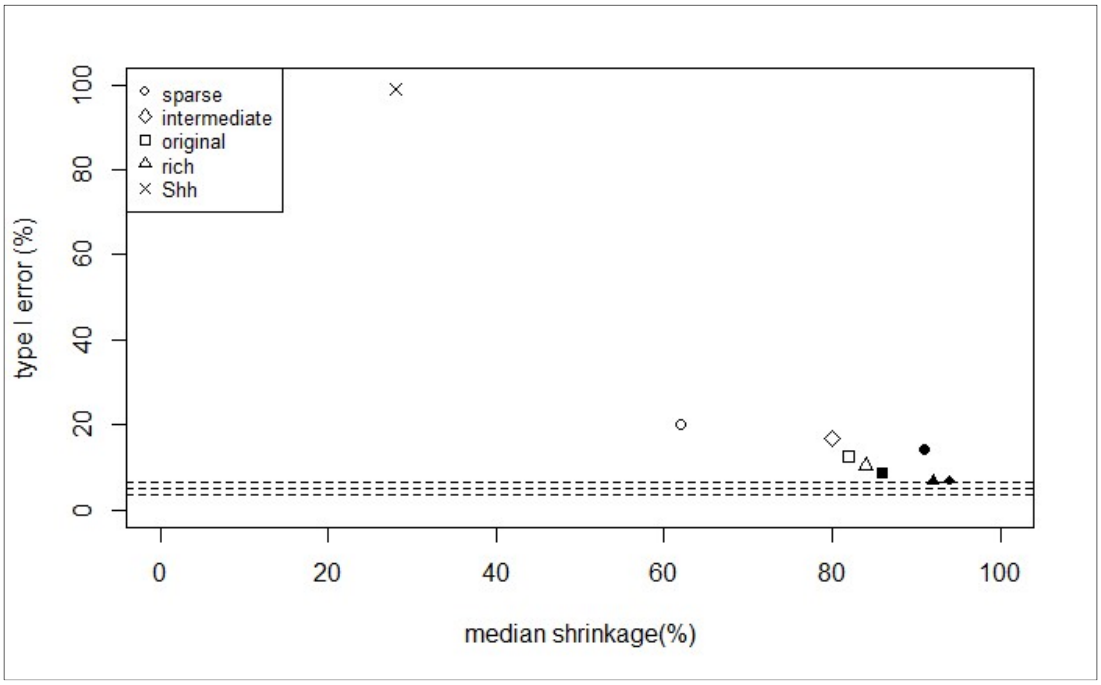


Figure 4.6: The Type-I-error vs shrinkage for the different designs and variability settings.

In Figure 4.6, the filled and non-filled symbols represent the S_{hl} and S_{ll}, respectively. The symbol × represents the intermediate design simulated with the S_{hh} variability setting.

Apart from S_{hh}, the median shrinkage is high in all designs and across the variability settings. There is a relationship between the median shrinkage and variability settings. In Figure 4.6 the shrinkage for S_{hl} is higher than that of S_{ll} for all designs. Unexpectedly, the shrinkage is 28% for S_{hh}. This value is small when compared to the other variability settings and considering that, the residual variability is high and individual data is limited. For S_{ll} the median shrinkage increases with an increasing number of samples per subject, and with the decreasing global Type-I-error. The shrinkage is high (above 20%) for all designs and variability settings.

4.5.3 The Bias and Root Mean Square Error Assessment

The NCA and the NLMEM estimates were compared based on the performance of the NLE model used to analyse the $\ln AUC_{0-\infty}$ data. As described in Section 3.10, a better method must have a trivial bias and RMSE. The plots of the RMSE for each design and each variability setting for the NCA and NLMEM methods are respectively, presented in Figure 4.7 and Figure 4.8 below. The values of the bias and associated CI for each design and variability setting are given in Table 4.25.

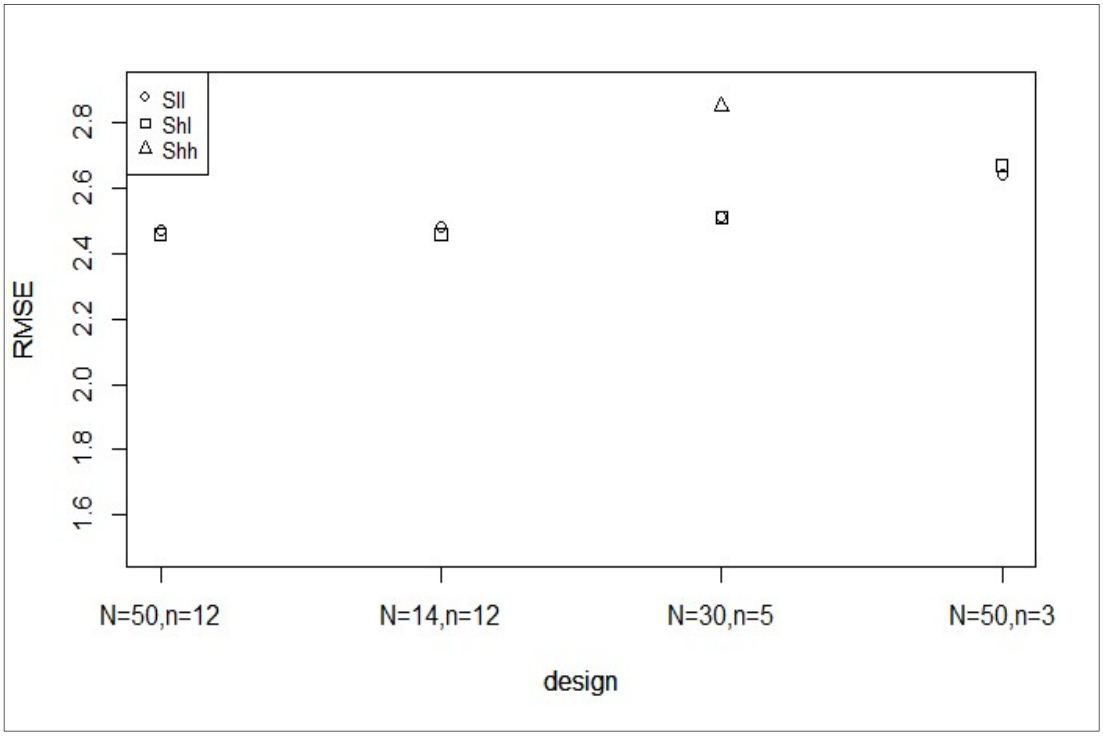


Figure 4.7: The RMSE for each design and each variability settings for the NCA estimates.

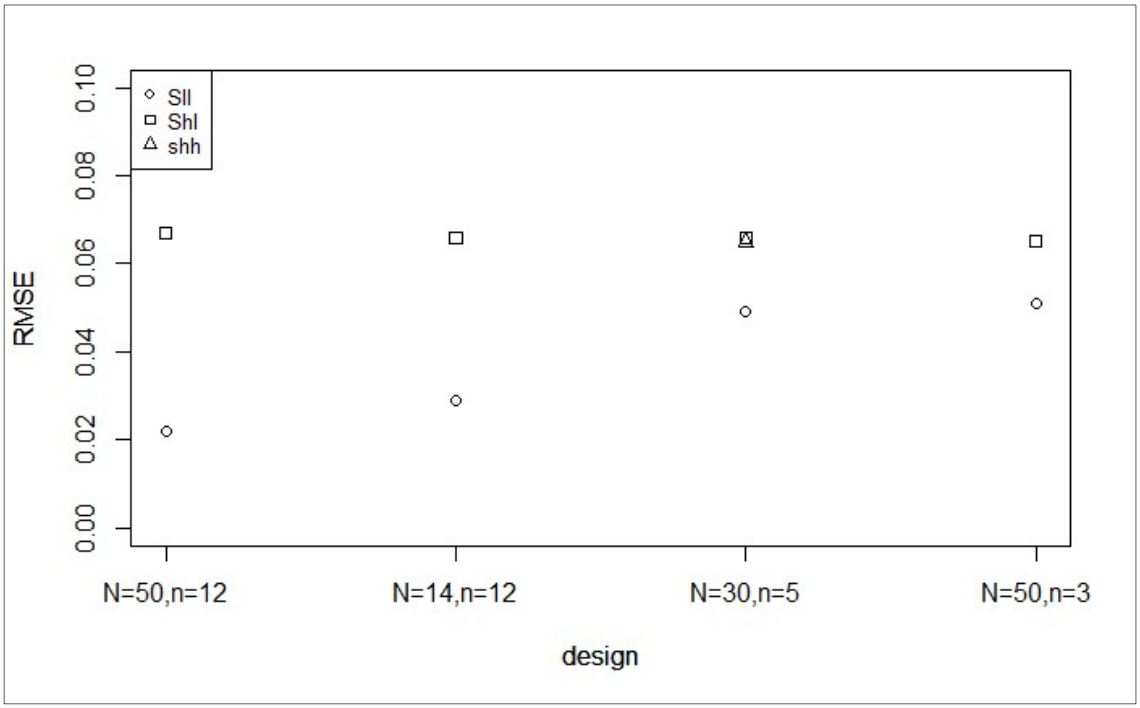


Figure 4.8: The RMSE for each design and each variability settings for the NLMEM estimates.

Similar to Figure 4.5, the combination of *N* and *n* in Figure 4.7 and Figure 4.8 represents a design.

For the NCA estimates (Figure 4.7), the RMSE is relatively similar for S_{ll} and S_{hl} in each design. The RMSE is slightly higher for S_{hh} compared to S_{ll} and S_{hl}. The RMSE increases when the number of samples per subject decreases.

The RMSE is small for the NLMEM estimates (Figure 4.8) as compared to that of the NCA estimates. It can be observed from Figure 4.8 that the RMSE of the NLMEM estimates is higher for S_{hh} and S_{hl} compared to S_{ll}. In Figure 4.8 for S_{ll}, the RMSE slightly increases when the number of samples per subject decreases. For S_{hl}, the RMSE is similar for the four designs.

Table 4.25: The Bias for the mean of $\ln AUC_{0-\infty}$.

Methods		Sparse design (N=50, n=3)	Original design (N=14, n=12)	Rich design (N=50, n=12)	Intermediate design (N=30, n=5)
		Bias (95% CI)	Bias (95% CI)	Bias (95% CI)	Bias (95% CI)
S_{ll}	NCA	2.64 (2.61;2.67)	2.47 (2.40;2.55)	2.47 (2.43;2.51)	2.51 (2.45;2.56)
	NLMEM	1.04E-17 (-3.01E-16; 3.22E-16)	8.25E-19 (-3.38E-16; 3.40E-16)	3.73E-18 (-4.12E-16; 4.049E-16)	-1.30E-18 (-3.94E-16; 3.91E-16)
S_{hl}	NCA	2.65 (2.55;2.75)	2.43 (2.25;2.61)	2.43 (2.34; 2.53)	2.48 (2.35;2.61)
	NLMEM	9.12E-18 (-5.19E-16; 5.37E-16)	1.98E-17 (-4.81E-16; 5.21E-16)	9.77E-18 (-5.050E-16; 4.86E-16)	3.94E-18 (-4.99E-16; 5.071E-16)
S_{hh}	NCA				2.62 (2.21;3.03)
	NLMEM				-2.035E-18 (-4.33E-16; 4.29E-16)

For all the designs and all the variability settings, the bias of the NCA estimates is significantly high, i.e. all of the 95% CI's of the biases did not include zero. The bias decreases when the number of samples per subject increases. However, the bias remained constant across variability settings. For the sparse design, the bias is 2.64, and 2.65 for S_{ll} and S_{hl}, respectively. For the intermediate design, the bias is 2.51, 2.48 and 2.62 for S_{ll}, S_{hl} and S_{hh}, separately. For the original design, the bias is 2.47and 2.43 for S_{ll} and S_{hl}, discretely. For the rich design, the bias is 2.47 and 2.43 for S_{ll} and S_{hl}, correspondingly. The NCA bias results are similar for S_{ll} for the original design and the rich design. The NCA bias results for S_{hl} are also similar for the original design and the rich design.

For the NLMEM estimates, the bias is small for all the designs and all the variability settings. The bias of the NLMEM is not significantly different from zero. That is the 95% CI of the NLMEM estimates includes zero.

4.6 Summary of the Results

The objective of the study was to compare the bioequivalence tests based on the NCA and NLMEM estimates. In line with this, the study captures and demonstrates cases wherein employing the NLMEM method or the NCA method was more appropriate and beneficial. A correlation analysis was conducted utilising simulated data wherein the results of the ABE test based on the NCA or the NLMEM method were compared based on the Type-I-error, RMSE and bias.

Similar results of ABE were observed for the fourteen healthy pigs when using the NCA and NLMEM methods. ABE was concluded using $\ln C_{\max}$, which is independent of the two methods ($\ln C_{\max}$ is observed directly from the data). Furthermore, PBE was accepted when using $\ln C_{\max}$ and for both estimation methods when using $\ln AUC_{0-\infty}$. Hence, a subject can safely and effectively be introduced to either the R or the T formulation. Based on results obtained from the original data of fourteen healthy pigs, the NCA and the NLMEM estimates yielded similar conclusions of ABE and PBE which was also achieved using C_{max}.

Table 4.24 detailed the Type-I-error rate of the TOST conducted on the simulated trials wherein the trials were simulated under two null hypotheses $H_{0;80\%}$ and $H_{0;125\%}$. The Type-I-error with the NCA estimates was similar and close to 5% for all designs simulated with S_{ll} and S_{hl} variability settings, for both hypotheses. On the other hand, the Type-I-errors with the NLMEM estimates were different for the two hypotheses and in most instances the estimates were outside the 95% prediction interval of the 5% Type-I-error. If the tests were done on the true values of AUC the Type-I-error would be identical for the two null hypotheses, since the two hypotheses are symmetric. On account of estimates, the Type-I-error for both hypotheses might not be indistinguishable. Subsequently, the largest Type-I-error (global Type-I-error) is used for comparison purposes.

From Figure 4.5, it was evident that for S_{ll} the global Type-I-error increases with a decreasing number of samples per subject, for both NCA and NLMEM estimates. This implies that when the variabilities are low, the residual error is low tests based on the NCA, and NLMEM estimates improve as the number of samples per subject increases. The global Type-I-error is

highly inflated for S_{hh} when using the NLMEM estimates and very low for the NCA estimates. That is when the variabilities and the residual error are high the TOST procedure perform poorly on the NLMEM estimates.

With the NCA method, the global Type-I-error was low and in most cases close to 5%. On the other hand, for the test based on the NLMEM estimates, the global Type-I-error was inflated and fell outside of the 95% prediction interval of the 5% Type-I-error. Therefore, it is concluded that based on the Type-I-error comparison, better TOST results are achieved when using the NCA estimates as compared to the NLMEM estimates.

Moreover, the NLMEM estimates are known to be affected by shrinkage (reference). When the shrinkage is high (>20%) the evaluation based on the NLMEM estimates may lack the power to detect the exposure-response relationship (Bertrand, et al., 2009). Indeed, from Figure 4.6 it was observed that the shrinkage was high and for all designs the shrinkage was greater than 20%.

Figure 4.7 and Figure 4.8 show the RMSE obtained from fitting the model specified in (3.3) on the $\ln AUC_{0-\infty}$ of the R formulation for NCA and NLMEM methods, respectively. Table 4.25 contains the bias and its 95% CI for the R formulation for each design and each variability settings. The bias of the NCA estimates is large and significantly different from zero, whereas the bias of the NLMEM estimates is negligible. The RMSE of the NCA estimates is large in comparison to the NLMEM estimates. For the NCA estimates, the RMSE decreases when the number of samples per subject increases. Furthermore, the RMSE is slightly higher for S_{hh} compared to S_{ll} and S_{hl} variability settings. With the NLMEM estimates, the RMSE is higher for S_{hh} and S_{hl} compared to S_{ll}.

CHAPTER 5 CONCLUSION AND RECOMMENDATION

5.1 Introduction

This study comprised five chapters wherein it compared two methods of estimating the bioequivalence parameter (ln*AUC*), the general non-compartment method (NCA) and the non-linear mixed effect model (NLMEM). In the case of NCA approach, ln*AUC* was estimated using the linear trapezoidal rule. In the case of NLMEM method, the parameters of the NLMEM were estimated using the SAEM algorithm implemented in the Monolix 2.4 software and ln*AUC* was then obtained using the NLMEM parameter estimates. Furthermore, the study demonstrated how bioequivalence can be tested using both the NCA estimates and the NLMEM estimates. This study also compared the two methods using simulated data by comparing the Type-I-error of the TOST, bias, and RMSE of the model used to analyse ln*AUC* data. For the simulated data only the ABE tests are used to compare the two methods.

This final chapter presents conclusions and recommendations in line with the objectives of the study.

5.2 Conclusion

The NCA estimates are widely used in bioequivalence analysis (Panhard and Mentre, 2005). However, the drawback of this is that in studies with limited observations, the NCA estimates tend to be biased. Bias estimates may lead to incorrect conclusions of bioequivalence of the two formulations. Researchers have suggested the use of the NLMEM estimates for bioequivalence analysis, with the use of the SAEM algorithm to improve the precision and accuracy of the PK parameter estimates. Because of the scarcity of real data in this field, researchers have suggested the use of simulated data to compare the NCA and the NLMEM methods.

Dubois et al. (2010) and Panhard and Mentre (2005), noted that the bias and precision for ln*AUC* obtained using the NCA method are dependent on the individual level information, and for the sparse design the bias is high. This study corroborates the findings of Dubois, et al. (2010) and Panhard and Mentre (2005), the NLMEM estimates provided satisfactory results in terms of bias and RMSE, for a study with sparsely sampled subjects. Furthermore, for all the

designs (sparse, intermediate, original and rich design) the NLMEM estimates provided the results with smaller RMSE and smaller bias. In this study, the accuracy (RMSE) of the NCA estimates are better when the sampling intensity for each subject increased. On the other hand, RMSE of the NLMEM estimates increased slightly with a decreasing number of samples per subject. However, both RMSE and bias were found negligible.

Xu et al. (2012) state that EBEs are widely accepted for estimating posterior random effects for individuals. However, the trade-off of this approach is its tendency to shrink estimates towards a population mean, which lead to the poor performance of the EBEs' estimator. In addition, high shrinkage may lead to lower power to detect an exposure–response relationship. As a result, the NLMEM estimates have higher Type-I-errors than tests based on the NCA estimates when the shrinkage is high. In other words, the bioequivalence is obtained more easily with the NLMEM estimates when the shrinkage is high. Indeed, in this study, the shrinkage in $\ln AUC$ is high and consequently, the Type-I-error of the TOST was inflated with the NLMEM estimates. For the test based on the NLMEM estimates the global Type-I-error was above 10%, and the shrinkage above 60%. These results were also inline with results obtained by Bertrand et al. (2009). In that study, they evaluated the ANOVA results conducted on EBEs to test the effect of a single nucleotide polymorphism on a PK parameter. They demonstrated the influence of the shrinkage on the power of ANOVA. The power is reduced when the shrinkage increases. Henceforth, it was difficult to detect the existing group differences if the shrinkage was high.

The study confirmed that for $S_{l,l}$ the global Type-I-error decreases with an increasing number of samples per subject for the NCA estimates. In other words, when the number of samples per subject is increasing, and the variability is low the NCA estimates perform better. These results are consistent with the study of Dubois et al. (2010). In that work, they demonstrated that when the sampling intensity for each subject is high and the variability is low, tests conducted on the NCA estimates yielded satisfactory outcomes and remain favourable due to the simplicity of this method.

It was evident that for the intermediate design simulated with S_{hh} the Type-I-error is 0% for the NCA estimates, but the RMSE is the highest. Though for the tests based on the NLMEM estimates, the Type-I-error is at the highest 99%, the RMSE is the smallest amongst other

design and variability combinations. This is expected when the number of samples per subject is small, the residual variability is high. Both the NCA and the NLMEM estimates may lead to ambiguous results.

In summary, the test based on the NCA provided satisfactory global Type-I-errors for the sparse, original and rich design. However, the bias and RMSE in these estimates were high. The RMSE of the NCA estimates got better as the number of sample per subjects increased. Both the global Type-I-error and shrinkage of the NLMEM estimates were high. The shrinkage increased when the between-subject and within-subject variability was high, and the residual variability was low. The global Type-I-error with the NLMEM estimates decreased when the between-subject and within-subject variability increased.

5.3 Limitations

This study has several limitations. Firstly, the simulation methodology used is based on a one-compartment PK model. However, a multi-compartmental model can be considered to improve the results. Secondly, in this study, the period effect analysis was not considered, as it is not the main goal of the bioequivalence analysis. However, they are often tested and removed from the analysis. Thirdly, this study relies on the sampling scheme provided by Panhard and Mentre (2005). It is important to note that, for simulation studies, the sampling scheme can be optimised for a particular drug. This is done to better characterise the change in the distribution and clearance of the drug. Such an exercise is beyond the scope of this study. Fourthly, for simplicity, only the parametric methods were considered for the bioequivalent tests. Lastly, the study did not explicitly demonstrate how small the sample per subject should be for NCA method to fail.

5.4 Recommendations

The NCA method provided satisfactory bioequivalence test results. However, the results must be scrutinized when the information at the individual level is limited. This study therefore recommends that when the shrinkage is high the NLMEM approach for estimating PK parameters for bioequivalence analysis should be used with caution.

REFERENCES

Act 21CFR Part 320.1. (2009) Available at: https://www.govinfo.gov/content/pkg/CFR-2009-title21-vol5/pdf/CFR-2009-title21-vol5-sec320-1.pdf (Accessed: 24 June 2019)

Anderson, S. & Hauck, W. W. (1990) Consideration of individual bioequivalence. *Journal of Pharmacokinetics and Biopharmaceutics,* vol.18, no.3, pp. 259-273.

Berger, R.L. & Hsu, J.C. (1996) Bioequivalence trials, intersection-union tests and equivalence confidence sets. *Statistical Science*, vol.11, no.4, pp. 283-319.

Bergstrand, M., Hooker, A.C., Wallin, J.E. & Karlsson, M.O. (2011) Prediction-corrected visual predictive checks for diagnosing nonlinear mixed-effects models. *The AAPS Journal*, vol.13, no.2, pp. 143-151.

Bertrand, J., Treluyer, J.M., Panhard, X., Tran, A., Auleley, S., Rey, E., Salmon-Céron, D., Duval, X., Mentré, F. & COPHAR2-ANRS 111 study group. (2009) Influence of pharmacogenetics on indinavir disposition and short-term response in HIV patients initiating HAART. *European journal of clinical pharmacology*, vol.65, no.7, pp. 667-678.

Birkett, D. J. (2003) Generics-equal or not? *Australian Prescriber,* vol.26, no.4, pp. 85-86.

Chen, M. L., Patnaik, R., Hauck, W. W., Schuirmann, D. J., Hyslop, T. & Williams, R. (2000) An individual bioequivalence criterion: regulatory considerations. *Statistics in medicine*, vol. 19, no.20, pp. 2821-2842.

Chow, S. C., Shao, J. & Wang, H. (2003) Statistical tests for population bioequivalence. *Statistica Sinica,* vol.13, no.2, pp. 539-554.

Chow, S.C. & Liu, J.P. (2004) *Design and Analysis of Bioavailability and Bioequivalence Studies.* 3rd ed. Durham: CRC Press, 760 pages.

Chow, S.C., Wang, H. & Shao, J. (2007) *Sample size calculations in clinical research.* 2nd ed. CRC Press, 480 pages.

Comets, E. & Mentré, F. (2001) Evaluation of tests based on individual versus population modelling to compare dissolution curves. *Journal of biopharmaceutical statistics*, vol. 11, no. 3, pp.107-123.

Comets, E., Brendel, K. & Mentré, F. (2008) Computing normalised prediction distribution errors to evaluate nonlinear mixed-effect models: the npde add-on package for R. *Computer methods and programs in biomedicine*, vol.90, no.2, pp. 154-166.

Comets, E., Brendel, K. & Mentré, F. (2010) Model evaluation in nonlinear mixed effect models, with applications to pharmacokinetics. *Journal de la Société Française de Statistique*, vol.151, no.1, pp.106-128.

Davidian, M. & Giltinan, D.M. (1995) *Nonlinear models for repeated measurement data.* vol.62. London: CRC Press, 360 pages.

Department of Health (2006) *Guidelines for Good Practice in the Conduct of Clinical Trials with Human Participants in South Africa*. Department of Health: Pretoria, South Africa.

Delyon, B., Lavielle, M. & Moulines, E. (1999) Convergence of a stochastic approximation version of the EM algorithm. *Annals of statistics*, vol. 27, no. 1, pp. 94-128.

Dempster, A.P., Laird, N.M. & Rubin, D.B. (1977) Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 39, no.1, pp. 1-38.

Dubois, A., Gsteiger, S., Pigeole, E. & Mentré, F. (2010) Bioequivalence Tests Based on Individual Estimates Using Non-compartmental or Model-Based Analyses: Evaluation of Estimates of Sample Means and Type I for Different Designs. *Pharmaceutical research,* vol.27, no.1, pp. 92-104.

FDA (2001) *Guidance for Industry: Statistical Approaches to Bioequivalence*, US Department of Health and Human Services Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Rockville, Maryland pp. 1-48.

FDA (2003) *Guidance for Industry: Bioavailability and Bioequivalence Studies for Oral Administered Drug Products-General Considerations*, US Department of Health and Human Services Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Rockville, Maryland, pp. 1-22.

FDA. (2013) *Guidance for Industry Bioequivalence Studies with Pharmacokinetic Endpoints for Drugs Submitted Under an ANDA,* New Hampshire Ave.: Food and Drug Administration.

Fleiss, J. L. (1989) A critique of recent research on the two-treatment crossover design. *Controlled Clinical Trials,* vol.10, no.3, pp. 237-243.

Fox, J. (1997) *Applied Regression Analysis, Linear Models, and Related Methods*. California: Sage Publications, 306 pages.

Friedman, L.M., Furberg, C., DeMets, D.L., Reboussin, D. & Granger, C.B. (2015) *Fundamentals of clinical trials*. 5th ed. New York: Springer, 550 pages.

Gabrielsson, J. & Weiner, D. (2001) *Pharmacokinetic and pharmacodynamic data analysis: concepts and applications.* 3rd ed. Stockholm: CRC Press, 924 pages.

Greenblatt, D.J., Von Moltke, L.L., Harmatz, J.S. and Shader, R.I. (1998) Pharmacokinetics, Pharmacodynamics, and Drug Disposition. In:Davis KL, Charney D, Coyle JT, Nemeroff C, editors. *Neuropsychopharmacology: the Fifth Generation of Progress Philadelphia: Lippincott, Williams and Wilkins*, pp. 507-524.

Hu, C., Moore, K.H., Kim, Y.H. & Sale, M.E. (2004) Statistical issues in a modelling approach to assessing bioequivalence or PK similarity with presence of sparsely sampled subjects. *Journal of pharmacokinetics and pharmacodynamics*, vol.31, no.3, pp. 321-339.

Hyslop, T., Hsuan, F. & Holder, D.J. (2000) A small sample confidence interval approach to assess individual bioequivalence. *Statistics in Medicine*, vol.19, no.20, pp. 2885-2897.

Jones, B. & Kenward, M. G. (2003) *Design and analysis of crossover trials.* 3rd ed. London: CRC Press, 438 pages.

Korkmaz, S. & Orman, M.N. (2012) The Use of Nonlinear Mixed Effects Models in Bioequivalence Studies: A Real Data Application. *Department of Biostatistics and Medical Informatics, Faculty of Medicine, Trakya University, Merkez, Edirne, Turkey,* vol.1, no.11, pp.1-4.

Kuhn, E. & Lavielle, M. (2005) Maximum likelihood estimation in nonlinear mixed effects models. *Computational Statistics & Data Analysis*, vol. 49, no. 4, pp. 1020-1038.

Lavielle, M. & Mentré, F. (2007) Estimation of population pharmacokinetic parameters of saquinavir in HIV patients with the MONOLIX software. *Journal of pharmacokinetics and pharmacodynamics*, vol.34, no.2, pp .229-249.

Lavielle, M. (2014) *Mixed effects models for the population approach: models, tasks, methods and tools*. Boca Raton: CRC Press, 383 pages.

Lindstrom, M.J & Bates, D.M. (1990) Nonlinear mixed effects models for repeated measures data. *Biometrics*, vol. 46, no.3, pp. 673-687.

Liu, J.P. & Weng, C.S. (1995) Bias of two one-sided tests procedures in assessment of bioequivalence. *Statistics in Medicine*, vol.14, no.8, pp.853-861.

Martinez, M.N. & Jackson, A.J. (1991) Suitability of Various Noninfinity Area Under the

Plasma Concentration–Time Curve (AUC) Estimates for Use in Bioequivalence Determinations: Relationship to AUC from Zero to Time Infinity (AUCO–INF). *Pharmaceutical research*, vol.8, no.4, pp. 512-517.

Medicines Control Council of South Africa (2003) Guideline on Application and Information Requirements-Veterinary Medicines, pp. 1-47.

Mentré, F., Dubruc, C. & Thénot, J. P. (2001) Population pharmacokinetic analysis and optimization of the experimental design for mizolastine solution in children. *Journal of pharmacokinetics and pharmacodynamics,* vol.28, no.3, pp. 299-319.

Metzler, C. M. & Huang, D. C. (2009) Statistical methods for bioavailability and bioequivalence. *Clinical Research Practices and Drug Regulatory Affairs,* vol.1, no.2, pp. 109-132.

Morais, J. A. G. A. G. & Lobato, M. d. R. (2010) The New European Medicines Agency Guideline on the Investigation of Bioequivalence. *Basic & Clinical Pharmacology & Toxicology,* vol.106, no.3, pp. 221-225.

Mould, D.R. & Upton, R.N. (2013) Basic concepts in population modeling, simulation, and model-based drug development—part 2: introduction to pharmacokinetic modeling methods. CPT: *pharmacometrics & systems pharmacology*, vol.2, no.4, pp.1-14.

Niazi, S.K. (2014) Handbook *of bioequivalence testing*. 2nd ed. Denver: CRC Press, 1007 pages.

Oh, H. S., Ko, S.G. & Oh, M.S. (2003) A Bayesian approach to assessing population bioequivalence in a 2 x2 x 2 crossover design. *Journal of Applied statistics,* vol.30, no.8, pp. 881-891.

Panhard, X. & Mentre, F. (2005) Evaluation by simulation of tests based on non-linear mixed-effects models in pharmacokinetic interaction and bioequivalence crossover trials. *Statistics in medicine,* vol.24, no.10, pp. 1509-1524.

Panhard, X., Taburet, A.M., Piketti, C. & Mentré, F. (2007) Impact of modelling intra-subject variability on tests based on non-linear mixed-effects models in cross-over pharmacokinetic trials with application to the interaction of tenofovir on atazanavir in HIV patients. *Statistics in medicine*, vol.26, no.6, pp.1268-1284.

Panhard, X. & Samson, A. (2009) Extension of the SAEM algorithm for nonlinear mixed

models with 2 levels of random effects. *Biostatistics*, vol.10, no.1, pp. 121-135.

Patterson, S. D. & Jones, B. (2005) *Bioequivalence and statistics in clinical pharmacology.* 1st ed. New York: CRC Press, 400 pages.

Pineiro, G., Perelman, S., Guerschman, J.P. & Paruelo, J.M. (2008) How to evaluate models: observed vs. predicted or predicted vs. observed?. *Ecological Modelling*, vol.216, no.3, pp.316-322.

Rani, S. & Pargal, A. (2004) Bioequivalence: An overview of statistical concepts. *Indian* J*ournal of Pharmacology*, vol.36, no.4, pp. 209-216.

Rowland, M. & Tozer, T. N. (2005) Clinical pharmacokinetics/pharmacodynamics 4th ed. Philadelphia: Lippincott Williams and Wilkins Philadelphia, 864 pages.

Samson, A., Lavielle, M. & Mentre, F. (2007) The SAEM algorithm for group comparison tests in longitudinal data analysis based on non-linear mixed-effects model. *Statistics in medicine*, vol.26, no.27, pp. 4860-4875.

Schall, R. (1995) Assessment of Individual and Population Bioequivalence Using the Probability That Bioavailabilities Are Similar. *Biometrics,* vol.51, no.2, pp. 615-626.

Schuirmann, D. J. (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,* vol.15, no.6, pp. 657–680.

Schwinghammer, T.L. and Kroboth, P.D. (1988) Basic concepts in pharmacodynamic modeling. *The Journal of Clinical Pharmacology*, vol.28, no.5, pp.388-394.

Sheiner, L. B. & Steimer, J. L. (2000) Pharmacokinetic/Pharmacodynamic modelling in drug development. *Annual Review of Pharmacology and Toxicology,* vol.40, no.1, pp. 67-95.

Vaida, F. & Blanchard, S. (2005) Conditional Akaike information for mixed-effects models. *Biometrika*, vol.92, no.2, pp. 351-370.

Vonesh, E. & Chinchilli, V.M. (1996) *Linear and nonlinear models for the analysis of repeated measurements*. CRC Press, 560 pages.

Walker, R. B., Kanfer, I. & Skinner, M.F. (2006) Bioequivalence assessment of generic products: an innovative South African approach. *Clinical Research and Regulatory Affairs*, vol. 23, no. 1, pp. 11-20.

Westlake, W. J. (1979) Statistical aspects of comparative bioavailability trials. *Biometrics,* pp. 273-280.

Wu, H. & Wu, L. (2002) Identification of significant host factors for HIV dynamics modelled by non-linear mixed-effects models. *Statistics in Medicine,* vol.21, no.5, pp. 753-771.

Xu, X.S., Yuan, M., Karlsson, M.O., Dunne, A., Nandy, P. & Vermeulen, A. (2012) Shrinkage in nonlinear mixed-effects population models: quantification, influencing factors, and impact. *The AAPS journal*, vol.14, no.4, pp.927-936.

Yen, K. C. & Kwan, K. C. (1978) A comparison of numerical integrating algorithms by trapezoidal, Lagrange, and spline approximation. Journal of pharmacokinetics and biopharmaceutics. *Journal of pharmacokinetics and biopharmaceutics,* vol.6, no.1, pp. 79-98.

APPENDICES

Appendix A The SAEM Algorithm

Introduction

The parameters of the NLMEM will be estimated using the stochastic approximation expectation-maximisation, let $\theta = \left(\mu, \beta, \Omega, \psi, \sigma^2\right)$, The objective of the SAEM is to determine by maximising the likelihood of the observations. Let $\tilde{b}_i = \mu + b_i,\ y = (y_{ijk}, 1 \le i \le N, 1 \le j \le n_{ik})$ and $\phi = (\phi_{ik}, 1 \le i \le N, k = R, T),$ where n_{ik} is the number of subjects in treatment *k* and *N* is the sample size. The likelihood of the observed is as follows:

$$p\left(y;\theta\right) = \int p\left(y,\phi,\tilde{b};\theta\right) d(\phi,\tilde{b}) \qquad \text{(A.1)}$$

Where $p\left(y,\phi,\tilde{b};\theta\right)$ denote the likelihood of the full data set $(y,\phi,\tilde{b})$. Since f and the random effects ϕ_{ik} have a non-linear relationship the likelihood has no closed form (Panhard and Samson, 2009). Hence the SAEM algorithm is used.

The SAEM Algorithm Procedure

The EM algorithm proposed by Dempster et al. (1977) is an ideal method for estimating parameters of models with missing data. In the case of NLMEM, the random effects $(\phi,\tilde{b})$ are non-observed data and $(y,\phi,\tilde{b})$ are the observed data. The EM algorithm maximises the function $\mathbb{Q}(\theta \mid \theta') = E(\ln P(y,\phi,\tilde{b};\theta))$, the condition expectation under the posterior distribution $p\left(\phi,\tilde{b} \mid y;\theta\right)$.

At the ℓ iteration, the EM consists of two steps:

E step: evaluate $\mathbb{Q}(\theta \mid \hat{\theta}_\ell)$.

M step: update $\hat{\theta}_\ell$ by maximising $\mathbb{Q}(\theta \mid \hat{\theta}_\ell)$.

It is important to show that $\mathbb{Q}(\theta \mid \hat{\theta}_\ell)$ can be reduced in the case of non-linear mixed effect model.

Firstly, given that $p(\tilde{b} \mid y, \phi; \theta) = p(\tilde{b} \mid \phi; \theta)$, when implementing the Bayes theorem yields the following expression:

$$p(\phi, \tilde{b} \mid y; \theta) = p(\tilde{b} \mid \phi; \theta)\, p(\phi \mid y; \theta) = \prod_{i=1}^{N} p(\tilde{b}_i \mid \phi_i; \theta)\, p(\phi_i \mid y_i; \theta). \quad \text{(A.2)}$$

Secondly, because of the linearity form of the individual parameters in Equation (2.5), the posterior distribution of the subject i is explicit, hence $p(\tilde{b} \mid y, \phi; \theta)$, is normally distributed with a mean $m(\phi_i, \theta)$ and variance $V(\theta)$ where;

$$m(\phi_i, \theta) = V(\theta)\left(\psi^{-1} \sum\nolimits_{k=T,R} (\phi_{ik} - \beta_k) + \Omega^{-1}\mu\right), \quad \text{(A.3)}$$

$$V(\theta) = (\Omega^{-1} + K\psi^{-1})^{-1}$$

Where K is the number of treatment, in this study K=2.

Because of the factorisation presented in equation (A.2), $\mathbb{Q}$ can be rewritten as:

$$\mathbb{Q}(\theta \mid \theta') = \int \left(\int \ln\left(y, \phi, \tilde{b}; \theta\right) p\left(\tilde{b} \mid \phi; \theta'\right) d\tilde{b} \right) p\left(\phi \mid y; \theta'\right) d\phi\,. \quad \text{(A.4)}$$

The results in equation (A.3) allows the conditional expectation to be dived in two integrals: first, the integral is evaluated with respect to the posterior distribution of $\tilde{b}$. Secondly, it is evaluated with respect to ϕ.

The first integral $R(y, \phi, \theta, \theta')$ is computed as follows:

$$R(y, \phi, \theta, \theta') = \int \ln\, p(y, \phi, \tilde{b}; \theta)\, p(\tilde{b} \mid \phi; \theta') d\tilde{b}. \quad \text{(A.5)}$$

Because $\ln\, p(y, \phi, \tilde{b}; \theta)$ has the analytic form then so is equation (A.5), the log-likelihood is evaluated as:

$$\ln p(y,\phi,\tilde{b};\theta)=\frac{1}{2}\sum_{k=R,T}\sum_{i=1}^{N}\sum_{j=1}^{n_{ik}}\ln\left(2\pi\sigma^2 g^2\left(t_{ijk},\phi_{ik}\right)\right)$$

$$-\frac{1}{2}\sum_{k=R,T}\sum_{i=1}^{N}\sum_{j=1}^{n_{ik}}\frac{\left(y_{ijk}-f(t_{ijk},\phi_{ik})\right)}{\sigma^2 g^2\left(t_{ijk},\phi_{ik}\right)}-\frac{nK}{2}\ln(2\pi\det\psi)$$

$$-\frac{1}{2}\sum_{k=R,T}\sum_{i=1}^{N}\left(\phi_{ik}-\tilde{b}_i-\beta_k\right)'\psi^{-1}\left(\phi_{ik}-\tilde{b}_i-\beta_k\right) \tag{A.6}$$

$$-\frac{n}{2}\ln(2\pi\det\Omega)-\frac{1}{2}\sum_{i=1}^{N}\left(\tilde{b}_i-\mu\right)'\Omega^{-1}\left(\tilde{b}_i-\mu\right)'.$$

Since $\ln p(\phi\,|\,y;\theta)$ can be computed hence we have:

$$R(y,\phi,\theta,\theta')=\frac{1}{2}\sum_{i,j,k}\ln\left(2\pi\sigma^2 g^2\left(t_{ijk},\phi_{ik}\right)\right)\frac{1}{2}\sum_{i,j,k}\frac{\left(y_{ijk}-f\left(t_{ijk},\phi_{ik}\right)\right)^2}{\sigma^2 g^2\left(t_{ijk},\phi_{ik}\right)}$$

$$-\frac{NK}{2}\ln\left(2\pi\det\psi\right)-\frac{NK}{2}\psi^{-1/2}V\left(\theta'\right)\psi^{-1/2}$$

$$-\frac{1}{2}\sum_{i,k}\left(\phi_{ik}-m\left(\phi_i,\theta'\right)-\beta_k\right)'\psi^{-1}\left(\phi_{ik}-m\left(\phi_i,\theta'\right)-\beta_k\right) \tag{A.7}$$

$$-\frac{N}{2}\ln\left(2\pi\det\Omega\right)-\frac{N}{2}\Omega^{-1/2}V\left(\theta'\right)\Omega^{-1/2}$$

$$-\frac{1}{2}\left(m\left(\phi_i,\theta'\right)-\mu\right)'\Omega^{-1}\left(m\left(\phi_i,\theta'\right)-\mu\right).$$

Then the task of evaluating $\mathbb{Q}$ is minimised to a problem of evaluating the second integral under $p(\phi\,|\,y;\theta)$ as:

$$\mathbb{Q}(\theta|\theta')=\int R(y,\phi,\theta,\theta')p(\phi|y;\theta')d\phi. \tag{A.8}$$

Since f has a non-linear relationship with ϕ, $p(\phi|y;\theta')$ can not be computed explicitly and thus, $\mathbb{Q}(\theta|\theta')$ in equation (A.2) is intractable. Panhard and Samson (2009), proposed the use of the SAEM algorithm which evaluates $\mathbb{Q}(\theta|\theta')$ using a stochastic procedure.

The SAEM algorithm for the two levels NLMEM can be detailed as follows:

Firstly, the function $R(y,\phi,\theta,\theta')$ is the member of the regular curved exponential family, and $R(y,\phi,\theta,\theta')$ may be presented in a form:

$$R(y,\phi,\theta,\theta') = -\Lambda(\theta) + \langle S(y,\phi,\theta'), \Phi(\theta)\rangle. \quad \text{(A.9)}$$

With Λ and Φ, the functions that are twice continuously and differentiable on Θ, $\langle .,. \rangle$ indicate a scalar product and $S(y,\phi,\theta')$ is the known sufficient statistics of a complete model.

Hence $\mathbb{Q}(\theta|\theta')$ is simplified to:

$$\mathbb{Q}(\theta|\theta') = -\Lambda(\theta) + \langle (\int S(y,\phi,\theta')p(\phi|y;\theta')d\phi), \Phi(\theta)\rangle. \quad \text{(A.10)}$$

It follows that, at the ℓth step, the SAEM algorithm starts with:

Simulation-step: Under conditional distribution, $p\left(\phi|y;\hat{\theta}_\ell\right)$ simulate $\left({\phi_i}^{\ell}\right)_i$ (the missing data),

Stochastic approximation-step: compute $s_{\ell+1} = s_\ell + \Upsilon_\ell\left(S\left(y,\phi^\ell,\hat{\theta}_\ell\right) - s_\ell\right)$, this step is for the stochastic aproximation of $E\left(S\left(y,\phi,\hat{\theta}_\ell\right)|y;\hat{\theta}_\ell\right) = \int S\left(y,\phi,\hat{\theta}_\ell\right)p\left(\phi|y;\hat{\theta}_\ell\right)d\phi$, using Υ_ℓ, $\ell \geq 0$, decending sequence of positive numbers,

Maximisation-step: update $\hat{\theta}_{\ell+1}$, according to

$$\hat{\theta}_{\ell+1} = \arg\max_{\theta\epsilon\Theta}(-\Lambda(\theta) + \langle s_{\ell+1}, \Phi(\theta)\rangle).$$

The stochastic approximation-step updates $S(y,\phi,\theta')$ as follows:

$$s_{1,i,\ell+1} = s_{1,i,\ell} + \gamma_\ell\left(\sum_{k=R,T}\phi_{ik}^{(\ell)} - s_{1,i,\ell}\right), \qquad i = 1,2,\ldots,N,$$

$$s_{2,k,\ell+1} = s_{2,k,\ell} + \gamma_\ell\left(\sum_{i=1}^{N}\phi_{ik}^{(\ell)} - s_{2,k,\ell}\right), \qquad k = R,T,$$

$$s_{3,\ell+1} = s_{3,i,\ell} + \gamma_\ell \left(\sum_{i=1}^{N} m\left(\phi_{ik}^{(\ell)}, \hat{\theta}_\ell\right)' m\left(\phi_i^{(\ell)}, \hat{\theta}_\ell\right) - s_{3,\ell} \right),$$

$$s_{4,\ell+1} = s_{4,\ell} + \gamma_\ell \left(\sum_{k=R,T} \sum_{i=1}^{N} \left(\phi_{ik}^{(\ell)} - m\left(\phi_i^{(\ell)}, \hat{\theta}_\ell\right)\right)' \left(\phi_{ik}^{(\ell)} - m\left(\phi_i^{(\ell)}, \hat{\theta}_\ell\right)\right) - s_{3,\ell} \right)$$

$$s_{5,\ell+1} = s_{5,\ell} + \gamma_\ell \left(\sum_{i,j,k} \left(\frac{y_{ijk} - f\left(t_{ijk}, \phi_{ik}^{(\ell)}\right)}{g\left(t_{ijk}, \phi_{ik}^{(\ell)}\right)} \right)^2 - s_{5,\ell} \right)$$

Where $s_{1,i,\ell+1}, s_{4,\ell+1}, s_{5,\ell+1}, \ s_{3,\ell+1}, s_{2,k,\ell+1}$ are the steps of the iteration procedure. Pharmacokinetic parameters are then estimated as follows:

$$\hat{\mu}_{\ell+1} = \mathrm{V}\left(\hat{\theta}_\ell\right)\hat{\psi}_\ell^{-1} \left(\frac{1}{N} \sum_{i=1}^{N} s_{1,i,\ell+1} - \sum_{k=R,T} \hat{\beta}_{k,\ell} \right) + \mathrm{V}\left(\hat{\theta}_\ell\right)\widehat{\Omega}_\ell^{-1} \hat{\mu}_\ell,$$

$$\hat{\beta}_{k,\ell+1} = \frac{s_{2,k,\ell+1}}{N} - \hat{\mu}_{\ell+1}, \qquad k = R, T,$$

$$\widehat{\Omega}_{\ell+1} = \mathrm{V}\left(\hat{\theta}_\ell\right) + \frac{s_{3,\ell+1}}{N} - (\hat{\mu}_{\ell+1})'\hat{\mu}_{\ell+1},$$

$$\hat{\psi}_{\ell+1} = \mathrm{V}\left(\hat{\theta}_\ell\right) + \frac{s_{4,\ell+1}}{NK} - \frac{1}{K} \sum_{k=R,T} \left(\hat{\beta}_{k,\ell+1}\right)' \hat{\beta}_{k,\ell+1},$$

$$\widehat{\sigma^2}_{\ell+1} = \frac{s_{5,\ell+1}}{\sum_{i=1}^{N} \sum_{k=R,T} n_{ik}}.$$

Where, $s_{i,k}$ is the $p-$ vector that approximates $E\left(S\left(y, \phi, \hat{\theta}_\ell\right)|y; \hat{\theta}_\ell\right)$

Appendix B Modelling Codes

SAS Codes

```
data work02;
set work01;
by subject Treatment;
subjectid=Compress(Subject||Treatment);
*Xtime=Actual_Time;
*Yvalue=Conc;
*keep Xtime Yvalue subjectid;

if Treatment="R" then tmt=1;
if Treatment="T" then tmt=2;
run;
proc sort data= work02;
by subjectid subject;
run;
*Gplot settings;
SYMBOL1 V=NONE C=GREEN I=NEEDLE;
SYMBOL2 V=DOT C=RED I=JOIN;
PROC GPLOT DATA =work02;
by subjectid subject;
PLOT conc*Scheduled_time conc*Scheduled_time / OVERLAY FRAME VREF=4.0
CVREF=BLUE
LVREF=34 HMINOR=0 VMINOR=0 VAXIS=0 TO 2 BY 0.1 HAXIS=0.0 TO 40 BY 5;
TITLE 'Sample Data Plot: Metabolic Values vs. Time';
LABEL conc='concetration' Scheduled_time='Relative Time (min)';
RUN;
```

TOST Analysis for AUC

```
proc sort data=mydata;
by treatment AUC;
```

```
run;
data mydata2;
set mydata;
if treatment="R";
ref=AUC_;
run;
data mydata3;
set mydata;
if treatment="T";
test=AUC_;
run;
data mydata4;
merge mydata3 mydata2;
keep ref test;
run;
ods graphics on;
proc ttest data= mydata4 dist=lognormal tost(0.8, 1.25);
  paired Test*Ref;
run;
ods graphics off;
```

***AUC* Calculation Using the Trapezoidal Approach (NCA Method)**

```
data work03;
set work02;
*by subjectid subject;
drop lagconc lagtime;
lagconc=lag(Conc);
lagtime=lag(Actual_Time);
if Actual_Time=0 or Conc=0
then do ;
lagconc=0;
lagtime=0;
```

```
sumAUC=0;
AUC=0;
end;
AUC=(Conc+lagConc)/2*(Actual_Time-lagtime);
sumAUC+AUC;
*if subjectid="10T";
*keep xtime yvalue;
run;

proc sort data= work03;
by subjectid subject Actual_Time;
run;
* keep the last observation for each AUCsum;
data work04a;
set work03;
by subjectid;
if last.subjectid;
AUC_=sumAUC;
if Subject=26 then delete;
drop Actual_Time conc scheduled_time AUC sumAUC;
run;
```

Anova Analysis for ln*AUC*

```
proc glm data=work04 outstat=tanova;
by design;
class tmt period subject;
model log_AUC_=subject period tmt/ss1 ss2 ss4 solution p;
output out=pred p=pred r=resid student=stresid;
means tmt period;
lsmeans tmt period/stderr pdiff out=lsmeans1;
run;
```

Anova analysis for *Cmax*

```
proc glm data=work04 outstat=tanova;
by design;
class tmt period subject;
model AUC_=subject period tmt/ss1 ss2 ss4 solution p;
output out=pred p=pred r=resid student=stresid;
means tmt period;
lsmeans tmt period/stderr pdiff out=lsmeans1;
run;
```

Anova analysis for ln*Cmax*

```
proc glm data=work04 outstat=tanova;
by design;
class tmt period subject;
model AUC_=subject period tmt/ss1 ss2 ss4 solution p;
output out=pred p=pred r=resid student=stresid;
means tmt period;
lsmeans tmt period/stderr pdiff out=lsmeans1;
run;
```

Estimation of AUC based on NLMEM Method.

Model Fitted in Monolix

```
[LONGITUDINAL]
input = {V,Cl}
EQUATION:
Cc = pkmodel(V,Cl)
OUTPUT: output = Cc
```

SAS Code for Estimating AUC from the Model's Parameter Estimates.

```
Data individual estimates1;
```

```
Length design $30;
Set individual estimates;
*design='sparse_80hl';
sd_cl=___Cl_sd_;
subjectid=Compress(Subject||Trt);
d=5;
if ___strt_T_R=1 then seq=1;
else seq=2;
if ___OCC_2=1 then period=2;
else period=1;
if ___trt_T=1 then trt='T';
else trt='R';
ID1=compress(upcase(_____ID));
ID2=substr (id1,1,2);
if ID2 in('1#' '2#' '3#' '4#' '5#' '6#' '7#' '8#' '9#') then subject=substr(id2,1,1);
else subject=ID2;
V=____V_mode;
cl=___Cl_mode;
sd_v=___V_sd_;
AUC_=d/cl;
Keep design subject trt seq period v cl sd_v sd_cl D;
run;
```

The TOST and ANOVA procedure are similar to that of the AUC estimated by NCA method.

Appendix C Plotted Results

Individual Concentration Profiles

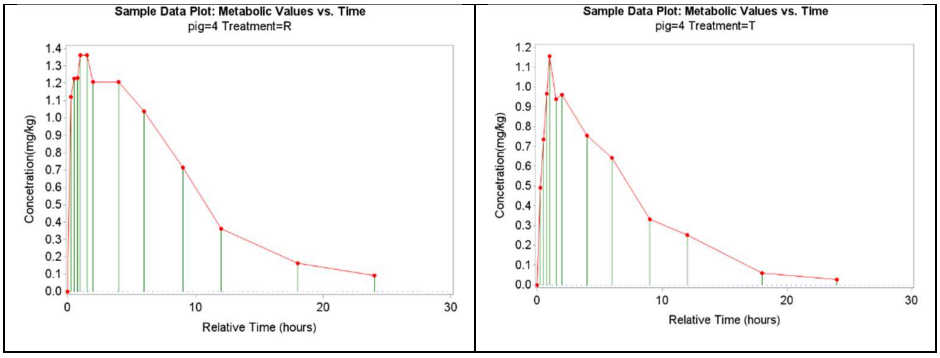


Figure A.1: The concentration-time curve for subject 1(pig=4)

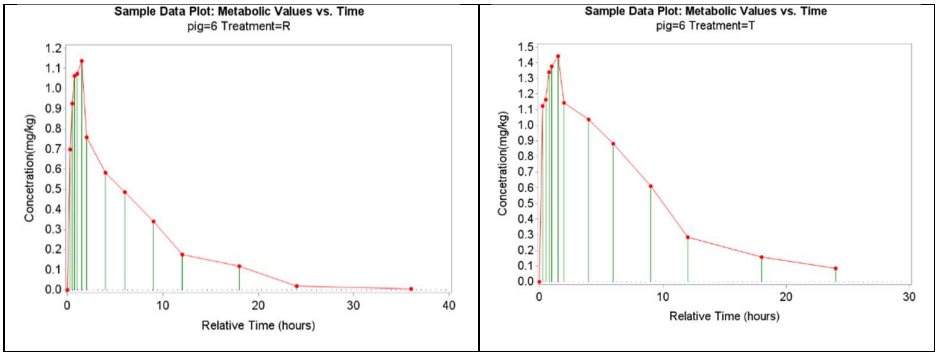


Figure A.2: The concentration-time curve for subject 2 (pig=6)

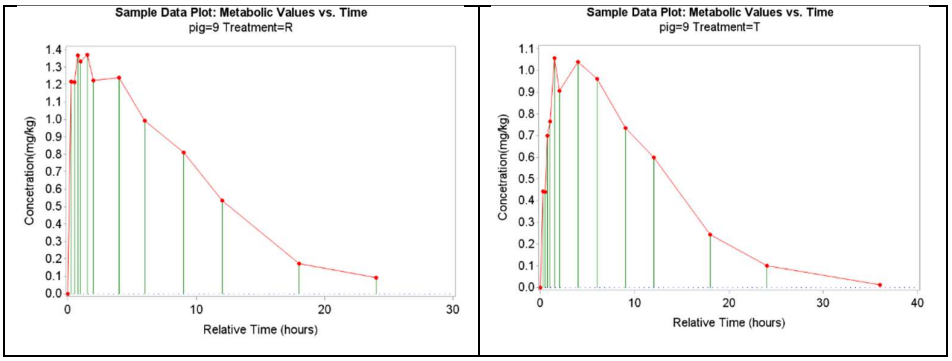


Figure A.3: The concentration-time curve for subject 3 (pig=9)

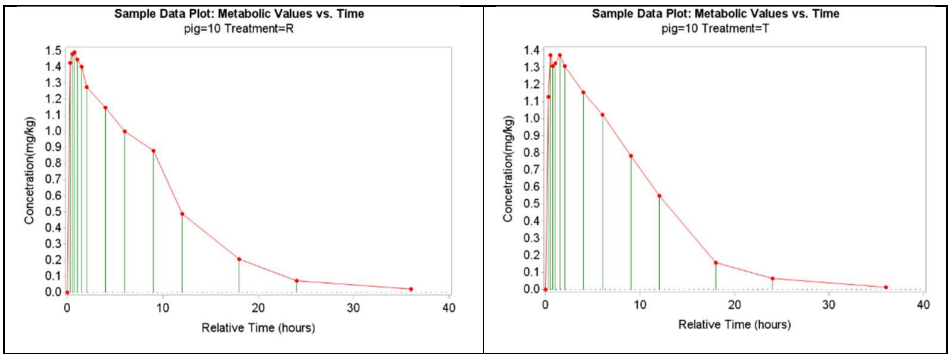


Figure A.4: The concentration-time curve for subject 4 (pig=10)

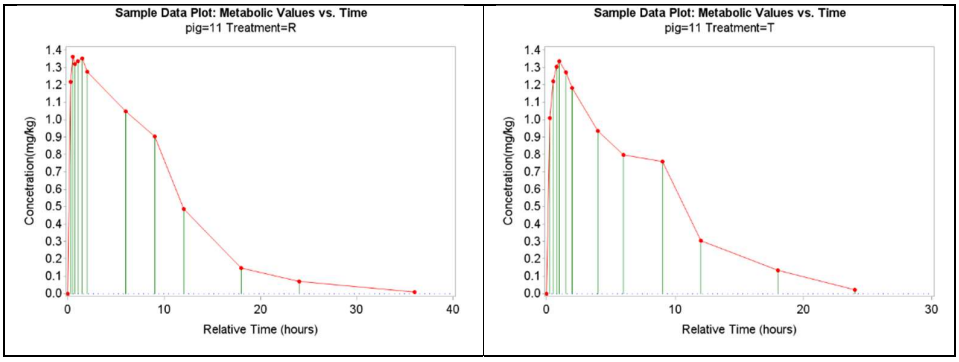


Figure A.5: The concentration-time curve for subject 5 (pig=11)

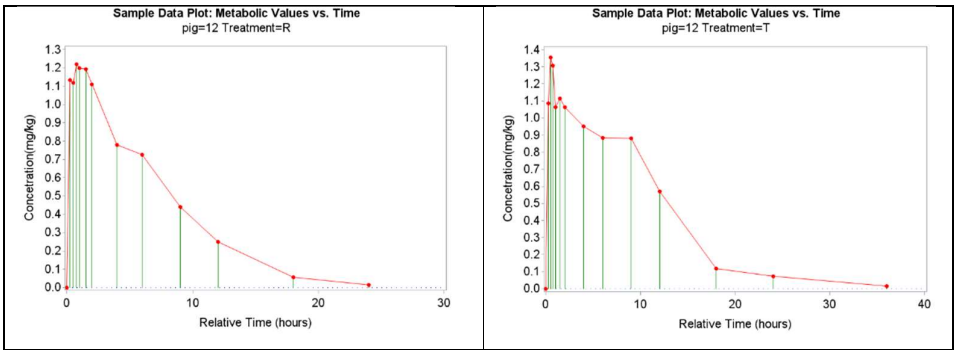


Figure A.6: The concentration-time curve for subject 6 (pig=12)

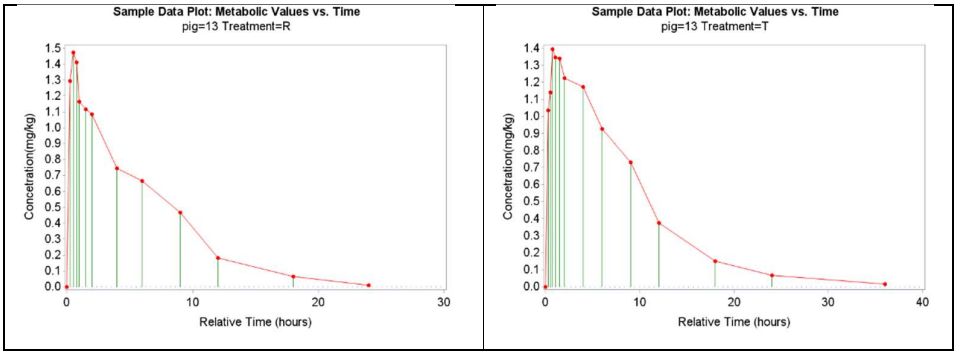


Figure A.7: The concentration-time curve for subject 7 (pig=13)

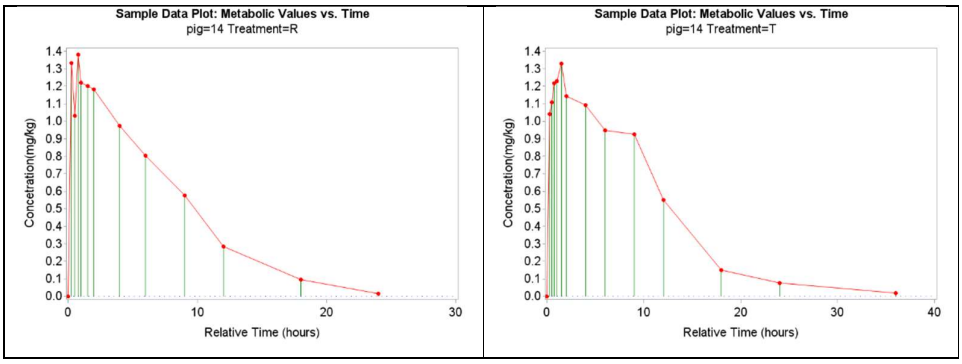


Figure A.8: The concentration-time curve for subject 8 (pig=14)

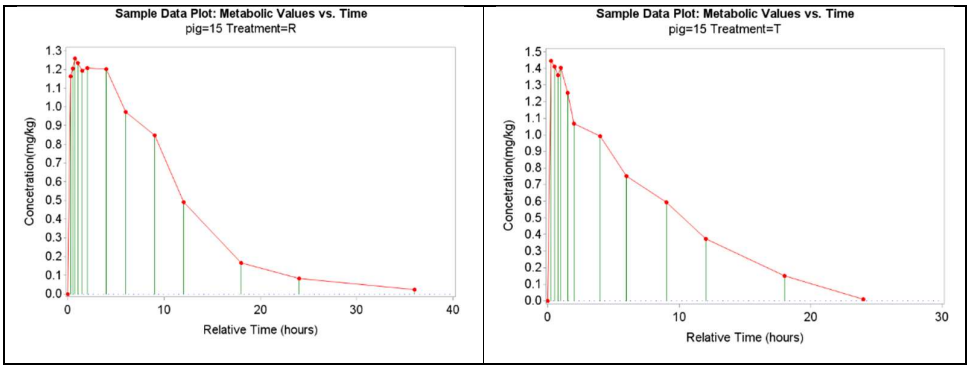


Figure A.9: The concentration-time curve for subject 9 (pig=15)

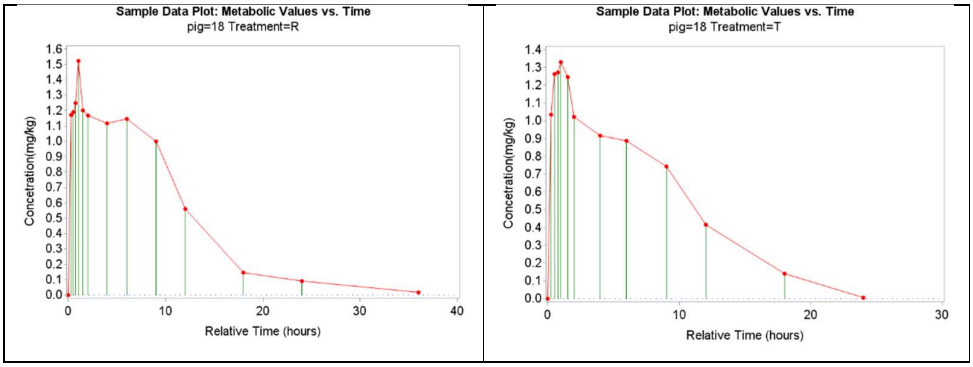


Figure A.10: The concentration-time curve for subject 10 (pig=18)

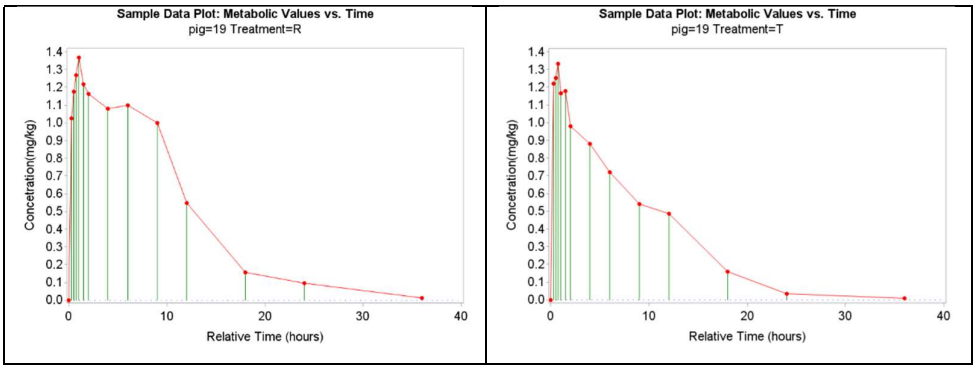


Figure A.11: The concentration-time curve for subject 11 (pig=19)

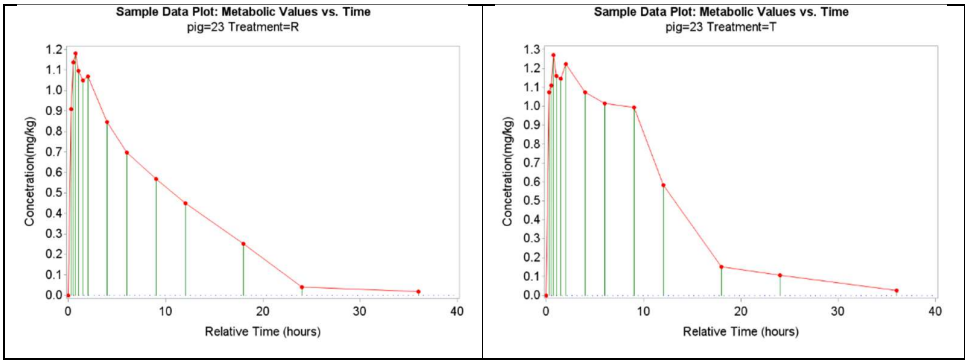


Figure A.12: The concentration-time curve for subject 12 (pig=23)

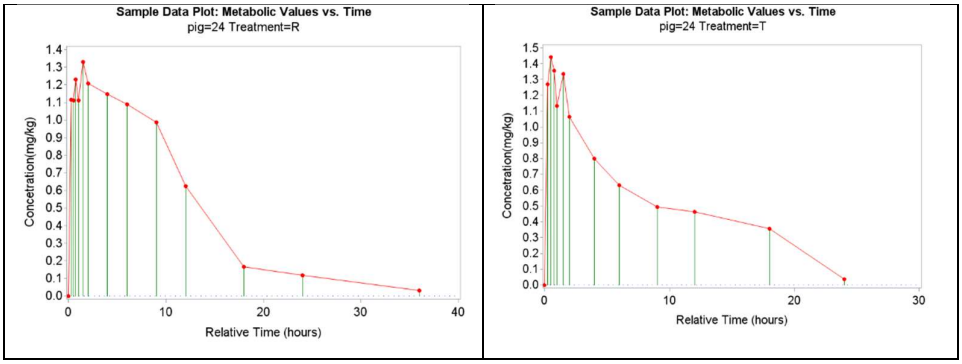


Figure A.13: The concentration-time curve for subject 13 (pig=24)

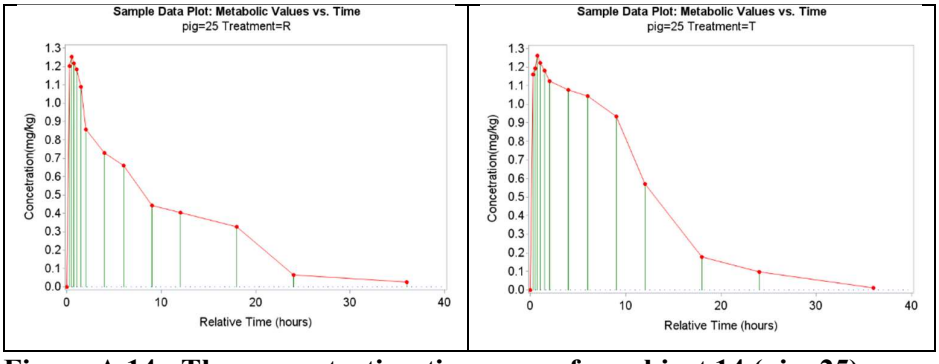
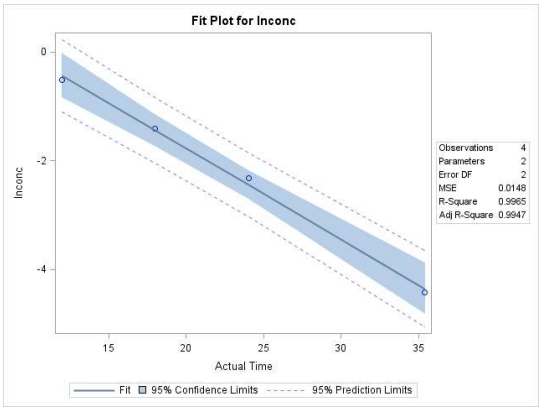
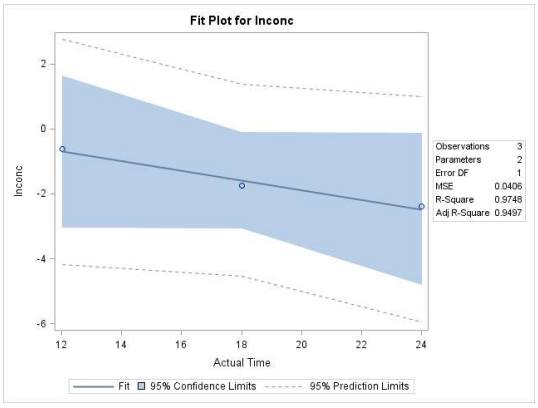


Figure A.14: The concentration-time curve for subject 14 (pig=25).

Regression Diagnostic for Calculation lnConc

T treatment	R treatment





Parameter Estimates

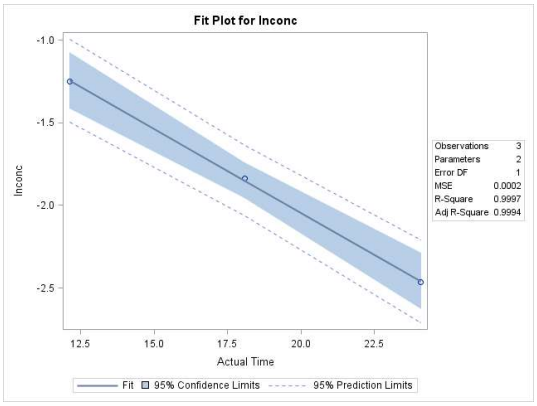
Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	1.07662	0.44420	2.42	0.2491
Actual_Time	Actual Time	1	-0.14815	0.02380	-6.22	0.1014

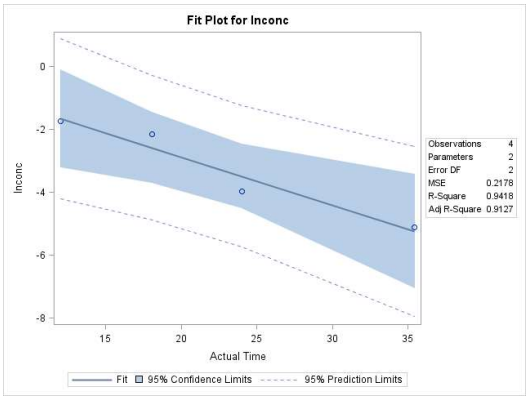
Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	1.07662	0.44420	2.42	0.2491
Actual_Time	Actual Time	1	-0.14815	0.02380	-6.22	0.1014

Figure A.15: Fit Plot for ln(Conc) for pig=9.

T treatment	R treatment





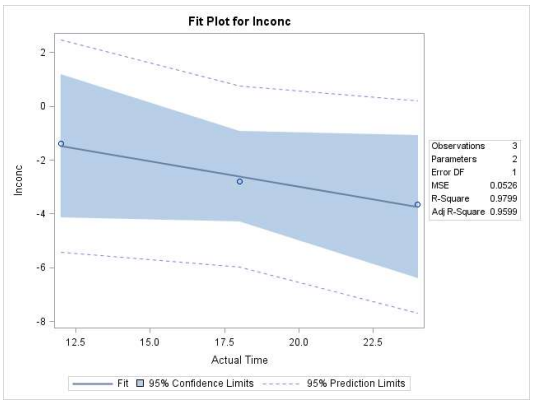
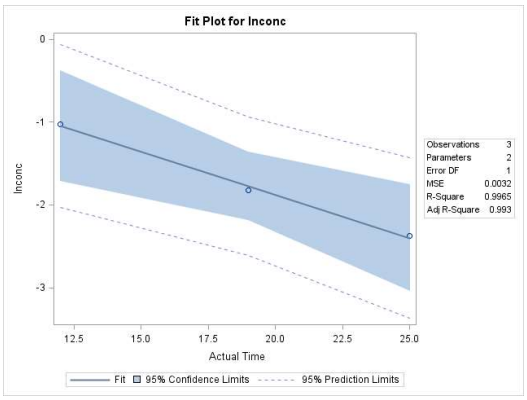
Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	-0.01720	0.03236	-0.53	0.6890
Actual_Time	Actual Time	1	-0.10140	0.00173	-58.73	0.0108

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	0.18923	0.64676	0.29	0.7974
Actual_Time	Actual Time	1	-0.15338	0.02696	-5.69	0.0295

Figure A.16: Fit Plot for ln(Conc) for pig=6.

T treatment	R treatment
	

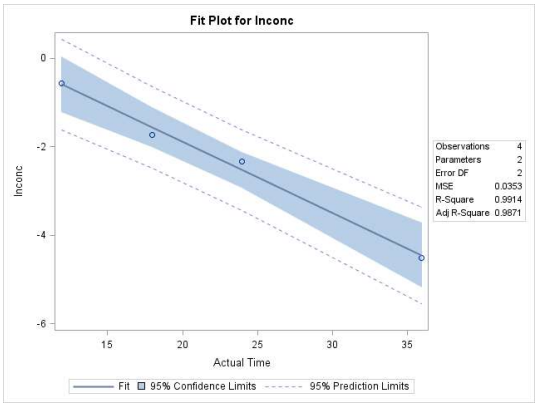
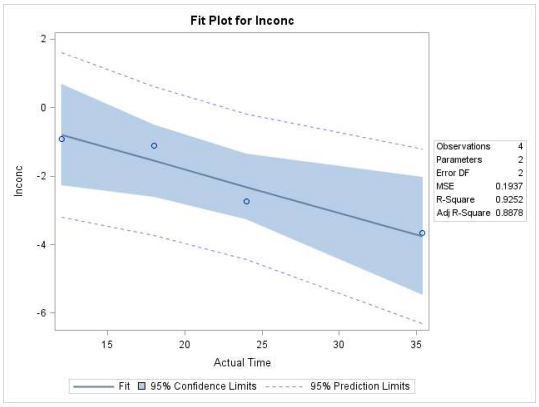
Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	0.79142	0.50423	1.57	0.3611
Actual_Time	Actual Time	1	-0.18895	0.02703	-6.99	0.0905

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	0.20789	0.11981	1.74	0.3329
Actual_Time	Actual Time	1	-0.10439	0.00618	-16.90	0.0376

Figure A.17: Fit Plot for In(Conc) for pig=4.

T treatment	R treatment
	

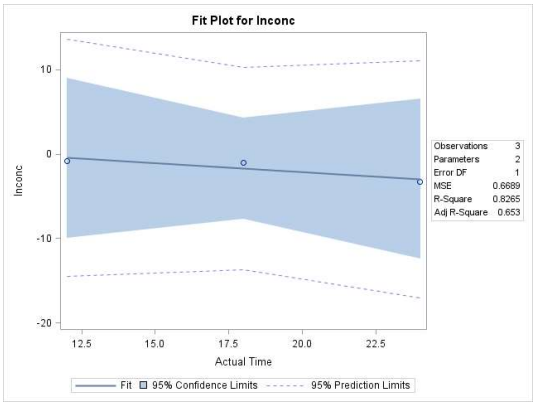
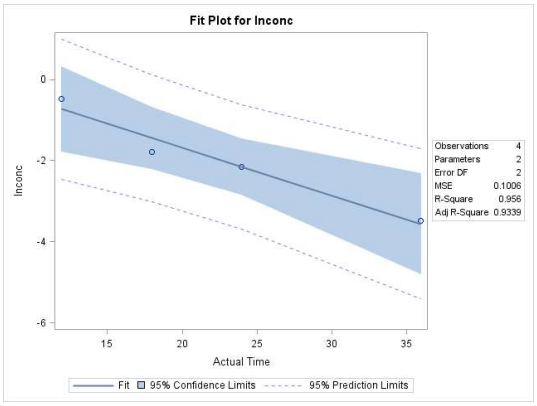
Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	1.32934	0.25552	5.20	0.0350
Actual_Time	Actual Time	1	-0.16100	0.01059	-15.20	0.0043

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	0.72984	0.61061	1.20	0.3545
Actual_Time	Actual Time	1	-0.12676	0.02548	-4.97	0.0381

Figure A.18: Fit Plot for In(Conc) for pig=25.

T treatment	R treatment
	

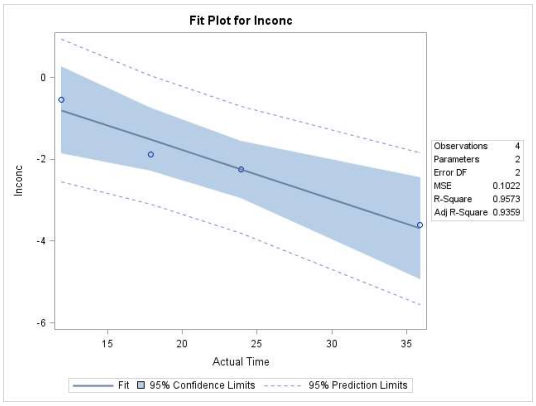
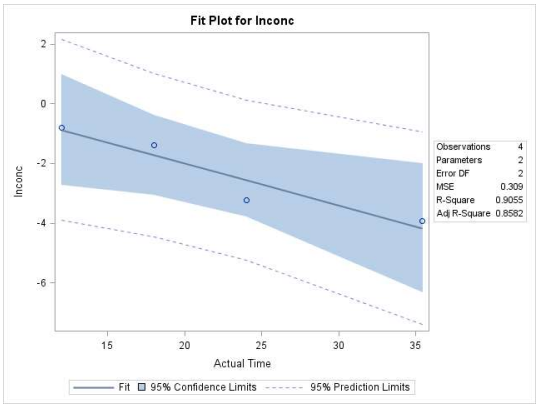
Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	2.08635	1.79809	1.16	0.4528
Actual_Time	Actual Time	1	-0.21039	0.09639	-2.18	0.2735

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	0.67123	0.43130	1.56	0.2599
Actual_Time	Actual Time	1	-0.11775	0.01787	-6.59	0.0223

Figure A.19: Fit Plot for In(Conc) for pig=24.

T treatment	R treatment
	

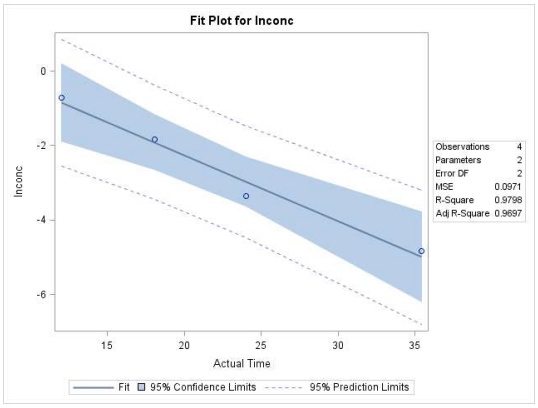
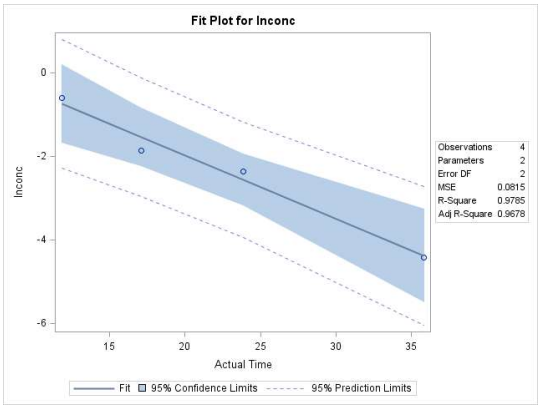
Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	0.63353	0.43448	1.46	0.2822
Actual_Time	Actual Time	1	-0.12064	0.01803	-6.69	0.0216

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	0.81750	0.77046	1.06	0.3999
Actual_Time	Actual Time	1	-0.14066	0.03214	-4.38	0.0484

Figure A.20: Fit Plot for In(Conc) for pig=23.

T treatment	R treatment
	

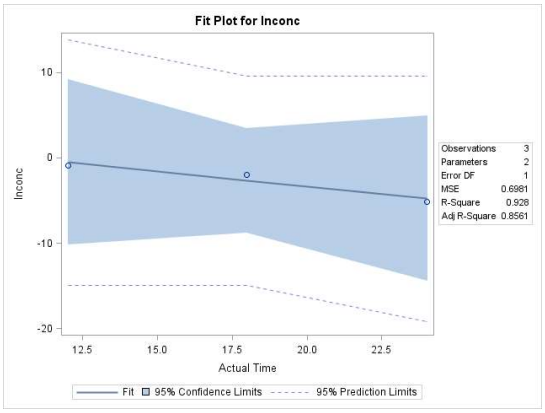
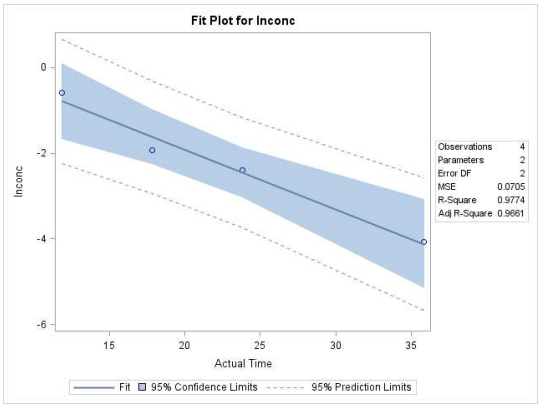
Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	1.28398	0.43182	2.97	0.0969
Actual_Time	Actual Time	1	-0.17747	0.01801	-9.85	0.0101

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	1.04936	0.38007	2.76	0.1100
Actual_Time	Actual Time	1	-0.15176	0.01590	-9.54	0.0108

Figure A.21: Fit Plot for ln(Conc) for pig=19.

T treatment	R treatment
	

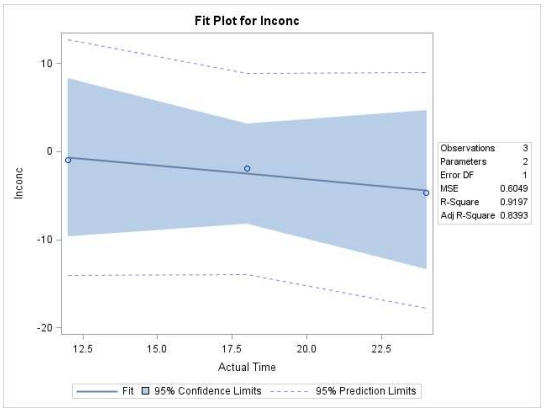
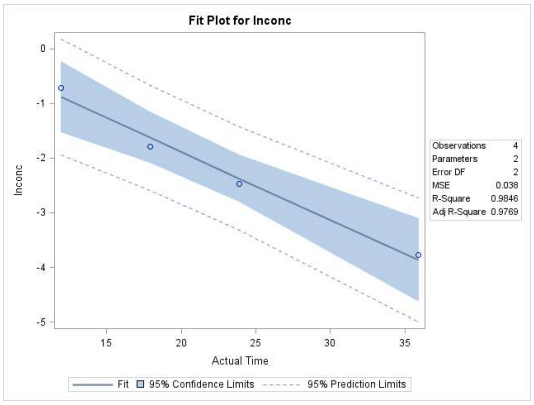
Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	3.71650	1.83798	2.02	0.2924
Actual_Time	Actual Time	1	-0.35360	0.09847	-3.59	0.1729

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	0.85636	0.35939	2.38	0.1401
Actual_Time	Actual Time	1	-0.13900	0.01495	-9.30	0.0114

Figure A.22: Fit Plot for ln(Conc) for pig=18.

T treatment	R treatment
	

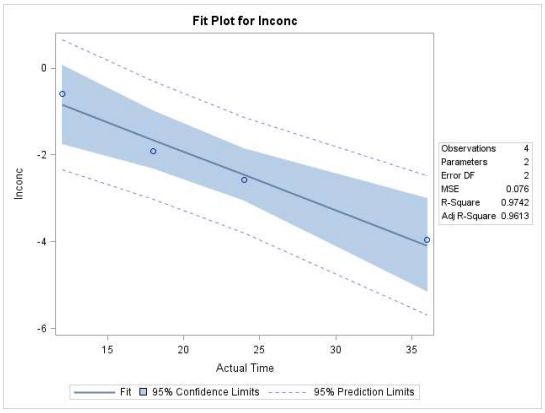
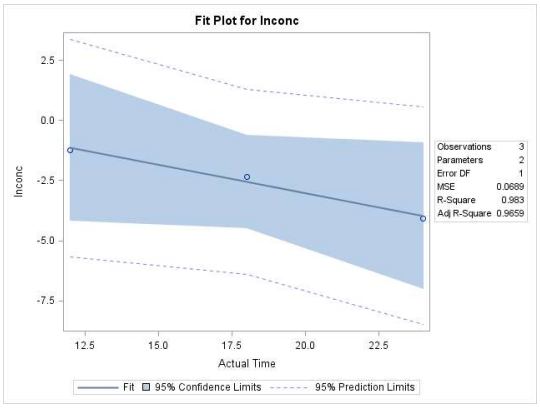
Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	3.05033	1.70993	1.78	0.3253
Actual_Time	Actual Time	1	-0.31014	0.09166	-3.38	0.1829

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	0.60013	0.26515	2.26	0.1519
Actual_Time	Actual Time	1	-0.12431	0.01100	-11.30	0.0077

Figure A.23: Fit Plot for ln(Conc) for pig=15.

T treatment	R treatment
	

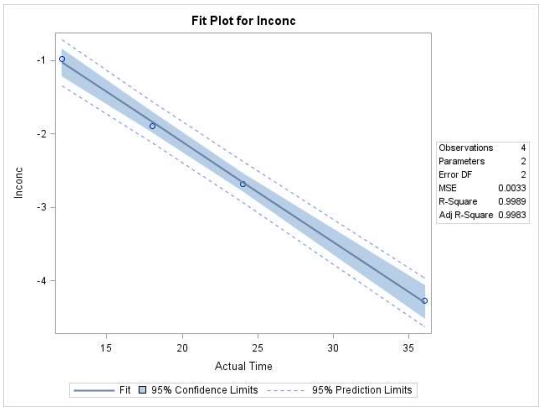
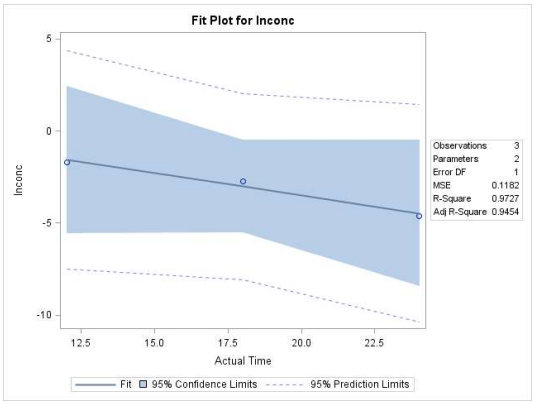
Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	0.77727	0.37574	2.07	0.1745
Actual_Time	Actual Time	1	-0.13496	0.01553	-8.69	0.0130

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	1.67118	0.57710	2.90	0.2117
Actual_Time	Actual Time	1	-0.23494	0.03094	-7.59	0.0833

Figure A.24: Fit Plot for ln(Conc) for pig=14.

T treatment	R treatment
	

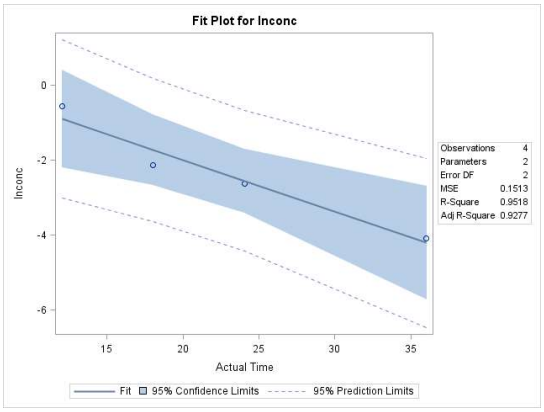
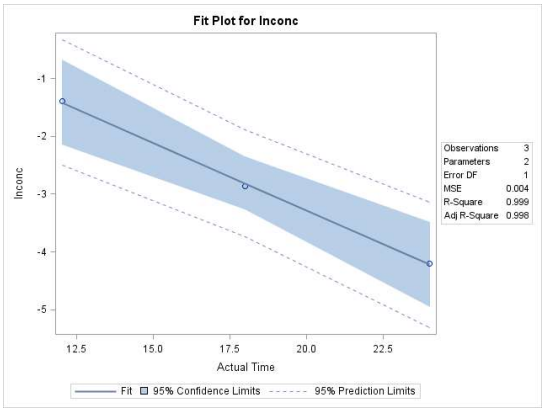
Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	0.60741	0.07818	7.77	0.0162
Actual_Time	Actual Time	1	-0.13602	0.00323	-42.13	0.0006

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	1.34207	0.75653	1.77	0.3268
Actual_Time	Actual Time	1	-0.24179	0.04052	-5.97	0.1057

Figure A.25: Fit Plot for ln(Conc) for pig=13.

T treatment	R treatment
	

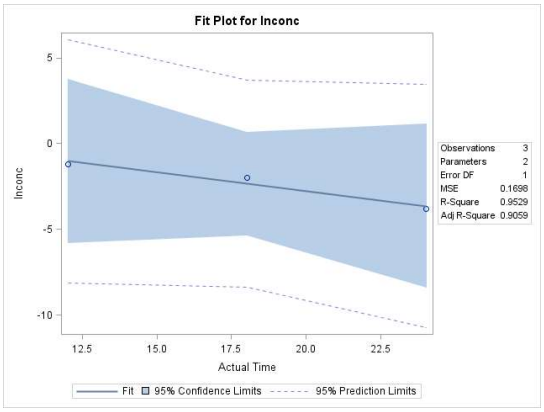
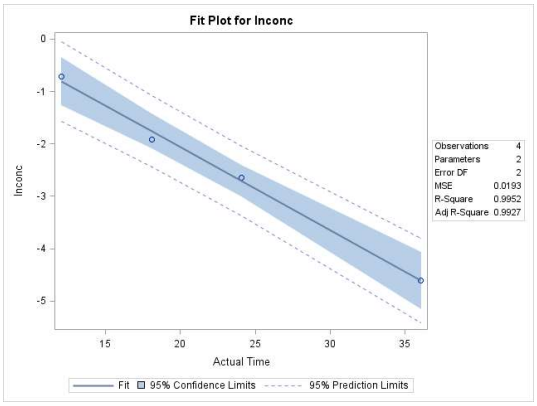
Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	0.75642	0.53087	1.42	0.2902
Actual_Time	Actual Time	1	-0.13775	0.02192	-6.29	0.0244

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	1.41267	0.13955	10.12	0.0627
Actual_Time	Actual Time	1	-0.23476	0.00747	-31.41	0.0203

Figure A.26: Fit Plot for ln(Conc) for pig=12.

T treatment	R treatment
	

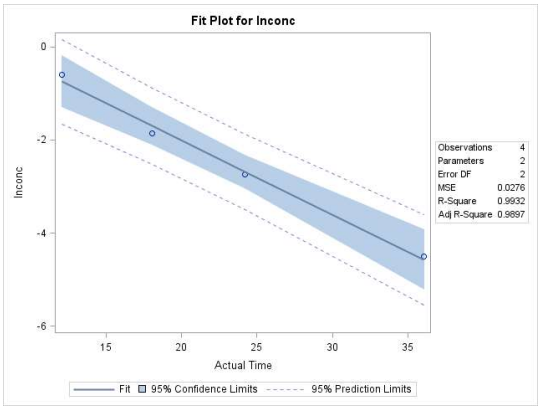
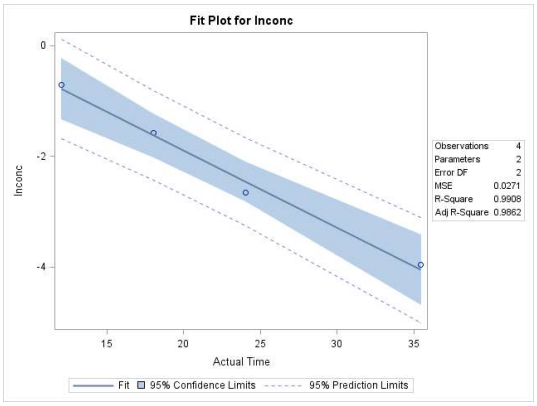
Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	1.59620	0.90610	1.76	0.3287
Actual_Time	Actual Time	1	-0.21822	0.04850	-4.50	0.1392

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	1.10216	0.18952	5.82	0.0283
Actual_Time	Actual Time	1	-0.15854	0.00782	-20.27	0.0024

Figure A.27: Fit Plot for In(Conc) for pig=11.

T treatment	R treatment
	

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	1.18320	0.22753	5.20	0.0350
Actual_Time	Actual Time	1	-0.15973	0.00937	-17.04	0.0034

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	0.89188	0.22800	3.91	0.0596
Actual_Time	Actual Time	1	-0.13947	0.00950	-14.69	0.0046

Figure A.28: Fit Plot for In(Conc) for pig=10.

Simulated Data

Four trial designs were used to simulate the concentration data as detailed in Section 3.5.2 (the rich design, the original design, the sparse design and the intermediate design). With each

design two variability settings were used ($S_{l,l}$ and $S_{h,l}$) but for the intermediate design an extra variability $S_{h,h}$ setting was added. To introduce a difference in bioavailability between the R and T formulations, the trials for the T formulation were simulated under two hypotheses the $H_{0;80\%}$ and $H_{0;125\%}$, as detailed in Section 3.5.1. The plots of the simulated trials are presented in Figure A.29 to Figure A.42.

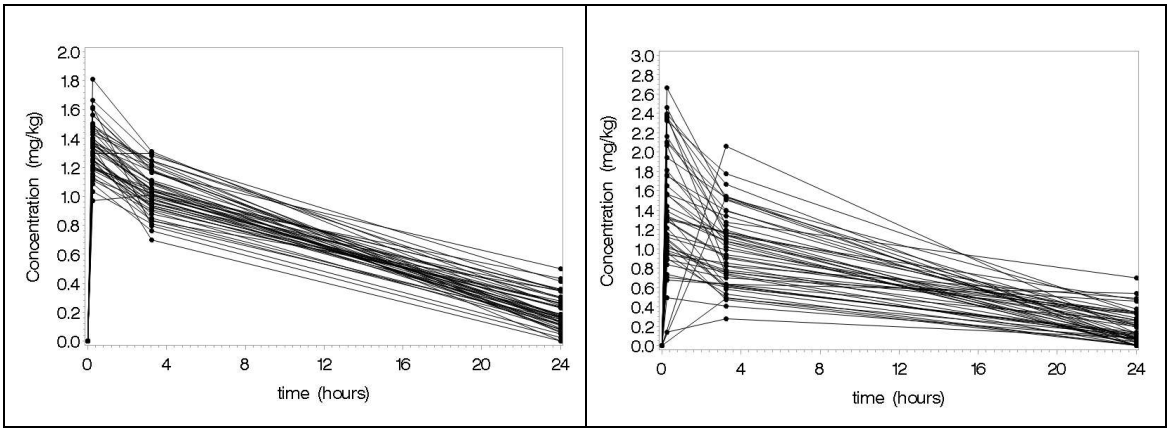


Figure A.29: The simulated concentration of the R formulation for the sparse design (N=50, n=3) left with (S_{ll}) and right with (S_{hl}).

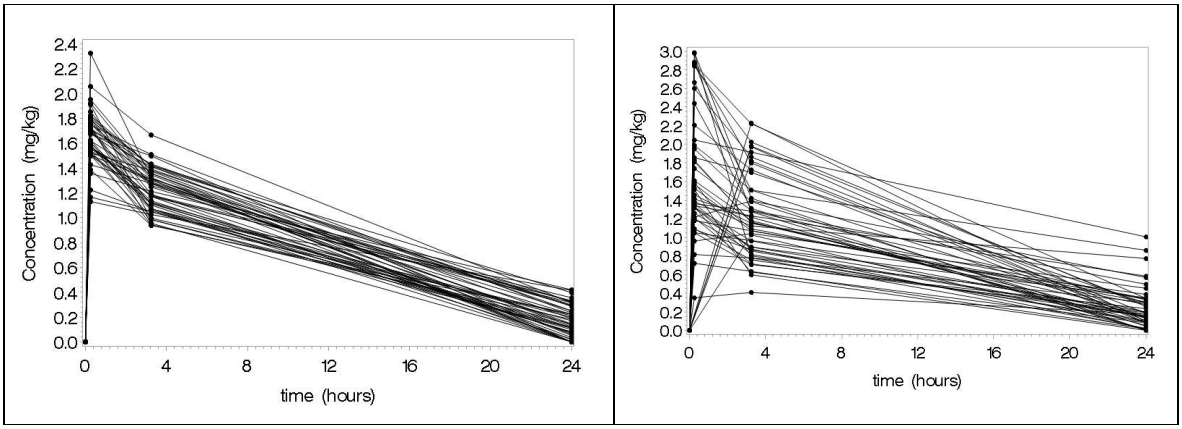


Figure A.30: The simulated concentration of the T formulation for the sparse design (N=50, n=3) under $H_{0;\,80\%}$ hypothesis left with (S_{ll}) and right with (S_{hl}).

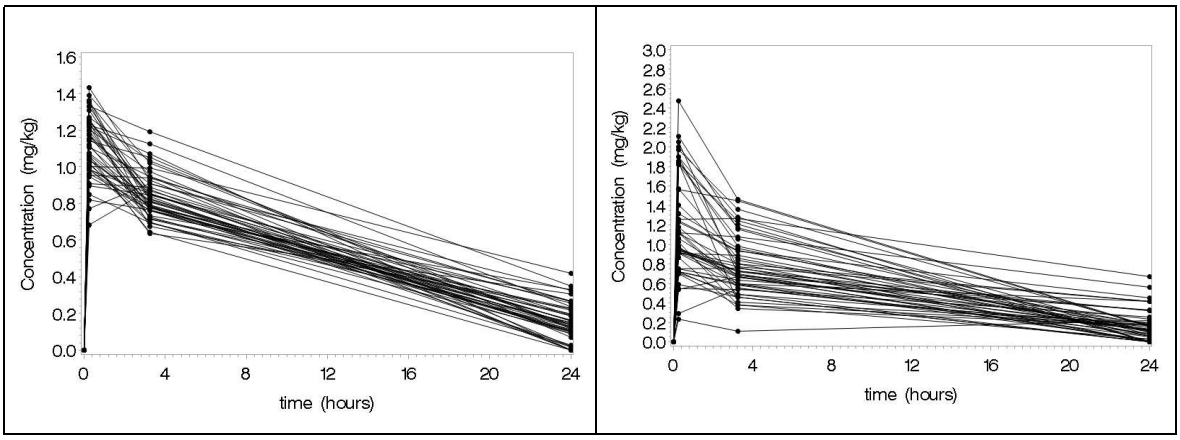


Figure A.31: The simulated concentration of the T formulation for the sparse design (N=50, n=3) under $H_{0;\,125\%}$ hypothesis left with (S_{ll}) and right with (S_{hl}).

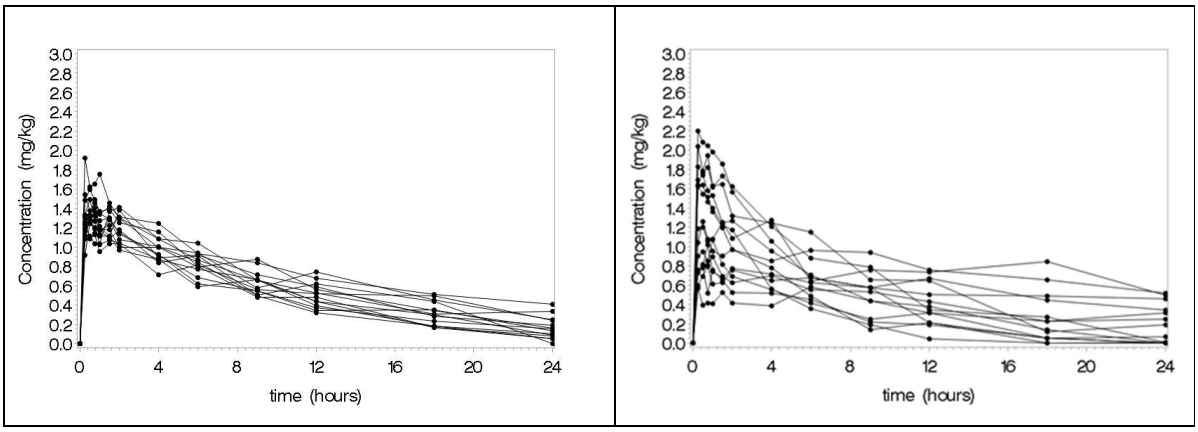


Figure A.32: The simulated concentration of the R formulation for the original design original design (N=14, n=12) left with (S_{ll}) and right with (S_{hl}).

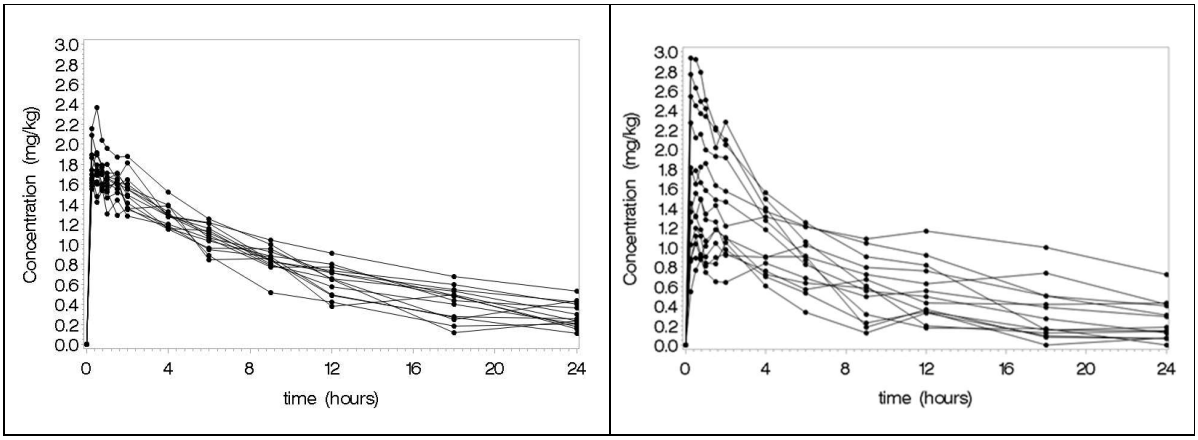


Figure A.33: The simulated concentration of the T formulation for the original design (N=14, n=12) under $H_{0;\,80\%}$ hypothesis left with (S_{ll}) and right with (S_{hl}).

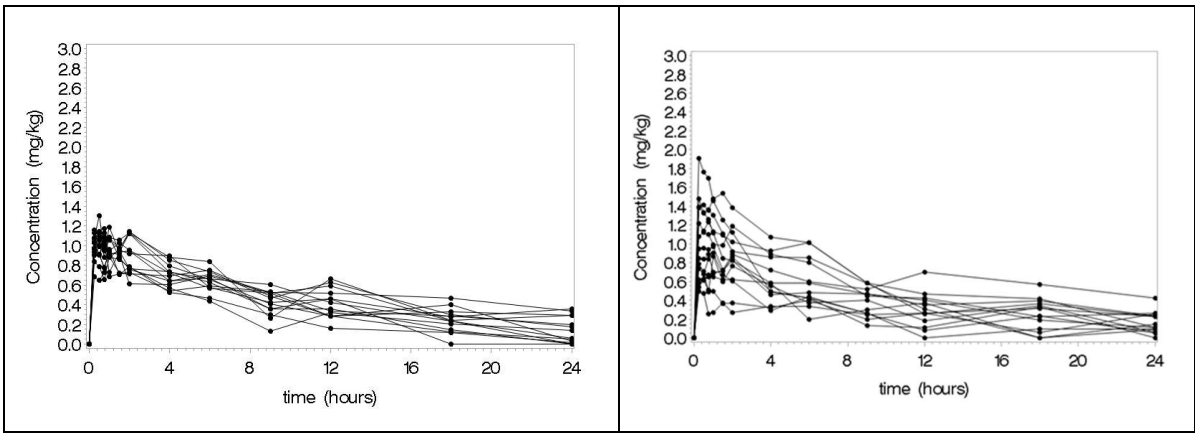


Figure A.34: The simulated concentration of the T formulation for the original design (N=14, n=12) under $H_{0;\,125\%}$ hypothesis left with (S_{ll}) and right with (S_{hl}).

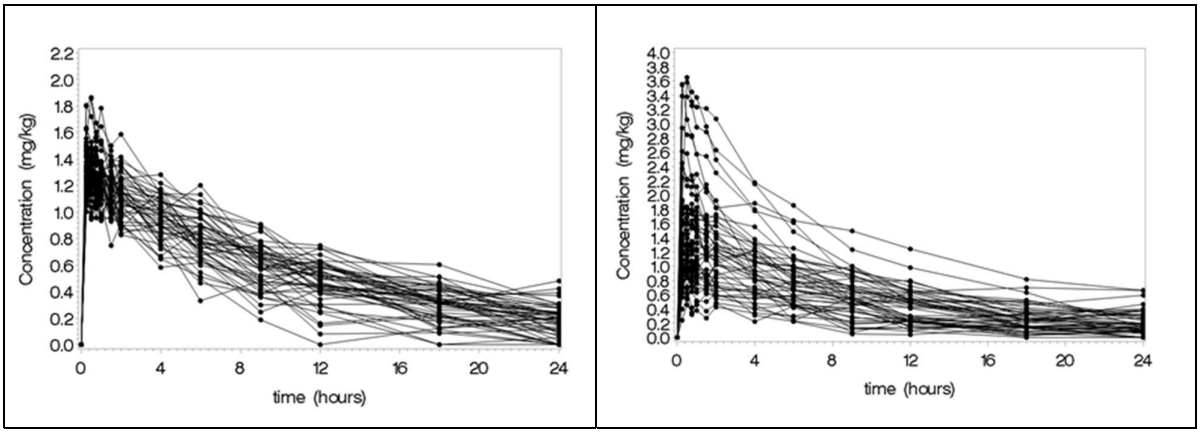


Figure A.35: The simulated concentration of the R formulation for the rich design (N=50, n=12) left with (S_{ll}) and right with (S_{hl}).

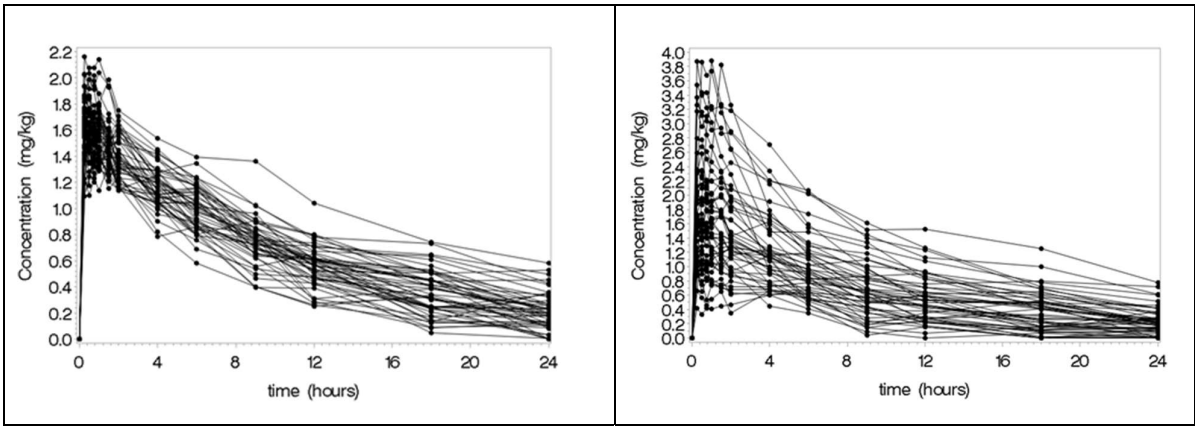


Figure A.36: The simulated concentration of the T formulation for the rich design (N=50, n=12) under $H_{0;\,80\%}$ hypothesis left with (S_{ll}) and right with (S_{hl}).

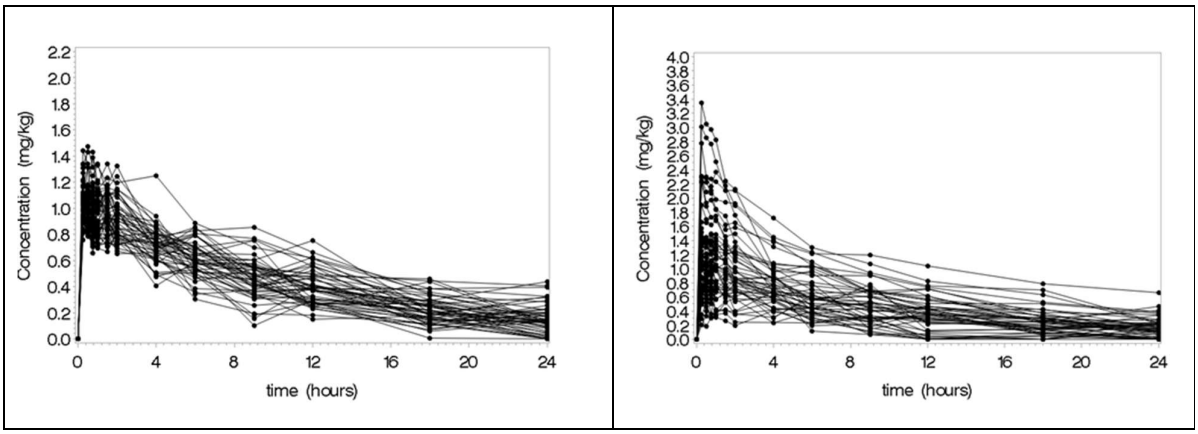


Figure A.37: The simulated concentration of the T formulation for the rich design (N=50, n=12) under $H_{0;\,125\%}$ hypothesis left with (S_{ll}) and right with (S_{hl}).

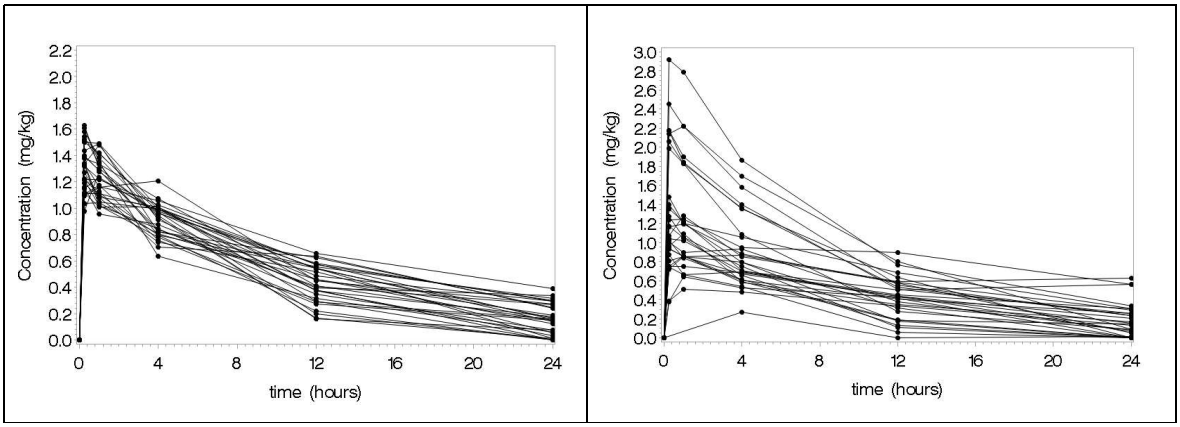


Figure A.38: The simulated concentration of the R formulation for the intermediate design (N=30, n=5) left with (S_{ll}) and right with (S_{hl}).

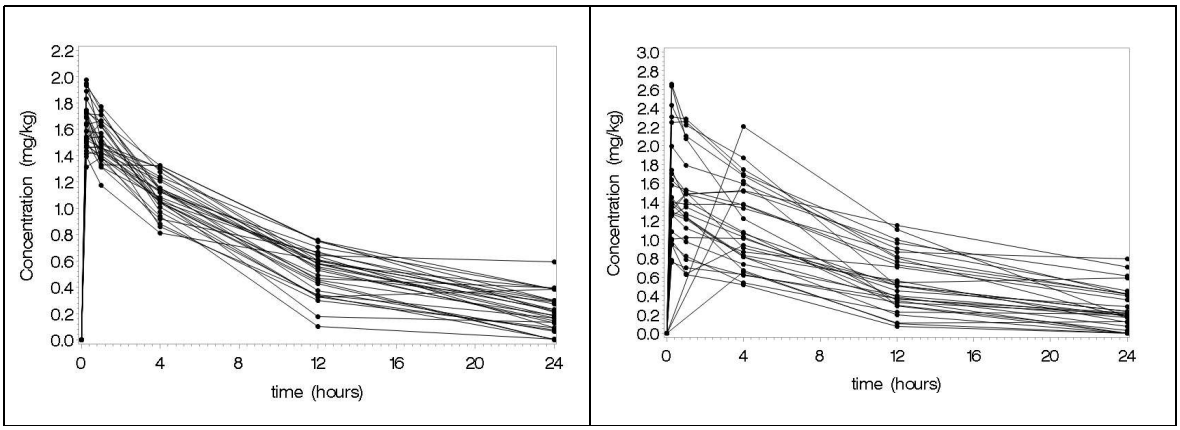


Figure A.39: The simulated concentration of the T formulation for the intermediate design (N=30, n=5) under $H_{0;\,80\%}$ hypothesis left with (S_{ll}) and right with (S_{hl}).

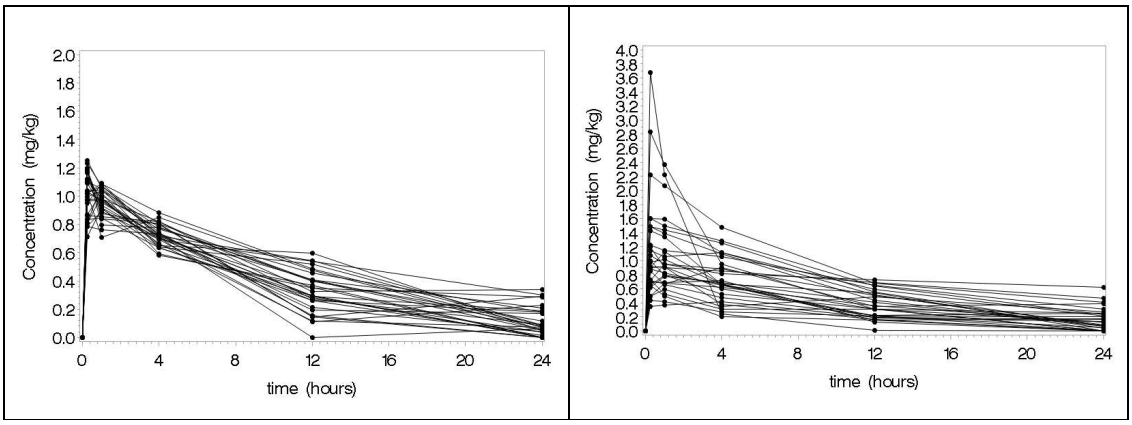


Figure A.40: The simulated concentration of the T formulation for the intermediate design (N=30, n=5) under $H_{0;\,125\%}$ hypothesis left with (S_{ll}) and right with (S_{hl}).

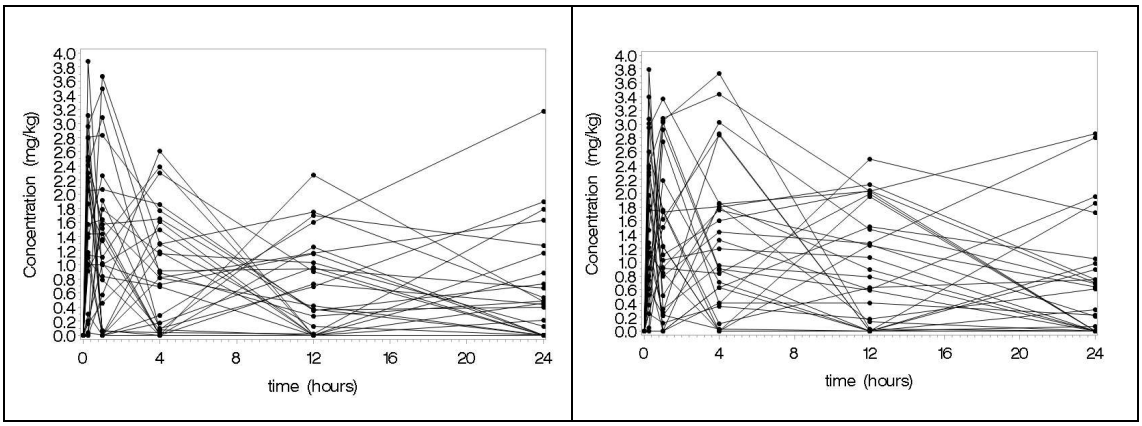


Figure A.41: The simulated concentration of the R and the T formulations for the intermediate design (N=30, n=5) under $H_{0;\,80\%}$ hypothesis, simulated with S_{hh}, left (R formulation) and right (T formulation).

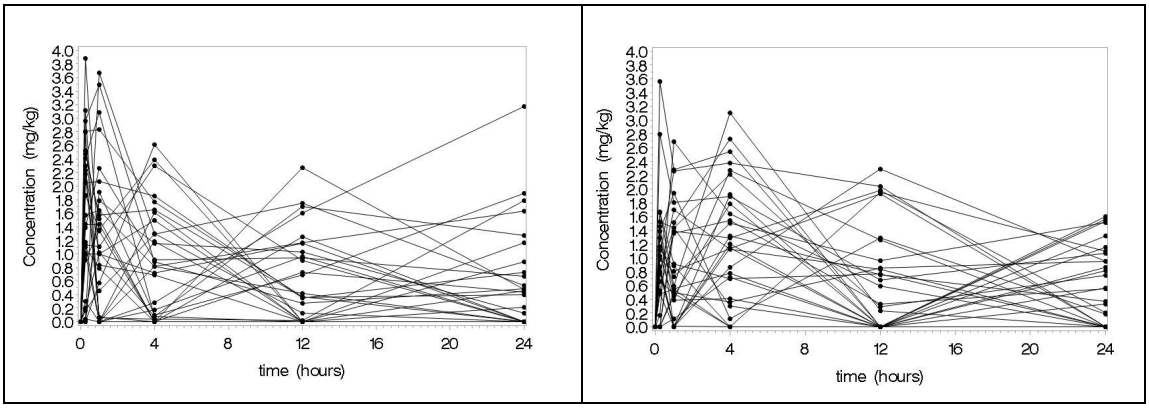


Figure A.42: The simulated concentration of the R and the T formulations for the intermediate design (N=30, n=5) under $H_{0;\,125\%}$ hypothesis simulated with S_{hh}, left (R formulation) and right (T formulation).

The SAEM Covergence

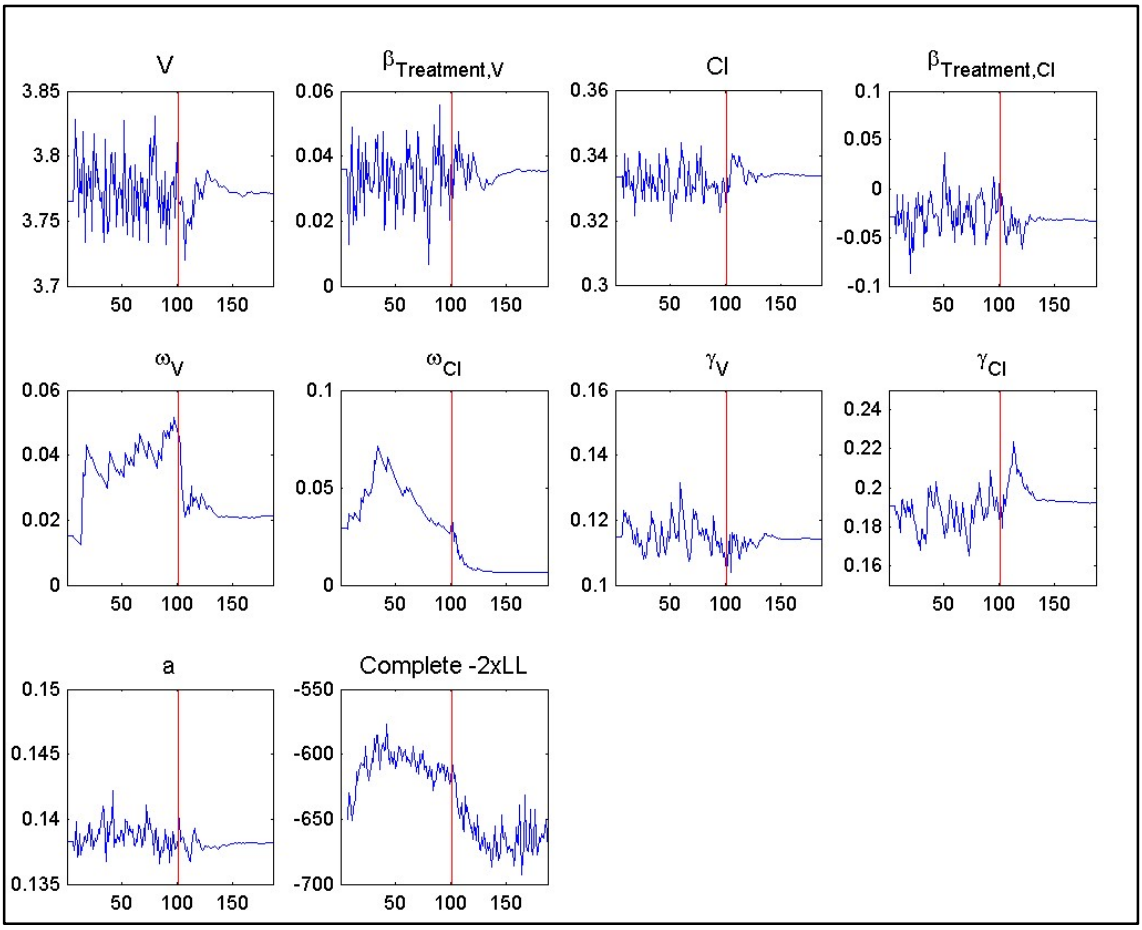


Figure A.43: The SAEM covergence.

Appendix D Tabulated Results

The ANOVA Analysis for Population Bioequivalence

Table B.7 and Table B.8 show the ANOVA results for $\ln AUC_{0-\infty}$ and $\ln AUC_{0-k}$ obtained by the NCA method. Table B.9 shows the ANOVA results for $\ln AUC_{0-\infty}$ obtained by the NLMEM method and Table B.10 presents the ANOVA results for $\ln C_{max}$. The values of UN (1, 1) and UN (2, 2) were used to compute θ_{PBE}, the PBE value, to be compared to θ_P, the PBE limit.

Table B.1: The Covariance Parameter Estimates for $\ln AUC_{0-k}$ using the NCA method.

Covariance Parameter Estimates									
Cov Parm	pig	Group			Estimate				
UN(1,1)	pig				0.01612				
UN(2,1)	pig				0.01268				
UN(2,2)	pig				0.006882				
Residual	pig	Formulation R			0.0115				
Residual	pig	Formulation T			0.0115				
Type 3 Tests of Fixed Effects									
Effect		Num DF		DF	F Value			Pr > F	
sequence		1		12	1.16			0.3027	
period		1		12	48.81			<.0001	
formulation		1		12	0.04			0.8371	
Least Squares Means									
Effect	form	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper
formulation	R	2.5482	0.04442	12	57.37	<.0001	0.1	2.469	2.6274
formulation	T	2.5563	0.03624	12	70.54	<.0001	0.1	2.4917	2.6209

Table B.2: The Covariance Parameter Estimates for $\ln AUC_{0-\infty}$ using the NCA method.

Covariance Parameter Estimates			
Cov Parm	Subject	Group	Estimate
UN(1,1)	pig		0.01623
UN(2,1)	pig		0.01146
UN(2,2)	pig		0.006296
Residual	pig	formulation R	0.01126
Residual	pig	formulation T	0.01126

Type 3 Tests of Fixed Effects				
Effect	Num DF	DF	F Value	Pr > F
sequence	1	12	1.25	0.285
period	1	12	47.42	<.0001
formulation	1	12	0.03	0.8741

Least Squares Means									
Effect	form	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper
formulation	R	2.5552	0.04431	12	57.66	<.0001	0.1	2.4762	2.6342
formulation	T	2.5616	0.03541	12	72.33	<.0001	0.1	2.4985	2.6247

Table B.3: The Covariance Parameter Estimates for $\ln AUC_{0-\infty}$ using the NLMEM method.

Covariance Parameter Estimates			
Cov Parm	Subject	Group	Estimate
UN(1,1)	pig		0.009787
UN(2,1)	pig		0.004787
UN(2,2)	pig		0.001087
Residual	pig	formulation R	0.005437
Residual	pig	formulation T	0.009787

Type 3 Tests of Fixed Effects				
Effect	Num DF	DF	*F* Value	Pr > *F*
sequence	1	11	3.48	0.089
period	1	11	53.32	<.0001
formulation	1	11	1.65	0.2253

Least Squares Means									
Effect	form	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper
formulation	R	2.7146	0.03432	11	79.09	<.0001	0.1	2.6529	2.7762
formulation	T	2.754	0.02247	11	122.58	<.0001	0.1	2.7137	2.7943

Note: The model did not converge with subject (pig no 4) in the data, hence it was omitted from the analysis.

Table B.4: The Covariance Parameter Estimates for lnC_{max}.

Covariance Parameter Estimates			
Cov Parm	Subject	Group	Estimate
UN(1,1)	Pig		0.003631
UN(2,1)	Pig		-0.00041
UN(2,2)	Pig		0.003992
Residual	Pig	formulation R	0.003812
Residual	Pig	formulation T	0.003812

Type 3 Tests of Fixed Effects				
Effect	Num DF	DF	F Value	Pr > F
sequence	1	12	0.03	0.8725
period	1	12	1.77	0.2079
formulation	1	12	0.09	0.7697

Least Squares Means									
Effect	form	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper
formulation	R	0.2864	0.02306	12	12.42	<.0001	0.1	0.2453	0.3275
formulation	T	0.2762	0.02361	12	11.7	<.0001	0.1	0.2341	0.3183

Median Shrinkage

Table B.5 contains the details of the median shrinkage of 1000 simulated trials for each design and each variability setting.

Table B.5: The shrinkage and Type-I-error for the different designs and variability settings.

Variability settings	Sparse design ($N = 50, n = 3$)		Original design ($N = 14, n = 12$)		Rich design ($N = 50, n = 12$)		Intermediate ($N = 30, n = 5$)	
	Type-I-error	shrinkage	Type-I-error	shrinkage	Type-I-error	shrinkage	Type-I-error	shrinkage
$S_{l,l}$	20	62	12.7	82	10.5	84	16.8	80
$S_{h,l}$	14.2	91	8.6	86	6.8	92	7	94
$S_{h,h}$							99	28

The RMSE for Each Design and Each Variability

Table B.6 presents the values of the RMSE for each design and each variability settings.

Table B.6: The RMSE for the mean of $\ln AUC_{0-\infty}$.

Variability settings	Sparse design ($N = 50, n = 3$)		Original design ($N = 14, n = 12$)		Rich design ($N = 50, n = 12$)		Intermediate design ($N = 30, n = 5$)	
	NCA	NLMEM	NCA	NLMEM	NCA	NLMEM	NCA	NLMEM
$S_{l,l}$	2.64	0.051	2.48	0.029	2.47	0.022	2.51	0.049
$S_{h,l}$	2.67	0.065	2.46	0.066	2.46	0.067	2.51	0.066
$S_{h,h}$							2.85	0.065

The ANOVA Analysis for Population Bioequivalence

Table B.7 and Table B.8 show the ANOVA results for $\ln AUC_{0-\infty}$ and $\ln AUC_{0-k}$ obtained by the NCA method. Table B.9 shows the ANOVA results for $\ln AUC_{0-\infty}$ obtained by the NLMEM method and Table B.10 presents the ANOVA results for $\ln Cmax$. The values of UN (1, 1) and UN (2, 2) were used to compute θ_{PBE}, the PBE value, to be compared to θ_P, the PBE limit.

Table B.7: The Covariance Parameter Estimates for $\ln AUC_{0-\infty}$ using the NCA method.

Covariance Parameter Estimates			
Cov Parm	pig	Group	Estimate
UN(1,1)	pig		0.01612
UN(2,1)	pig		0.01268
UN(2,2)	pig		0.006882
Residual	pig	Formulation R	0.0115
Residual	pig	Formulation T	0.0115

Type 3 Tests of Fixed Effects				
Effect	Num DF	DF	F Value	Pr > F
sequence	1	12	1.16	0.3027
period	1	12	48.81	<.0001
formulation	1	12	0.04	0.8371

Least Squares Means

Effect	form	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper
formulation	R	2.5482	0.04442	12	57.37	<.0001	0.1	2.469	2.6274
formulation	T	2.5563	0.03624	12	70.54	<.0001	0.1	2.4917	2.6209

Table B.8: The Covariance Parameter Estimates for $\ln AUC_{0-\infty}$ using the NCA method.

Covariance Parameter Estimates			
Cov Parm	Subject	Group	Estimate
UN(1,1)	pig		0.01623
UN(2,1)	pig		0.01146
UN(2,2)	pig		0.006296
Residual	pig	formulation R	0.01126
Residual	pig	formulation T	0.01126

Type 3 Tests of Fixed Effects				
Effect	Num DF	DF	F Value	Pr > F
sequence	1	12	1.25	0.285
period	1	12	47.42	<.0001
formulation	1	12	0.03	0.8741

Least Squares Means									
Effect	form	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper
formulation	R	2.5552	0.04431	12	57.66	<.0001	0.1	2.4762	2.6342
formulation	T	2.5616	0.03541	12	72.33	<.0001	0.1	2.4985	2.6247

Table B.9: The Covariance Parameter Estimates for $\ln AUC_{0-\infty}$ using the NLMEM method.

Covariance Parameter Estimates			
Cov Parm	Subject	Group	Estimate
UN(1,1)	pig		0.009787
UN(2,1)	pig		0.004787
UN(2,2)	pig		0.001087
Residual	pig	formulation R	0.005437
Residual	pig	formulation T	0.009787

Type 3 Tests of Fixed Effects				
Effect	Num DF	DF	F Value	Pr > F
sequence	1	11	3.48	0.089
period	1	11	53.32	<.0001
formulation	1	11	1.65	0.2253

Least Squares Means									
Effect	form	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper
formulation	R	2.7146	0.03432	11	79.09	<.0001	0.1	2.6529	2.7762
formulation	T	2.754	0.02247	11	122.58	<.0001	0.1	2.7137	2.7943

Note: The model did not converge with subject (pig no 4) in the data, hence it was omitted from the analysis.

Table B.10: The Covariance Parameter Estimates for lnC_{max}.

Covariance Parameter Estimates			
Cov Parm	Subject	Group	Estimate
UN(1,1)	Pig		0.003631
UN(2,1)	Pig		-0.00041
UN(2,2)	Pig		0.003992
Residual	Pig	formulation R	0.003812
Residual	Pig	formulation T	0.003812

Type 3 Tests of Fixed Effects				
Effect	Num DF	DF	F Value	Pr > F
sequence	1	12	0.03	0.8725
period	1	12	1.77	0.2079
formulation	1	12	0.09	0.7697

Least Squares Means									
Effect	form	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper
formulation	R	0.2864	0.02306	12	12.42	<.0001	0.1	0.2453	0.3275
formulation	T	0.2762	0.02361	12	11.7	<.0001	0.1	0.2341	0.3183

Calculation of $AUC_{0-\infty}$ from the Original Data

Table B.11: Calculation of $AUC_{0-\infty}$ for sequence 1.

Sequence	Pig	Treatment	AUC_{0-k}	β_1	C_k	$\lambda = -2.303 * \beta_1$	$AUC_{0-\infty}$
1	6	R	7.6844	-0.1534	0.0060	0.3532	7.7014
	6	T	12.1904	-0.1014	0.0850	0.2335	12.5544
	10	R	15.4897	-0.1395	0.0190	0.3212	15.5489
	10	T	15.0854	-0.1597	0.0110	0.3679	15.1153
	12	R	9.5135	-0.2348	0.0150	0.5407	9.5413
	12	T	14.0283	-0.1378	0.0170	0.3173	14.0819
	13	R	9.1711	-0.2418	0.0100	0.5568	9.1891
	13	T	12.3953	-0.1360	0.0140	0.3133	12.4400
	14	R	11.0338	-0.2349	0.0170	0.5411	11.0652
	14	T	14.9046	-0.1350	0.0190	0.3108	14.9657
	23	R	12.2230	-0.1407	0.0200	0.3240	12.2847
	23	T	15.6595	-0.1206	0.0270	0.2778	15.7567
	25	R	11.9016	-0.1268	0.0260	0.2919	11.9906
	25	T	15.4633	-0.1610	0.0110	0.3708	15.4930

Table B.12: Calculation of $AUC_{0-\infty}$ for sequence 2.

Sequence	Pig	Treatment	AUC_{0-k}	β_1	C_k	$\lambda = -2.303 * \beta_1$	$AUC_{0-\infty}$
2	4	R	13.8473	-0.1044	0.0930	0.2404	14.2341
	4	T	8.3514	-0.1890	0.0260	0.4351	8.4112
	9	R	14.7908	-0.1482	0.0910	0.3412	15.0575
	9	T	14.1151	-0.1671	0.0120	0.3848	14.1463
	11	R	15.2143	-0.1585	0.0100	0.3651	15.2417
	11	T	11.9056	-0.2182	0.0220	0.5026	11.9494
	15	R	14.9325	-0.1243	0.0230	0.2863	15.0128
	15	T	11.8040	-0.3101	0.0090	0.7143	11.8166
	18	R	15.8373	-0.1390	0.0170	0.3201	15.8904
	18	T	12.3218	-0.3536	0.0060	0.8143	12.3291
	19	R	15.4373	-0.1518	0.0120	0.3495	15.4717
	19	T	11.9164	-0.1775	0.0080	0.4087	11.9359
	24	R	16.4227	-0.1178	0.0310	0.2712	16.5370
	24	T	12.4213	-0.2104	0.0370	0.4845	12.4977

Example of the TOST Results for Simulated Data

Table B.13: TOST results for the sparse design with S_{II} under $H_{0;\,80\%}$ $AUC_{0-\infty}$ data.

N	Geometric Mean	Coefficient of variation	Minimum	Maximum
50	0.7844	0.0445	0.721	0.8363

Geometric Mean	95% CL Mean		Coefficient of variation	95% CL CV	
0.7844	0.7745	0.7944	0.0445	0.0372	0.0555

TOST Level 0.05 Equivalence Analysis

Geometric Mean	Lower Bound	90% CL Mean		Upper Bound	Assessment
0.7844	0.8	0.7761	0.7927	1.25	Not equivalent

Test	Null	DF	t Value	P-Value
Upper	0.8	49	-3.13	0.9985
Lower	1.25	49	-74.01	<.0001
Overall				0.9985

The Example of the *AUC* Data Estimated from the Simulated Data

Table B.14: The Sparse design simulated with S_{II} under $H_{0;\,80\%}$ $AUC_{0-\infty}$ data.

Sequence	Pig	Treatment	AUC_{0-k}	β_1	λ	$AUC_{0-\infty}$	sequence	Pig	Treatment	AUC_{0-k}	β_1	λ	$AUC_{0-\infty}$
													19.6883
	2	R	18.1401	-0.0786	0.18105	19.4071		27	R	16.7097	-0.046	0.10584	20.1776
	3	R	15.9113	-0.0886	0.20405	16.7385		28	R	17.2252	-0.0564	0.12993	19.5775
	4	R	13.624	-0.5189	1.19492	13.6241		29	R	16.1964	-0.1026	0.23628	16.7126
	5	R	18.085	-0.1091	0.25128	18.5162		30	R	17.3292	-0.1119	0.25767	17.7425
	6	R	20.4069	-0.054	0.12424	23.4862		31	R	13.1258	-0.5142	1.18414	13.1258
	7	R	19.7974	-0.0482	0.111	23.4559		32	R	17.2552	-0.0629	0.14477	19.2192
	8	R	14.7251	-0.0602	0.13859	16.7602		33	R	16.7276	-0.0673	0.15504	18.5325
	9	R	15.6427	-0.0928	0.2137	16.3142		34	R	17.7112	-0.0949	0.21852	18.4563
	10	R	14.4446	-0.0962	0.22152	15.0779		35	R	16.448	-0.0833	0.19181	17.4684
	11	R	19.5927	-0.053	0.12202	22.9377		36	R	12.1601	-0.0921	0.21201	12.7496
1	12	R	19.9681	-0.0792	0.18234	21.2212	2	37	R	15.9405	-0.0801	0.18438	17.0014
	13	R	12.1056	-0.2308	0.53163	12.1171		38	R	15.4289	-0.0759	0.17486	16.5532
	14	R	13.7785	-0.1039	0.23925	14.2183		39	R	19.7083	-0.048	0.11042	23.3586
	15	R	15.706	-0.1206	0.27777	15.9891		40	R	15.1754	-0.1262	0.29053	15.4315
	16	R	17.0181	-0.0934	0.21513	17.7705		41	R	13.7932	-0.1087	0.25021	14.1749
	17	R	14.4451	-0.1938	0.44634	14.4799		42	R	19.3779	-0.0564	0.12993	22.0108
	18	R	11.4875	-0.1234	0.28421	11.7074		43	R	13.4475	-0.1436	0.33062	13.5627
	19	R	18.861	-0.0671	0.15462	20.795		44	R	16.7229	-0.0566	0.13028	19.0626
	20	R	15.142	-0.0585	0.13461	17.0386		45	R	19.5202	-0.0813	0.18723	20.7691
	21	R	14.2994	-0.5198	1.19716	14.2994		46	R	16.5792	-0.1323	0.30467	16.7989
	22	R	15.7687	-0.0676	0.15571	17.2706		47	R	17.491	-0.1171	0.26967	17.8547
	23	R	14.0326	-0.0687	0.15812	15.3441		48	R	15.6043	-0.0761	0.17526	16.6791

Sequence	Pig	Treatment	AUC_{0-k}	β_1	λ	$AUC_{0-\infty}$	sequence	Pig	Treatment	AUC_{0-k}	β_1	λ	$AUC_{0-\infty}$
1	24	R	19.9762	-0.0496	0.1142	23.7027	2	49	R	17.9401	-0.0632	0.14543	19.7072
	25	R	16.628	-0.0737	0.16961	18.0179		50	R	15.6248	-0.0664	0.15292	17.3244
	1	T	13.233	-0.0613	0.14124	14.8984		26	T	13.239	-0.0794	0.18275	14.0753
	3	T	9.7907	-0.2088	0.48089	9.8065		28	T	14.8299	-0.0565	0.1301	16.8747
	4	T	10.9726	-0.1092	0.2515	11.2856		29	T	9.8984	-0.1438	0.33125	9.9865
	5	T	15.2071	-0.0416	0.09586	19.1372		30	T	15.1337	-0.0925	0.21292	15.7359
	6	T	14.4075	-0.0801	0.18442	15.4001		31	T	11.3629	-0.0926	0.21326	11.8813
	7	T	16.2952	-0.0583	0.13423	18.2483		32	T	11.5558	-0.0994	0.22881	11.9697
	8	T	15.1131	-0.1	0.23029	15.6256		33	T	10.3261	-0.5093	1.1728	10.3261
	9	T	9.4809	-0.4991	1.14947	9.4809		34	T	13.5527	-0.1835	0.42254	13.5943
	10	T	12.7396	-0.067	0.15439	14.0749		35	T	9.7312	-0.1134	0.26111	9.9923
	11	T	10.8729	-0.091	0.20965	11.4204		36	T	12.2805	-0.0854	0.19668	12.957
	12	T	14.0487	-0.0869	0.20019	14.8019		37	T	12.8831	-0.0764	0.17603	13.8736
	13	T	13.8565	-0.0666	0.15343	15.1783		38	T	15.137	-0.0568	0.13071	17.1706
	14	T	14.6057	-0.0657	0.15137	16.0709		39	T	14.6215	-0.1004	0.23117	15.1117
	15	T	10.9646	-0.5125	1.18032	10.9646		40	T	10.9351	-0.1318	0.30363	11.0955
	16	T	14.0863	-0.0715	0.16471	15.2546		41	T	10.4887	-0.5089	1.17197	10.4887
	17	T	10.9535	-0.1312	0.30221	11.1049		42	T	15.1591	-0.052	0.11974	17.7744
	18	T	9.4011	-0.5086	1.17138	9.4011		43	T	12.7908	-0.0957	0.2205	13.2876
	19	T	11.1269	-0.0854	0.19671	11.8091		44	T	10.4624	-0.0479	0.11034	12.6583
	20	T	16.8844	-0.0544	0.12517	19.2376		45	T	15.4903	-0.0695	0.1601	17.0436
	21	T	13.2221	-0.0963	0.22166	13.6846		46	T	14.8021	-0.0497	0.11437	17.7343
	22	T	11.162	-0.1045	0.24061	11.5276		47	T	16.1124	-0.0634	0.14591	17.8918
	23	T	11.6072	-0.0653	0.15036	12.827		48	T	9.7047	-0.0796	0.18322	10.3976
	24	T	16.4032	-0.0551	0.1269	18.7687		49	T	11.0382	-0.0595	0.13692	12.4236
	25	T	14.3017	-0.0634	0.1461	15.7428		50	T	13.378	-0.083	0.19113	14.0743