Pharmacometrically Driven Optimisation of Dose Regimens in Clinical Trials

Kabir Soeny

School of Mathematical Sciences
Queen Mary University of London
United Kingdom

Submitted in partial fulfilment of the requirements
of the Degree of Doctor of Philosophy

July 2016

To mother and father
with humility and gratitude.

Declaration of Originality

I, Kabir Soeny, confirm that the research included within this thesis is my own work or that where it has been carried out in collaboration with, or supported by others, that this is duly acknowledged below and my contribution indicated. Previously published material is also acknowledged below.

I attest that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge break any UK law, infringe any third party's copyright or other Intellectual Property Right, or contain any confidential material.

I accept that the College has the right to use plagiarism detection software to check the electronic version of the thesis.

I confirm that this thesis has not been previously submitted for the award of a degree by this or any other university.

Signed: Kabir Soeny

July 20, 2016

Details of collaboration and publications:

The work in this thesis was throughout supervised by Dr Barbara Bogacka, QMUL and Prof Byron Jones, Novartis Pharma AG.

Our joint publication, Soeny, Bogacka, Jones, and Bouillon (2016), is primarily based on Chapter 4. The problem presented in this chapter was discussed by all the authors, the algorithm was developed by myself.

The example of the kill rate curves (shown in Section 7.2) was suggested by Dr Thomas Bouillon as a part of our collaboration with Novartis. The solution to the given problem was developed by myself.

Abstract

The dose regimen of a drug gives important information about the dose sizes, dose frequency and the duration of treatment. Optimisation of dose regimens is critical to ensure therapeutic success of the drug and to minimise its possible adverse effects. The central theme of this thesis is the Efficient Dosing (ED) algorithm - a computation algorithm developed by us for optimisation of dose regimens. In this thesis, we have attempted to develop a quantitative framework for measuring the efficiency of a dose regimen for specified criteria and computing the most efficient dose regimen using the ED algorithm. The criteria considered by us seek to prevent over- and under-exposure to the drug. For example, one of the criteria is to maintain the drug's concentration around a desired target level. Another criterion is to maintain the concentration within a therapeutic range or window. The ED algorithm and its various extensions are programmed in MATLAB®. Some distinguishing features of our methods are: mathematical explicitness in the optimisation process for a general objective function, creation of a theoretical base to draw comparisons among competing dose regimens, adaptability to any drug for which the PK model is known, and other computational features. We develop the algorithm further to compute the optimal ratio of two partner drugs in a fixed dose combination unit and the efficient dose regimens. In clinical trials, the parameters of the PK model followed by the drug are often unknown. We develop a methodology to apply our algorithm in an adaptive setting which enables estimation of the parameters while optimising the dose regimens for the typical subject in each cohort. A potential application of the ED algorithm for individualisation of dose regimens is discussed. We also discuss an application for computation of efficient dose regimens for obliteration of a pre-specified viral load.

Contents

List of Figures

List of Tables

Acknowledgements

"When the disciple is ready, the guru appears."
My passion and love for statistics which blossomed during my BSc and MSc days at the University of Delhi and a deep yearning to make a contribution to the discipline, took a momentous turn when I was awarded a PhD studentship at Queen Mary University of London. My supervisors - Dr Barbara Bogacka and Prof Byron Jones have been my friends, philosophers and preceptors during this long but enjoyable journey. The excellent blend of academic guidance, motivation and constructive criticism I experienced under their supervision has propelled me to accomplish milestones I had doubted in the beginning. Without them, this project would have certainly not seen the light of the day. I am more grateful to them than the words at my command can express.

Dr Thomas Bouillon, formerly at Novartis Pharma AG, must be credited for introducing the clinical problem of dose optimisation to me and for providing an industry perspective to my doctoral work. I have thoroughly enjoyed working with him remotely and during my visits to Novartis in Basel, Switzerland. I am also grateful to the members of the Statistical Methodology group at Novartis for guiding me in increasing the relevance of my research to the industry. My presentations at Novartis have enabled me to interact with some of the best minds working in the field of pharmacometrics.

For someone who had never studied outside his hometown before, studying in a new continent initially seemed like an intimidating prospect. QMUL and the School of Mathematical Sciences left no stone unturned in ensuring that I have a comfortable and memorable time in London. I have some beautiful memories of my stay in the UK to cherish throughout my life. The conducive environment enabled me to fully concentrate on my research while the regular series of seminars and talks helped me to look beyond my domain. Thanks to Dr Steve Coad and Dr Hugo Maruri-Aguilar, my assessors at the annual review meetings, for giving me useful feedback on my work and overseeing my progression formalities with interest and encouragement.

My PhD studentship was funded jointly by QMUL and Novartis. The PGR fund of QMUL made it possible for me to present my research at the 36^{th} annual meeting of the Society for Clinical Trials (2015) at Arlington, USA. The School of Mathematical Sciences, QMUL sponsored me to present my research at the 23^{rd} annual meeting of PAGE (2014) at Alicante, Spain. My visits to Novartis and participation in the PODE 2014 meeting at Roche Pharma AG, Basel were financially supported by Novartis. I am grateful for all the financial assistance rendered to me.

In addition, I would like to express my gratitude to my examiners Prof Thomas Jaki and Dr Alexander Donev for their helpful comments and suggestions which have led to clarification and appreciable extension of the results of this thesis.

Heartfelt thanks to my parents for their continuous support, love and patience during my studies.

Lastly, thanks to the Almighty for being the guiding light in everything.

Chapter 1

Introduction

Paracelsus (1493 - 1546) made this timeless maxim on toxicology: "*All things are poison and nothing is without poison; only the dose makes a thing not a poison.*"

For most illnesses, multiple doses of the therapy are prescribed for a successful treatment. Generally, with an increase in the drug's concentration in the blood, the efficacy and the magnitude of adverse effects also increase. The concentration should therefore be maintained in a range that minimises any under- and overexposure to the drug. The science of pharmacokinetics explains that the concentration of the drug in the blood depends on, among other factors, the associated *dose regimen*, that is on various aspects of administration of the drug to the patient.

Some of the factors that need to be considered for designing a dose regimen are:

- the dose sizes,
- number of doses to be administered,
- the duration of the treatment,
- the time interval between two successive doses or the dose frequency,
- the combination ratio(s) in case multiple drugs are to be administered together in a single dosing unit,
- the route of administration of the drug,
- whether a single dose regimen would suffice for the entire population of patients or different dose regimens need to be designed for subpopulations.

The first five factors are quantitative variables while the other two are qualitative.

The *dose size* is the amount of the drug that is administered at a time. The time for which a subject is exposed to the drug is called the *duration of treatment*. For drugs which are quickly eliminated from the body, the duration could be the time interval between the first and the last dose. For drugs which are retained longer in the body, the exposure continues even after the last dose has been administered. The dose frequency, like the dose size, is an important variable as the concentration levels of the drug in blood depend on it and thus, the dose frequency affects the therapeutic outcome.

Fixed dose combinations consist of two (or more) drugs that are constrained to be administered together at the same times and in the same ratio, with the help of a dosing unit such as a tablet or a capsule. According to WHO (2005), combination units have several advantages such as lower risk of side effects, decreased manufacturing cost, higher patient adherence and simpler logistics of distribution. Such combinations are particularly useful for treatment of HIV/AIDS, malaria, tuberculosis and other serious infectious diseases. Determination of the optimal ratio for combining the partner drugs is, therefore, a crucial problem.

There are various channels to introduce a drug into the body. Some of the common routes are: gastrointestinal which includes oral and rectal administration, intravenous which involves administration into a vein, intramuscular and nasal. Each route has its advantages and drawbacks and the properties of the drug, feasibility, cost and patient convenience are some of the factors which determine the most appropriate route. For instance, while intravenous administration of the drug can introduce the drug quickly into the blood stream by bypassing the gastrointestinal tract, it is not as convenient as an oral administration which can also be had at home.

The design of the dose regimen will affect the resultant concentration of the drug in the blood which, in turn, will affect the treatment outcome. For example, to optimise the therapeutic levels of vitamin D from supplementation, Chao et al. (2013) found that the dose size, frequency and the duration of supplementation are important factors.

Once a candidate drug has been clinically proven to have a therapeutic effect and an acceptable level of safety, the design of the dose regimen is an important problem. An ideal dose regimen achieves the twin objectives of patient safety and therapeutic efficacy. Dose regimen optimisation is about choosing the optimal values of variables such as dose sizes, dose frequencies, combination ratio (if applicable) and the treatment duration. The objective of optimisation is generally to prevent over- and underexposure to the drug. This optimisation can be done at two levels - population level and patient level.

Since dose regimens are usually prescribed for a population or subpopulations of patients, they are designed for the typical patient in the target group. If a concentration value at which the drug has the best balance of efficacy and toxicity is known, then the dose regimen should be designed in such a way that the concentration in the blood plasma is maintained as close as possible to this value, throughout the duration of treatment. For some drugs, it might be desirable to restrict the concentration profile within a *therapeutic range* which consists of a lower and an upper value of blood concentration of the drug. Best results are obtained from the treatment when the concentration of the drug is maintained within this range. In all such cases, optimisation of the dose regimen has to be done at the population level.

Some drugs have a very narrow therapeutic range, for example, some antibiotics and analgesics which means that their concentration in blood should be maintained within the therapeutic range in every patient. This may also be necessary in sensitive patients such as infants and pregnant women. For the reasons of safety and efficacy, the dose regimens for such drugs have to be personalised or individualised to every patient's profile as one dose regimen cannot be relevant for the entire population. That is, the optimisation for these drugs has to be done at the patient level.

1.1 Motivation for our Work

While optimisation of dose regimens for a given target is not a new problem, generally the focus is on the optimisation of dose sizes, ignoring the other aspects of a dose regimen such as the number of doses or the combination ratio. In practice, optimisation of dose regimen is performed on a case-to-case basis for a particular drug or a class of drugs using simulations or with the help of clinical studies. For example, Kuti et al. (2003) use Monte Carlo simulations to compare seven plausible dose regimens of the antibiotic meropenem for their ability to achieve predefined target exposure of the drug in the body. For some drugs, clinical studies are performed to choose the best dose regimen from a number of candidate regimens. For example, Dickinson and Evans (2002) compare three dose regimens of intravaginal misoprostol for second-trimester pregnancy termination by observing three groups of volunteers, each assigned to one of the three dose regimens. Novel solutions are sometimes found for some specific optimisation problems. For example, to improve the image quality in FDG-PET scan of obese patients, Groot et al. (2013) developed a quadratic model for the relationship between bodyweight and the dose size of the radioactive compound fluorodeoxyglucose (FDG). The optimised dose regimen of FDG maintained the image quality, irrespective of the patients' bodyweights.

Some authors, for example Mehvar (1998) or Thompson (1992), use pharmacoki-

netic equations and steady-state concentrations to find optimum loading and maintenance dose sizes for constraining the concentration profile within a given therapeutic range. However, they do not discuss any method to quantify the extent of deviations of the resultant profile from the target concentration. Nor do they discuss the case of optimising the combination ratio and the dose regimen in case of combination therapy.

A mathematically formal and holistic approach for explicit optimisation of dose regimens and a general method of computing optimal dose regimens and combination ratios was found lacking in the statistical literature reviewed by us. A theoretical base for comparing competing dose regimens on their degree of adherence to the target was also found undefined. Our work is a step towards filling this gap.

Our contribution to the subject can be described as follows. Firstly, we propose a mathematical expression for dose regimen which is composed of some of the variables that can be controlled in a clinical trial. This enables us to express the optimisation problem formally. We consider some practical constraints on the dose regimen and incorporate them into the optimisation problem. We then present a computational algorithm we developed, called the *Efficient Dosing (ED) algorithm*, which is able to solve these types of optimisation problems. The criteria of optimisation of the dose regimens we discuss in this thesis aim at preventing over- and underexposure to the drug by maintaining the concentration profile around a target concentration or within a therapeutic range. The algorithm is flexible for a variety of practical and computational constraints. For example, our algorithm offers the choice of having discretised dose sizes or letting them assume real values. The problem of finding the best ratio to combine two drugs in a fixed dose unit is solved by extending the ED algorithm to optimise a linear function of two objective functions. We demonstrate that the ED algorithm can also be used for individualisation of dose regimens. Finally, we propose a solution for optimisation of the dose regimen for obliteration of a target bacterial load.

Apart from the problems discussed in this thesis, the ED algorithm has the potential to be adapted for solving other variants of the objective function. The basic assumption that is made is that the mechanistic model which describes the concentration time relationship of the drug is known. As far as the parameters of the model are concerned, we consider the optimisation problem in both settings - when the model parameters are known and when they are unknown but some initial values are available.

To implement the ED algorithm and its extensions and applications, we wrote programs in MATLAB® software. These are presented at the end of the thesis in Appendices D.1 - D.6. These programs are helpful in not just solving the optimisation problems but also in gaining insights about the optimisation process.

1.2 Outline of the Thesis

The thesis is organised as follows.

In Chapter 2 (Introduction to Pharmacometrics) we introduce the field of pharmacometrics. We explain the basic pharmacokinetic terms and derive the commonly used compartmental models. We then derive the multiple dosing formula for some of the compartmental models. Pharmacodynamics is also introduced and some basic models are discussed. The compartmental models introduced in this chapter are extensively used throughout the thesis to exemplify our methodologies.

In Chapter 3 (Optimisation of Dose Regimens: The Problem) we introduce the problem of optimisation of a dose regimen. Current methods are explained along with their merits and drawbacks. We then explain the medical needs and benefits of optimisation of dose regimens. We end the chapter by a mathematical expression of the general optimisation problem for dose regimens. In further chapters of this thesis, we present some possible solutions to this optimisation problem.

In Chapter 4 (The Efficient Dosing Algorithm for the Case of Known Model Parameters) we introduce the Efficient Dosing (ED) Algorithm developed by us to optimise a dose regimen with respect to the dose sizes. We lay down a few pharmacokinetic criteria and illustrate how the ED algorithm can compute the optimised dose regimen for them. We discuss several practical constraints and their solutions. We also show how the algorithm can be extended to compute the optimal ratio in which two or more drugs are to be combined into a single dosing unit. The ED algorithm and its extensions have been programmed in MATLAB® . We use only mechanistic models in this chapter.

In Chapter 5 (Non-linear Mixed Effects Models - Estimation and Design) we introduce the theory of non-linear mixed effects models. This is essential for application of the ED algorithm to models whose parameters are treated as random variables. We discuss application of the theory to the PK and PD models to account for both inter- and intra-individual variability in the patient population. We then briefly explain the methods of estimation of the parameters, in particular, the methods included in the software MATLAB® . We briefly discuss them and contrast their properties. We then introduce the theory of optimal design of experiments. We discuss the different optimality criteria, the Fisher information matrix and briefly discuss some computational methods to determine the optimal design.

In Chapter 6 (The ED Algorithm for the Case of Unknown Parameters) we consider the problems discussed in Chapter 4 with the difference that the PK parameters of the model are assumed to be unknown and the objective is not just dose regimen

optimisation, but also estimation of the model parameters. We accomplish this by an adaptive approach where information derived from every preceding cohort of patients in a clinical trial is used to improve the dose regimen administered to the next cohort. For this, we extensively apply the theory described in Chapter 5. We validate our methodology by the means of a simulation study. Sensitivity analyses are undertaken to study how the choice of input values affects the results from the methodology. The simulation studies have been programmed in MATLAB® . Unlike Chapter 4 where only mechanistic models are discussed, we use statistical models in this chapter. The results from this chapter demonstrate that the ED algorithm can be applied even without the precise knowledge of the PK parameters.

In Chapter 7 (Potential Applications of the ED Algorithm) we present two further potential applications of the ED algorithm. The first application illustrates the use of the algorithm for dose individualisation. The theory of non-linear mixed effects models is useful here as well. In the second application, we use pharmacodynamic models to optimise the dose regimen for anti-infective drugs where instead of aiming for a target concentration, we modify the methodology so as to optimise the dose regimen for a target reduction in the bacterial load.

In Chapter 8 (Discussion) we summarise our research presented in the thesis and discuss some directions in which we would like to take this work forward.

In the appendices, we present some derivations related to the compartmental models and the MATLAB® programs for implementing our methodologies described in Chapters 4, 6 and 7.

Chapter 2

Introduction to Pharmacometrics

The Food and Drug Administration (FDA or USFDA) defines pharmacometrics in FDA (2015) as:

> "Pharmacometrics is an emerging science defined as the science that quantifies drug, disease and trial information to aid efficient drug development and/or regulatory decisions."

The field of pharmacometrics rests on three pillars:

- Drug models which consist of two types: Pharmacokinetic (PK) models which describe the relationship between the drug's exposure (concentration in blood) and time and Pharmacodynamic (PD) models which describe the drug's desired and undesired effects as a function of the exposure.

- Disease models which study naturally occurring or experimentally induced diseases in animals whose pathological processes are mostly similar to that of humans. For example, Pandey and Nichols (2011) use the common fruit fly, *Drosophila melanogaster*, to discuss models of human diseases such as cardiovascular disease, cancer and diabetes for therapeutic discovery. This is based on the belief that nearly 75% of human disease-causing genes have a functional homologue in the fly.

- The trial models which describe the inclusion/exclusion criteria for a clinical trial and patient compliance to the clinical trial protocol. For example, Kenna et al. (2005) present models for patient adherence to the assigned therapy in clinical trials.

However, the major thrust of pharmacometrics is on the drug models and for our work as well, we focus on pharmacokinetics and pharmacodynamics. They are discussed in detail in this chapter.

2.1 Pharmacokinetics

The term pharmacokinetics (PK) is derived from the Greek words *pharmakon* and *kinetikos* which mean 'drug' and 'in motion'. PK is the study of how the body processes a substance, usually a drug, from the point at which it is administered until its elimination from the body. A drug, when administered in a subject, goes through the four phases of absorption, distribution, metabolism, and excretion (ADME). Absorption refers to the movement of the drug into the bloodstream. The rate and extent of the absorption depend on the particular drug and also on the route of administration. Distribution of the drug refers to the reversible transfer of the drug from one site to another within the body. Metabolism is the biotransformation or the breakdown of the drug into its metabolites by the body. The metabolites are responsible for the pharmacological action of the drug. Finally, the drug is eliminated from the body; the main organs involved in this are the liver (for metabolism) and the kidneys (for elimination).

Gibaldi and Levy (1976) describe pharmacokinetics as:

> "Pharmacokinetics is concerned with the study and characterization of the time course of drug absorption, distribution, metabolism and excretion, and with the relationship of these processes to the intensity and time course of therapeutic and adverse effects of drugs. It involves the application of mathematical and biochemical techniques in a physiologic and pharmacologic context."

The scope of PK is not just limited to drugs but can be extended to any substance in an organism such as toxins, nutrients and hormones. PK studies typically involve collection of data related to concentration of a substance in the body over time. Thus, for example, blood samples can be collected over time to give values of the plasma concentration of a drug. The data can be used to build models that can describe the time course of the drug's concentration in the body.

Pharmacodynamics, on the other hand, relates the therapeutic effect of a drug to its concentration in the body. Thus, Pharmacokinetics may be simply defined as what the body does to the drug, as opposed to Pharmacodynamics which may be defined as what the drug does to the body, Benet (1984).

2.1.1 Pharmacokinetic Parameters

We now define some common PK parameters and some of the empirical relationships between them as given in Rowland and Tozer (1994):

t_{max} is the *time of maximum concentration* achieved by the drug in plasma.

$C(t_{max})$ or C_{max} is the *maximum concentration* achieved by the drug in plasma.

Bioavilability (F) is the fraction of the dose actually absorbed into the systemic circulation. Even for the same drug, F can vary widely depending on how it is administered into the body. Intravenous (i.v.) bolus administration ensures a bioavailability of 1 as it bypasses the barriers to absorption, but all other modes of administration have bioavailability less than 1. The bioavailability may also be affected by such factors as food or other drugs taken at the same time.

Absorption Rate Constant (K_a) is the rate at which the drug is absorbed into the systemic circulation. It has units of inverse time.

Elimination Rate Constant (K_e) is the rate at which the drug is eliminated from the body. It has units of inverse time.

Volume of distribution (V) is the apparent volume into which the drug distributes in the body at equilibrium. Thus, if X is the amount of drug in the body at plasma concentration C then, $V = \frac{X}{C}$. In general, the volume of distribution and the true volume of the compartment are different as the drug is often bound in the tissues or the drug solution is not homogeneous.

Clearance (CL) is the volume of plasma cleared of the drug per unit time. CL is the proportionality factor that relates the rate of elimination of the drug from the body to its concentration. The units of Clearance are volume/time. The total body clearance is equal to the sum of renal, intrinsic or hepatic (i.e. by liver) and lung clearance.

Area Under the Curve (AUC) is the area bound under the concentration-time curve. If the function $C(t)$ is known, the AUC can be computed from the integral of the concentration-time curve, i.e.,

$$AUC = \int_0^\infty C(t)\,\mathrm{d}\,t.$$

AUC gives an idea of the total exposure of a drug to an individual and relates the effect of a drug on the body to its concentration. Thus, an excessively large AUC can have undesired toxic effects while even a highly effective drug may fail to provide therapeutic relief if the AUC is below the threshold level.

Elimination Half Life ($t_{1/2}$) is the time taken for the drug concentration in the body to fall by one-half.

2.1.2 Additional Parameters

For patients who are more sensitive to the adverse effects of excessive levels of the treatment (for example neonates or pregnant women), care should be taken that they are not exposed to unnecessary levels of treatment. Nor should they be under-exposed as lack of sufficient therapeutic intervention could be ineffective. Thus, there is a certain level at which the concentration is desired to be maintained. Alternatively, there might be a therapeutic range - a lower and upper concentration levels between which the profile of a patient's concentration must be maintained. For this purpose we define the following parameters:

Target Concentration (C_{tgt}) is the concentration of the drug which is required to be maintained in the blood plasma such that concentration greater than C_{tgt} is unnecessary and unwanted and concentration lower than C_{tgt} is not desired for therapeutic reasons.

Lower limit of the therapeutic range (C_{tgt}^{-}) is the concentration level below which the therapy is rendered ineffective and it is desirable to maintain the concentration profile above this level.

Maximum target concentration (C_{tgt}^{+}) is the concentration level above which the therapy is potentially toxic and it is desirable to restrict the concentration profile below this level.

We will use these parameters in the later chapters.

2.1.3 Pharmacokinetic Models

Compartmental models are the most popular models in pharmacokinetics. In compartmental analysis the body is thought to be made up of several 'compartments'. These compartments have no physiological meaning; they just depict the different rates at which the drug is distributed in these compartments. For example, organs more perfused with blood like heart, liver and kidney could form one compartment and organs less perfused with blood like skin, muscle and peripheral organs could form the second compartment. The drug flow takes place between the various compartments and it is mathematically modelled.

We describe some common compartmental models below. All these models assume first order kinetics. Concentration is proportional to the dose administered and the rate of elimination of the drug is proportional to its concentration, c.f. Rowland and Tozer (1994).

These models are mechanistic, that is, the relationship between the variables in these models is specified in terms of the underlying biological processes. The parameters involved in such models have biological interpretation. Furthermore, we assume in this chapter that these models are deterministic, that is a given input will always produce the same output and there is no random variation acting on any of the variables. In Chapter 6, we will work with statistical models, which seek to best describe the relationship between the variables in the data.

One Compartment Models

In one compartment models, the body is seen as one system in which the drug distributes uniformly at the same rate. It is assumed that the concentration of the drug in the organism is the same everywhere in the body at any given time. However, the kinetics of absorption of the drug in the compartment can vary depending on how the drug is administered in the body.

One Compartment Model with Zero Order Absorption

In this model, the drug is assumed to get absorbed instantaneously in the compartment. For example, in case of an intravenous bolus injection, absorption time of the drug is negligible. The elimination of the drug from the body is modelled by a first order differential equation. Thus, if $X_1(t)$ is the amount of drug in the compartment at time t and K_e is the elimination rate constant then,

$$\frac{\mathrm{d}}{\mathrm{d}\,t}X_1(t) = -K_e X_1(t).$$

Figure 2.1 shows the scheme of this model.

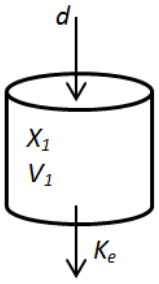


Figure 2.1: Scheme of a one compartment model with zero order absorption. d is the initial dose, K_e is the elimination rate constant and V_1 is the volume of the compartment.

If the initial dose is d, we assume that $X_1(0) = d$. Solving the above equation for this

initial condition we get

$$X_1(t) = de^{-K_e t}. \tag{2.1}$$

Dividing by V_1, the volume of the main compartment, we get the concentration $C(t)$ at time t as

$$C(t) = \frac{de^{-K_e t}}{V_1}. \tag{2.2}$$

Figure 2.2 shows the concentration-time course for this model for different values of K_e. It can be seen from the figure that a higher value of K_e causes quicker elimination of the drug from the body.

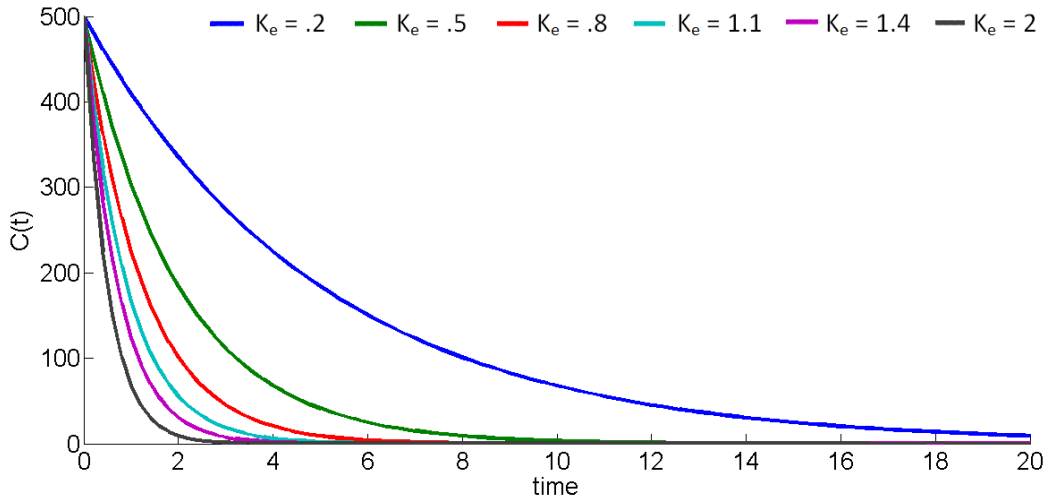


Figure 2.2: Plot of $C(t)$ (mg/L) vs t (h) for different values of K_e for a single dose of $d = 500$ mg. V_1 is taken to be 1 L. Units of K_e are h^{-1}.

One Compartment Model with First Order Absorption

First order absorption means that instead of the drug being absorbed into the compartment instantaneously, as was the case with the previous model, the rate at which the drug is absorbed is proportional to the amount not yet absorbed. Figure 2.3 shows the scheme of such a model.

Thus, if K_a is defined as the absorption rate constant and $X_2(t)$ is the amount not yet absorbed, the differential equations representing this case can be written as

$$\frac{\mathrm{d}}{\mathrm{d}\,t}X_1(t) = K_a X_2(t) - K_e X_1(t),$$

$$\frac{\mathrm{d}}{\mathrm{d}\,t}X_2(t) = -K_a X_2(t),$$

with initial conditions: $X_2(0) = d$ and $X_1(0) = 0$.

The solution of the second equation applied into the first equation gives a first

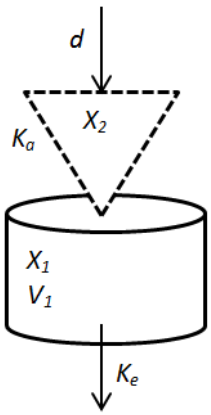


Figure 2.3: Scheme of a one compartment model with first order absorption. d is the initial dose, K_a is the absorption rate constant, K_e is the elimination rate constant and V_1 is the volume of the compartment.

order linear differential equation. The solution of this equation divided by the volume V_1 gives the concentration of the drug in the body at time t as

$$C(t) = \frac{X_1(t)}{V_1} = \frac{dK_a}{V_1(K_a - K_e)}(e^{-K_e t} - e^{-K_a t}). \tag{2.3}$$

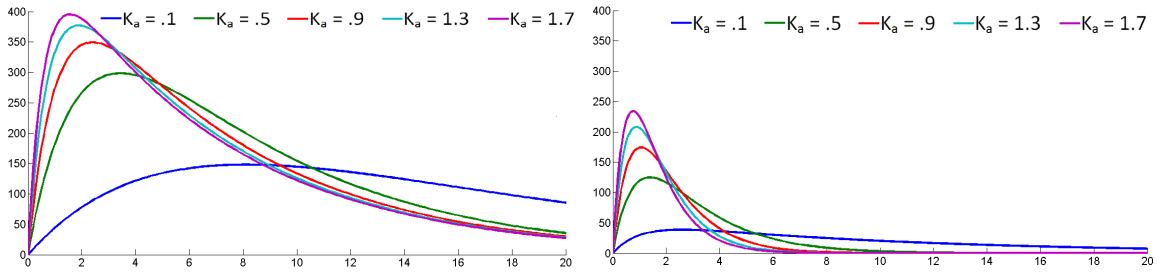
More often than not, the drug is not absorbed completely in the body. A small portion of the drug is egested without being metabolised. To account for this, the dose d is multiplied by F, the bioavailability, in the above equation to get:

$$C(t) = \frac{FdK_a}{V_1(K_a - K_e)}(e^{-K_e t} - e^{-K_a t}). \tag{2.4}$$

Note that for a bolus injection $F = 1$ and the absorption of the drug is complete and instantaneous. Therefore, letting $K_a \to \infty$ in the above equation, we get the equation for the previous model. Figure 2.4 shows the concentration-time course for the one compartment first order absorption model for different values of K_a and K_e. As can be seen in the figure, a higher K_a results in faster absorption of the drug. Consequently, for a fixed K_e, higher K_a results in higher peak concentrations. On the other hand, for a fixed K_a, higher K_e gives lower total exposure to the drug usually measured as the area under the PK curve.

It is pragmatic to say that a patient will be administered multiple doses of the same drug over a certain time period or for the rest of her life. If we assume that the same dose d is administered at each of the occasions and let the fixed time period between two successive doses be τ, then the concentration $C(t; n, \tau)$ at time t after the n^{th} dose is

$$C(t; \tau, n) = \frac{FdK_a}{V_1(K_a - K_e)}\left(\frac{1 - e^{-nK_e\tau}}{1 - e^{-K_e\tau}}e^{-K_e t} - \frac{1 - e^{-nK_a\tau}}{1 - e^{-K_a\tau}}e^{-K_a t}\right). \tag{2.5}$$



(a) $K_e = 0.2$/h (b) $K_e = 1.0$ /h

Figure 2.4: Plot of $C(t)$ vs t for a one compartment model with first order absorption for different values of K_a and K_e. The other parameters are: $F = 1$, $d = 500$ mg and $V_1 = 1$ L.

The derivation of this equation is given in Appendix A.

As the number of doses administered to a subject increases, the concentration settles to what is known as the *steady state concentration*. In steady state phase, concentration fluctuates between a constant maximal and minimal level. Letting $n \to \infty$ in Equation (2.5), we get the steady state concentration as:

$$C(t; \tau, \infty) = \frac{FdK_a}{V_1(K_a - K_e)} \left(\frac{e^{-K_e t}}{1 - e^{-K_e \tau}} - \frac{e^{-K_a t}}{1 - e^{-K_a \tau}} \right). \tag{2.6}$$

Figure 2.5 shows the case when $n = 6$ doses of a drug are administered at different values of τ. It can be seen that increasing the dose frequency results in higher peak concentrations of the drug. However, for a given n, longer dosing intervals mean that the drug remains in the system for a longer time but reaching a lower maximum concentration. It may be mentioned here that τ may not always be the same for all

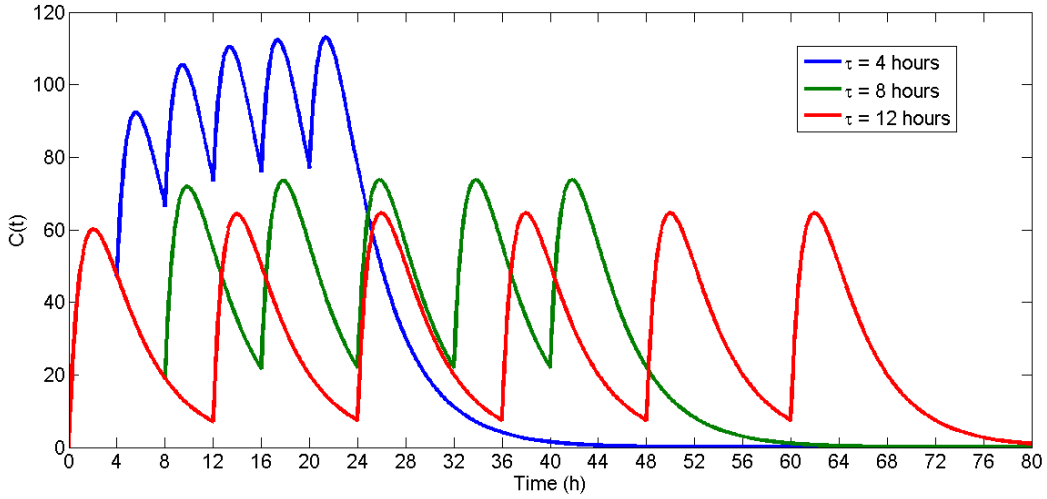


Figure 2.5: Plot of $C(t)$ (mg/L) vs t (h) for different values of τ for $n = 6$ doses of $d = 100$ mg each. The drug follows one compartment model with first order absorption and the PK parameters are: $K_a = .85$ /h, $K_e = .25$ /h, $F = 1$ and $V_1 = 1$ L.

the doses. In this thesis, we consider the general case when the doses are administered at the time points $\boldsymbol{t} = (t_1, ..., t_n)^T$ which are not necessarily equidistant.

Two Compartment Model with First Order Absorption

Next, we describe the two compartment model. As the name suggests, this model assumes that the body is made up of two compartments having different rates of drug distribution. The dose d is absorbed into the 'central compartment' at the rate of K_a. From the central compartment, the drug gets absorbed into the 'peripheral compartment' at the rate of K_{12} and it gets re-absorbed back at the rate of K_{21}.

The drug can be eliminated only through the central compartment which takes place at the rate of K_e. The volume of the blood plasma that is cleared off of the drug per unit time is called clearance. For elimination from the central compartment, the clearance is denoted by CL and for exchange of the drug between the two compartments, Q is the inter-compartmental clearance. The volume of the peripheral compartment is denoted by V_2. Figure 2.6 shows the scheme of this model.

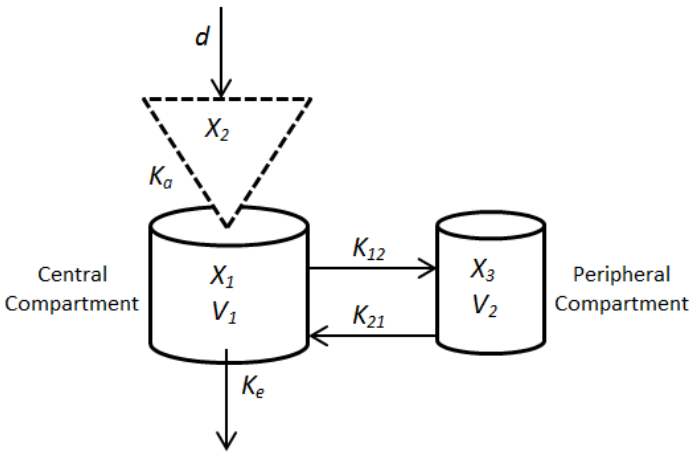


Figure 2.6: Scheme of a two compartment model with first order absorption. d is the initial dose, K_a is the absorption rate constant, K_e is the elimination rate constant, V_1 and V_2 are the volumes of the central and the peripheral compartment. The drug gets absorbed into the peripheral compartment from the central compartment at the rate of K_{12} and gets re-absorbed back at the rate of K_{21}.

Then, the concentration of the drug $C(t)$ at time t is given as

$$C(t) = Ae^{-\lambda t} + Be^{-\mu t} - (A + B)e^{-K_a t}, \tag{2.7}$$

where the coefficients A and B depend on the PK parameters as follows:

$$A = \frac{FdK_a(K_{21} - \lambda)}{V_1(K_a - \lambda)(\mu - \lambda)} \quad \text{and} \quad B = \frac{FdK_a(K_{21} - \mu)}{V_1(K_a - \mu)(\lambda - \mu)},$$

where $\lambda = \frac{1}{2}(S+R)$, $\mu = \frac{1}{2}(S-R)$, $S = K_{12}+K_{21}+K_e$, $R = \sqrt{S^2 - 4K_{21}K_e}$, $K_{12} = Q/V_1$, $K_{21} = Q/V_2$, and $K_e = CL/V_1$.

The derivation of Equation (2.7) is given in Appendix B.1. Figure 2.7 gives an idea of the concentration-time course of this model for different values of the inter-compartmental clearance Q. It can be seen from this figure that as Q increases, the concentration of the drug in the central compartment decreases. This is on account of the absorption of the drug in the peripheral compartment. In the two compart-

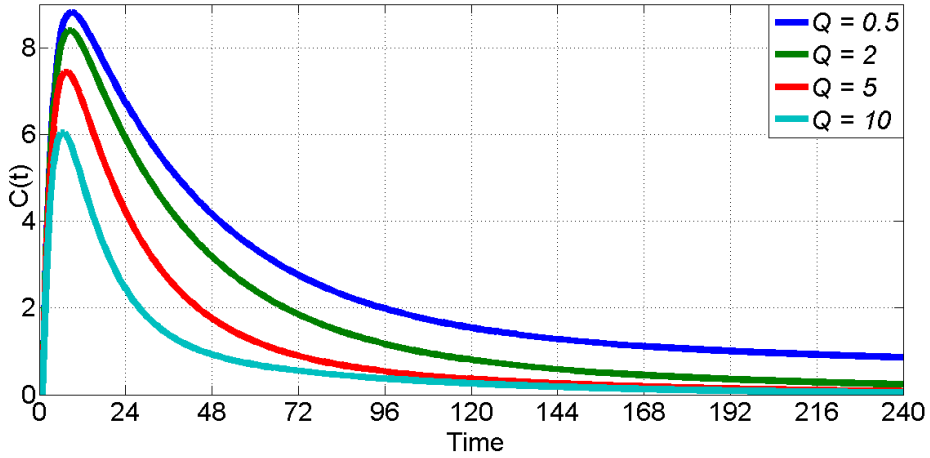


Figure 2.7: Plot of C(t) vs t for a two compartment model for different values of the inter-compartmental clearance Q. The dose size is $d = 500$. The values of the other PK parameters are: $CL = 10$ L/h, $V_1 = 100$ L, $K_a = 1.5$ /h and $V_2 = 50$ L and $F = 1$.

ment model, the decline in $C(t)$, the concentration in the central compartment, after reaching C_{max} is initially rapid but then it reduces to 0 gradually.

This biphasic behaviour reflects the physiological context on which the two compartment model has been defined and the two phases are called the α-phase and the β-phase. The decline of plasma concentration in the α-phase is mainly attributable to the distribution of the drug from the central compartment to the peripheral compartment. In β-phase, however, the decline in concentration in the central compartment is mainly because of metabolism and elimination of the drug from the body. This can be seen in Figure 2.7 as well.

For a multiple dose two compartment model with first order absorption, the principle of superposition can be applied to get the expression for the concentration at time t after the n^{th} dose as follows,

$$C(t;\tau,n) = A\frac{1-e^{-n\lambda\tau}}{1-e^{-\lambda\tau}}e^{-\lambda t} + B\frac{1-e^{-n\mu\tau}}{1-e^{-\mu\tau}}e^{-\mu t} - (A+B)\frac{1-e^{-nK_a\tau}}{1-e^{-K_a\tau}}e^{-K_a t}, \quad (2.8)$$

where A and B are same as before and τ is the fixed time interval between two successive doses. The derivation of Equation (2.8) is given in Appendix B.2.

The steady state concentration is:

$$C(t;\tau,\infty) = A\frac{e^{-\lambda t}}{1-e^{-\lambda\tau}} + B\frac{e^{-\mu t}}{1-e^{-\mu\tau}} - (A+B)\frac{e^{-K_a t}}{1-e^{-K_a\tau}}. \tag{2.9}$$

Three Compartment Models

The compartmental models described above are among the simplest and the most commonly used models. However, models exist which consider three or more compartments. For example, the concentration of remifentanil, an analgesic drug, is described by a three compartment model, Cascone et al. (2013). Since we do not use such models in this thesis, we do not go into further details. A comprehensive library of PK models is given in Bertrand and Mentré (2008).

Physiologically Based Pharmacokinetic Models

The compartmental models described above simplify drug disposition in an organism by describing the drug flow between 'compartments' which have no unique physiological meaning. Physiologically based pharmacokinetic models, on the other hand, attempt to describe the complex processes of absorption, distribution, metabolism and elimination of the drug using actual biological functions that regulate these processes in the body. Instead of having imaginary compartments based on how the drug distributes in them, the compartments in physiologically based PK models are specific body organs and tissues, such as heart, kidneys, liver, brain, skin, lungs. Thus, they are multi-compartment models with the compartments having clearly defined physiological and biological meaning. Figure 2.8 gives a graphical representation of a physiologically based PK model.

A set of differential equations makes up the structure of the model. The parameters represent physiological measures such as the blood flow rates and organ volumes. Information about these is usually available from the scientific literature. Integrating this system of differential equations gives the concentration of drug in each compartment as a function of time and dose, c.f., Bois et al. (2010).

Interestingly, the very first PK model described in the literature was a physiologically based PK model, Teorell (1937). However, lack of computing resources resulted in it being given up in favour of simpler models which had analytical solutions. With the advent of powerful computers, interest in these models has been renewed.

In this thesis, we are applying the simple (not physiologically based) compartmental models in all our methods. However, other ways of modelling the concentration

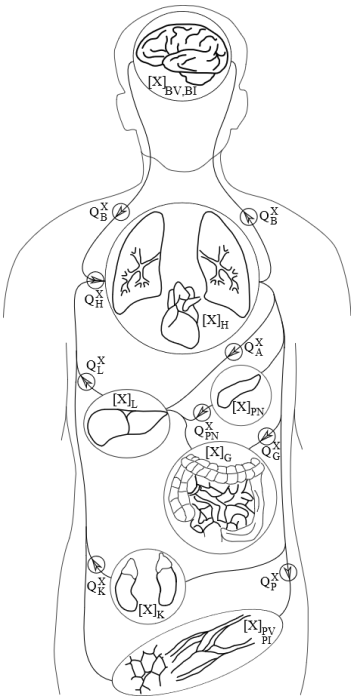


Figure 2.8: A graphical representation of a physiologically based PK model, Wikipedia (2011). The body is divided into 7 compartments - brain, lungs and heart, pancreas, liver, gut, kidney, and adipose/muscle tissue. Q denotes the blood flow between the compartments and X is the drug's concentration.

could also be implemented in our methodology. We plan to incorporate them in our algorithms in the near future.

Non-Compartmental Approach

As the name suggests, in non-compartmental approach (NCA), the PK parameters of a drug are determined without assuming a particular compartmental model. In this case, the PK parameters for a drug, such as AUC, C_{max}, t_{max}, CL, $t_{1/2}$ are indirectly determined by the data derived from PK sampling from a subject. PK sampling entails measuring the drug's concentration in the blood samples drawn at designated time points from the body. Using these samples, AUC can be estimated by, for example, the linear trapezoidal rule. Clearance is computed from the estimated AUC by the relation $CL = dose/AUC$. The accuracy of these techniques depends on the richness of the sampling points.

The main advantage of NCA over model-based inference is that development and validation of a true model is not required. Non-linear mixed effects models, in particular, are complicated and the methods for estimation of parameters have several computational drawbacks. Gabrielsson and Weiner (2012) discuss some methodologies used for NCA whereas Jaki and Wolfsegger (2012) present NCA methods that can be used in sparse sampling situations. PK is an R package that applies non-compartmental

theory for computation of PK parameters, Jaki and Wolfsegger (2011).

The main disadvantage of NCA is that it is based on approximations and also, it is unable to predict the concentration at any given time since no model is fitted to the data. In this thesis, we work with the compartmental models.

2.2 Pharmacodynamics

While pharmacokinetics describes the concentration-time relationship of a drug, pharmacodynamics studies the effect-concentration behaviour of a drug. Pharmacodynamic models relate the effect that a drug produces to the concentration of the drug in the body.

2.2.1 Sigmoid E_{max} Model

The most commonly used pharmacodynamic model is the sigmoid E_{max} model, also called the Hill equation, Hill (1910). Apart from pharmacology, the equation has wide applications in biochemistry.

The general sigmoid model relates the effect E to the concentration C as follows:

$$E(C) = \frac{E_{max} \times C^H}{EC_{50}^H + C^H}, \tag{2.10}$$

where H is the sigmoidicity coefficient, also called the Hill coefficient, E_{max} is the maximum effect, EC_{50} is the concentration required for 50% of E_{max} to take place.

For $H = 1$, the sigmoid E_{max} model is simply referred to as the E_{max} model.

Figure 2.9 plots the Hill equation for $E_{max} = 5$, $EC_{50} = 2$ and different values of H.

Properties

From Equation 2.10, we get

$$E(C) = \frac{E_{max}}{EC_{50}^H/C^H + 1}. \tag{2.11}$$

Therefore, as the concentration increases, the effect converges to E_{max}. The Hill coefficient, H, determines the steepness of the curve. Higher the value of H, the

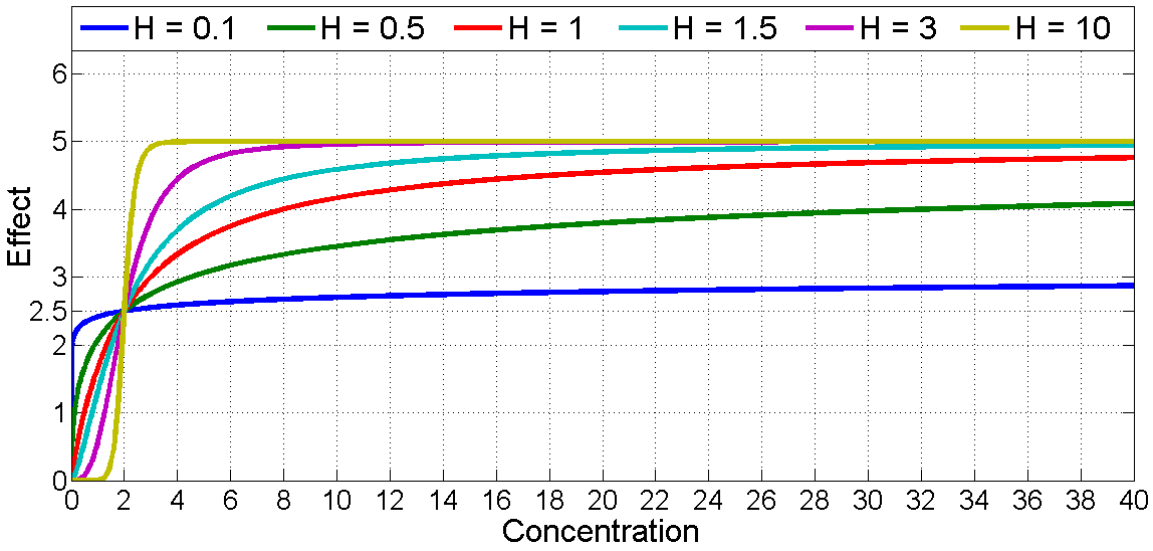


Figure 2.9: The Sigmoid E_{max} model for various values of the Hill coefficient H. The parameter values are: $E_{max} = 5$ and $EC_{50} = 2$.

faster is the convergence. This can also be observed in Figure 2.9.

Furthermore, the curve passes through (EC_{50}, $E_{max}/2$) irrespective of the value of H. This can be verified from the figure by observing that all curves pass through the point (2, 2.5). This also follows from the definition of EC_{50}.

If a baseline effect E_o is present, the intercept term can be added to the model. Then,

$$E(C) = E_o + \frac{E_{max} \times C^H}{EC_{50}^H + C^H}. \tag{2.12}$$

The drug's response can also be studied as a percentage of E_{max}, in the form of $E_p(C)$ as,

$$E_p(C) = \left(\frac{E(C)}{E_{max}}\right) \times 100 = \left(\frac{C^H}{EC_{50}^H + C^H}\right) \times 100. \tag{2.13}$$

The general sigmoid model is one of the most commonly used PD models and we use it in Chapter 7 to describe the effect of anti-microbials on extermination of parasites. Some other PD models are presented in, for example, Rowland and Tozer (2011).

Minimum Inhibitory Concentration

In this thesis, we will frequently use a PD term called the Minimum Inhibitory Concentration (MIC). The MIC is defined as the lowest concentration of an anti-infective or an antimicrobial required to inhibit the visible growth of a parasite after overnight incubation. Determination of the MIC is done mainly for two reasons. Firstly, it

is used in diagnostic laboratories to confirm resistance (if any) to the antimicrobial. Secondly, it is used as a research tool to determine the activity of a new antimicrobial, Andrews (2001).

2.3 The Population Approach

The ADME process described in the beginning of this chapter is not the same in every patient. A drug will have different PK and PD profiles in any two patients. This variation can be ascribed to mainly two sources: the inter-individual variability and the intra-individual variability.

The compartmental models discussed in this chapter are deterministic in the PK parameters. In patient populations, the form of the concentration-time function is the same for everyone because of the same underlying mechanistic model. However, the parameters of the model differ from subject to subject giving rise to different concentration profiles. The difference in the parameters is attributed to the inter-individual variability. The non-linear mixed effects models are often used to model the inter-individual variability through an assumption of the parameters being random variables.

The second source of variation in the measured drug's concentration is due to some unobserved individual variation. We assume here that this variation is due to observational errors. Various models have been proposed to incorporate this variability. They are discussed in Chapter 5.

To assess the population variability and estimate the parameters of the mechanistic model, the so called *population approach* is often used. The population approach is defined by FDA (1999) as:

> "Population pharmacokinetics is the study of the sources and correlates of variability in drug concentrations among individuals who are the target patient population receiving clinical relevant doses of a drug of interest."

In population analysis, instead of fitting a individual profile to every subject, which will call for a dense sampling scheme, parameters are assumed to be realizations from probabilistic distributions. This enables the investigator to assess the mean profile of the drug's concentration and its variability in the population of patients taking a minimal number of samples from a subject, which is preferable for economical and ethical reasons.

L.B. Sheiner was one of the earliest advocates of the population approach, Sheiner

and Beal (1980) and Sheiner and Beal (1981). He is also one of the developers of the software NONMEM® which is the industry standard software for analysing population PK data, ICON plc (2015). Regulatory agencies also endorse the population approach, especially for drugs for children, EMEA-CHMP (2006b).

The basic tenets of population pharmacokinetics are as follows: collect a certain number of blood samples at designated time points from everyone in the cohort. Based on the measured drug concentration in these samples, estimate the means of the parameters of the underlying mechanistic model and their variances. The average PK parameters give the concentration profile of a typical subject in the population.

The precision of the estimated parameters depends on the time points at which the blood samples are collected. This necessitates the use of theory of optimal design of experiments. The population models (non-linear mixed effects models) are presented in Chapter 5, along with the methods of estimation and computation of an optimal design.

Patient Covariates

A part of the inter-individual variability can be explained on the basis of the covariates, such as age, gender, body weight, comorbidities, which define the physiological profile of a patient. The covariates affect every phase of the ADME process and thus the dose needs to be carefully determined so that the effect of the covariates is properly incorporated into the PK model. Inclusion of covariates has the twin benefits of increasing the predictive power of the model and opening up the possibility of dose individualisation based on the patients' specific covariates. This will be discussed in the next chapter.

Figure 2.10 gives an idea of how some covariates affect the organ functions. For instance, intrinsic clearance of an individual is a function of age, body surface area (a function of height and weight) and genetic profile.

The inclusion of covariates in the PK model introduces new parameters, which need to be efficiently estimated along with the PK parameters. This may make the process of parameters estimation complicated. Therefore, only the relevant covariates which are found to be significantly correlated with a PK parameter of the model should be selected into the model. We discuss this later in the thesis.

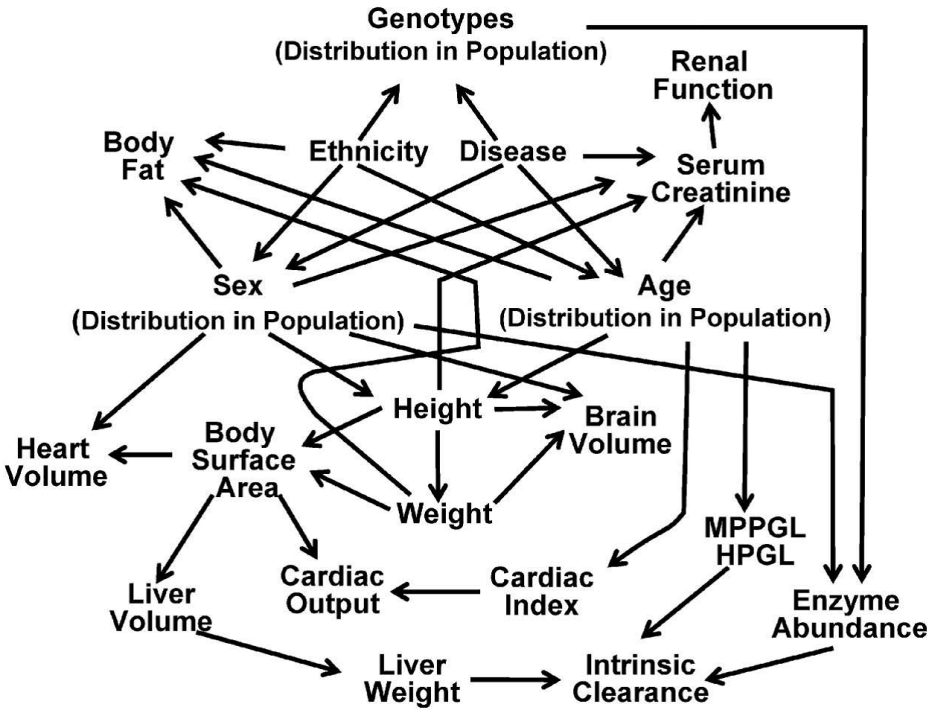


Figure 2.10: An illustration from Jamei et al. (2009) which shows the effect of different covariates on the drug's absorption, distribution, metabolism, and elimination (ADME).

The Role of a Dose Regimen

In the compartmental models discussed in this chapter, the concentration of the drug in the body is directly proportional to the administered dose, d. In a multiple dose model, the time points at which the doses are administered, $t_1, ..., t_n$, also affect the resultant concentration. The exposure to the drug also depends on n, the number of doses which are administered. In case of a fixed dose combination unit, where two drugs are constrained to be administered in the same ratio and at the same dosing time points, the combination ratio is an important determinant of individual exposure to the two partner drugs.

As discussed, there are generally no beneficial effects accrued from a drug beyond a certain level of exposure. However, excessive amount of a drug in the body can increase the risk of toxicity. Therefore, to get the best outcomes from a therapy, that is, to get the maximum possible efficacy and the least toxicity, the choice of the dose regimen is an important consideration. In the next chapter, we elaborate on this and propose a mathematical formulation for a dose regimen which enables us to define the associated optimisation problem. The optimisation problems that we solve in this thesis assume the knowledge of the PK model associated with the drug. The models discussed in this chapter are helpful in illustrating the optimisation algorithm that we introduce in Chapter 4.

Chapter 3

Optimisation of Dose Regimens: The Problem

In this chapter, we highlight the importance of optimisation of dose regimens. We discuss the methods currently used for dose regimen optimisation, along with the methods of extrapolating the optimum doses from those used for other populations, e.g., from adult's doses to children's doses. Dose regimen optimisation is an important problem, especially for certain therapies such as antimicrobials, analgesics, anaesthetics and anticancer drugs. Such optimisation problems are usually attempted for individual drugs by the means of simulation studies. There is a need for a more general approach to the problem, independent of any particular drug or therapy. In this chapter, we propose a mathematical formulation of dose regimen and use it to explicitly express the problem of dose regimen optimisation. In the next chapter, we introduce an algorithm to solve some of the so-defined problems. We firstly discuss the need for optimised dose regimens.

3.1 Why Should a Dose Regimen be Optimised?

As discussed in the previous chapter, the therapeutic effect of a drug is a function of its concentration in the body. However, increase in the concentration beyond a limit increases the risk of toxicity. Therefore, a good dose regimen should be able to keep the concentration within the therapeutic range. This calls for optimisation of the dose regimen which, for most drugs, is generally done for the typical individual in the population. For certain drugs, the optimisation has to be done for each patient, a process called dose individualisation. We discuss both these situations in this chapter.

3.1.1 Toxicity Reduction and Enhancing Efficacy

Following the correct selection of a therapy based on the indication, an optimal dose regimen is the most important determinant of therapeutic success of the therapy. Since health authorities may lack sufficient pharmacometric capacity to check the dose information provided by the pharmaceutical industry and most physicians rely on the prescription information for choosing dose regimens, the ethical responsibility for supplying an optimal dose regimen remains largely with pharmaceutical companies. In the past, the pharmaceutical industry was predominantly interested in achieving a new drug approval based on differentiation of the drug against placebo or an active comparator, neglecting the optimisation of dose regimens. The maximal safe dose was often chosen for confirmatory phase III trials and having been initially approved, administered to the general patient population.

Nowadays, health authorities actively challenge the dose regimen suggested by the industry, usually by stating that the doses are unnecessarily high and leave the burden of proof with the company applying for approval or marketing, DiNicolantonio and Serebruany (2013) and Sacks et al. (2014). Provision of the minimal clinical effective dose, the maximum safe dose and the optimal dose by indication not only improves the chances of a successful approval, but can be considered state of the art, Martinez et al. (2012).

There are several disease areas (e.g., oncology, infectious diseases, advanced analgesic therapies) where the optimal dose regimen consists of co-administration of several drugs with different pharmacokinetic and pharmacodynamic properties ideally at the same time. For some indications such as HIV, malaria and tuberculosis, combination therapies are mandatory to prevent the spread and development of resistance to the single components of the regimen. The most stringent way to enforce co-administration is co-formulation, i.e., several drugs are contained in a single dosing unit. In this case, the physician can only choose to administer fixed multiples of two (or more) doses, the dose ratio having been selected by the manufacturer of the drugs. Therefore, it is crucial that the optimal dose ratio has been identified during the development process.

All dose regimens should be optimised for the best possible treatment outcome, but there are certain drugs for which the optimisation could make a difference between therapeutic success or failure. Antibiotics are the best examples of this. Based on the main pharmacodynamic parameters of these drugs, which are strongly correlated with clinical outcome and are used as predictors of clinical efficacy, they can be divided into the following types, Wise (2003):

- Type 1. The treatment goal is to maximise the drug's concentration relative to the minimum inhibitory concentration (MIC). The PD parameter to maximise is the ratio C_{max}/MIC. For example, aminoglycosides which are used to treat serious life-threatening infections such as that of the abdomen and urinary tract.

- Type 2. The treatment goal is to maximise the duration of exposure to the drug. The PD parameter to maximise is the time for which the blood concentration exceeds the MIC. For example, penicillin, one of the earliest discovered antibiotics. It is a broad spectrum antibiotic for treatment of infections, for example, those caused by the bacteria staphylococci and streptococci.

- Type 3. The treatment goal is to maximise the exposure to the drug. The PD parameter to maximise is the ratio AUC/MIC. For example, azithromycin is an antibiotic which has a long half life and is used in combination therapies for the treatment of infections like respiratory and gastrointestinal.

Therefore, for these antibiotics to be able to eliminate the infection effectively, the relevant criterion for that type of antibiotic must be met as closely as possible. For example, for the Type 2 antibiotics, the dose regimen should be such that the drug's concentration remains above the MIC for the duration of the treatment. This can be accomplished by either giving 'large' doses at regular intervals of time or making the dosing 'more frequent'. For accurate values of the optimal doses or the dosing time points, a quantitative approach to the optimisation problem becomes a necessity.

Optimisation of dose regimens is important in oncology as well. For example, an important finding that has been made from breast cancer trials is that increasing the dose beyond a certain limit does not accrue any additional benefits in reducing the tumour burden. However, more frequent dosing, called the 'dose-dense' regimen, has been found to be more potent in killing cancer cells, Citron et al. (2003). This is because of the fact that growth of most cancer cells follows non-exponential Gompertzian kinetics which means that the regrowth of cancer cells between the drug cycles is rapid. This necessitates optimisation of the dose regimen for such drugs.

Thus, to balance the beneficial and the harmful effects of a drug, the design of its dose regimen is an important consideration. Aronson (2005) gives some general guidance for adjusting dose regimens to balance the beneficial and harmful effects of the drug.

3.1.2 Optimisation for Dose Individualisation

Dose personalisation or individualisation is adaptation of a therapy to the physiological profile of a patient. Individualisation of a therapy to a patient is not just choosing

the appropriate dose levels or dose schedules, but also examining the patient's profile to check if the particular therapy will benefit the patient at all. This is explained below.

Therapeutic Drug Monitoring

Therapeutic Drug Monitoring (TDM) is only used for certain drugs which exhibit the following characteristics: a narrow therapeutic range, large PK variability, existence of a direct relationship between the serum concentration and the toxic or beneficial effects of the drug, Kang and Lee (2009) and Aarnoutse et al. (2003). For example, TDM is a standard of care for certain antimicrobial agents such as gentamicin and vancomycin and it appears that TDM of β-lactam antibiotics may be clinically beneficial for critically ill patients, Felton et al. (2014). TDM, as the name suggests, consists in measuring the drug concentrations in an individual at regular intervals of time. Depending on the serum concentrations, the dose is adjusted to maintain it within the target range. A person's PK profile may keep changing over time owing to change in physical attributes like age, body weight and emergence of comorbidites. Also non-compliance with the advised dose regimen may result in departure from the target.

TDM strives to take care of all these factors by periodic adjustment of the dose regimen. It entails administering an initial dose regimen to the patient based on the covariates, comorbidites and any concomitant drug therapies. Also taken into account are the past PK/PD data available for the drug, demographic data and physiological profile of the patient. This is referred to as a priori TDM. Blood samples are then collected at designated intervals and based on the results, the dose regimen is adjusted so that the blood concentration remains within the desired range, Burton (2006).

The dose adjustment for TDM requires clinical and computational resources. A variety of software are available for adjustment of the dose regimen based on the collected blood samples, of which MwPharm®, MM-USC*PACK® and TCIworks are commonly used. A comparison of 12 such software including the ones mentioned above can be found in Fuchs et al. (2013).

Below we discuss, as an example, the software MwPharm®, Proost and Meijer (1992). It contains a database of PK information on 180 drugs. The clinician needs to select from this database the drug whose doses are to be individualised to a patient, and needs to enter the patient's covariates like age, weight, gender. Based on the past PK information about the drug already stored in the software, the desired target range and the patient covariates, the software determines the best loading dose to be administered to the patient. After some time, a PK sample is drawn from the

patient and the measured concentration in entered into the software. Using Bayesian techniques, the PK parameters' estimates of the drug are revised on the basis of this observation, Thomson and Whiting (1992). In other words, the revised estimates are the PK parameters' estimates of the patient for whom the software computes the next dose to be administered. Since TDM requires careful monitoring of the drug's concentration in the patient, regular samples are drawn at designated intervals of time. Every time a sample is drawn, the next dose will be computed by the software by taking into account the concentration values obtained from all the previous samples. This process will continue as long as necessary to achieve the desired outcome.

Therapeutic Concentration Intervention

Therapeutic Concentration Intervention (TCI), another dose individualisation technique, has been suggested as an alternative conceptual strategy to TDM, Holford (1999). TCI is a variation of TDM in the sense that instead of optimising the dose regimen for a range of dose levels, a target concentration is sought to be achieved. This target concentration is computed on the basis of a target effect desired to be achieved. The two main arguments given by the author against TDM are the following. Firstly, TDM aims to maintain the concentration within a therapeutic range and thus gives rise to a range of possible doses from which one dose needs to be chosen by the clinician. TDM gives no guidance to make such a choice. Secondly, TDM assumes that all concentrations within a range are equally desirable. Intuitively, this appears to be a sub-optimal strategy as within the therapeutic range as well, a patient may have varying response to the therapy. Based on these two advantages and the fact that TCI uses PD models to determine the target concentration, the author advocates this technique over TDM.

For example, if estimates of EC_{50}, E_{max} and H are available, the target concentration can be calculated from the general sigmoid model described in Equation 2.10 as,

$$C_{tgt} = \left[\frac{\widehat{EC}_{50}^{\widehat{H}} \times E_{tgt}}{\widehat{E}_{max} - E_{tgt}} \right]^{\frac{1}{\widehat{H}}},$$

where E_{tgt} denotes the desired target effect.

Once the target concentration has been determined, the rest of the steps are the same as TDM. Software used for maintaining a target range by TDM can also be used to maintain a target concentration by TCI, as the basic steps of computing patients' PK parameters using sparse data remain the same.

Pharmacogenetic Studies

Pharmacogentic studies are used to predict the response of an individual to a particular therapy by studying his/her genetic profile. A well known example is the drug trastuzumab which is used to treat breast cancer in patients whose cancer cells have over-expression of the HER2/neu gene, Vogel et al. (2002). Patients whose cancer cells lack this over-expression derive no benefit from this drug and thus should not be administered it. Other examples include dasatinib (used for treating myelogenous leukemia, Philadelphia chromosome-positive acute lymphoblastic leukemia and advanced prostrate cancer) and detuximab (used for treating metastatic colorectal cancer and metastatic non-small cell lung cancer with epidermal growth factor receptor (EGFR)-expressing cells).

Genetic variability could also result in some patients showing heightened adverse reactions to a drug which are not seen in the general population, Roden and George Jr (2002), Ma and Lu (2011).

However, pharmacogenetic studies are currently available for a limited number of drugs and more research is needed before using them as an alternative or even complementary to dose regimen individualisation methods, Gervasini et al. (2010). For our work in this thesis, we do not make use of such studies.

3.1.3 Optimisation for Randomized Concentration - Controlled Trials

The Randomized Dose-Controlled Trials are one of the most commonly used clinical designs. They involve random assignment of the subjects to predetermined dose levels of the candidate drug.

In contrast, the Randomized Concentration-Controlled Trials (RCCTs) involve random assignment of the subjects to predetermined levels of drug concentrations or their ranges. It is important to understand that RCCT is a clinical trial design and unlike TDM and TCI, no dose adjustments take place for a subject.

This scheme accounts for, to a large extent, the variability between the subjects due to different rates of absorption, elimination and distribution. Some studies have shown that RCCTs on an average require smaller sample sizes as compared to RDCTs, and therefore could be more powerful, Reeve (1996) and Sanathanan and Peck (1991). The main drawback of RCCT is that the protocol becomes more tedious and the adherence to a specified concentration or its range may be difficult to achieve in practice.

To bring every subject to the specified target concentration, an adaptive procedure is followed in which a covariate such as the subject's body weight and subsequent doses are adjusted depending on the results obtained by repeated blood sampling. Optimisation of the dose regimen is, therefore, an essential requirement for conducting RCCTs.

3.1.4 Optimisation by Dose Scaling

Clinical trials for infants and children present ethical and legal challenges which often result in sparse PK/PD data for this population. In the absence of proper PK/PD studies in children, extrapolation from adult data to determine the appropriate doses for children is a common practice. If the adult dose regimen happens to be optimised for given set of values of adult covariates, then a sensible method of scaling should give optimised dose regimen for children's covariates as well.

Size is the primary covariate used to scale adult's doses for children, although age based models also exist. We discuss here briefly the three size models as given in Johnson (2008) and Anderson and Meakin (2002). Age based mechanistic models can be found in Anderson and Holford (2008).

Body Weight Model

The body weight (BW) dose scaling model to determine the appropriate dose for a paediatric patient is given as

$$Dose_P = Dose_A \left(\frac{BW_P}{BW_A} \right),$$

where $Dose_P$ and $Dose_A$ are the paediatric and adult doses respectively. BW_P and BW_A are the paediatric and adult values of BW respectively. Usually, BW_A is taken as 70 Kg as body weight of a typical adult.

The fundamental assumption of this model is the linear relationship between the body weight and the dose, which in general, is not found to be true. It has been observed that children require a larger dose when expressed in mg per kg of weight than adults. Despite of these limitations, this model is commonly used for dose scaling because of its simplicity.

Body Surface Area Model

The basic assumption of this method is that children and adults are geometrically similar. Since it is practically difficult to measure the body surface area (BSA) directly, the Du Bois and Du Bois height-weight formula by D. Du Bois and E. F. Du Bois (1916), one of the earliest formulas to estimate the BSA, is commonly used to estimate it as shown below:

$$BSA = Weight^{0.425} \times Height^{.725} \times 0.007184,$$

where weight is given in kilograms and height in centimetres.

The estimated surface area can be used for dose scaling as follows:

$$Dose_P = Dose_A \left(\frac{BSA_P}{BSA_A} \right),$$

where BSA_P and BSA_A are the paediatric and adult values of BSA respectively. Usually BSA_A is taken as 1.73 m^2 as a typical adult's BSA.

The body surface area model is generally more accurate than the body weight model as development of physiological systems is represented more by the changing body surface area than the change in body weight, Johnson (2008). However it is not very useful for dose scaling for infants as their bodies are morphologically different having shorter legs and relatively larger heads and trunks, Anderson and Meakin (2002).

The Allometric 3/4 Power Model

Galileo in 1637 discussed the relationship of skeletal size to the body weight. It has since been scientifically established that the plot of the log of the basal metabolic rate (the amount of energy required by the body to keep functioning at rest) in almost all species including humans, produces a straight line with a slope of 3/4. This gives the model its name, although, it is used for scaling of many physiological parameters and the power may not necessarily be 3/4 every time. For instance, time related parameters like heart rate, respiratory rate are scaled using the body weight with the power 1/4.

For the i^{th} individual with body weight BW_i the clearance parameter is scaled as,

$$CL_i = CL_{std} \left(\frac{BW_i}{BW_{std}} \right)^{3/4},$$

where CL_{std} is the clearance parameter of a standardized individual with weight BW_{std}.

The model for physiological volume, for body weight BW_i has power parameter 1,

$$V_i = V_{std}\left(\frac{BW_i}{BW_{std}}\right),$$

where V_{std} is the volume parameter of a standardized individual with weight BW_{std}.

For scaling of the loading dose, which is a function of the volume V, the model is:

$$Dose_i = Dose_{stdL}\left(\frac{BW_i}{BW_{std}}\right),$$

where $Dose_{stdL}$ is the loading dose for a standardized individual with weight BW_{std}.

Maintenance dose, being a function of the clearance CL, can accordingly be scaled as:

$$Dose_i = Dose_{stdM}\left(\frac{BW_i}{BW_{std}}\right)^{3/4},$$

where $Dose_{stdM}$ is the maintenance dose for a standardized individual with weight BW_{std}.

Traditionally, doses suitable for adults have been scaled for children using simple rules based mainly on the age, weight or surface area of the child. However, scaling of adult data to children's is not as simple as it appears. Infants and young children have immature organ systems and the resultant PK profile may be erratic and unpredictable. As noted by Halpern (1988), "Paediatrics does not deal with miniature men and women, with reduced doses and the same class of disease in smaller bodies, but ... has its own independent range and horizon."

Kearns et al. (2003) emphasize that human growth is not a linear process and age associated changes in body composition and organ functions are dissonant during the first decade of life. The situation for neonates is the most complicated. Their body composition and organ functions change rapidly during the first few weeks. Extrapolation from adult or even children's data to determine doses for neonates can be very misleading. Over-dosing may have an adverse effect on a neonate whose body may lack the mechanism to tackle drug induced toxicity and under-dosing may all together defeat the purpose of therapeutic intervention. Detailed regulatory guidance on the impact of kidney, liver and cardiovascular immaturity of neonates on dose selection is available, EMEA (2004), EMEA-CHMP (2005) and EMEA-CHMP (2006a).

Johnson (2008) found that the precision and bias of these methods in determining the appropriate dose depend on the age group of the paediatric patients. The author further highlights the need of having more robust methods of dose scaling which are set up on physiologically based pharmacokinetic models (PBPK) which were discussed in Chapter 2. For example, PBPK models were formulated to study the PK profiles of

theophylline and caffeine in adults and neonates, Ginsberg et al. (2004). As compared to adults, faster metabolism and clearance of theophylline relative to caffeine were observed in neonates.

It may be mentioned here that dose scaling is not limited to children applications only. The standard way to prescribe a dose is usually per kilogram of the body weight. Traditionally, the dosage of chemotherapy drugs has been determined according to the BSA of the patient, estimated using the height and the weight in the Du Bois and Du Bois formula. However, there have been concerns about whether the BSA method increases the risk of under-dosing (over-dosing is more readily recognised). Although a better method than the BSA approach is not yet available, research has urged clinicians not to place full reliance on the BSA but to simultaneously look at other factors like possible interactions with other drugs and to assess the drug elimination profile of the patient, Gurney (2002).

Thus, for obtaining the best outcome it is necessary for the therapy's exposure to be of just the right magnitude. As discussed in the previous chapter, the exposure is determined by not just the dose sizes, but also by the number and frequency of the administered doses. In case of a fixed dose combination unit, the two partner drugs are constrained to be administered at the same dosing time points and in the same ratio. Variables such as dose sizes, number of doses, frequency of dose administration and the combination ratio (if applicable) are design variables which form the dose regimen of a drug and the choice of values of these variables eventually influences the therapeutic outcome from the treatment. Indeed, optimisation of the dose regimen amounts to determining the optimum values of the design variables like these. To facilitate explicit optimisation of dose regimens, it is imperative that the concept of a dose regimen is explained in mathematical terms. In the next section, we propose such a form which will be useful for our work in this thesis.

3.2 Dose Regimen - Mathematical Formulation

A dose regimen $\mathcal{R}$ of dimension n is defined as the following collection of design variables,

$$\mathcal{R} = \{n, \boldsymbol{D}, \boldsymbol{t}, T, \theta\},$$

where

n is the number of doses to be administered;

$\boldsymbol{D} = (d_1, d_2, ..., d_n)^T$ is a n-dimensional column vector of the doses. $\boldsymbol{D} \in \mathcal{D} \subset \mathbb{R}_+^n$, where $\mathcal{D} = [r_1, s_1] \times [r_2, s_2] \times ... \times [r_n, s_n]$, for some positive constants r_i and s_i such

that $r_i < s_i$, $i = 1, ..., n$. $\boldsymbol{D}$ will be referred to as the dose vector in this thesis;

$\boldsymbol{t} = (t_1, t_2, ..., t_n)^T$ is a n-dimensional column vector of the dosing time points; we take $t_1 = 0$ and $t_i \in \mathbb{R}_{>0}$ for $i = 2, ..., n$;

T is the duration of the treatment; $T \in \mathbb{R}_{>0}$;

θ is the ratio of drug A to drug B in the fixed dose combination unit, $\theta \in [e, f] \subset \mathbb{R}_{\geq 0}$ for some positive constants e and f. In case of a single drug, we set $\theta = 0$.

$n, \boldsymbol{D}, \boldsymbol{t}$, T and θ are referred to as the *components* of the dose regimen. In case of a combination therapy, $\boldsymbol{D}$ refers to the dose regimen of drug A. The dose regimen for drug B will be given by $\theta \times \boldsymbol{D}$.

Many drug therapies, e.g., for diabetes and hypertension, are taken for a long time over regular intervals, possibly for the rest of the life. For such drugs n is not specified and the time of the treatment is assumed to be $T = \infty$. Then, the dose regimen can be expressed as $\mathcal{R}_\infty = \{\boldsymbol{D}, \boldsymbol{t}, \infty, \theta\}$.

For such regimens, $\boldsymbol{D} = (d_1, d_2, ...)^T$ and $\boldsymbol{t} = (t_1, t_2, ...)^T$. All the doses of the dose vector $\boldsymbol{D}$ can be the same for drugs used to treat chronic diseases like diabetes and hypertension, while if the drug is being administered under the TDM set-up, in which the doses are regularly adapted, the individual doses could be different.

3.3 The Optimisation Problem

By optimisation of a dose regimen, we mean determination of $\mathcal{R}$ which results in maximisation or minimisation of a function $\vartheta(\mathcal{R})$, called the objective function. As discussed before, a therapy can have beneficial as well as adverse effects. We formulate $\vartheta(\mathcal{R})$ as a function of the components of the dose regimen to quantitatively measure the beneficial and the adverse effects of the therapy. In most cases, as the components $(n, \boldsymbol{D}, T)$ of the dose regimen increase, the beneficial effects of the therapy will increase only upto a point. Beyond that, the risk of toxicity will increase without any increase in the beneficial effects. It will be ideal to determine that point at which the benefits are maximum and the risk of toxicity minimal.

Owing to manufacturing and other practical reasons, the components of the dose regimen will have certain constraints imposed on them. These constraints must be taken into account when optimising the dose regimen. In this thesis, we focus on optimisation of $\vartheta(\mathcal{R})$ with respect to one or two components at a time, keeping the others fixed. The main reasons for this are clinical practicalities and computational limitations. We consider two types of problems:

Type 1. Optimisation with respect to $\boldsymbol{D}$ and θ.

Here, for given values of n, $\boldsymbol{t}$ and T, we optimise the dose vector $\boldsymbol{D}$ for a single drug and if the therapy is a combination of two drugs, we optimise $\boldsymbol{D}$ and θ. The objective function is expressed as a function of $\mathcal{R}$ given $n, \boldsymbol{t}$ and T. The optimisation problem can then be stated as:

$$\text{minimise } \vartheta(\mathcal{R}|n, \boldsymbol{t}, T),$$

that is, choose $\boldsymbol{D} \in \mathcal{D}$ and $\theta \in [e, f]$ to minimise function ϑ for given values of n, $\boldsymbol{t}$ and T.

Type 2. Optimisation with respect to $\boldsymbol{D}$, T and n.

The second type of problem we attempt is the optimisation of $\vartheta(\mathcal{R})$ for given values of $\boldsymbol{t}$ and θ. This is done in Chapter 7 where we find the best dose regimen for a target reduction in the viral load. The optimisation problem in this case is:

$$\text{minimise } \vartheta(\mathcal{R}|\boldsymbol{t}, \theta),$$

that is, choose $\boldsymbol{D} \in \mathcal{D}$, n and T to minimise function ϑ for given values of $\boldsymbol{t}$ and θ.

Feasible Region for the Dose Vector $\boldsymbol{D}$

For the vector of dose sizes $\boldsymbol{D}$, we consider two cases. The first is the case of doses whose values may be any real numbers in the admissible range $r_i \leq d_i \leq s_i$, $i = 1, ..., n$. That is, the dose vector is an element of a hyper-rectangle $\mathcal{D}$. In the next chapter we assume that $r_i = 0$, $i = 1, ..., n$ and $s_i = d_{max}$ for all i, where d_{max} is the maximum dose which can be administered at a given time.

Then, the solution space for $\mathcal{D}$ is the hypercube $[0, d_{max}]^n$. The second case is of discretised doses, where it is assumed that the doses can be administered in multiples of a constant, κ. The solution space in this case is a discretised hypercube $\{0, \kappa, 2\kappa, ..., d_{max}\}^n$, assuming that d_{max} is a multiple of κ. Such discretised doses are useful to consider when the drug is available only in fixed single units.

A Note on the Objective Function ϑ

In the above optimisation problems, the objective function ϑ can be formulated according to the desired clinical criteria. In this thesis, we consider criteria which seek

to minimise over- and under-exposure to the drug. Therefore, in the problems that we attempt, ϑ measures the magnitude of deviation from the desired level of exposure. For example, for maintaining the drug's concentration around a target concentration of C_{tgt}, the objective function ϑ must be able to measure the deviations of the expected concentration profile from this target.

Function ϑ will depend on the mechanistic model, say $C(\boldsymbol{D}, \boldsymbol{t})$, which describes the concentration-time relationship of the drug. Generally, ϑ will be a complicated function of the components of $\mathcal{R}$ and the parameters of the mechanistic model of the drug and therefore, optimisation of the objective function will not be a straightforward problem. This motivated us to develop an algorithm to find the optimal dose regimen $\mathcal{R}$.

Assuming that the form of the mechanistic model C is known, we introduce the notion of *efficient dose regimens* in the next chapter, which defines the goodness of a dose regimen by its ability to keep the serum concentrations close to the target concentration or within a therapeutic range. However, the concept of efficiency extends to all possible cases where a dose regimen is being optimised by minimisation or maximisation of ϑ. In the same chapter, we present the *Efficient Dosing algorithm* developed by us, to compute the efficient dose regimens for a given target concentration or range. We consider several constraints like discrete dose levels, skipped doses and propose solutions for them. We also extend the algorithm for the case of combination therapies where we present a method of computing the optimal ratio in which two drugs should be combined and the efficient dose regimen for the combination unit.

We later discuss in Chapter 6 the case when the parameters of the assumed PK model are unknown and simultaneous estimation of the parameters and optimisation of the dose regimen are to be done. For this, the theory of non-linear mixed effects models would be indispensable, which has been briefly explained in Chapter 5.

Chapter 4

The Efficient Dosing Algorithm for the Case of Known Model Parameters

In this chapter, we introduce the Efficient Dosing (ED) algorithm to determine the optimal dose regimen for the two types of optimisation problems introduced in Section 3.3. We handle the Type 1 problem in Chapters 4 and 6 while the Type 2 problem is discussed in Chapter 7.

The Type 1 problem pertains to optimisation of $\vartheta(\mathcal{R})$ with respect to $\boldsymbol{D}$ in case of a single drug and with respect to $\boldsymbol{D}$ and θ for a combination therapy, for given values of n, $\boldsymbol{t}$, and T. We assume that the mechanistic model which describes the concentration-time relationship for the drug(s) is known. In this chapter, we make a further assumption that the parameters of the model are known as well. In this chapter, we work with mechanistic models which are assumed to be deterministic.

4.1 Efficient Dose Regimens

We denote by $C(t, d)$ the concentration of the drug at time t after a single dose d is administered. An example of the dependence on time t and on a single dose d in a one-compartment parametric model of a drug concentration is shown in Figure 4.1. The five different dose levels clearly impact the concentration which after achieving a maximum value, decreases to zero in its elimination phase. The model parameters are the same for all these cases.

Suppose a target concentration, denoted by C_{tgt}, is to be achieved and maintained during the treatment period for T hours. Knowledge of the target concentration,

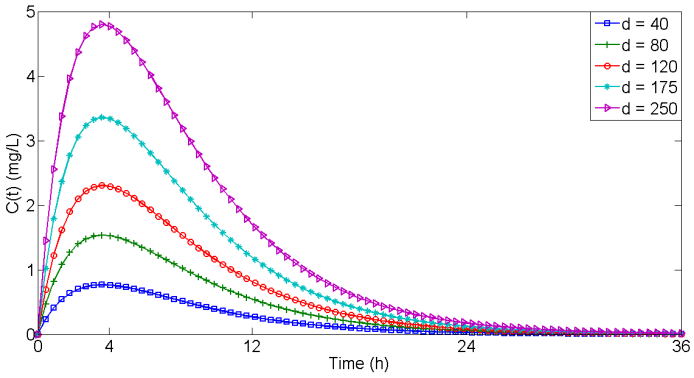


Figure 4.1: The function $C(t, d)$ for five different single dose levels, where $C(t, d)$ is a one compartment first order absorption model with parameter estimates $\widehat{K}_a = .37\,\mathrm{h}^{-1}$, $\widehat{K}_e = 0.2\,\mathrm{h}^{-1}$, $\widehat{V} = 24$ L and $\widehat{F} = .95$.

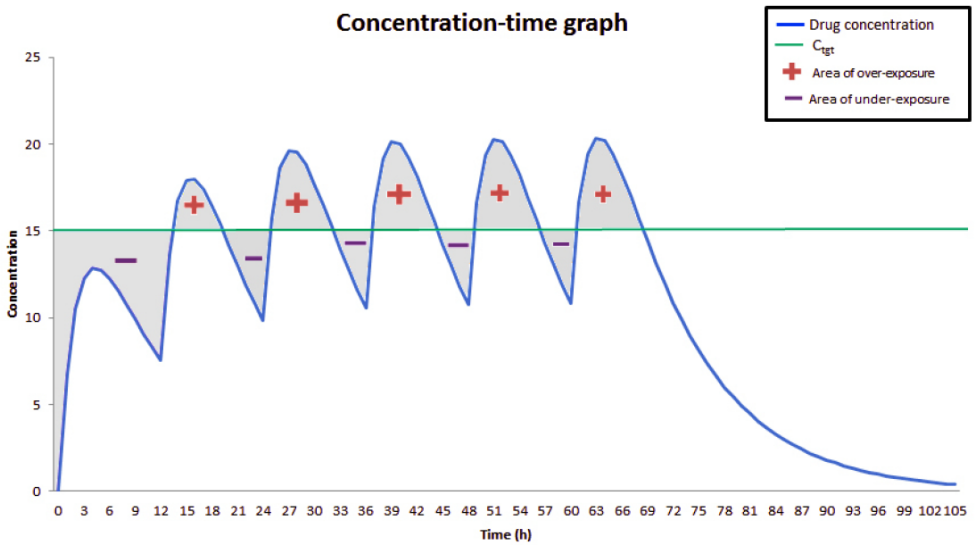


Figure 4.2: The problem is to find a dose regimen which minimises under- and over-exposure, that is the shaded area.

C_{tgt}, is a prerequisite to the application of the ED algorithm.

Then, Figure 4.2 shows the problem to be solved: for given n, $\boldsymbol{t}$ and T, how to compute a dose vector $\boldsymbol{D} = (d_1, ..., d_n)^T$ which minimises the areas denoted by the '+' and '-' signs. It is assumed that the mechanistic model followed by the drug is known along with the estimates of the pharmacokinetic parameters.

Underexposure occurs when the concentration is below the target level (C_{tgt}) and overexposure occurs when the concentration is above the target level. As a measure of the under- and overexposure, we consider the area between the drug's concentration and the assumed target. This can be calculated by integration of $|C(t, d) - C_{tgt}|$ over the appropriate intervals of time. Let t_j be the dose time point for the j^{th} dose, where $j = 1, ..., n$ and $t_1 = 0$. The number of doses n and the times t_j's should be chosen in such a manner that T is covered. Let $\tau_j = t_{j+1} - t_j$, $j = 1, .., n-1$, and $\tau_n = T - t_n$.

We define a function $\Delta_i : \mathbb{R}^i_{\geq 0} \mapsto \mathbb{R}_{\geq 0}$ such that,

$$\Delta_1(d_1) = \int_0^{\tau_1} |C(t, d_1) - C_{tgt}|dt,$$
$$\Delta_2(d_1, d_2) = \int_0^{\tau_2} |C(\tau_1 + t, d_1) + C(t, d_2) - C_{tgt}|dt,$$
$$\Delta_3(d_1, d_2, d_3) = \int_0^{\tau_3} |C(\tau_1 + \tau_2 + t, d_1) + C(\tau_2 + t, d_2) + C(t, d_3) - C_{tgt}|dt,$$
$$\vdots$$
$$\Delta_n(d_1, d_2, ..., d_n) = \int_0^{\tau_n} |C(\tau_1 + ... + \tau_{n-1} + t, d_1) + ... + C(t, d_n) - C_{tgt}|dt.$$

The Δ-functions measure the areas of under- and overexposure around the target concentration. Δ_1 measures the areas in the interval $[0, t_2]$ as a function of d_1. Δ_2 measures the areas in the interval $[t_2, t_3]$ as a function of (d_1, d_2) and so on. Δ_1 is generally the largest as the concentration increases from 0 and it may take a few doses to stabilise around the target concentration. In the next section we define various efficiency criteria based on these measures for optimisation of the dose regimen.

4.2 Criteria of Efficiency

Let $\boldsymbol{\Delta} = (\Delta_1, \Delta_2, ..., \Delta_n)^T$. We consider real valued functionals $\varphi : \mathbb{R}^n_{\geq 0} \mapsto \mathbb{R}_{\geq 0}$ of $\boldsymbol{\Delta}$. Dose regimens that minimise $\varphi(\boldsymbol{\Delta})$ will be called φ-efficient. We minimise the function $\varphi(\boldsymbol{\Delta})$ over the sequences of doses $\widetilde{\boldsymbol{D}} = \{(d_1), (d_1, d_2), \ldots, (d_1, d_2, \ldots, d_n)\}$.

The last element of the sequence is a dose vector $\boldsymbol{D}$ and it contains all other elements of the sequence. Hence, for brevity, we will be saying that $\boldsymbol{D}$ minimises the function $\varphi(\boldsymbol{\Delta})$. It is possible to have $d_i = 0$ for any $i = 1, ..., n$. This may correspond to intended or unintended skipping of the dose. We discuss these situations later. Function $\varphi(\boldsymbol{\Delta})$ can also be expressed using the notation introduced in Chapter 3 as follows:

$$\varphi(\boldsymbol{\Delta}) = \vartheta(\mathcal{R}|n, T, \boldsymbol{t}, \theta = 0). \tag{4.1}$$

Since here the objective function is being optimised with respect to the dose vector $\boldsymbol{D}$ only, the terms optimum dose regimen and optimum dose vector will be used interchangeably.

Definition 1. *Let $\mathfrak{D}$ be the class of all dose vectors $\boldsymbol{D}$ defined on $[0, d_{max}]^n$, where d_{max} is the maximum dose which can be administered.*

The choice of φ depends on many factors. Below we propose some possible effi-

ciency measures.

Definition 2 (φ_A-efficiency)**.** *A regimen $\boldsymbol{D}^* \in \mathfrak{D}$, $\boldsymbol{D}^* = (d_1^*, \ldots, d_n^*)^{\mathrm{T}}$ is called φ_A-efficient if the function*

$$\varphi_A(\boldsymbol{\Delta}(\boldsymbol{D})) = \frac{1}{n}\sum_{i=1}^{n}\Delta_i$$

is minimised by $\boldsymbol{D}^*$ or equivalently

$$\sum_{i=1}^{n}\Delta_i^* \leq \sum_{i=1}^{n}\Delta_i$$

for all $\boldsymbol{D} \in \mathfrak{D}$, where $\Delta_i^* = \Delta_i(d_1^*, ..., d_i^*)$, $i = 1, ..., n$.

φ_A-efficiency is the first criterion which comes to mind to achieve the objective of ensuring closeness of $C(t)$ and C_{tgt}. However, arithmetic mean as a measure of location is quite sensitive to extreme observations. Generally, values of Δ_1 will be quite large as compared to subsequent values because the concentration increases from 0 to hit C_{tgt} for the first time. This would lead to selection of a very high loading dose which although may be prescribable, may not be desirable. To overcome this problem, one could use the φ_G-efficiency criterion which is based on the geometric mean, a measure of location relatively less influenced by extreme observations.

Definition 3 (φ_G-efficiency)**.** *A regimen $\boldsymbol{D}^* \in \mathfrak{D}$ is called φ_G-efficient if the function*

$$\varphi_G(\boldsymbol{\Delta}) = (\Pi_{i=1}^{n}\Delta_i)^{\frac{1}{n}}$$

is minimised by $\boldsymbol{D}^*$ or equivalently

$$\Pi_{i=1}^{n}\Delta_i^* \leq \Pi_{i=1}^{n}\Delta_i$$

for all $\boldsymbol{D} \in \mathfrak{D}$.

Alternatively, one could take a weighted mean of the Δ_is so that initial doses are given lesser importance as compared to subsequent ones. We can thus define φ_C-efficiency as follows.

Definition 4 (φ_C-efficiency)**.** *Let $\boldsymbol{C} = (c_1, \ldots, c_n)^T$ be a vector of real numbers defined on $[0, 1]^n$ such that $\sum_{i=1}^{n} c_i = 1$. A regimen $\boldsymbol{D}^*$ is called φ_C-efficient if the function*

$$\varphi_C(\boldsymbol{\Delta}) = \boldsymbol{C}^T\boldsymbol{\Delta} = \sum_{i=1}^{n} c_i\Delta_i$$

is minimised by $\boldsymbol{D}^*$ *or equivalently*

$$\boldsymbol{C}^T\boldsymbol{\Delta}^* \leq \boldsymbol{C}^T\boldsymbol{\Delta}$$

for all $\boldsymbol{D} \in \mathfrak{D}$*, where* $\boldsymbol{\Delta}^* = (\Delta_1^*, ..., \Delta_n^*)^T$.

Another possibility is the harmonic mean. It is less affected by large outliers than the arithmetic mean and thus leads to a smaller loading dose in the optimised dose vector. This makes the dose sizes in the optimised dose vector more uniform.

Definition 5 (φ_H-efficiency)**.** *A regimen* $\boldsymbol{D}^*$ *is called* φ_H*-efficient if the function*

$$\varphi_H(\boldsymbol{\Delta}) = \frac{1}{\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\Delta_i}}$$

is minimised by $\boldsymbol{D}^*$ *or equivalently*

$$\sum_{i=1}^{n}\frac{1}{\Delta_i^*} \geq \sum_{i=1}^{n}\frac{1}{\Delta_i}$$

for all $\boldsymbol{D} \in \mathfrak{D}$.

The harmonic mean is also used in the science of drug discovery to identify the active components in a mixture of compounds, Santos et al. (2011).

A potential problem with φ_G and φ_H criteria is that they can not be used if any of the functions Δ_i, $i = 1, ..., n$, are zero. In this thesis, we mainly use the φ_A criterion.

In the next section we present the Efficient Dosing (ED) algorithm for solving some of the optimisation problems we proposed in Section 3.3.

4.3 The Efficient Dosing Algorithm

We propose the following iterative algorithm for finding the efficient dose regimens for the criteria defined previously. The optimal doses so obtained will be real numbers which, for practical or manufacturing reasons, can be rounded to the nearest dose levels possible. Alternatively, we propose to perform the optimisation over discrete set of possible dose levels. Such a case is discussed later.

Let $L_i^k = \{d_{i1}^k, d_{i2}^k, d_{i3}^k\}$ be the set of 3 possible doses that can be administered at the ith occasion, $i = 1, ..., n$, and k-th iteration, $k = 1, 2, ..., w$. These sets will be called the *dose sets*. Thus, the number of possible dose regimens in this case is 3^n.

Iteration No. 1

Set $L_i^1 = \{10\% \text{ of } d_{max},\ 50\% \text{ of } d_{max},\ d_{max}\}$ for all $i = 1, ..., n$, e.g., if $d_{max} = 1000$, $L_i^1 = \{100, 500, 1000\}$ for all $i = 1, ..., n$. Any three values can be selected from $(0, d_{max}]$ to form L_i^1, though the algorithm will generally converge quicker if the values are uniformly spread in the interval $(0, d_{max}]$. Zero should not be used for a reason which will be apparent later. For each of the 3^n dose regimens, we compute $\varphi(\boldsymbol{\Delta})$. Let $\boldsymbol{D}^1$ be the dose regimen which minimises $\varphi(\boldsymbol{\Delta})$ under the desired criterion. This will be the most efficient regimen at the 1st iteration.

Iteration No. k

In iteration k the dose sets L_i^k are reconstituted based on the doses in $\boldsymbol{D}^{k-1}$, that is the doses minimising $\varphi(\boldsymbol{\Delta})$ in the previous iteration. For example, if d_{i3}^{k-1} was selected for i-th occasion in iteration $k-1$, the new L_i^k will be:

$$L_i^k = \left\{ \delta \times d_{i3}^{k-1},\ d_{i3}^{k-1},\ \min\left(\frac{d_{i3}^{k-1}}{\delta}, d_{max} \right) \right\}, \tag{4.2}$$

where $\delta \in (0, 1)$ is a fixed constant, called the *resolution*. What value of δ to use is discussed later. If d_{i1}^{k-1} or d_{i2}^{k-1} were selected, L_i^k will be constituted in the similar way.

Terminal Iteration

To terminate the algorithm, a suitable convergence criterion is required. The algorithm is terminated at w^{th} iteration if

$$d_i^w = d_i^{w-1} \ \forall\ i = 1, \ldots, n, \tag{4.3}$$

where d_i^k denotes the optimal dose to be administered at the i^{th} occasion, at the k^{th} iteration. Thus, the algorithm terminates when for the given resolution, no further improvement is possible.

The Efficient Dosing (ED) algorithm was programmed in MATLAB® version 7.13.0.564 and the code is given in Appendix D.1.

Figure 4.3 presents the algorithm schematically.

By choosing an appropriate resolution, $\boldsymbol{D}^w$ can be driven as close as required to the most efficient dose regimen $\boldsymbol{D}^*$. This is stated in the following theorem.

Theorem 1. *The ED Algorithm converges to the true unknown φ-efficient dose regimen $\boldsymbol{D}^*$ when the resolution tends to 1, that is*

$$\delta \to 1 \ \Rightarrow \boldsymbol{D}^w \to \boldsymbol{D}^*.$$

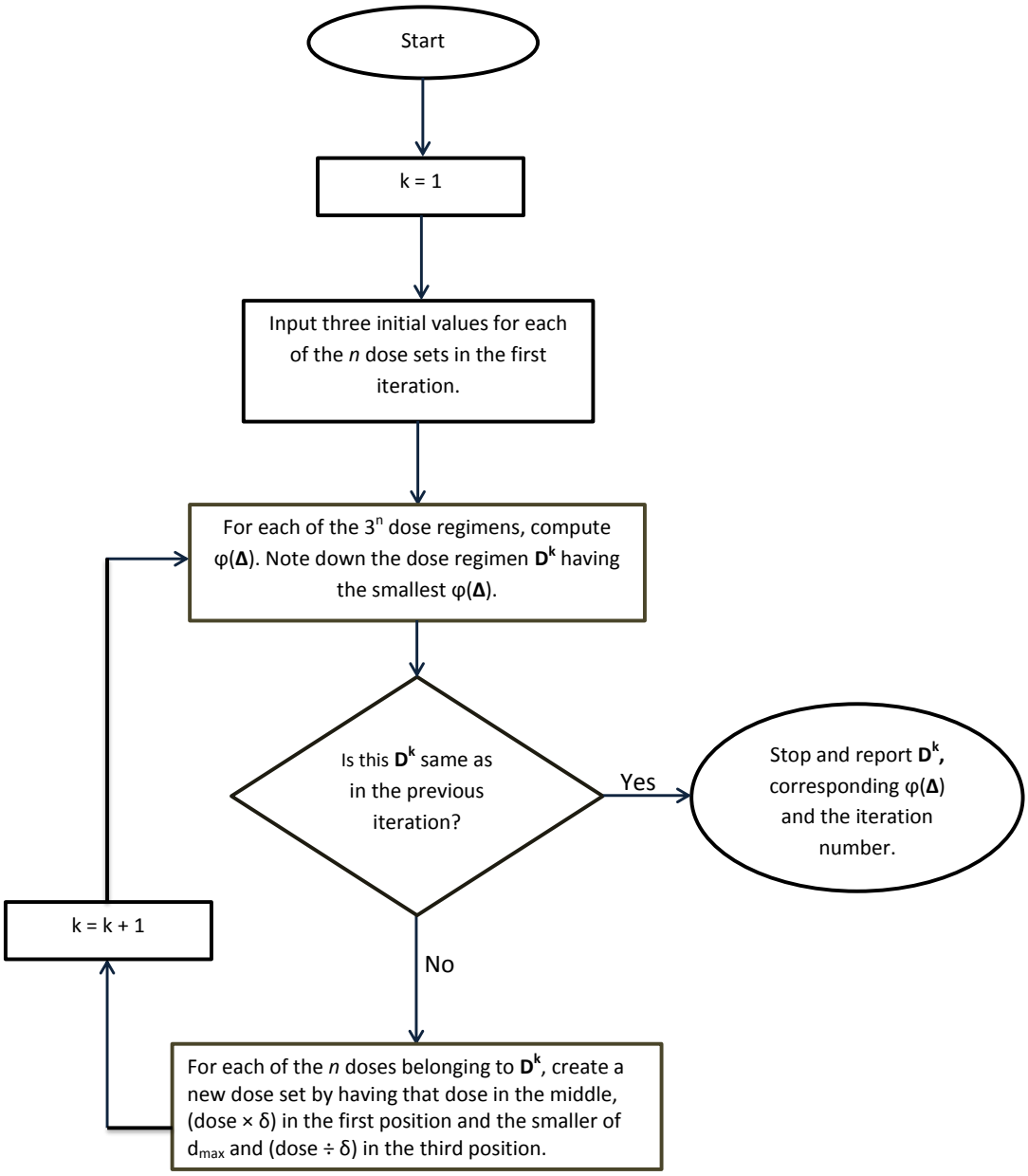


Figure 4.3: Flowchart showing the working of the ED algorithm. The 'position' in a dose set is the order in which the doses appear. So for $L_i = \{a, b, c\}$, a is in the first position, b in the middle and c in the third position.

Proof. At each iteration, the algorithm selects the best dose regimen - the one having minimum $\varphi(\boldsymbol{\Delta})$, amongst the 3^n dose regimens available at that iteration. For this dose regimen, the n doses are then *matched* to their corresponding dose sets L. Matching means that the positions of the selected doses in their respective dose sets are identified by the algorithm. This matching enables the algorithm to find the direction in which the optimal dose is located. For example, if the first member d_{i1} of a dose set $\{d_{i1}, d_{i2}, d_{i3}\}$ gets selected, that means the optimal dose is located in the interval $[0, d_{i2})$. Dose d_{i2} will be selected when the optimal dose lies in $(d_{i1}, d_{i2}] \cup [d_{i2}, d_{i3})$. If d_{i3} gets selected, it means that the optimal dose is in the interval (d_{i2}, ∞). However, if the optimal dose lies beyond $d_{\max}$, the algorithm selects $d_{\max}$ as the optimal dose for that occasion. We deem $d_{\max}$ to be the optimal dose in that case.

In the next iteration, for all the n doses, new dose sets are constituted in the respective directions, around the doses selected in the previous iteration. Thus, at each iteration, the dose sets move towards the optimal doses d_i^*, $i = 1, ..., n$. The algorithm is terminated at the w^{th} iteration when, for the given resolution δ, the same dose as in the previous iteration is selected for each occasion according to the stopping rule (4.3). This rule is equivalent to

$$|d_i^w - d_i^*| < |\delta d_i^w - d_i^*| \text{ and } |d_i^w - d_i^*| < \left| \min\left(\frac{d_i^w}{\delta}, d_{max}\right) - d_i^* \right|, \tag{4.4}$$

where the unknown true optimal doses d_i^* lie in the respective intervals

$$\left(\delta d_i^w, \min\left\{\frac{d_i^w}{\delta}, d_{max}\right\}\right), \tag{4.5}$$

whose lower and upper limits tend to d_i^w when $\delta \to 1$, that is, the solution tends to the optimum. This means that $\boldsymbol{D}^w \to \boldsymbol{D}^*$ when $\delta \to 1$. □

Equation (4.5) gives the interval in which the true optimal dose would lie. Obviously, one would want this interval to be as narrow as possible. The only way to do this is to choose a value of δ very close to 1. However, a downside of having δ close to 1 is slower convergence of the algorithm. If the distance between the optimal dose and the dose at current iteration is large, it will take many iterations for that interval to move towards and encompass the optimal dose. The idea behind choosing the initial values in iteration 1 is to cover the range of doses - $(0, d_{max}]$ in the three points. The optimal doses so found, are a function of the initial values chosen in iteration 1. As the resolution is increased, this dependence on initial values diminishes and as $\delta \to 1$, the algorithm is indifferent to the initial values chosen. We discuss in an example later on how the resolution affects the precision and the rate of convergence. Henceforth, $\varphi(\boldsymbol{\Delta}^*|\delta)$ will denote the optimal value of $\varphi(\boldsymbol{\Delta})$ obtained at resolution δ.

Alternatives to having 3 doses in a dose set

Let l be the number of dose levels in a dose set. We consider two alternatives to $l = 3$:

- $l = 2$. Dose sets in successive iterations can be constituted as (assuming d_{i1}^k is selected in the current iteration),

$$L_i^k = \left\{\delta d_{i1}^k, \min\left(\frac{d_{i1}^k}{\delta}, d_{max}\right)\right\}.$$

  The number of possible dosing regimens in this case is 2^n which is significantly less than 3^n. This will enable an iteration to be completed faster than when

$l = 3$. However, if d_{max} is large, the number of iterations required for the neighborhood to be formed around the optimal doses will be large and thus any gain in efficiency will be substantially outweighed. However, if d_{max} is small, this could be more efficient than taking $l = 3$.

- $l = 4$. We define two types of dose sets here.

  Type 1 dose sets are those which in successive iterations can be constituted as (assuming d_{i1}^k is selected in the current iteration),

  $$L_i^k = \left\{ \delta^2 d_{i1}^k,\, \delta d_{i1}^k,\, d_{i1}^k,\, \min\left(\frac{d_{i1}^k}{\delta},\, d_{max} \right) \right\}$$

  Type 2 dose sets are those which in successive iterations can be constituted as (assuming d_{i1}^k is selected in the current iteration),

  $$L_i^k = \left\{ \delta d_{i1}^k,\, d_{i1}^k,\, \min\left(\frac{d_{i1}^k}{\delta},\, d_{max} \right),\, \min\left(\frac{d_{i1}^k}{\delta^2},\, d_{max} \right) \right\}$$

  The number of dose regimens in this case is 4^n which will considerably increase the time required to complete a single iteration. However, the number of iterations required in case of $l = 4$ will be smaller as the optimal doses d_i^* will be quickly captured. This could be more efficient if d_{max} is very large. In the next section, we shall explore the effect of the choice of l on the performance of the ED algorithm with the help of an example.

Practical Constraints and Solutions

We discuss below some constraints which may arise in practical application of the methods discussed above and how to resolve them.

Discretisation of doses: It may not be always possible to administer doses of any quantity. For example, a tablet is manufactured at only one weight say, 20 mg and multiple tablets can be taken as required. Thus the doses which can be administered to the patient would be in multiples of 20. The problem then is to find the most efficient dose regimen in such a case. One way could be to run the ED algorithm over $\mathfrak{D} = [0, d_{max}]^n \subset \mathbb{R}^n$ and round the optimal doses found to the nearest multiples of 20. However, this will be a sub-optimal method as changing one dose affects all subsequent doses.

The ED algorithm can be adapted to find the most efficient dose regimen from a discrete set of dose levels. Let κ be the quantity at which a drug is manufactured

and so doses can only be in multiples of κ. We also assume that d_{max} is a multiple of κ. Let $\mathfrak{D}_\kappa = \{0, \kappa, 2\kappa, ..., d_{max}\}^n$ be the class of all dose regimens in this case. The objective then is to find $\boldsymbol{D}^*_\kappa \in \mathfrak{D}_\kappa$ such that $\varphi(\boldsymbol{\Delta}_\kappa|\kappa)$ is minimised, where the vector $\boldsymbol{\Delta}_\kappa$ contains the Δ-functions for the case of discretised doses.

The ED algorithm can now be adapted for this situation by making some changes. Firstly, the initial dose sets should have doses which are multiples of κ. They can be chosen in such a way that the interval $(0, d_{max}]$ is covered uniformly in three points of the initial dose sets. Secondly, the re-constitution of the new dose sets in successive iterations will be different than what was done previously. Instead of multiplying and dividing by δ, we add and subtract κ. For example Equation (4.2), in the adapted algorithm, will become

$$L_i^k = \left\{\max(0, d_{i3}^{k-1} - \kappa),\ d_{i3}^{k-1},\ \min\left(d_{i3}^{k-1} + \kappa, d_{max}\right)\right\}.$$

As a result, at each iteration, the possible doses will remain as multiples of κ. The max function is necessary to ensure non-negativity of the doses.

One might be interested to know the loss in efficiency in this case as compared to the case of real valued dose levels. Let $\varphi(\boldsymbol{\Delta}^*|\delta)$ be the φ-efficient regimen at resolution δ and let $\varphi(\boldsymbol{\Delta}^*|\kappa)$ be the efficient dose regimen for the case of discrete dose levels. We define a measure of relative efficiency as

$$\chi = \frac{\varphi(\boldsymbol{\Delta}^*|\delta)}{\varphi(\boldsymbol{\Delta}^*|\kappa)} \times 100. \tag{4.6}$$

In Section 4.4 we present an example and calculate the efficiency to see the effect of discretising the dose levels' space.

Skipped or partially administered doses: There might be some deviations from the intended dose regimens because of a variety of reasons. Whenever a departure takes place, this information should be entered, before the next occasion, in the ED algorithm to get the revised dose regimens. For example, if a dose is skipped, the dose set at that time point should be set to 0 and the ED algorithm should be re-run to get the updated subsequent doses. Another instance could be when only a part of the intended dose is administered. The dose sets at such occasions could be retrospectively set to the amount which got administered. Then, re-running the ED algorithm would compute the updated dose regimen. The algorithm can thus be used to quantify the effect of patient non-compliance on the efficiency of the administered dose regimen.

All doses constrained to be equal: If there is an additional constraint on the dose vector that the doses must be equal, i.e, $d_i = d$ for all $i = 1, ..., n$, then this can be incorporated in the algorithm. In every iteration, the n dose sets are taken to be equal to each other, i.e., $L_i^k = L^k$ for all $i = 1, ..., n$ and $k = 1, ..., w$. This results in just $l = 3$ distinct dose vectors in each iteration from which the algorithm selects that dose vector which minimises φ. In the next iteration, the dose selected in the previous iteration is used to form n equal dose sets as shown in Equation 4.2. This continues until convergence is obtained.

4.4 Applications and Extensions

We now illustrate the ED algorithm by means of a numerical example. We also give an example of how the algorithm can be extended to find the best dose regimen and the combination ratio in case of two drugs in a fixed dose combination unit. In these examples, we assume the knowledge of the mechanistic model followed by the drug's concentration and the model parameters as well.

Example 1: Dose Regimen for a Single Drug

We consider the concentration profile of a one-compartment model which, after a single dose d is the following function of time t:

$$C(t) = \frac{FdK_a}{V(K_a - K_e)}(e^{-K_e t} - e^{-K_a t}), \tag{4.7}$$

where K_a denotes the absorption rate constant, K_e denotes the elimination rate constant, V is the volume of distribution and F is the bioavailability. For the calculations we take the following values of the parameters as their estimates: $\widehat{K}_a = .37$ h^{-1}, $\widehat{K}_e = 0.2$ h^{-1}, $\widehat{V} = 24$ L and $\widehat{F} = .95$. We assume that $C_{tgt} = 3$ mg/L, $d_{max} = 240$ mg and we consider $n = 7$ occasions to administer the drug with $\tau_j = 6$ h, $j = 1, ..., 7$. Furthermore, we assume that it is desired to maintain C_{tgt} for $T = 42$ h. We take the resolution of the algorithm as $\delta = 0.99$.

For each of the seven dosing time points, the corresponding initial dose sets to choose from are shown in Table 4.1. The values in italic minimise $\varphi_A(\mathbf{\Delta}|0.99)$ over the 3^7 dose regimens. They are used in the next iteration as the middle points of the new dose sets. To exemplify the ED algorithm we show the dose sets of iterations two and three (Tables 4.2 and 4.3).

t_1	t_2	t_3	t_4	t_5	t_6	t_7
40.00	40.00	*40.00*	40.00	40.00	*40.00*	40.00
120.00	*120.00*	120.00	*120.00*	*120.00*	120.00	*120.00*
240.00	240.00	240.00	240.00	240.00	240.00	240.00

Table 4.1: The seven dose sets in the first iteration, represented by the columns. We chose the same initial dose sets for all the occasions. The rows represent the positions of the doses in their respective dose sets.

t_1	t_2	t_3	t_4	t_5	t_6	t_7
118.80	*118.80*	39.60	*118.80*	*118.80*	39.60	*118.80*
120.00	120.00	40.00	120.00	120.00	40.00	120.00
121.21	121.21	*40.40*	121.21	121.21	*40.40*	121.21

Table 4.2: The seven dose sets in the second iteration.

t_1	t_2	t_3	t_4	t_5	t_6	t_7
120.00	*117.61*	40.00	*117.61*	*117.61*	40.00	*117.61*
121.21	118.80	40.40	118.80	118.80	40.40	118.80
122.44	120.00	*40.81*	120.00	120.00	*40.81*	120.00

Table 4.3: The seven dose sets in the third iteration.

The algorithm converged at the 85^{th} iteration and the following φ_A-efficient dose regimen in $\mathfrak{D} = [0, d_{max}]^7$, at resolution of $\delta = 0.99$, was obtained:

$$\boldsymbol{D}_A^\star = (163.87,\ 69.04,\ 92.12,\ 87.00,\ 87.88,\ 88.49,\ 87.88)^T .$$

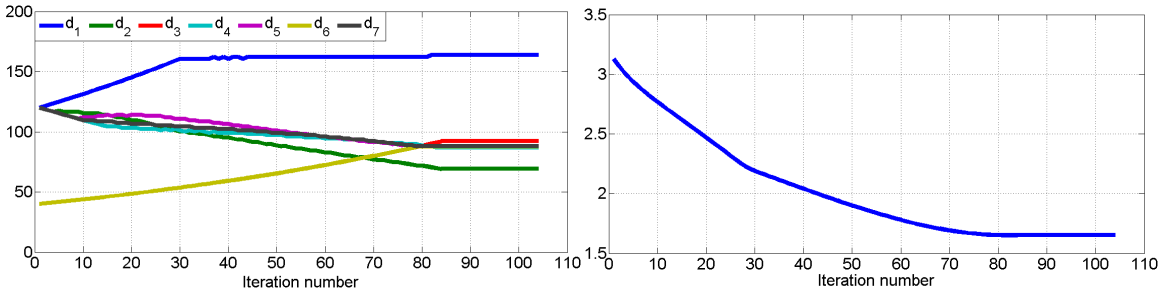
The loading dose is the largest and after dose 3, the dose levels stabilise at about 88 mg/L. Figure 4.4 shows the output from the ED algorithm. Plots 4.4a and 4.4b show the property of convergence of the algorithm. Although convergence was achieved at the 85^{th} iteration, we allowed the algorithm to run for another 20 iterations to demonstrate convergence.

For the φ_A criterion, we obtain $\varphi_A(\boldsymbol{\Delta}^\star|0.99) = 1.6494$ (mg/L)×h. Plot 4.4c shows the concentration profile obtained for the efficient dose regimen for the assumed parameter values. It oscillates about the target concentration.

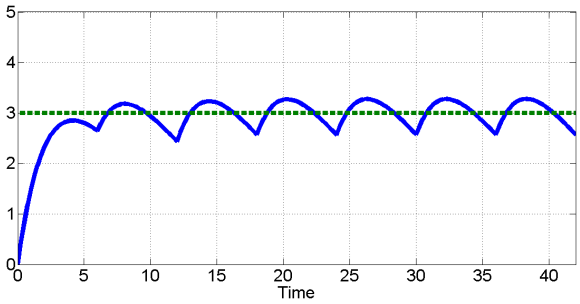
On the other hand, the φ_H-efficient dose regimen is

$$\boldsymbol{D}_H^\star = (159.00,\ 72.00,\ 91.20,\ 87.00,\ 88.76,\ 87.60,\ 87.88)^T .$$

Figure 4.5 shows the output from the ED algorithm. For this criterion, we obtain $\varphi_H(\boldsymbol{\Delta}^*|0.99) = 1.5190$. It can be seen that the initial loading dose is lower for the φ_H-efficient dose regimen in comparison to the φ_A-efficient dose regimen. The reason for this is that as Δ_1 is the largest in $\boldsymbol{\Delta}$, the arithmetic mean is more influenced by its value than the harmonic mean is.

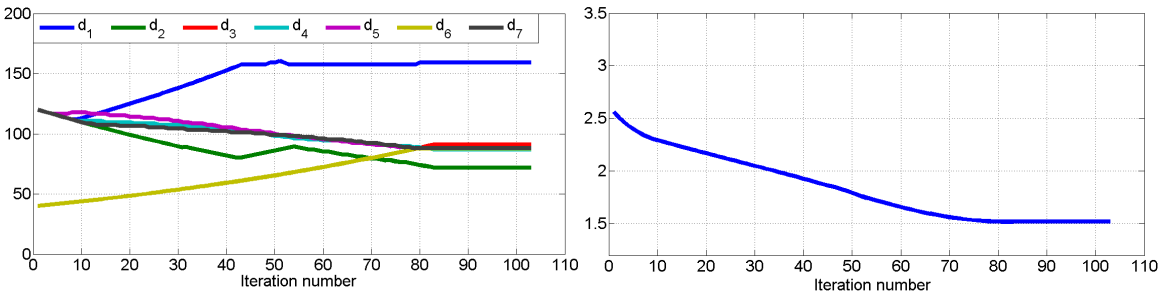


(a) Convergence of $\boldsymbol{D}_A = (d_1, ..., d_7)^T$. (b) Convergence of $\varphi_A(\boldsymbol{\Delta}|0.99)$.

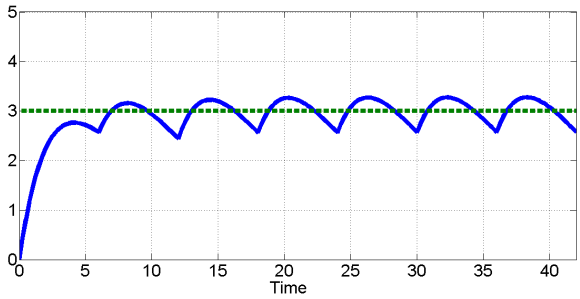


(c) Concentration profile resulting from $\boldsymbol{D}^\star$.

Figure 4.4: Output from the ED algorithm for φ_A-efficiency criterion. The Algorithm Converged at the 85^{th} iteration, with $\varphi_A(\boldsymbol{\Delta}^*|0.99) = 1.6494$, but was allowed to run for another 20 iterations to demonstrate convergence.



(a) Convergence of $\boldsymbol{D} = (d_1, ..., d_7)^T$. (b) Convergence of $\varphi_H(\boldsymbol{\Delta}^*|.99)$.



(c) Concentration profile resulting from $\boldsymbol{D}^*$.

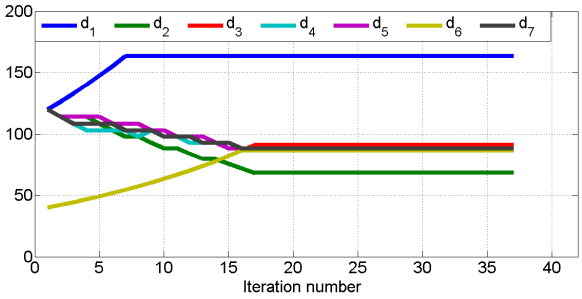
Figure 4.5: Output from ED algorithm for φ_H-efficiency criterion. The Algorithm Converged at 84^{th} iteration, with $\varphi_H(\boldsymbol{\Delta}^*|0.99) = 1.5190$, but was allowed to run for another 20 iterations to demonstrate convergence.

To compare the results with the case of having a lower resolution, the above com-

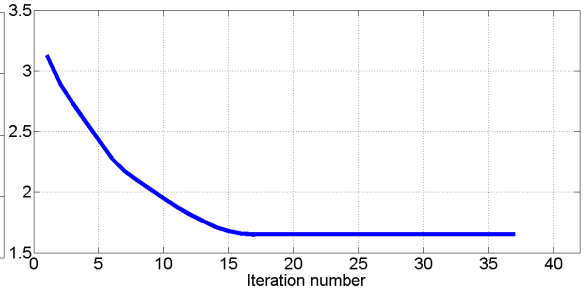
putations were repeated for resolution 0.95. The φ_A-efficient dose regimen is

$$\boldsymbol{D}_A^{\star} = (163.24,\ 68.26,\ 90.88,\ 88.21,\ 88.21,\ 86.34,\ 88.21)^T .$$

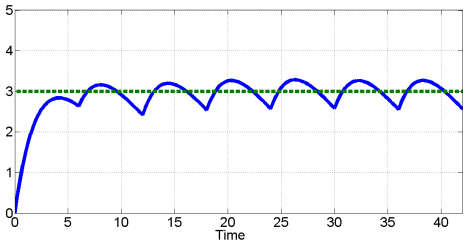
The algorithm converged in just 18 iterations and $\varphi_A(\boldsymbol{\Delta}^*|.95) = 1.6511$. It took 97 seconds while that for $\delta = 0.99$ took 421 seconds, which may be an important gain when the algorithm needs to be repeatedly used as part of a larger simulation study. Thus, there is no significant loss in precision but substantial gain in computational efficiency. One may, therefore, use a resolution of 0.95 for practical purposes. Figure 4.6 displays the output from the ED algorithm.



(a) Convergence of $\boldsymbol{D}_A = (d_1, ..., d_7)^T$



(b) Convergence of $\varphi_A(\boldsymbol{\Delta}^*|.95)$



(c) Concentration profile resulting from $\boldsymbol{D}^*$.

Figure 4.6: Output from ED algorithm for φ_A-efficiency criterion. The Algorithm Converged at 18^{th} iteration with, $\varphi_A(\boldsymbol{\Delta}^*|0.95) = 1.6511$, but was allowed to run for another 20 iterations to demonstrate convergence.

To explore the dependence of the algorithm on the initial values, we took 5 different sets of starting values as shown in Table 4.4. It is clear that at a resolution of 0.99, the choice of starting values does not significantly affect the solution.

We also considered different sizes of dose sets. Table 4.5 presents the results. The optimal dose regimens are almost the same for all choices of the dose set size. There is a slight difference in the case of Type 2 dose set of size four. The φ_A values marginally improve with increase in the value of l. Since the number of possible dose regimens in each iteration is l^n, this small improvement is because of the increased refinement of the search grid. As discussed earlier, the number of iterations decrease with increase in the value of l.

Let us now consider the case of discretisation of the dose space. Let us take

d_1^1	d_2^1	d_3^1	$d_1^\star$	$d_2^\star$	$d_3^\star$	$d_4^\star$	$d_5^\star$	$d_6^\star$	$d_7^\star$	$\varphi_A^\star$
60	80	160	163.25	69.50	92.09	86.70	88.46	87.57	88.46	1.6492
75	150	225	162.56	69.90	91.70	87.18	88.08	88.08	88.08	1.6495
50	100	150	162.56	69.64	92.30	86.87	87.75	87.75	87.75	1.6495
30	130	250	163.81	69.02	91.54	86.97	87.84	87.93	87.84	1.6494
25	125	250	163.97	69.09	92.33	87.05	87.93	87.81	87.93	1.6495

Table 4.4: Dependence of the ED algorithm on the initial values chosen chosen for the dose sets.

l	d_1	d_2	d_3	d_4	d_5	d_6	d_7	φ_A	Iterations	Time
2	163.81	68.83	91.20	87.60	87.85	87.60	87.60	1.6497	100	34
3	163.87	69.04	92.12	87.00	87.88	88.49	87.88	1.6494	85	421
4 (I)	163.25	69.50	92.09	86.7	88.46	87.57	88.46	1.6492	16	511
4 (II)	160.00	71.63	91.17	87.57	87.57	88.46	87.57	1.6485	18	526

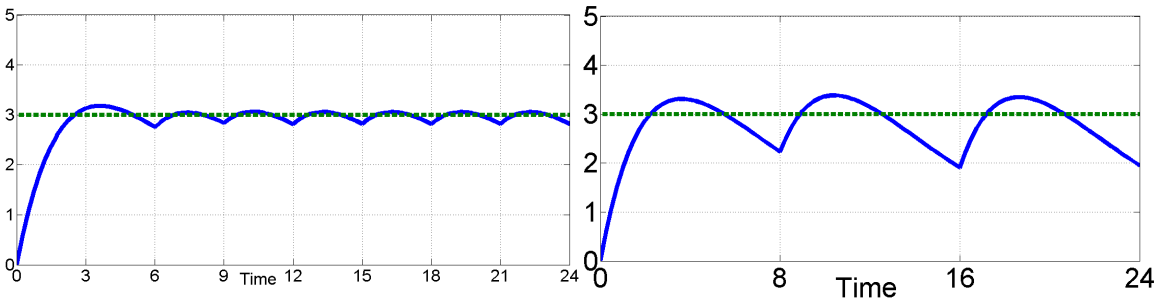
Table 4.5: The effect of different dose set sizes on the convergence properties of the ED algorithm. The time in seconds taken by the algorithm shown in the last column.

$\kappa = 10$ mg. For this value of κ, $\varphi_A(\boldsymbol{\Delta}_{10}^*) = 1.6676$. For the case of continuous doses with $\delta = 0.99$, $\varphi_A(\boldsymbol{\Delta}^*|0.99)$ was found to be 1.6494. Comparing the two values we obtain $\chi = 1.6494/1.6676 = 98.91\%$. The loss in efficiency is therefore not that much. However, when $\kappa = 20$ mg, we obtain $\varphi_A(\boldsymbol{\Delta}_{20}^*) = 1.7217$. Comparing it with the continuous doses case we get $\chi = 1.6494/1.7217 = 95.80\%$. As expected, χ decreases when κ is increased. The dose regimens are given in Table 4.6.

κ	Efficient Dose Regimen	φ_A
0	$(163.87, 69.04, 92.12, 87.00, 87.90, 88.50, 87.90)^T$	1.6493
10	$(160, 70, 90, 90, 90, 90, 90)^T$	1.6676
20	$(160, 80, 80, 100, 80, 100, 80)^T$	1.7217

Table 4.6: Effect of Discretisation on the Efficient Dose Regimens ($\kappa = 0$ describes the case of no discretisation).

Another observation regarding the criterion φ is that, generally, $\varphi(\boldsymbol{\Delta}^*)$ increases with increasing dosing time τ. Maintenance around C_{tgt} becomes difficult if τ is large, whereas, for example in the case of i.v. infusion, for which τ is very small, we will obtain $\varphi(\boldsymbol{\Delta}^*)$ close to zero. This is clear from Figure 4.7.



(a) $C(t)$ for $\tau = 3$ h. $\varphi_A(\boldsymbol{\Delta}^*|.99) = 0.5284$. (b) $C(t)$ for $\tau = 8$ h. $\varphi_A(\boldsymbol{\Delta}^*|.99) = 3.4515$.

Figure 4.7: Graphs showing how the course of concentration about the target depends on the dosing interval τ.

4.4.1 Misspecification of the PK Model

A practical issue that can be studied in the context of the ED algorithm is the occurrence of a mild misspecification of the PK model. That is, if the PK parameters of the patient are slightly different from those assumed to be true, how inferior will the administered dose regimen be for the patient. This is also important in all the trials where cohorts of patients are treated with the same dose regimen obtained from the population mean model parameters.

In the example discussed above, for maintenance of the concentration around $C_{tgt} =$ 3 mg/L, the φ_A-efficient dose regimen and the optimal value of the objective function were computed when the parameters of the PK model were assumed to be:

$$\boldsymbol{\beta}_o = (K_{ao}, K_{eo}, V_o, F_o)^T = (.37, .2, 24, .95)^T.$$

Further, in this section we drop index A in φ_A to simplify the notation. Let $\boldsymbol{D}_o^*$ and $\varphi_o(\boldsymbol{D}_o^*)$ be the optimum dose regimen and the optimal value of the objective function corresponding to the assumed vector of parameters $\boldsymbol{\beta}_o$ and $\boldsymbol{D}_o^*$.

$\boldsymbol{D}_o^*$ is the dose regimen which is administered to the patient. As shown in the above example, for resolution $\delta = 0.95$. We have,

$$\boldsymbol{D}_o^* = (163.24,\ 68.26,\ 90.88,\ 88.21,\ 88.21,\ 86.34,\ 88.21)^T \quad \text{with} \quad \varphi_o(\boldsymbol{D}_o^*) = 1.6511.$$

Now, suppose that the true vector of parameter values is $\boldsymbol{\beta}_t$ and not the assumed vector $\boldsymbol{\beta}_o$. The robustness of the ED algorithm can be ascertained by analysing the magnitude of deviations of:

- the optimal dose regimen, $\boldsymbol{D}_t^*$, based on $\boldsymbol{\beta}_t$, from the administered dose regimen $\boldsymbol{D}_o^*$ and
- $\varphi_t(\boldsymbol{D}_o^*)$ from $\varphi_t(\boldsymbol{D}_t^*)$, where $\varphi_t(.)$ is the average over- and under-exposure when

the true PK parameters of the model are $\boldsymbol{\beta}_t$. Since the dose regimen $\boldsymbol{D}_t^*$ is optimal for $\boldsymbol{\beta}_t$, therefore, $\varphi_t(\boldsymbol{D}_o^*) \geq \varphi_t(\boldsymbol{D}_t^*)$. Let us denote by φ_{dev} the percentage by which $\varphi_t(\boldsymbol{D}_o^*)$ exceeds $\varphi_t(\boldsymbol{D}_t^*)$, i.e.,

$$\varphi_{dev} = \frac{\varphi_t(\boldsymbol{D}_o^*) - \varphi_t(\boldsymbol{D}_t^*)}{\varphi_t(\boldsymbol{D}_t^*)} \times 100.$$

The greater the magnitude of φ_{dev}, the higher is the inefficiency of the administered dose regimen $\boldsymbol{D}_o^*$ for the patient. The range of φ_{dev} is $[0, \infty)$.

We can also measure the relative efficiency, φ_{ref}, as

$$\varphi_{ref} = \frac{\varphi_t(\boldsymbol{D}_t^*)}{\varphi_t(\boldsymbol{D}_o^*)} \times 100.$$

φ_{ref} measures the relative efficiency of $\boldsymbol{D}_o^*$, the dose regimen which is actually administered, to the dose regimen which should have been administered, i.e., $\boldsymbol{D}_t^*$. The range of φ_{ref} is $[0, 100]$. A value of φ_{ref} close to 100 signifies that the administered dose regimen was not very inferior to the true optimal dose regimen.

- φ_{ref} and φ_{dev} are related as follows:

$$\varphi_{dev} = 100 \left(\frac{100}{\varphi_{ref}} - 1 \right).$$

We explore these concepts by means of a simulation study as described below.

We simulate $N_{sim} = 1000$ vectors of true PK parameters, $\boldsymbol{\beta}_t^{(k)}$, $k = 1, ..., N_{sim}$, from normal distributions centred at the assumed values of the parameters and having variances as a certain percentage of the respective assumed values. That is, ${\boldsymbol{\beta}_t}^{(k)} \sim \mathcal{N}_4(\boldsymbol{\beta}_o, \boldsymbol{\Omega})$ where $\boldsymbol{\beta}_o = (.37, .2, 24, .95)^T$ and $\boldsymbol{\Omega} = \text{diag}(.0056, .003, .36, .0143)$. These variances are 1.5% of the respective values of the parameters.

The simulated concentration profiles, for the administered dose regimen $\boldsymbol{D}_o^*$, are shown in Figure 4.8 to see the extent of variability generated in $C(t)$. Most simulated profiles deviate mildly from the concentration profile based on the assumed parameters $\boldsymbol{\beta}_o$ (shown in Figure 4.6c). However, several profiles show large departure from the assumed model. We analyse this behaviour later in this section.

For every simulated vector $\boldsymbol{\beta}_t^{(k)}$, the optimal dose regimen ${\boldsymbol{D}_t^*}^{(k)}$, the percentage deviation, $\varphi_{dev}^{(k)}$ and the relative efficiency $\varphi_{ref}^{(k)}$ are computed, $k = 1, ..., N_{sim}$.

To understand the distributions of the deviations, we compute the following statistics:

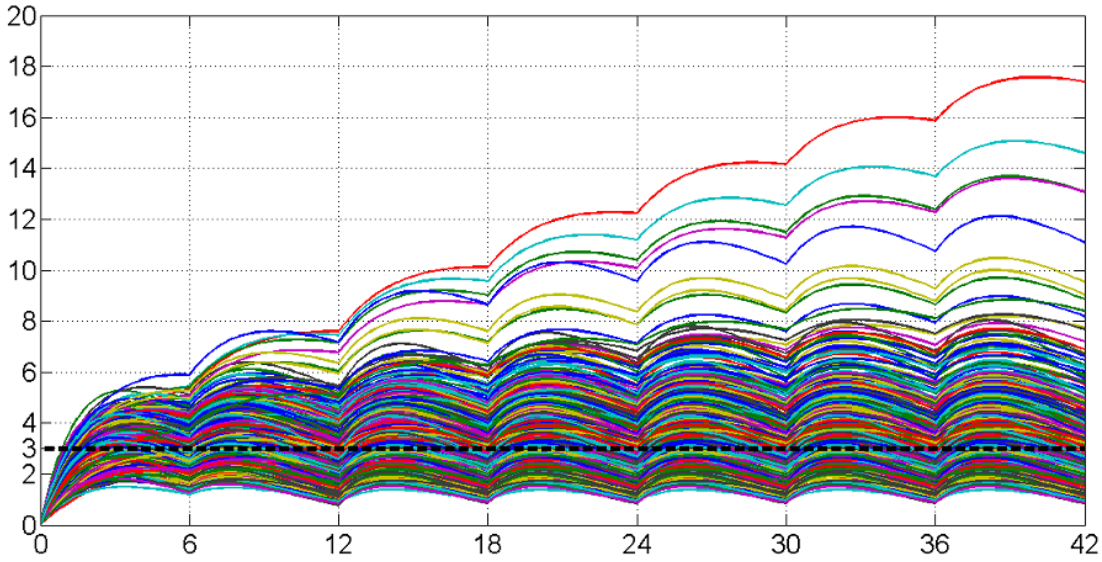


Figure 4.8: Concentration profiles generated by PK parameters simulated from normal distributions having variance of 1.5% of the assumed parameter values, $\boldsymbol{\beta}_o$, for $\boldsymbol{D}_o^* = (163.24, 68.26, 90.88, 88.21, 88.21, 86.34, 88.21)^T$.

- $\overline{\boldsymbol{D}_t^*} = \frac{1}{N_{sim}} \sum_{k=1}^{N_{sim}} \boldsymbol{D}_t^{*(k)}$, which gives the average of the optimal dose regimens based on the simulated parameters. Similarly, $\overline{\varphi}_{dev}$ and $\overline{\varphi}_{ref}$ give the average percentage deviation and the average relative efficiency.

- $\widetilde{\boldsymbol{D}}_t^*$ is the median of the N_{sim} optimal dose regimens. Similarly, $\widetilde{\varphi}_{dev}$ and $\widetilde{\varphi}_{ref}$ are the median percentage deviation and the median relative efficiency.

- The coefficient of variation (CV) is the percentage ratio of the standard deviation of the simulated values ($\boldsymbol{D}_t^{*(k)}$, $\varphi_{dev}^{(k)}$ and $\varphi_{ref}^{(k)}$, $k = 1, ..., N_{sim}$) to their arithmetic mean.

Results

Figure 4.9 presents the distribution of the optimal dose regimens based on the simulated vector of parameter values $\boldsymbol{\beta}_t^{(k)}$. The administered dose regimen in every simulation is $\boldsymbol{D}_o^* = (163.24, 68.26, 90.88, 88.21, 88.21, 86.34, 88.21)^T$.

The distributions of the individual optimal doses seem to be centred around the administered doses. The data presented in Table 4.7 also confirm this observation. The mean and median optimal dose regimens are almost equal to the administered dose regimen.

The symmetric distributions of the simulated optimal dose regimens could be attributed to the use of a symmetric, i.e., normal distribution for generation of the PK parameters.

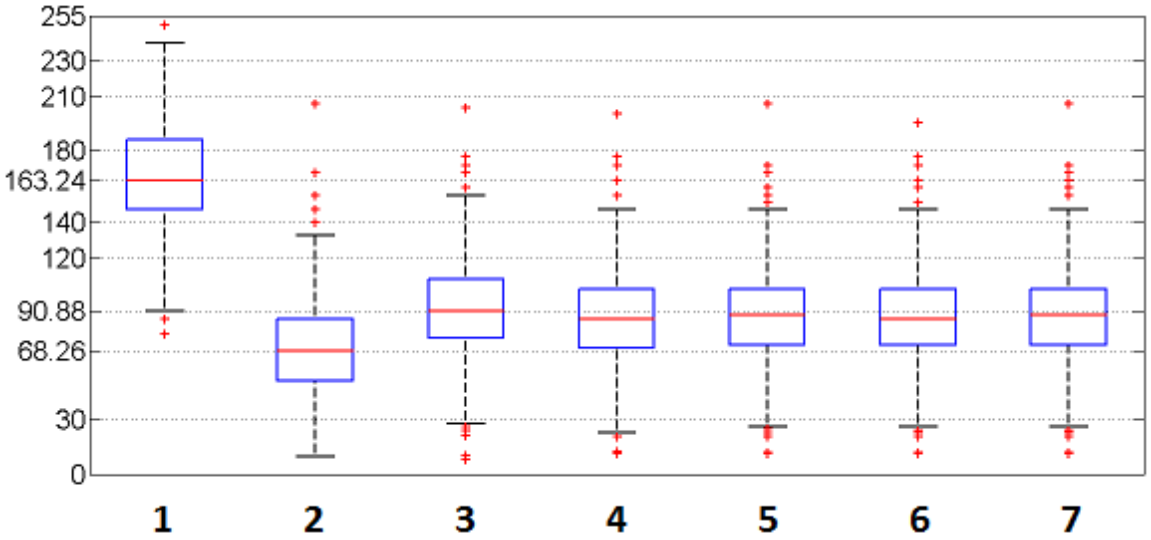


Figure 4.9: Distributions of the optimal doses d_{ti}^*, $i = 1, ..., 7$, based on the simulated parameters $\boldsymbol{\beta}_t^{(k)}$, $k = 1, ..., 1000$.

Statistic	d_1	d_2	d_3	d_4	d_5	d_6	d_7
$\overline{\boldsymbol{D}_t^*}$	167.26	69.57	92.58	87.87	88.91	88.80	88.82
$\widetilde{\boldsymbol{D}}_t^*$	163.24	68.26	90.88	86.34	88.21	86.34	88.21
CV	19.02	38.32	28.22	29.52	29.09	29.34	29.18
$\boldsymbol{D}_o^*$	163.24	68.26	90.88	88.21	88.21	86.34	88.21

Table 4.7: Mean, median and the CV of the doses in the optimal dose regimens based on the simulated parameters $\boldsymbol{\beta}_t^{(k)}$. The administered dose regimen, $\boldsymbol{D}_o^*$, has been added to facilitate comparisons.

Table 4.8 presents the average, median, CV and the IQR for φ_{dev} and φ_{ref}. From the table, it can be seen that the average relative efficiency of the administered dose regimen on account of model misspecification is about 50%. In other words, on account of model misspecification described above, the average under- and over-exposure experienced by a typical subject is almost twice of what is experienced when the model is specified correctly.

Figure 4.10 presents the percentage distribution of φ_{ref} for the N_{sim} simulations. It can be seen from the figure that in about 13% of the simulations, the administered dose regimen is at least 90% as efficient with respect to the optimal dose regimen whereas in about 5% of the simulations, the relative efficiency is less than 10%.

Now, let us consider three scenarios for further analysis of the simulated data. According to the relative efficiency measure, φ_{ref}, we define Scenarios I, II and III as follows: Scenario I is when the relative efficiency is at least 80%, i.e., $\varphi_{ref} \geq 80\%$,

Deviations		Efficiency	
Statistic	Value	Statistic	Value
$\overline{\varphi}_{dev}$	202.61	$\overline{\varphi}_{ref}$	52.53
$\widetilde{\varphi}_{dev}$	102.85	$\widetilde{\varphi}_{ref}$	49.30
CV	175.92	CV	51.15
IQR	188.93	IQR	44.95

Table 4.8: Average, median, CV and the IQR of φ_{dev} and φ_{ref} for the N_{sim} simulations of $\boldsymbol{\beta}_t^{(k)} \sim \mathcal{N}_4(\boldsymbol{\beta}_o, \boldsymbol{\Omega})$ where $\boldsymbol{\beta}_o = (.37, .2, 24, .95)^T$ and $\boldsymbol{\Omega} = \text{diag}(.0056, .003, .36, .0143)$.

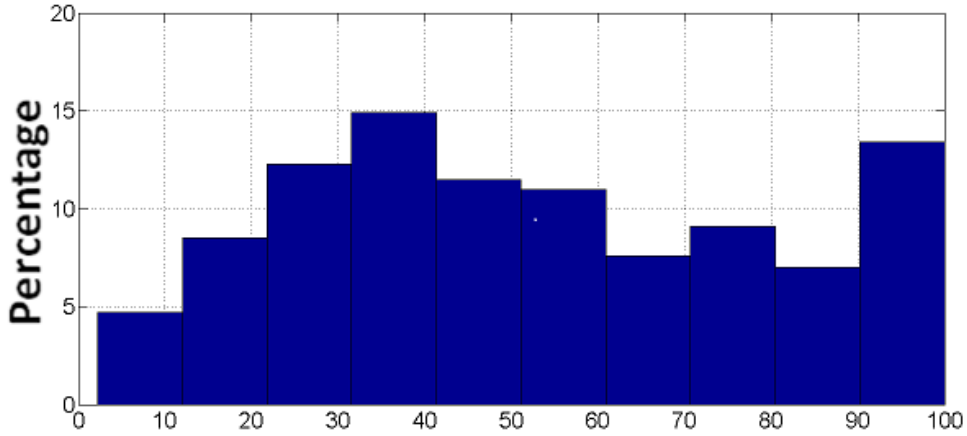


Figure 4.10: Percentage distribution of $\varphi_{ref}^{(k)}$, $k = 1, ..., N_{sim}$.

Scenario II is such that $50\% \leq \varphi_{ref} < 80\%$ and Scenario III when $\varphi_{ref} < 50\%$.

Table 4.9 shows the percentages of the simulations for which the scenarios occurred, average, median and the CV of φ_{ref} for these three scenarios.

Statistic	I	II	III
Percent of N_{sim}	20.8	28.2	51.0
$\overline{\varphi}_{ref}$	91.74	64.37	30.0
$\widetilde{\varphi}_{ref}$	92.46	63.03	31.58
CV	6.21	13.87	40.12

Table 4.9: Average, median and the CV of φ_{ref} for the three scenarios. Scenario I is when $\varphi_{ref} \geq 80\%$, Scenario II is such that $50\% \leq \varphi_{ref} < 80\%$ and Scenario III is when $\varphi_{ref} < 50\%$.

As expected, the average relative efficiency of the administered dose regimen is highest for Scenario I followed by Scenario II and Scenario III. Figure 4.11 presents the distribution of φ_{ref} for the three scenarios.

The PK model considered in this section depends on four parameters. The dose regimen $\boldsymbol{D}_o^*$ was determined to be optimal for the parameters contained in $\boldsymbol{\beta}_o$. It

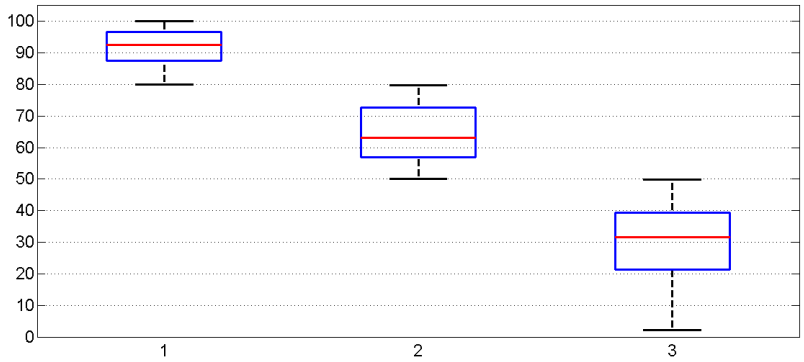


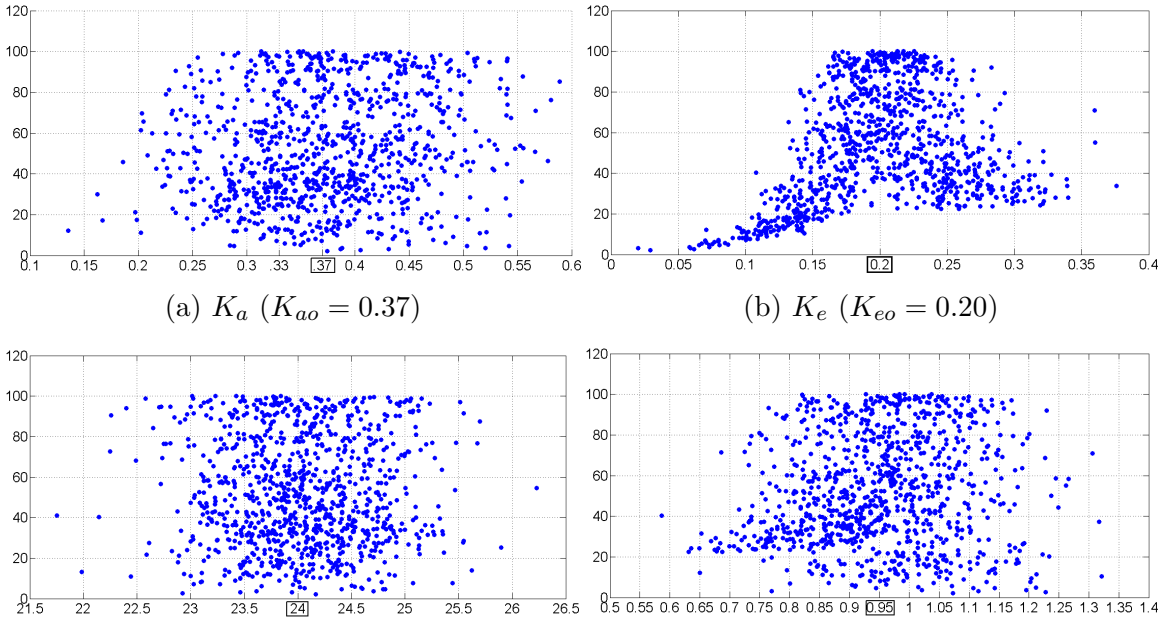
Figure 4.11: Distribution of φ_{ref} for the three scenarios.

is obvious that if a single parameter is misspecified, $\boldsymbol{D}_o^*$ will cease to be optimal and greater the degree of misspecification, the larger will be the deviation from the optimal dose regimen. However, when all the four parameters are misspecified to varying degrees, the effect on optimality is not straightforward to understand. This is what we explore now.

Since we are sampling $\boldsymbol{\beta}_t$ from a four-dimensional normal distribution, it is not obvious which combinations of the parameter values give large and which give small values of the inefficiency. Also, the induced variability in the PK parameters results in large variation in the simulated concentration profiles. To examine the effects of the parameter misspecification, we first evaluate the relationship between the efficiency of the administered dose regimen and the misspecification in individual PK parameters. For this, in Figure 4.12, we plotted φ_{ref} against each of the simulated true parameters contained in $\boldsymbol{\beta}_t$. The parameters contained in $\boldsymbol{\beta}_t$ are drawn from the respective normal distributions as explained earlier.

The plots give interesting insights into how the values assumed by the four parameters influence φ_{ref}. We would expect that large values of K_a, the absorption rate constant, result in small values of φ_{ref} on account of the over-exposure that occurs because of rapid absorption of the drug. However, it can be seen from the figure that even for some large values of K_a, the relative efficiency is quite high. This behaviour can also be seen in the plots for the other three parameters. This points to the possibility that the effect of misspecification in one parameter may get counterbalanced by misspecification in other parameters thereby resulting in minimal net effect on the relative efficiency of the administered dose regimen. Out of the four parameters, values of the elimination rate constant, K_e show the most definitive relationship with φ_{ref}.

It is, therefore, pertinent to study the joint effect of the parameters' values on the relative efficiency. Let us first consider the parameters pairwise. Now, the parameters



(a) K_a ($K_{ao} = 0.37$) (b) K_e ($K_{eo} = 0.20$)

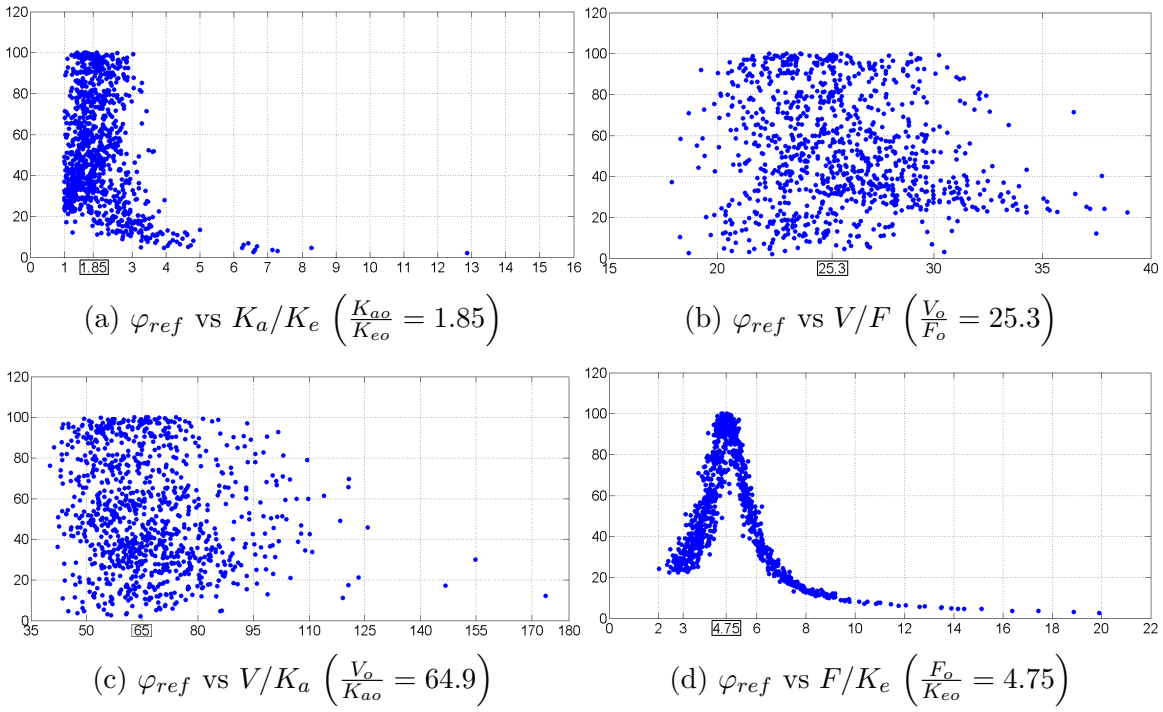
(c) V ($V_o = 24$) (d) F ($F_o = 0.95$)

Figure 4.12: Plots of φ_{ref} (y - axis) and the simulated PK parameters.

V and F behave like scaling parameters in the assumed PK model, as can be seen in Equation (4.7). Large value of F, which comes in the numerator, increases exposure to the drug while large values of V, which comes in the denominator, decrease the exposure to the drug. Therefore, a small ratio of V/F should, in general, result in excessive exposure to the drug while a large ratio should cause under-exposure. Also, the concentration achieved generally increases with increase in K_a and decreases with an increase in K_e. Figure 4.13 plots the relative efficiency against four ratios of the parameters: K_a/K_e, V/F, V/K_a and F/K_e.

These plots are more informative and helpful than the plots in Figure 4.12 in explaining the effect of the parameter values on the relative efficiency of the administered dose regimen. For example, in Figure 4.13a, the relative efficiency can take almost any value if the ratio of K_a to K_e is around the assumed value of 1.85. However, it declines rapidly with increasing values of (K_a/K_e). Similar inference can be drawn from Figures 4.13b and 4.13c, but there is less influence of high ratios shown on the efficiency values.

In Figure 4.13d, however, the pair (F, K_e) seem to explain the variability in the relative efficiency of the simulated dose regimens quite well. For values of the ratio F/K_e less than the assumed value (4.75), the relative efficiency is low, because of under-exposure to the drug due to small bioavailability and slow rate of elimination. The relative efficiency is highest near the assumed value of the ratio and then declines with increasing values on account of over-exposure to the drug due to high bioavailability and slow rate of elimination. This figure reinforces importance of K_e already seen in Figure 4.12.



(a) φ_{ref} vs K_a/K_e $\left(\frac{K_{ao}}{K_{eo}} = 1.85\right)$

(b) φ_{ref} vs V/F $\left(\frac{V_o}{F_o} = 25.3\right)$

(c) φ_{ref} vs V/K_a $\left(\frac{V_o}{K_{ao}} = 64.9\right)$

(d) φ_{ref} vs F/K_e $\left(\frac{F_o}{K_{eo}} = 4.75\right)$

Figure 4.13: Plots of φ_{ref} (y - axis) and ratios of simulated PK parameters.

To further assess the effects of parameter values on the relative efficiency of the administered dose regimen, it is helpful to see the dependence of φ_{ref} on two ratios simultaneously. This is done in Figure 4.14, which depicts the dependence of φ_{ref} on the ratios F/K_e and V/K_a.

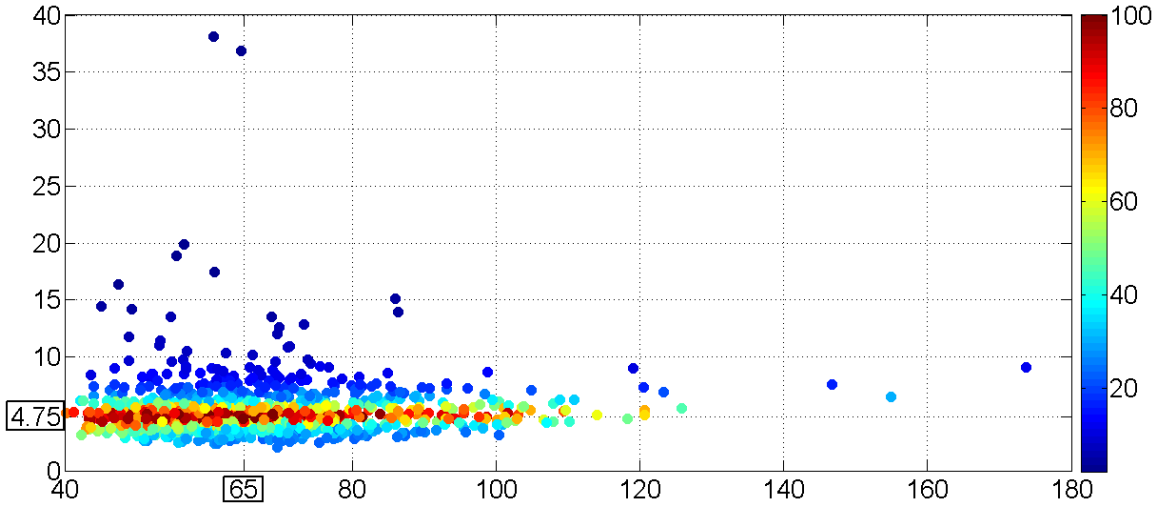


Figure 4.14: Dependence of φ_{ref} on the ratios F/K_e (y-axis) and V/K_a (x-axis) - complete graph. The colour scheme represents φ_{ref}.

For better clarity of the plot, Figure 4.15 presents the same data but over truncated ranges of the ratios F/K_e and V/K_a, since the bulk of the data points lie in these ranges.

Interesting observations can be made from this plot. Most dose regimens for which

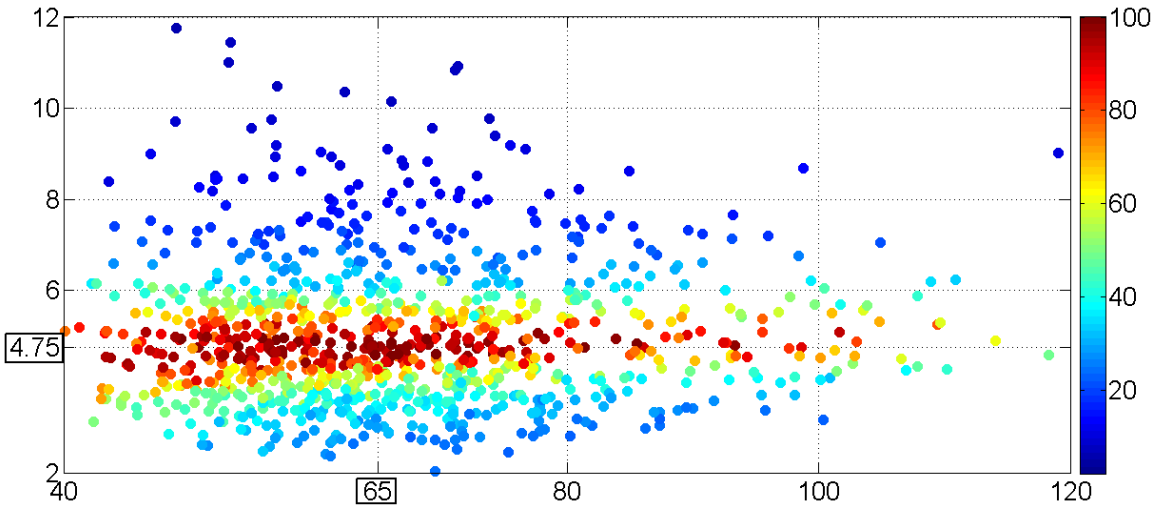


Figure 4.15: Dependence of φ_{ref} on the ratios F/K_e (y-axis) and V/K_a (x-axis) - truncated graph. The colour scheme represents φ_{ref}.

the relative efficiency, φ_{ref}, is at least 80% have the ratio F/K_e close to the assumed value of 4.75. For dose regimens having the ratio F/K_e far from the assumed value, the relative efficiency is quite poor, in many cases even below 20%. Furthermore, even if the ratio F/K_e for a dose regimen is close to the assumed value, for very large departures of the ratio V/K_a from the respective assumed value, the relative efficiency can still be low, since a large V/K_a ratio will result in significant under-exposure to the drug. However, small ratios V/K_a do not affect the efficiency so much as long as F/K_e is close to the assumed one. The somewhat parallel pattern of the efficiencies indicates high sensitivity of φ_{ref} to misspecification of F/K_e irrespective of the value of V/K_a. From the figure, it is clear that for mild misspecification in the ratio F/K_e, the relative efficiency remains high (up to 80%) for a large range of values of V/K_a.

Therefore, the optimal dose regimen computed by the ED algorithm is reasonably robust against mild misspecification of the PK model parameters. Moderate to major misspecification, however, may result in poor efficiency of the administered dose regimen. As observed in this example, the misspecification in parameters may sometimes balance out so that the net effect on the relative efficiency of the administered dose regimen is negligible. This means that even major misspecification in the parameters may lead to the administered dose regimen being reasonably efficient whereas, in some cases, even moderate misspecification can severely reduce the administered dose regimen's efficiency. The robustness of the ED algorithm is, therefore, dependent not only on the degree of misspecification but also on the parameters which get misspecified. Careful analysis of the dependencies may help in understanding the sensitivity of the drug concentration to patient's characteristics represented by model parameters. The PK model is, therefore, an important determinant of how the misspecification in parameters will affect the efficiency of the administered dose regimen.

4.4.2 Efficient Dose Regimen for a Drug Combination Unit

We now consider the case of a fixed dose combination unit (e.g., a tablet or a capsule) which consists of two drugs: A and B. Each of the drugs will have their own PK profile and target concentration but will be administered in a fixed proportion together in one tablet. Let the ratio of drug B to drug A in the tablet be θ. The problem then is two-fold: Find the optimal ratio θ^* in which the drugs should be combined and find the most efficient dose regimen for the tablet.

Let C_{tgt}^A and C_{tgt}^B be the target concentrations for drug A and drug B and let $\boldsymbol{D}^A = (d_1^A, ..., d_n^A)^T$ and $\boldsymbol{D}^B = (d_1^B, ..., d_n^B)^T$ be the dose vectors of drug A and drug B such that $d_i^B/d_i^A = \theta$, for all $i = 1, ..., n$, i.e., the dosing unit always has the same ratio of drug B and drug A. The dose vector for the tablet will be $\boldsymbol{D}^A + \boldsymbol{D}^B = \boldsymbol{D}^A(1+\theta)$. The treatment duration of both the drugs is same and it is denoted by T. However, in case the drugs need to be maintained for different duration of time, the algorithm can still be applied.

The algorithm assumes that the drugs A and B do not interact with each other, i.e., when administered together, their pharmacological action is independent of each other. It is possible to adapt the algorithm when interaction effect is present. The example we discuss in this thesis, however, assumes no interaction between the partner drugs.

We minimise a linear combination denoted by $\varphi^C(\boldsymbol{\Delta})$ of efficiency criterion for drug A ($\varphi^A(\boldsymbol{\Delta})$) and of drug B ($\varphi^B(\boldsymbol{\Delta})$), that is

$$\vartheta(\mathcal{R}|n, \boldsymbol{t}, \boldsymbol{T}) = \varphi^C(\boldsymbol{\Delta}) = \omega\varphi^A(\boldsymbol{\Delta}) + (1-\omega)\varphi^B(\boldsymbol{\Delta}),$$

where ω is a weighting constant which can be used to adjust the importance to be attached to one drug in comparison to the other.

If the concentration profiles of the two drugs are not in the same units, the difference in the magnitude can lead to one drug dominating the other. For example, if C_{tgt}^A is 10 μg/L and C_{tgt}^B is 5 mg/L, the combined equation should be adjusted as $\varphi^C(\boldsymbol{\Delta}) = 1000 \times \omega\varphi^A(\boldsymbol{\Delta}) + (1-\omega)\varphi^B(\boldsymbol{\Delta})$ (as 1 mg/L = 1000 μg/L) so that the areas of under- and over-exposure of the two drugs are comparable.

The extension of the ED algorithm for this problem is called the Extended ED (EED) algorithm and is explained below.

The EED algorithm extends the ED algorithm to find optimal ratio θ^* in which the two drugs should be combined and the optimal dose regimen of the tablet. In each iteration of the EED algorithm, a *ratio set* for θ is created, the ED algorithm

is run and optimal doses found for each of the three values of the ratio in the ratio set. The ratio which results in the least value of $\varphi^C(\boldsymbol{\Delta})$ is then used for creating a new ratio set around it. The working of the EED algorithm is quite similar to the ED algorithm, however, at each iteration of the EED algorithm, the ED algorithm is run completely.

Let θ_{max} and θ_{min} denote the maximum and the minimum permissible values of the ratio θ. Furthermore, let $\Theta^k = \{\theta_1^k, \theta_2^k, \theta_3^k\}$ be the ratio set of possible values of the combination ratio, where k is the iteration number, $k = 1, 2, ..., w$. Below we present the iterations of the EED algorithm.

Iteration No. 1

We assume that some preliminary information about a possible value of θ is available. We use this information to choose the initial values θ_i^1, $i = 1, 2, 3$. In case no information is available about θ, we let $\Theta^1 = \{\theta_{min}, 1, \theta_{max}\}$.
For each of the θ_i^1s, the ED algorithm is run to find best dose vector $\boldsymbol{D}^A$ of drug A.

At each iteration of the ED algorithm,

- We set $\boldsymbol{D}^B = \theta_i^1 \boldsymbol{D}^A$.
- We compute $\varphi^C(\boldsymbol{\Delta}) = \omega\varphi_A^A(\boldsymbol{\Delta}) + (1-\omega)\varphi_A^B(\boldsymbol{\Delta})$.
- We check the stopping rule given in Equation (4.3) on $\boldsymbol{D}^A$ to find the optimum dose regimen, $\boldsymbol{D}^{A^*}$ conditional on θ_i^1.

The optimum dose regimen for drug B, conditional on θ_i^1, is then $\boldsymbol{D}^{B^*} = \theta_i^1 \boldsymbol{D}^{A^*}$.

The ratio θ_i^1 for which $\varphi^C(\boldsymbol{\Delta}) = \omega\varphi_A^A(\boldsymbol{\Delta}) + (1-\omega)\varphi_A^B(\boldsymbol{\Delta})$ is the smallest is then used to create a ratio set around it in the next iteration of the EED algorithm.

Iteration No. k

In iteration k, θ selected in the previous iteration is matched with its ratio set. The ratio set $\boldsymbol{\Theta}^k$ is constituted based on this matching. For example, if θ_3^{k-1} was selected, $\boldsymbol{\Theta}^k$ will be:

$$\Theta^k = \left\{ \max\left(\theta_{min}, \delta \times \theta_3^{k-1}\right), \theta_3^{k-1}, \min\left(\frac{\theta_3^{k-1}}{\delta}, \theta_{max}\right) \right\}, \tag{4.8}$$

where $\delta \in (0, 1)$ is the resolution. The steps in iteration no. 1 are then repeated using Θ^k in place of Θ^1.

Terminal iteration

The EED algorithm is terminated at the w^{th} iteration if $\theta^w = \theta^{w-1}$, i.e., the algorithm terminates when for the given resolution, no further improvement in θ is possible. The final combination ratio and the efficient dose regimen associated with it are then given as the output from the algorithm.

The EED algorithm was programmed in MATLAB® and the code is given in Appendix D.2.

Example 2: Dose Regimen for Coartem®

Coartem®/Riamet® is made up of two components - Artemether (A) and Lumefantrine (L) in a 1:6 ratio. Artemether is rapidly absorbed and eliminated from the body while lumefantrine is slowly absorbed and retained longer in the body. Artemether rapidly clears the parasites from the body while lumefantrine ensures extermination of residual parasites and prevents recrudescence.

The PK parameters of A and L in adult malaria patients which we use in this example are adapted from Ezzet et al. (1998). The pharmacokinetics of artemether are described by a one compartment model (4.7) with the following parameter values $\widehat{K}_a = .37\,/h$, $\widehat{K}_e = 0.829\,/h$, $\widehat{V} = 217$ L, $\widehat{CL} = 180$ L/h and $\widehat{F} = 0.7$.

The pharmacokinetics of lumefantrine are described by a two compartment model according to the following equation:

$$C(t) = \alpha e^{-\lambda t} + \beta e^{-\mu t} - (\alpha + \beta) e^{-K_a t}, \tag{4.9}$$

where K_a denotes the absorption rate and the constants α, β, λ and μ depend on dose, clearance, volume of distribution, bio-availability and rate constants as shown in Equation 2.7.

The parameter values for lumefantrine are $\widehat{CL} = 15$ L/h, $\widehat{V}_1 = 215$ L, $\widehat{K}_a = .13$ /h, $\widehat{Q} = 13.4$ L/h, $\widehat{V}_2 = 1043$ L and $\widehat{F} = 1$. Here a time dependent change of bioavailability for both drugs and auto-induction of the metabolism of artemether were not taken into account. The standard dose regimen of Coartem® is as follows: it is administered over a period of 3-days for a total of $n = 6$ doses. The dosing time points are: $\boldsymbol{t} = (0, 8, 24, 36, 48, 60)^T$, that is, $\tau_1 = 8$ h, $\tau_2 = 16$, $\tau_3 = \tau_4 = \tau_5 = 12$h. Assuming $T = 72$ h, we have $\tau_6 = 12$ h. Dose vector of artemether is $\boldsymbol{D}^A = (80, 80, 80, 80, 80, 80)^T$ and of lumefantrine is $\boldsymbol{D}^L = (480, 480, 480, 480, 480, 480)^T$ and the combination ratio is $\theta = 6$.

The current dose sizes and the ratio between artemether and lumefantrine were chosen in the 1990s and cure rates exceeding 95% have been observed in multiple

studies including several thousands of patients.

Therefore, for artemether having a short half life, the existing regimen was used to determine the target concentration, C^A_{tgt}, as the average concentration over the entire dosing period of 72 h (τ_6 = 12 h added to the last time of dosing at 60 h). It was calculated from cumulative dose (D_{cum}) which for artemether is 6 × 80 = 480, clearance (CL) and the duration of the treatment period as $C^A_{tgt} = D_{cum}/(72 \times CL)$ = .037 mg/L.

For lumefantrine, which has a longer half life, the target concentration was computed in the same way, but over a time period of 7 days (168 h) as C^L_{tgt} = 1.14 mg/L. Lumefantrine's efficacy is known to depend on its concentration at day 7 being above a threshold level, Ezzet et al. (1998). This motivated us to take the time period for maintenance of the average concentration of L, C^L_{tgt}, as seven days (168 h).

This example should serve as a positive control for our approach, since the dose regimen used to derive the target concentrations is expected to be returned by the extended ED algorithm.

This is an illustrative example, using PK models published in the literature. It should not be interpreted as suggesting that the current dose regimen and the proportions of A and L in Coartem® need to be modified in any way to achieve the clinical cure rates in malaria as observed today.

As derived above, we use C^A_{tgt} = .037 mg/L which is to be maintained for 72 h and C^L_{tgt} = 1.14 mg/L, which is to be maintained for 168 h. The resolution of the algorithm is set at $\delta = 0.99$ and ω at 0.5. Owing to the difference in magnitude of the concentrations of the drugs, we compute the objective function $\varphi^C(\mathbf{\Delta})$ as:

$$\varphi^C(\mathbf{\Delta}) = 1000 \times \omega\varphi^A_A(\mathbf{\Delta}) + (1-\omega)\varphi^L_A(\mathbf{\Delta}).$$

Figure 4.16 shows the concentration profiles resulting from the standard dose regimen along with the respective average concentrations (the targets used in the ED algorithm).

In the same graph, we have also plotted the ratio of concentration of A to the concentration of L. It is interesting to see that although the concentration ratio follows a pattern similar to that of the concentration profile of A, it is markedly higher immediately after the first dose is administered. This is on account of their respective rates of absorption in the body. As mentioned earlier, A is rapidly absorbed while L is absorbed more gradually and has a slower elimination rate than A. Therefore, as L accumulates in the body, the ratio of A to L decreases. The different rates of disposition of A and L is one of the reasons of partnering them into a successful

combination unit. While A rapidly kills parasites after administration, L prevents recrudescence by being retained longer in the body.

A major difference between the dose ratio, θ, and the concentration ratio is that while θ remains constant throughout the treatment, the concentration ratio is a function of time.

For the standard dose regimen, $\varphi^C(\boldsymbol{\Delta})$ defined above is computed as 136.16.

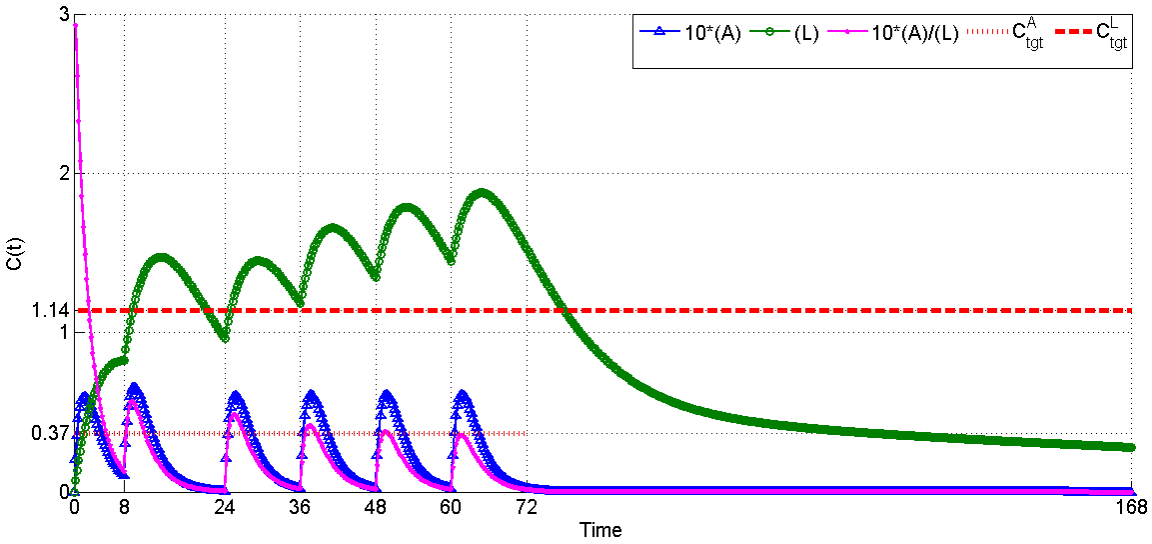


Figure 4.16: Concentration profiles of A and L for the standard dose regimen. The concentration of A and the concentration ratio have been multiplied with 10 to enhance the legibility of the graph.

The EED algorithm was applied to this problem with resolution $\delta = 0.99$. The optimal ratio θ^* was found to be 6.60 with $\boldsymbol{D}_A^* = (65.68,\ 80.00,\ 61.84,\ 61.22,\ 61.22,\ 61.22)^T$, $\boldsymbol{D}_L^* = (424.24,\ 516.74,\ 399.42,\ 395.42,\ 395.42,\ 395.42)^T$ and $\varphi^C(\boldsymbol{\Delta}^*) = 123.98$. The resulting concentration profiles and the concentration ratio of A to L are shown in Figure 4.17. The plot of the concentration ratio is similar to the one corresponding to the standard dose regimen.

Next, the algorithm was run again with one additional constraint: the six doses of each drug were enforced to be same. The optimal ratio θ^* was found to be 6.46 with

$$\boldsymbol{D}_A^* = (61.22,\ 61.22,\ 61.22,\ 61.22,\ 61.22,\ 61.22)^T,$$

$$\boldsymbol{D}_L^* = (403.45,\ 403.45,\ 403.45,\ 403.45,\ 403.45,\ 403.45)^T,$$

and $\varphi^C(\boldsymbol{\Delta}^*) = 124.96$. The resulting concentration profiles and the concentration ratio of A to L are shown in Figure 4.18.

It is interesting to observe that although a larger first dose of L would have further

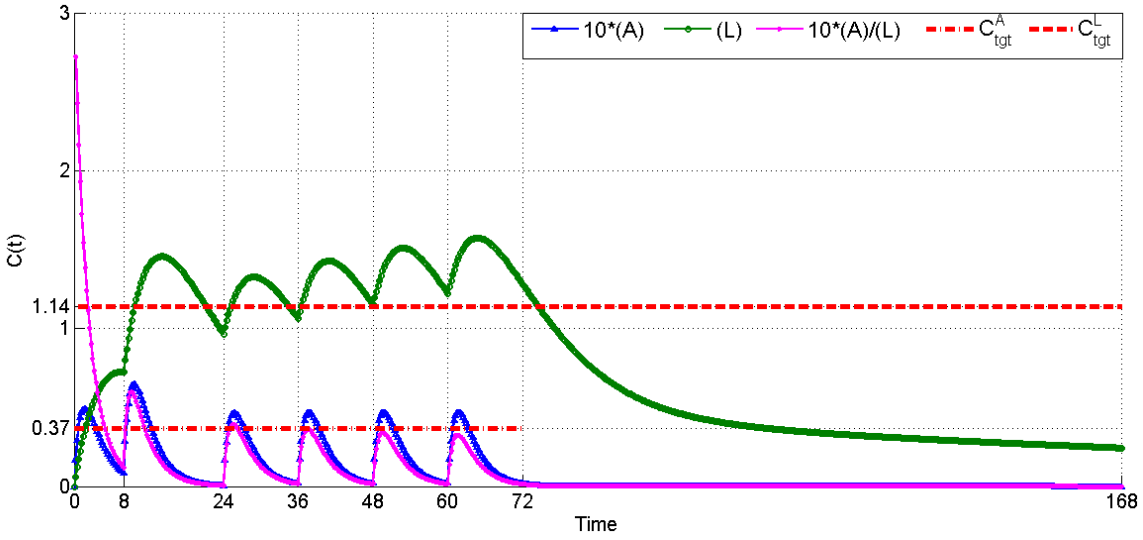


Figure 4.17: Concentration profiles of A and L for the dose regimen computed by the EED algorithm and the raio of concentration of A to that of A. The concentration of A and the concentration ratio have been multiplied with 10 to enhance the legibility of the graph.

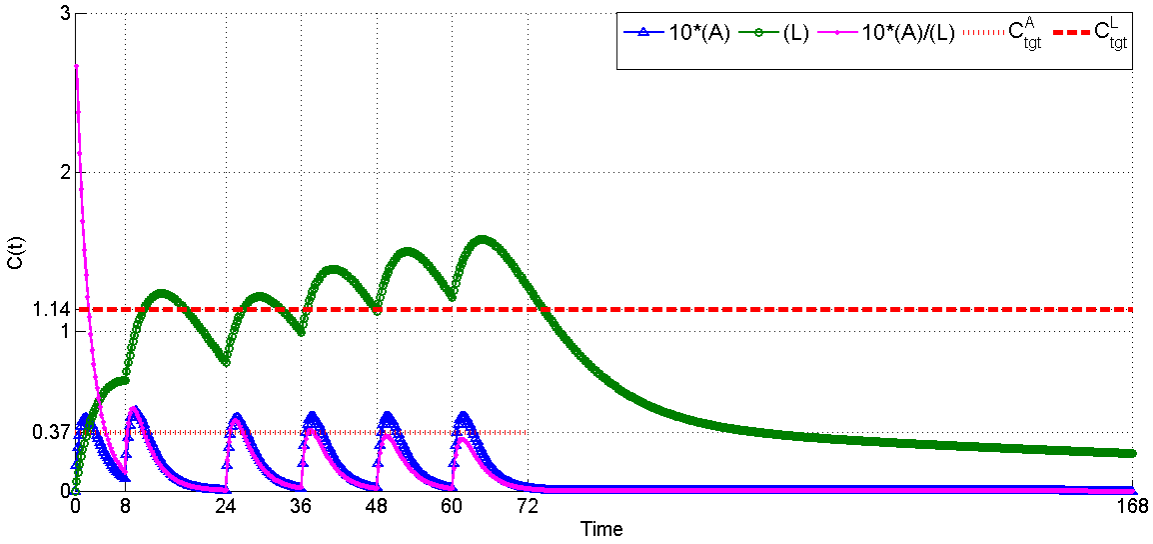


Figure 4.18: Concentration profiles of A and L for the dose regimen computed by the EED algorithm (all doses enforced to be equal). The concentration of A and the concentration ratio have been multiplied with 10 to enhance the legibility of the graph.

decreased $\varphi_A^L(\boldsymbol{\Delta})$, the simultaneous optimisation with A's doses prevents that from taking place.

Similarly, smaller fifth and sixth doses of L would also have decreased $\varphi_A^L(\boldsymbol{\Delta})$ but the simultaneous optimisation ensures that the ratio of doses of A to L, i.e., θ is maintained for all the n doses. It may appear that these doses of L are excessive but they are indeed optimal as a reduction in these doses of L would decrease the exposure to A. Since A and L have different rates of elimination, L tends to accumulate in the

body, in contrast with A which is quickly eliminated.

The standard dose regimen of Coartem® may actually aim at slightly higher targets than what we calculated. This explains the difference between the standard dose regimen and the one returned by the EED algorithm when the doses are constrained to be equal. This example is for illustrative purpose and it underscores the role our methodology can play in setting the combination ratio of combination therapies.

To the best of our knowledge, our approach is one of the first, which extends dose finding into optimisation of the dose ratio for fixed dose combinations. This is of pivotal importance for the development of fixed dose combinations, as preferred for the therapy of malaria, HIV, and tuberculosis. Moreover, the quantitative assessment of the effect of different ratios can help in analysing 'what-if' scenarios. The features of the ED algorithm such as discretisation of doses can be leveraged to accommodate other practical constraints.

4.4.3 Dose Regimens for a Therapeutic Range

Sometimes the interest is in restricting the concentration profile of the drug to a therapeutic range rather then around a target concentration, i.e., instead of C_{tgt}, we have C^+_{tgt} and C^-_{tgt} within which the profile must be maintained. Adherence to one limit may be more important than the other. For example, in case of malaria, the concentration of the drug in the body must exceed a specified C^-_{tgt} for effective extermination of the parasites. Figure 4.19 shows the problem to be solved: find a dose regimen that minimises the exposure to the drug outside the therapeutic range (C^-_{tgt}, C^+_{tgt}) as much as possible.

The ED algorithm can be adapted for this situation as well. The adapted algorithm penalises that dose regimen for which the concentration falls below C^-_{tgt} or increases beyond C^+_{tgt}. The penalty can be equal for both limits or weights can be attached if adherence to one is of greater importance.

The algorithm in this case is the same as the ED algorithm with the difference that the Δ-functions in this case will be defined as follows:

$$\Delta^{\pm}_i(d_1, ..., d_i) = \nu\Delta^-_i(d_1, ..., d_i) \; + \; (1-\nu)\Delta^+_i(d_1, ..., d_i)\,, \quad i = 1, ..., n,$$

where,

$$\Delta^-_1(d_1) = \int_0^{\tau_1} \max(0, C^-_{tgt} - C(t, d_1)) \; \mathrm{d}\, t,$$

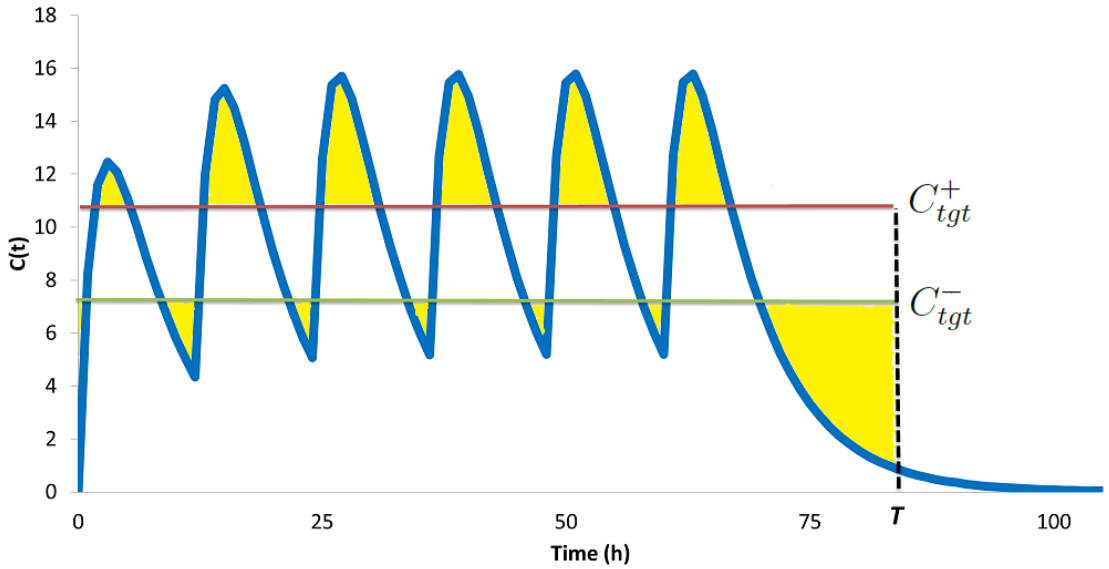


Figure 4.19: The problem is to find a dose regimen which minimises the area of underexposure (below C^-_{tgt}) and overexposure (above C^+_{tgt}) for duration T so that the concentration remains within the therapeutic range of (C^-_{tgt}, C^+_{tgt}) as much as possible.

$$\Delta^+_1(d_1) = \int_0^{\tau_1} \max(0, C(t, d_1) - C^+_{tgt})\ \mathrm{d}\,t,$$

$$\vdots \qquad\qquad \vdots \qquad\qquad \vdots$$

$$\Delta^-_n(d_1, ..., d_n) = \int_0^{\tau_n} \max(0, C^-_{tgt} - C(\tau_1 + ... + \tau_{n-1} + t, d_1) - ... - C(t, d_n))\ \mathrm{d}\,t,$$

$$\Delta^+_n(d_1, ..., d_n) = \int_0^{\tau_n} \max(0, C(\tau_1 + ... + \tau_{n-1} + t, d_1) + ... + C(t, d_n) - C^+_{tgt})\ \mathrm{d}\,t,$$

and $\nu \in (0, 1)$ is a chosen weighing constant. Constant ν close to 1 will ensure that the concentration remains above C^-_{tgt} while ν close to 0 will enable it to be maintained below C^+_{tgt}.

The rationale of the $\Delta^{\pm}$-function defined above is as follows: as long as the drug concentration remains above C^-_{tgt}, the max-function selects 0 and the contribution of this term to the $\Delta^{\pm}$-function is 0. But for those values of t where C^-_{tgt} exceeds the drug concentration, it contributes to the $\Delta^{\pm}$-function, the quantum of which can be controlled by the weighing constant ν. Similarly, as long as the drug concentration remains below C^+_{tgt}, the max-function selects 0 and the contribution of this term to the $\Delta^{\pm}$-function is 0. But for those values of t where the drug concentration exceeds C^+_{tgt}, it contributes positively to the $\Delta^{\pm}$-function.

The relationship between $\Delta^{\pm}$-function and the original Δ-function is expressed below.

Property 1. *If $C^+_{tgt} = C^-_{tgt} = C_{tgt}$ i.e. the therapeutic range is shrunk to a single*

target concentration, then

$$2\Delta_i^{\pm}(.) = \Delta_i(.)\,, \quad \forall \;\; i = 1, 2, \ldots n.$$

Proof. We have,

$$\Delta_1^{\pm}(d_1) = \nu\Delta_1^{-}(d_1) + (1-\nu)\Delta_1^{+}(d_1). \tag{4.10}$$

Given that $C_{tgt}^{+} = C_{tgt}^{-} = C_{tgt}$ and since in the case of a single target concentration, both under- and overexposure are equally minimised, $\nu = 0.5$.

Therefore, equation 4.10 can be written as,

$$\Delta_1^{\pm}(d_1) = \frac{1}{2}\int_0^{\tau_1} \max(0, C_{tgt} - C(t, d_1))\;\mathrm{d}\,t + \frac{1}{2}\int_0^{\tau_1} \max(0, C(t, d_1) - C_{tgt})\;\mathrm{d}\,t.$$

$$\Rightarrow 2\Delta_1^{\pm}(d_1) = \int_0^{\tau_1} \Big\{ \max(0, C_{tgt} - C(t, d_1)) + \max(0, C(t, d_1) - C_{tgt}) \Big\}\;\mathrm{d}\,t.$$

$$\Rightarrow 2\Delta_1^{\pm}(d_1) = \begin{cases} \int_0^{\tau_1} (C_{tgt} - C(t, d_1))\;\mathrm{d}\,t & \text{if } C(t, d_1) \leq C_{tgt} \\ \int_0^{\tau_1} (C(t, d_1) - C_{tgt})\;\mathrm{d}\,t & \text{if } C(t, d_1) \geq C_{tgt} \end{cases}$$

$$\Rightarrow 2\Delta_1^{\pm}(d_1) = \int_0^{\tau_1} |C(t, d_1) - C_{tgt}|\;\mathrm{d}\,t = \Delta_1(d_1).$$

Similarly, the result can be proved for $\Delta_i^{\pm}$, $i = 2, ..., n$. □

The ED algorithm can be applied to select a dose regimen which minimises a function of $\boldsymbol{\Delta}^{\pm}(\boldsymbol{D}) = (\Delta_1^{\pm}, \Delta_2^{\pm}, ..., \Delta_n^{\pm})^T$ in exactly the same way as it was done for a single target concentration.

Example 3: Maintenance Within a Therapeutic Range

Here we consider the drug discussed in Example 1 with $\tau = 6\,h$ and $n = 7$. Suppose instead of $C_{tgt} = 3$ mg/L, we have $C_{tgt}^{+} = 3.5$ mg/L and $C_{tgt}^{-} = 2.5$ mg/L. The MATLAB® code for implementing this method is given in Appendix D.3.

For φ_A-efficiency criterion of resolution 0.99 and $\nu = 0.5$, the dose regimen is

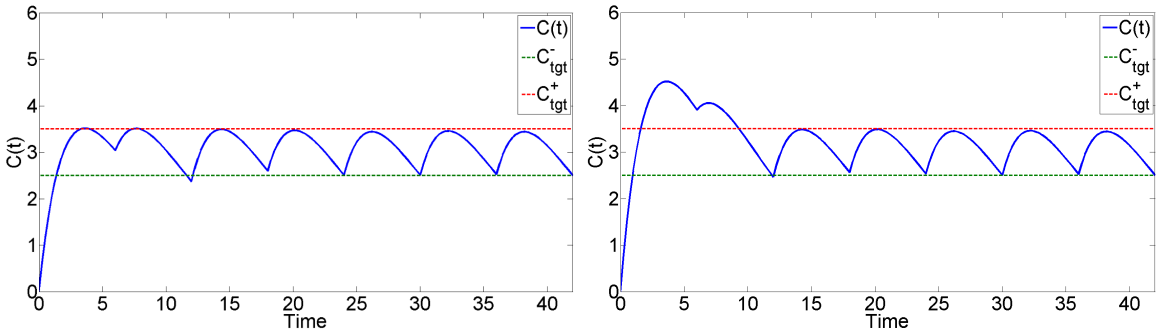
$$\boldsymbol{D} = (183.14,\; 67.71,\; 104.06,\; 91.54,\; 94.34,\; 95.06,\; 93.40)^T$$

and the corresponding concentration profile is shown in Figure 4.20a.

If maintaining a threshold concentration is more important, one might want to completely avoid the dose regimen for which the concentration drops below C^-_{tgt}. In this situation, ν can be increased to, say, 0.95 to reflect more importance attached to C^-_{tgt}. The most efficient dose regimen in this case is given as:

$$\boldsymbol{D} = (235.37,\ 52.07,\ 101.22,\ 94.34,\ 93.40,\ 95.30,\ 93.40)^T.$$

The corresponding concentration profile is shown in Figure 4.20b. As expected, the loading dose in this case is higher since more importance has been given to prevention of underexposure by taking ν close to 1. Because of the higher importance attached to maintenance of C^-_{tgt} by choosing a large value of ν, the larger loading dose causes the concentration profile to breach the upper limit of the target concentration, C^+_{tgt}.



(a) $\nu = 0.5$ (b) $\nu = 0.95$

Figure 4.20: Concentration profiles for different values of ν

From the figures it can be seen that the concentration profile can be maintained in the desired therapeutic range and the relative importance of the two constraints can be adjusted by the choice of ν.

In case of combination therapies, the criterion of maintenance of the concentration within the therapeutic range can be used for either or both of the partner drugs, and the extended ED algorithm can be used to optimise the dose vector and the combination ratio of the fixed dose combination unit.

We end this chapter with a discussion on the limitations of the ED algorithm, along with some possible remedies.

4.5 Limitations of the ED Algorithm and Possible Remedies

Prerequisites for the application of the algorithm are: knowledge of the PK model, availability of the model parameters in the target population, target concentration or therapeutic range for the drug and knowledge about the 'symmetry of risk', i.e., whether over- or under-dosing is more undesirable.

The limitations arise from the lack of knowledge about the correct target, from not accounting for underlying variability in the population and from not accounting for drug-drug interaction, when simultaneously optimising for two drugs. Any iterative or simulation based approach will carry forward the uncertainties of the input.

The user will have to make sure that the pharmacokinetic parameters of the target population match those of the source population, for example, that patient parameters are comparable to those of healthy volunteers, special sub-populations' (children, geriatric patients) PK parameters are comparable to those of general patient populations, or adequate corrections have been performed. The same is true for the target concentrations and therapeutic ranges, which also may differ between populations. In infectious diseases, the concentration range applicable in one geographic area may not be applicable in other areas due to local resistance. These limitations cannot be ameliorated by further work on the algorithm.

In the example related to combination therapies, we have assumed that the partner drugs behave independently of each other when administered together. Drug interactions, especially infra- or supra-additivity will severely impede the validity of the algorithm, when optimising for several drugs. In this case, a change in the dose ratio will lead to a change in the concentration ratio and therefore a different contribution of the interaction term to the overall effect. In infectious diseases, where near optimal concentration profiles of different drugs are targeted regardless of the combination partner (i.e. maximal proximity to EC90), this shortcoming is less relevant than in areas, where trade-offs for toxicity have to be made, i.e., co-administering doses of drugs, which are suboptimal for monotherapy, but effective in combination, in order to improve the adverse effects profile.

The parameter values which would generally be used in the algorithm will be those of a typical individual, i.e., the mean or the median concentration-time profile will be used for optimisation. It is expected that 50% of the population will be above and 50% of the population will fall below this value. It can be envisioned, that a different quantile will be needed to cover certain scenarios. For example, it will be entirely insufficient to have 50% of the population below the minimum inhibitory

concentration (MIC), should this threshold be targeted. Optimisation based on a different quantile of the distribution of concentration-time curves may be preferable, e.g., at least 95% of the concentration time courses above the MIC at any time.

In Chapter 6, we consider the case when the ED algorithm has to be applied without the knowledge of the PK parameters. This is done using an adaptive procedure which also allows assessment of the population variability. This will help in mitigating the limitation of lack of past knowledge to some extent. The foundation of the methodology is based on the theory of non-linear mixed effects models, which is discussed in the next chapter.

Chapter 5

Non-linear Mixed Effects Models - Estimation and Design

As most of the compartmental models describing the concentration-time relationship of a drug are non-linear functions of the PK parameters, the theory of parameter estimation and of design of experiments for non-linear models are particularly useful in PK/PD studies. If the parameters of the non-linear model are assumed to be random variables having a defined density function, the model is then called a 'non-linear mixed effects model'. It is assumed that the underlying mechanistic model, which explains the drug disposition process, is the same for all individuals. However, each individual has their own vector of parameters, which is a realization of a random vector. Such models are hierarchical and are generally described in two stages. The two stages refer to the modelling of the intra- and the inter-individual variability. Such models are commonly used for the population approach, as discussed in Section 2.3.

To extend the scope of the ED algorithm for the case of unknown parameters, we use non-linear mixed effects models. Before discussing the theory of such models, we firstly define a new expression for the drug concentration when multiple doses of a drug are administered. This will help us in synchronising the multiple dose compartmental model with the theory of non-linear mixed effects models.

5.1 Mechanistic Model for Multiple Doses of the Drug

The problem with Equations (2.5) and (2.8) is that for a pre-selected τ, they give the concentration at time t after the n^{th} dose is administered. In other words, the function

C becomes dependent on the dose number for which the subsequent concentration is sought. For computational purposes, it would be desirable to make the function C depend only on the time t at which the concentration is desired and not on a particular dose number. We define such a function as shown below. Let n doses of a drug be given with the corresponding dose vector $\boldsymbol{D} = (d_1, ..., d_n)^T$. Let t_k be the time point at which the k^{th} dose is administered, $k = 1, ..., n$, with the first dose being given at time 0, i.e., $t_1 = 0$. Let $C(t; d, \boldsymbol{\beta}_i)$ be the concentration at time t after a dose d is administered to a patient i, $i = 1, ..., N$, where the parameters $\boldsymbol{\beta}_i$ are unknown constants. When multiple doses are administered, the resulting concentration can be obtained by the principle of superposition as shown in Chapter 2. As discussed in Chapter 2, population approach comprises of drawing inferences about the PK parameters of a population based on the data collected from a cohort of patients. The data are the drug concentrations measured from the blood samples drawn from the subjects at pre-determined time points.

Let the j^{th} observation on subject i be collected at times T_{ij}, $i = 1, ..., N$ and $j = 1, ..., m$. We propose the following form to compute $C(T_{ij}; \boldsymbol{D}, \boldsymbol{\beta}_i)$, the concentration at any time T_{ij} when a dose vector $\boldsymbol{D}$ is administered to patient i:

$$C(T_{ij}; \boldsymbol{D}, \boldsymbol{\beta}_i) = \sum_{k=1}^{n} I_{\{T_{ij} \geq t_k\}} C(T_{ij} - t_k; d_k, \boldsymbol{\beta}_i), \tag{5.1}$$

where $I_{\{T_{ij} \geq t_k\}}$ is an indicator function defined as

$$I_{\{T_{ij} \geq t_k\}} = \begin{cases} 1 & \text{if } T_{ij} \geq t_k \\ 0 & \text{otherwise.} \end{cases} \tag{5.2}$$

Equation 5.1 is a computationally convenient expression for determining the concentration at time T_{ij} without specifying where T_{ij} occurs. If $C(t; d, \boldsymbol{\beta}_i)$ is differentiable with respect to $\boldsymbol{\beta}_i$, this expression preserves that in $C(T_{ij}; \boldsymbol{D}, \boldsymbol{\beta}_i)$, as sum of a finite number of differentiable functions is also differentiable.

If the samples are collected at the same times for all the N subjects in a group, then $T_{ij} = T_j$ for $i = 1, ..., N$ and $j = 1, ..., m$.

Example

Let $n = 4$ doses be administered at time points $\boldsymbol{t} = (t_1, t_2, t_3, t_4)^T = (0, 8, 24, 36)^T$ h and let $m = 3$ blood samples be collected at $(T_{i1}, T_{i2}, T_{i3}) = (6,\ 20,\ 40)$ h for $i = 1, ..., N$ subjects. Then, the drug concentration in the i^{th} subject at sampling time $T_{i1} = 6$ is $C(T_{i1}; \boldsymbol{D}, \boldsymbol{\beta}_i) = C(6; d_1, \boldsymbol{\beta}_i)$,

at $T_{2i} = 20$ it is $C(T_{i2}; \boldsymbol{D}, \boldsymbol{\beta}_i) = C(20; d_1, \boldsymbol{\beta}_i) + C(12; d_2, \boldsymbol{\beta}_i)$,
and at $T_{3i} = 40$ is

$$C(T_{3i}; \boldsymbol{D}, \boldsymbol{\beta}_i) = C(40; d_1, \boldsymbol{\beta}_i) + C(32; d_2, \boldsymbol{\beta}_i) + C(16; d_3, \boldsymbol{\beta}_i) + C(4; d_4, \boldsymbol{\beta}_i).$$

This follows the principle of superposition but unlike the usual multiple dose formula, it gives the drug's concentration at any time.

We now introduce non-linear mixed effects models which are commonly used in pharmacokinetic studies.

5.2 The General Statistical Model

Statistical models are a class of mathematical models which include assumptions regarding generation of sample data from a larger population. They are specified by a set of mathematical equations that relate random variables. Formally, a statistical model is defined as a pair $(\mathcal{S}, \mathcal{P})$, where $\mathcal{S}$ is the sample space, and $\mathcal{P}$ is a set of probability distributions on $\mathcal{S}$, McCullagh (2002). The probabilistic assumptions contained in a statistical model distinguishes it from deterministic models. Also, unlike mechanistic models, the objective of statistical models is to best describe the relationship between the variables without attempting to explain the reasons behind this relationship.

To discuss the ED algorithm in an adaptive trial setting in Chapter 6, we use statistical models, in contrast with Chapter 4 where we used only deterministic models. The statistical models are helpful in expressing not only the variability across the population but also the variability in the observations collected from a single subject. This is discussed below.

STAGE I (Intra-Individual Variability)

Let y_{ij} be the response variable, that is j^{th} observation on the i^{th} subject at the experimental setting x_{ij}. Some commonly used response models are:

Additive error model: $y_{ij} = \eta(x_{ij}, \boldsymbol{\beta}_i) + \epsilon_{ij}$

Exponential error model: $y_{ij} = \eta(x_{ij}, \boldsymbol{\beta}_i) \exp(\epsilon_{ij})$

Proportional error model: $y_{ij} = \eta(x_{ij}, \boldsymbol{\beta}_i)(1 + \epsilon_{ij})$

for $i = 1, ..., N$ and $j = 1, ..., m$ where,

ϵ_{ij} is the observational error associated with y_{ij} such that $\epsilon_{ij} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$,

$\boldsymbol{\beta}_i$ is the $q \times 1$ vector of parameters for the i^{th} subject,

x_{ij} is the experimental setting for the $(i,j)^{th}$ observation ,

m is the number of observations collected from each subject,

$M = Nm$ is the total number of observations collected from the N subjects,

$\eta(.)$ is the regression function, non-linear in some or all elements of the parameter vector $\boldsymbol{\beta}_i$.

For instance, in pharmacokinetics, $\eta(.)$ would be one of the compartmental models described previously and x_{ij} could be the time at which the j^{th} observation is collected on the i^{th} subject.

By a suitable log transformation, the last two models can also be expressed in the form of the additive error model and as such, we will be restricting ourselves to the additive error model throughout the thesis.

For an additive error model, the i^{th} subject's responses can be summarised in vector form as:

$$\boldsymbol{y_i} = \boldsymbol{\eta}(\boldsymbol{x_i}, \boldsymbol{\beta_i}) + \boldsymbol{\epsilon}_i\,, \quad i = 1, ..., N, \tag{5.3}$$

where, $\boldsymbol{y_i} = \begin{pmatrix} y_{i1} \\ \vdots \\ y_{im} \end{pmatrix}$, $\boldsymbol{x_i} = \begin{pmatrix} x_{i1} \\ \vdots \\ x_{im} \end{pmatrix}$, $\boldsymbol{\eta(x_i, \beta_i)} = \begin{pmatrix} \eta(x_{i1}, \boldsymbol{\beta}_i) \\ \vdots \\ \eta(x_{im}, \boldsymbol{\beta}_i) \end{pmatrix}$ and

$\boldsymbol{\epsilon_i} = \begin{pmatrix} \epsilon_{i1} \\ \vdots \\ \epsilon_{im} \end{pmatrix}$, $\boldsymbol{\epsilon}_i \sim \mathcal{N}_m(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$.

We assume in the above model that the variance of the errors, σ^2, is constant across subjects. The examples described in this thesis assume a homoscedastic error structure in this sense.

A heteroscedastic model would be when the variance of the errors is subject specific; however, we do not consider such situation in this work.

STAGE II (Inter-Individual Variability)

In the second stage of the hierarchical model, $\boldsymbol{\beta}_i$ is modelled as a function of mean PK parameters and random effects as follows:

$$\boldsymbol{\beta}_i = \boldsymbol{A}_i\boldsymbol{\beta} + \boldsymbol{B}_i\boldsymbol{b}_i, \;\; i = 1, ..., N, \tag{5.4}$$

where

$\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown constant population parameters,

$\boldsymbol{A}_i$ is a $q \times p$ dimensional matrix of constants or covariates associated with the population parameters relevant to the i^{th} subject,

$\boldsymbol{B}_i$ is a $q \times r$ dimensional matrix used to associate random effects to the relevant parameters,

$\boldsymbol{b_i}$ is the $r \times 1$ vector of random effects such that $\boldsymbol{b}_i \sim \mathcal{N}_r(\boldsymbol{0}, \sigma^2\boldsymbol{\Omega})$ assumed to be independent of the vector of the observational errors $\boldsymbol{\epsilon}_i$.

This notation makes it possible for some parameters of the population to be fixed across subjects by setting their corresponding variance and covariance entry in the matrix $\boldsymbol{\Omega}$ to 0.

The models for the N subjects can be combined in one as

$$\boldsymbol{y} = \boldsymbol{\eta}(\boldsymbol{x}, \boldsymbol{\phi}) + \boldsymbol{\epsilon}, \tag{5.5}$$

where $\boldsymbol{y} = \begin{pmatrix} \boldsymbol{y}_1 \\ \vdots \\ \boldsymbol{y}_N \end{pmatrix}$, $\boldsymbol{\eta(x, \phi)} = \begin{pmatrix} \boldsymbol{\eta}(\boldsymbol{x}_1, \boldsymbol{\beta}_1) \\ \vdots \\ \boldsymbol{\eta}(\boldsymbol{x}_N, \boldsymbol{\beta}_N) \end{pmatrix}$, $\boldsymbol{\phi} = \boldsymbol{A\beta} + \boldsymbol{Bb}$, $\boldsymbol{A} = \begin{pmatrix} \boldsymbol{A}_1 \\ \vdots \\ \boldsymbol{A}_N \end{pmatrix}$,

$\boldsymbol{B} = \text{diag}(\boldsymbol{B}_1, ..., \boldsymbol{B}_N)$, $\boldsymbol{b} = \begin{pmatrix} \boldsymbol{b}_1 \\ \vdots \\ \boldsymbol{b}_N \end{pmatrix}$ and $\boldsymbol{\epsilon} = \begin{pmatrix} \boldsymbol{\epsilon}_1 \\ \vdots \\ \boldsymbol{\epsilon}_N \end{pmatrix}$.

We assume that $\boldsymbol{\epsilon} \sim \mathcal{N}_M(\boldsymbol{0}, \sigma^2\boldsymbol{I})$ and $\boldsymbol{b} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2\widetilde{\boldsymbol{\Omega}})$, where $\widetilde{\boldsymbol{\Omega}} = \text{diag}(\boldsymbol{\Omega}, ..., \boldsymbol{\Omega})$.

The distribution of $\boldsymbol{y}$, conditional on the vector of random effects $\boldsymbol{b}$, can then be expressed as

$$\boldsymbol{y}|\boldsymbol{b} \sim \mathcal{N}_M(\boldsymbol{\eta}(\boldsymbol{x}, \boldsymbol{\phi}), \sigma^2\boldsymbol{I}), \tag{5.6}$$

i.e.,

$$\boldsymbol{y} - \boldsymbol{\eta}(\boldsymbol{x}, \boldsymbol{\phi}) \,|\, \boldsymbol{b} \sim \mathcal{N}_M(\boldsymbol{0}, \sigma^2 \boldsymbol{I}). \tag{5.7}$$

Another common form of expressing the model parameters is $\boldsymbol{\beta}_i = \boldsymbol{g}(\boldsymbol{\beta}, \boldsymbol{b}_i)$, where $\boldsymbol{g}$ is a vector of linear or non-linear functions of the population parameters and the random effects.

Example - Multiple Doses

We consider the above defined model in the context of the multiple dose model we defined in Equation 5.1.

We have a cohort of size n_c in which every subject is administered n doses of a particular drug. The form of the compartmental model is known but the model parameters are to be estimated. To accomplish this, m blood samples are taken from each of the n_c subjects at times T_j, $j = 1, ..., m$ (for all $i = 1, ..., n_c$, so index i is dropped). We consider the exponential error model. Then, stage 1 model is:

$$y_{ij} = \left(\sum_{k=1}^{n} I_{\{T_j \geq t_k\}} C(T_j - t_k; d_k, \boldsymbol{\beta}_i) \right) \exp(\epsilon_{ij}), \tag{5.8}$$

where $i = 1, ..., n_c$ and $j = 1, ..., m$. The error terms are independently distributed as $\epsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$. Consequently, the conditional distribution of y_{ij} given b_i is lognormal. This is consistent with the non-negativity of drug concentrations.

Taking natural logarithm of both sides we get:

$$\ln(y_{ij}) = \ln \left(\sum_{k=1}^{n} I_{\{T_j \geq t_k\}} C(T_j - t_k; d_k, \boldsymbol{\beta}_i) \right) + \epsilon_{ij}.$$

The above model is of the form of the additive error model. Here, however, for a given value of the random vector $\boldsymbol{b}_i$ the ln of the response variable is normally distributed.

Let us take

$$C(t; d, \boldsymbol{\beta}) = \frac{dK_a}{V(K_a - K_e)} (e^{-K_e t} - e^{-K_a t}). \tag{5.9}$$

Now, for the second stage model, we model the three PK parameters using a lognormal distribution for $\boldsymbol{\beta}_i = (K_{ai}, K_{ei}, V_i)^T$. We assume that $K_{a_i} = K_a \exp(b_{i1})$, $K_{e_i} = K_e \exp(b_{i2})$ and $V_i = V \exp(b_{i3}) w_i^{\beta_4}$, where w_i is the body weight of the i^{th} subject, the only covariate considered here. As discussed in Section 2.3, inclusion of covariates such as bodyweight or age in the model leads to a reduction in the residual

variability.

Taking natural logarithm on both sides we obtain:

$$\ln(K_{a_i}) = \ln(K_a) + b_{i1}, \tag{5.10}$$

$$\ln(K_{e_i}) = \ln(K_e) + b_{i2}, \tag{5.11}$$

$$\ln(V_i) = \ln(V) + \beta_4 \ln(w_i) + b_{i3}, \tag{5.12}$$

This system can be further expressed in the vector notation as:

$$\begin{pmatrix} \ln(K_{a_i}) \\ \ln(K_{e_i}) \\ \ln(V_i) \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & \ln(w_i) \end{pmatrix} \begin{pmatrix} \ln(K_a) \\ \ln(K_e) \\ \ln(V) \\ \beta_4 \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} b_{i1} \\ b_{i2} \\ b_{i3} \end{pmatrix}, \tag{5.13}$$

where $\boldsymbol{b}_i = (b_{i1}, b_{i2}, b_{i3})^T \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega})$ and $\boldsymbol{\Omega} = \text{diag}(\omega_1, \omega_2, \omega_3)$.

This is of the form: $\boldsymbol{\beta}_i = \boldsymbol{A}_i\boldsymbol{\beta} + \boldsymbol{B}_i\boldsymbol{b}_i$, $i = 1, .., n_c$, where vector $\boldsymbol{\beta}$ of the population parameters includes the log-transformed model parameters and the additional one corresponding to the covariate.

The distributions of the PK parameters in the above example are lognormal. For example, for the absorption rate constant K_{a_i}, since $\ln(K_{a_i}) \sim \mathcal{N}(\ln(K_a), \omega_1)$, we have $K_{a_i} \sim \ln\mathcal{N}(\ln(K_a), \omega_1)$.

Consequently,

$$\text{E}(K_{a_i}) = \exp\left\{\ln(K_a) + \frac{\omega_1}{2}\right\} = K_a e^{\omega_1/2} \text{ and}$$

$$\text{Var}(K_{a_i}) = \exp\{2\ln(K_a) + \omega_1\}\,(e^{\omega_1} - 1) = K_a^2 e^{\omega_1}\,(e^{\omega_1} - 1)\,.$$

Similarly, $K_{e_i} \sim \ln\mathcal{N}(\ln(K_e), \omega_2)$. We therefore have $\text{E}(K_{e_i}) = K_e e^{\omega_2/2}$ and $\text{Var}(K_{e_i}) = K_e^2 e^{\omega_2}\,(e^{\omega_2} - 1)\,.$

For the volume parameter, we have

$$\ln(V_i) \sim \mathcal{N}(\ln(V w_i^{\beta_4}), \omega_3) \iff V_i \sim \ln\mathcal{N}(\ln(V w_i^{\beta_4}), \omega_3).$$

Thus, $\text{E}(V_i) = V w_i^{\beta_4} e^{\omega_3/2}$ and $\text{Var}(V_i) = V^2 w_i^{2\beta_4} e^{\omega_3}\,(e^{\omega_3} - 1)\,.$

A usual practice when the bodyweight w_i is used in a PK model, as for example, in Equation (5.12) is to use (w_i/w_{std}) instead of w_i, where w_{std} is taken as the standardised weight such that the percentages of people having bodyweight greater than

and less than w_{std} are approximately equal. In adult trials, w_{std} is generally taken to be 70 Kg. This ratio is used to scale the subjects' bodyweights with reference to the standardised weight.

The lognormal distribution is the preferred choice for modelling pharmacokinetic parameters. There are mainly two reasons for this. Firstly, the PK parameters are non-negative. Secondly, the distributions of PK parameters, like of most biological data, have been observed to be positively skewed, thus making them suitable to be modelled by the log-normal distribution.

Having embedded the multiple dose model into the context of non-linear mixed effects models, our next problem is to estimate the model parameters as precisely as possible. This will require blood sampling at optimal time points. In the rest of this chapter, we discuss the methods and the software available to solve these twin problems of parameters' estimation and optimal design.

5.3 Estimation of the Model Parameters

We assume that the variance-covariance matrix, $\boldsymbol{\Omega}$, of the random effects is unknown. Let the column vector $\boldsymbol{\omega}$ consist of the unique elements of the variance-covariance matrix $\boldsymbol{\Omega}$. We are interested in estimation of the vector $\boldsymbol{\Psi} = (\boldsymbol{\beta}^T, \boldsymbol{\omega}^T, \sigma^2)^T$. In this section we present a brief overview of the most commonly used methods.

5.3.1 Method of Maximum Likelihood

The likelihood function for the N subjects, based on the assumption of subjects' independence, is given as:

$$L(\boldsymbol{\Psi}|\boldsymbol{y}) = \prod_{i=1}^{N} p(\boldsymbol{y}_i|\boldsymbol{x}_i; \boldsymbol{\Psi}), \tag{5.14}$$

where $p(\boldsymbol{y}_i|\boldsymbol{x}_i; \boldsymbol{\Psi})$ is the marginal distribution of $\boldsymbol{y}_i|\boldsymbol{x}_i$ obtained from integrating out the random effects $\boldsymbol{b}_i$, i.e.,

$$p(\boldsymbol{y}_i|\boldsymbol{x}_i; \boldsymbol{\Psi}) = \int p(\boldsymbol{y}_i, \boldsymbol{b}_i|\boldsymbol{x}_i; \boldsymbol{\Psi}) \, \mathrm{d}\, \boldsymbol{b}_i \tag{5.15}$$

$$= \int p(\boldsymbol{y}_i|\boldsymbol{b}_i, \boldsymbol{x}_i; \boldsymbol{\Psi}) p(\boldsymbol{b}_i|\boldsymbol{x}_i; \boldsymbol{\Psi}) \, \mathrm{d}\, \boldsymbol{b}_i. \tag{5.16}$$

We assume that the random effects $\boldsymbol{b}_i$ are independent of the design variables $\boldsymbol{x}_i$ which results in,

$$p(\boldsymbol{y}_i|\boldsymbol{x}_i;\boldsymbol{\Psi}) = \int p(\boldsymbol{y}_i|\boldsymbol{b}_i,\boldsymbol{x}_i;\boldsymbol{\Psi})p(\boldsymbol{b}_i;\boldsymbol{\Psi})\,\mathrm{d}\,\boldsymbol{b}_i. \tag{5.17}$$

The likelihood function is thus expressed as

$$L(\boldsymbol{\Psi}|\boldsymbol{y}) = \prod_{i=1}^{N}\int p(\boldsymbol{y}_i|\boldsymbol{b}_i,\boldsymbol{x}_i;\boldsymbol{\Psi})p(\boldsymbol{b}_i;\boldsymbol{\Psi})\,\mathrm{d}\,\boldsymbol{b}_i. \tag{5.18}$$

Further, let $\boldsymbol{\Omega}$ be parametrized by an $r \times r$ matrix $\boldsymbol{\gamma}$ such that $\boldsymbol{\Omega} = (\boldsymbol{\gamma}^T\boldsymbol{\gamma})^{-1}$. If $\boldsymbol{\Omega}$ is positive definite, such a $\boldsymbol{\gamma}$ will exist but may not be unique. One possible $\boldsymbol{\gamma}$ is the Cholesky factor of $\sigma^2\boldsymbol{\Omega}^{-1}$.

Now, as we have assumed that $\boldsymbol{b}_i \sim \mathcal{N}_r(\boldsymbol{0},\sigma^2\boldsymbol{\Omega})$, where $\boldsymbol{\Omega} = (\boldsymbol{\gamma}^T\boldsymbol{\gamma})^{-1}$, we can write

$$p(\boldsymbol{b}_i;\boldsymbol{\Psi}) = \frac{1}{(2\pi\sigma^2)^{r/2}|(\boldsymbol{\gamma}^T\boldsymbol{\gamma})^{-1}|^{1/2}}\exp\left[-\frac{1}{2\sigma^2}\boldsymbol{b_i}^T(\boldsymbol{\gamma}^T\boldsymbol{\gamma})\boldsymbol{b_i}\right]. \tag{5.19}$$

Also, under the assumption that $\boldsymbol{\epsilon}_i \sim \mathcal{N}_m(\boldsymbol{0},\sigma^2\boldsymbol{I})$, the distribution function of $\boldsymbol{y}_i$, conditional on the random effects $\boldsymbol{b}_i$, is given as,

$$p(\boldsymbol{y}_i|\boldsymbol{b}_i,\boldsymbol{x}_i;\boldsymbol{\Psi}) \;=\; \frac{1}{(2\pi\sigma^2)^{m/2}}\exp\left[-\frac{1}{2\sigma^2}(\boldsymbol{y_i}-\boldsymbol{\eta}(\boldsymbol{x_i},\boldsymbol{\beta_i}))^T(\boldsymbol{y_i}-\boldsymbol{\eta}(\boldsymbol{x_i},\boldsymbol{\beta_i}))\right], \tag{5.20}$$

where $\boldsymbol{\beta}_i = \boldsymbol{A}_i\boldsymbol{\beta} + \boldsymbol{B}_i\boldsymbol{b}_i$ expresses a subject's parameters as a function of the fixed effects, random effects and the covariates.

Finally, the marginal likelihood function is:

$$L(\boldsymbol{\Psi}|\boldsymbol{y}) = \prod_{i=1}^{N}\int \frac{1}{(2\pi\sigma^2)^{m/2}}\exp\left[-\frac{1}{2\sigma^2}(\boldsymbol{y_i}-\boldsymbol{\eta}(\boldsymbol{x_i},\boldsymbol{\beta_i}))^T(\boldsymbol{y_i}-\boldsymbol{\eta}(\boldsymbol{x_i},\boldsymbol{\beta_i}))\right]\times \frac{1}{(2\pi\sigma^2)^{r/2}|(\boldsymbol{\gamma}^T\boldsymbol{\gamma})^{-1}|^{1/2}}\exp\left[-\frac{1}{2\sigma^2}\boldsymbol{b_i}^T(\boldsymbol{\gamma}^T\boldsymbol{\gamma})\boldsymbol{b_i}\right]\mathrm{d}\,\boldsymbol{b}_i \tag{5.21}$$

$$= \frac{|\boldsymbol{\gamma}|^N}{(2\pi\sigma^2)^{(N(r+m))/2}}\prod_{i=1}^{N}\int \exp\left[-\frac{1}{2\sigma^2}(\boldsymbol{y_i}-\boldsymbol{\eta}(\boldsymbol{x_i},\boldsymbol{\beta_i}))^T(\boldsymbol{y_i}-\boldsymbol{\eta}(\boldsymbol{x_i},\boldsymbol{\beta_i}))\right]\times \exp\left[-\frac{1}{2\sigma^2}\boldsymbol{b_i}^T(\boldsymbol{\gamma}^T\boldsymbol{\gamma})\boldsymbol{b_i}\right]\mathrm{d}\,\boldsymbol{b}_i. \tag{5.22}$$

$$= \frac{|\boldsymbol{\gamma}|^N}{(2\pi\sigma^2)^{(N(r+m))/2}}\prod_{i=1}^{N}\int \exp\left[\frac{||\boldsymbol{y_i}-\boldsymbol{\eta}(\boldsymbol{x}_i,\boldsymbol{\beta_i})||^2+||\boldsymbol{\gamma}\boldsymbol{b}_i||^2}{-2\sigma^2}\right]\mathrm{d}\,\boldsymbol{b}_i, \tag{5.23}$$

where the symbol $||a||$ denotes the Euclidean norm of a vector $\boldsymbol{a} = (a_1, ..., a_n)^T$, that is $||\boldsymbol{a}|| = \sqrt{\boldsymbol{a}^T\boldsymbol{a}}$.

Therefore, maximising Equation (5.23) with respect to $\boldsymbol{\Psi} = (\boldsymbol{\beta}^T, \boldsymbol{\omega}^T, \sigma^2)^T$ would give us the estimates of the parameters. Unfortunately, the integral has no closed form solution, in general, and therefore must be evaluated numerically. Several methods have been proposed for a numerical solution. Pinheiro and Bates (2000) and Davidian and Giltinan (1995) give a review of these methods. For our work, we make use of the MATLAB® package *nlmefit*.

We briefly explain some of the estimation methods available for non-linear mixed effects models. We firstly express the function $\eta(\boldsymbol{x}_i, \boldsymbol{g}(\boldsymbol{\beta}, \boldsymbol{b}_i))$ in an approximate linearised form. This will also be useful when we discuss the theory of optimal designs for non-linear models.

5.3.2 Methods Based on Linearisation of the Response

Linearisation of $\eta(\boldsymbol{x}_i, \boldsymbol{g}(\boldsymbol{\beta}, \boldsymbol{b}_i))$

As mentioned before, the regression function $\eta(\boldsymbol{x}_i, \boldsymbol{g}(\boldsymbol{\beta}, \boldsymbol{b}_i))$ is non-linear in the parameters $\boldsymbol{\Psi}$. Let $\boldsymbol{\alpha}_i = (\boldsymbol{\beta}^T, \boldsymbol{b}_i^T)^T$. Using a first order Taylor series expansion, $\eta(\boldsymbol{x}_i, \boldsymbol{g}(\boldsymbol{\beta}, \boldsymbol{b}_i))$ can be linearised around $\boldsymbol{\alpha}_i$ at $\boldsymbol{\alpha}^o = (\boldsymbol{\beta}^{oT}, \boldsymbol{b}^{oT})^T$ as shown below:

$$\eta(\boldsymbol{x}_i, \boldsymbol{g}(\boldsymbol{\beta}, \boldsymbol{b}_i)) \approx \eta(\boldsymbol{x}_i, \boldsymbol{g}(\boldsymbol{\beta}^o, \boldsymbol{b}^o)) + \left(\frac{\partial \boldsymbol{\eta}(\boldsymbol{x}_i, \boldsymbol{g}(\boldsymbol{\beta}, \boldsymbol{b}_i))}{\partial \boldsymbol{\alpha}_i}\right)^T \Bigg|_{\boldsymbol{\alpha}^o} (\boldsymbol{\alpha} - \boldsymbol{\alpha}^o)$$

$$= \eta(\boldsymbol{x}_i, \boldsymbol{g}(\boldsymbol{\beta}^o, \boldsymbol{b}^o)) + \left(\frac{\partial \boldsymbol{\eta}(\boldsymbol{x}_i, \boldsymbol{g}(\boldsymbol{\beta}, \boldsymbol{b}_i))}{\partial \boldsymbol{\beta}}\right)^T \Bigg|_{\boldsymbol{\alpha}^o} (\boldsymbol{\beta} - \boldsymbol{\beta}^o) + \left(\frac{\partial \boldsymbol{\eta}(\boldsymbol{x}_i, \boldsymbol{g}(\boldsymbol{\beta}, \boldsymbol{b}_i))}{\partial \boldsymbol{b}_i}\right)^T \Bigg|_{\boldsymbol{\alpha}^o} (\boldsymbol{b}_i - \boldsymbol{b}^o),$$

$$= \boldsymbol{K}_i + \left(\frac{\partial \boldsymbol{\eta}(\boldsymbol{x}_i, \boldsymbol{g}(\boldsymbol{\beta}, \boldsymbol{b}_i))}{\partial \boldsymbol{\beta}}\right)^T \Bigg|_{\boldsymbol{\alpha}^o} \boldsymbol{\beta} + \left(\frac{\partial \boldsymbol{\eta}(\boldsymbol{x}_i, \boldsymbol{g}(\boldsymbol{\beta}, \boldsymbol{b}_i))}{\partial \boldsymbol{b}_i}\right)^T \Bigg|_{\boldsymbol{\alpha}^o} \boldsymbol{b}_i,$$

where $\boldsymbol{K}_i$ is a $m \times 1$ vector of constants. The non-linear mixed effects model can now be expressed in the form a linear mixed effects model as:

$$\boldsymbol{y}_i \approx \boldsymbol{K}_i + \boldsymbol{G}_i\boldsymbol{\beta} + \boldsymbol{H}_i\boldsymbol{b}_i + \boldsymbol{\epsilon}_i, \tag{5.24}$$

where the elements of the matrices $\boldsymbol{G}_i$ and $\boldsymbol{H}_i$ are given as:

$$(\boldsymbol{G}_i)_{jk} = \left.\frac{\partial \eta(x_{ij}, \boldsymbol{g}(\boldsymbol{\beta}, \boldsymbol{b}_i))}{\partial \beta_k}\right|_{\boldsymbol{\alpha}^o} \quad \text{and} \quad (\boldsymbol{H}_i)_{jk} = \left.\frac{\partial \eta(x_{ij}, \boldsymbol{g}(\boldsymbol{\beta}, \boldsymbol{b}_i))}{\partial b_k}\right|_{\boldsymbol{\alpha}^o},$$

for $j = 1, ..., m$ and $k = 1, ..., p$.

From Equation (5.24) we get an evaluation of the expectation vector and the dispersion matrix for $\boldsymbol{y}_i$, that is

$$\mathrm{E}(\boldsymbol{y}_i) = \boldsymbol{E}_i^o \approx \boldsymbol{K}_i + \boldsymbol{G}_i\boldsymbol{\beta},$$

and

$$\mathrm{var}(\boldsymbol{y}_i) = \boldsymbol{V}_i^o \approx \sigma^2(\boldsymbol{H}_i\boldsymbol{\Omega}\boldsymbol{H}_i^T + \boldsymbol{I}_m).$$

The expression of variance is valid only on the assumption of independence of $\boldsymbol{b}_i$ and $\boldsymbol{\epsilon}_i$. Further, as $\boldsymbol{\epsilon}_i \sim \mathcal{N}_m(\boldsymbol{0}, \sigma^2\boldsymbol{I})$, the approximate distribution of $\boldsymbol{y}_i$ is $\mathcal{N}_m(\boldsymbol{E}_i^o, \boldsymbol{V}_i^o)$. We now discuss some methods which are available for the estimation of non-linear mixed effects models.

First Order (FO)

The FO method is among the first methods proposed for estimation of population parameters by maximum likelihood. This method is due to Sheiner and Beal (1980) and consists in using the first-order Taylor series expansion of the model $\boldsymbol{\eta}$ around the current estimate, $\boldsymbol{\beta}$, of $\boldsymbol{\beta}$ and the expected value of the random effects, i.e., around $\mathrm{E}(\boldsymbol{b}) = \boldsymbol{0}$. The distribution of $\boldsymbol{y}_i$ is thus approximated as:

$$\boldsymbol{y}_i \dot{\sim} \mathcal{N}_m(\widehat{\boldsymbol{E}}_i, \widehat{\boldsymbol{V}}_i), \tag{5.25}$$

where $\dot{\sim}$ stands for 'approximately distributed as' and $\widehat{\boldsymbol{E}}_i$ and $\widehat{\boldsymbol{V}}_i$ are evaluated at $\widehat{\boldsymbol{\alpha}} = (\widehat{\boldsymbol{\beta}}^T, \boldsymbol{0}^T)^T$.

This simplification makes the FO method computationally less expensive than the Linear Mixed Effects method (discussed below) but may result in significantly biased estimates. An extended least squares method, as discussed in Sheiner and Beal (1985), can be used to estimate the population parameters from Equation (5.25) and these estimates can then be used to linearise $\boldsymbol{\eta}$ in the next iteration. These iterations continue until a pre-specified convergence criterion is met. This method has been

implemented in the software NONMEM, Beal and Sheiner (1989).

Linear Mixed Effects (LME)

We briefly describe here the Linear Mixed Effects (LME) method as given in Lindstrom and Bates (1990).

LME consists in using a Taylor series expansion of $\eta(\boldsymbol{x}_i, \boldsymbol{g}(\boldsymbol{\beta}, \boldsymbol{b}_i)$ about estimates of both $\boldsymbol{b}$ and $\boldsymbol{\beta}$. The estimates are then updated by minimisation of a penalised least squares function. It is a two-step algorithm as discussed below:

Step 1

Let the initial values of $\boldsymbol{\Omega}$ and σ^2 be $\boldsymbol{\Omega}^o$ and $(\sigma^o)^2$.

The penalised least squares function:

$$\sum_{i=1}^{N} \left[\frac{||\boldsymbol{y_i} - \boldsymbol{\eta}(\boldsymbol{x}_i, \boldsymbol{g}(\boldsymbol{\beta}, \boldsymbol{b}_i)||^2 + \boldsymbol{b}_i^T \boldsymbol{\Omega}^o \boldsymbol{b}_i}{(\sigma^o)^2} \right]$$

is then minimised with respect to $\boldsymbol{\beta}$ and $\boldsymbol{b}_i$. We denote the optimised values by $\widehat{\boldsymbol{\alpha}}_i = (\widehat{\boldsymbol{\beta}}^T, \widehat{\boldsymbol{b}}^T)^T$.

Step 2

First order Taylor series is then used to expand $\boldsymbol{\eta}(\boldsymbol{x}_i, \boldsymbol{g}(\boldsymbol{\beta}, \boldsymbol{b}_i))$ around $\widehat{\boldsymbol{\alpha}}_i$. The vector of response of the i^{th} subject can be expressed as

$$\boldsymbol{y}_i \approx \boldsymbol{\eta}(\boldsymbol{x}_i, \boldsymbol{g}(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{b}}_i)) + \left(\frac{\partial \boldsymbol{\eta}(\boldsymbol{x}_i, \boldsymbol{g}(\boldsymbol{\beta}, \boldsymbol{b}_i))}{\partial \boldsymbol{\beta}} \right)^T \Bigg|_{\widehat{\boldsymbol{\alpha}}_i} (\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}) + \left(\frac{\partial \boldsymbol{\eta}(\boldsymbol{x}_i, \boldsymbol{g}(\boldsymbol{\beta}, \boldsymbol{b}_i))}{\partial \boldsymbol{b}_i} \right)^T \Bigg|_{\widehat{\boldsymbol{\alpha}}_i} (\boldsymbol{b} - \widehat{\boldsymbol{b}}_i) + \boldsymbol{\epsilon}_i. \quad (5.26)$$

Let $\widehat{\boldsymbol{X}}_i = \left(\frac{\partial \boldsymbol{\eta}(\boldsymbol{x}_i, \boldsymbol{g}(\boldsymbol{\beta}, \boldsymbol{b}_i))}{\partial \boldsymbol{\beta}} \right)^T \Bigg|_{\widehat{\boldsymbol{\alpha}}_i}$ and $\widehat{\boldsymbol{Z}}_i = \left(\frac{\partial \boldsymbol{\eta}(\boldsymbol{x}_i, \boldsymbol{g}(\boldsymbol{\beta}, \boldsymbol{b}_i))}{\partial \boldsymbol{b}_i} \right)^T \Bigg|_{\widehat{\boldsymbol{\alpha}}_i}$.

We then have,

$$\mathrm{E}(\boldsymbol{y}_i) \approx \widehat{\boldsymbol{E}}_i = \boldsymbol{\eta}(\boldsymbol{x}_i, \boldsymbol{g}(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{b}}_i)) + \widehat{\boldsymbol{X}}_i(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}) - \widehat{\boldsymbol{Z}}_i \widehat{\boldsymbol{b}}_i,$$

and

$$\text{var}(\boldsymbol{y}_i) \approx \widehat{\boldsymbol{V}}_i = \sigma^2(\boldsymbol{I}_m + \widehat{\boldsymbol{Z}}_i \boldsymbol{\Omega} \widehat{\boldsymbol{Z}}_i^T)$$

Since $\boldsymbol{y}_i$ is now a linear function of $\boldsymbol{b}_i$ and $\boldsymbol{\epsilon}_i$, the distribution of $\boldsymbol{y}_i$ is approximately $\mathcal{N}_m(\widehat{\boldsymbol{E}}_i, \widehat{\boldsymbol{V}}_i)$. The log-likelihood function for this distribution is given as:

$$l_F(\boldsymbol{\beta}, \sigma, \boldsymbol{\Omega} | \boldsymbol{y}_i) \approx -\frac{1}{2}\ln(\widehat{\boldsymbol{V}}_i) - \frac{1}{2}(\boldsymbol{y}_i - \widehat{\boldsymbol{E}}_i)^T \widehat{\boldsymbol{V}}_i^{-1}(\boldsymbol{y}_i - \widehat{\boldsymbol{E}}_i).$$

This is obtained for a linear mixed effects model for which standard methods are available. Let $\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\Omega}}$ and $\widehat{\boldsymbol{\sigma}}^2$ be the values which maximise the approximate likelihood function generated by $\boldsymbol{y}_i$ above. Now setting $\boldsymbol{\Omega}^o = \widehat{\boldsymbol{\Omega}}$ and $(\sigma^o)^2 = \widehat{\sigma}^2$, the algorithm returns to Step 1 and minimises the updated penalised least squares function.

The algorithm iterates between these two steps until convergence is attained. The maximum likelihood estimates of $\boldsymbol{\beta}$, $\boldsymbol{\Omega}$ and σ^2 are obtained at the end along with the updated estimate of the vector of random effects $\widehat{\boldsymbol{b}}$. We will use the vector $\widehat{\boldsymbol{b}}$ in Chapter 7, where we present a new method of dose individualisation, by computing the subjects' individual PK parameters.

First Order Conditional Estimation (FOCE)

A major disadvantage of the FO method is that it assumes the expected vector of random effects to be equal to 0, which will produce imprecise estimates if there is large inter-individual variability in the data. The FOCE method is a refinement over the FO method and has also been implemented in NONMEM, Beal and Sheiner (1998). The primary difference between these two methods is the way the log-likelihood function is approximated. While in FO method the linearisation was done around the expected value of the random effect, i.e., $\text{E}(\boldsymbol{b}) = \boldsymbol{0}$, in the FOCE method, the linearisation is done around the current estimates of $\boldsymbol{b}$. The estimation algorithm for the parameters is similar to the two step algorithm proposed by Lindstrom and Bates (1990). A disadvantage of this method is that it is very sensitive to the initial estimates of the parameters.

Stochastic Expectation Maximisation Algorithms

The Expectation Maximisation (EM) algorithm is a popular tool and was first proposed by Dempster et al. (1977) to handle the issue of missing values in maximum likelihood problems. However, it has a wider scope and has several applications in statistics including computation of MLEs. The algorithm consists of the E-step in which a function for expectation of the log-likelihood is evaluated at the current

estimates of the parameters. Next, in the M-step, the parameters maximising the log-likelihood function found in the E-step are computed. These parameters are then used to find the log-likelihood function for the next E-step. The algorithm terminates with the convergence of the parameters. The EM algorithm can thus be used to find the MLEs when the equations cannot be solved directly. An advantage of this method is that no derivatives need to be computed for computation of the MLEs.

However, in the case of non-linear mixed effects models, the E-step is difficult to carry out due to the complexity of the integrals involved. To overcome this problem, many stochastic methods have been proposed. The two most commonly used methods are:

- Monte Carlo integration during the E-step with importance sampling around the current individual estimates as in Bauer and Guzy (2004). If x is a random variable with pdf $f(x)$, then Monte Carlo integration consists in estimating $\mathrm{E}[g(x)] = \int g(x)f(x)\,\mathrm{d}\,x$ as $\tilde{g}(\boldsymbol{x}_n) = \frac{1}{n}\sum_{i=1}^{n} g(x_i)$, where $\boldsymbol{x}_n = (x_1, ..., x_n)^T$ is a random sample drawn from f. $\tilde{g}(\boldsymbol{x}_n)$ is an unbiased and consistent estimator of $\mathrm{E}[g(x)]$. Importance sampling is a variance-reduction technique which modifies the above estimator to facilitate a lower variance, Tokdar and Kass (2010).

- The Stochastic Approximated EM (SAEM) method which makes use of stochastic approximation of the expected likelihood function, Delyon et al. (1999). Stochastic approximation comprises of optimisation methods which are used to find extrema of functions of random variables which cannot be computed directly.

Mentré and Lavielle (2008) discuss the performances of the Stochastic EM methods for PKPD analyses. They conclude that the SAEM algorithm has good statistical properties and overcomes some of the limitations of the other methods described above.

Other Methods for Estimation

Laplace approximation to integrals can be used to approximate the integral given in (5.23), Wolfinger (1993). Laplace approximation converts the problem of integration into a simpler problem of maximisation. Suppose we have to evaluate the integral $I = \int g(\boldsymbol{b})\,\mathrm{d}\,\boldsymbol{b}$ where $\boldsymbol{b}$ is a $r \times 1$ vector. Let $l(\boldsymbol{b})$ be the ln of the integrand and $\widehat{\boldsymbol{b}}$ be the point of maxima of $l(\boldsymbol{b})$. Then, the first derivative is $l'(\boldsymbol{b}) = 0$ and using second order Taylor Series expansion at the point $\widehat{\boldsymbol{b}}$ we have: $l(\boldsymbol{b}) \approx l(\widehat{\boldsymbol{b}}) + \frac{1}{2}(\boldsymbol{b} - \widehat{\boldsymbol{b}})^T l''(\widehat{\boldsymbol{b}})(\boldsymbol{b} - \widehat{\boldsymbol{b}})$, where $l''(\widehat{\boldsymbol{b}})$ denotes the second derivative evaluated at $\widehat{\boldsymbol{b}}$. Using this, I can be approximated as $I = \int \exp(l(\boldsymbol{b}))\,\mathrm{d}\,\boldsymbol{b} \approx \exp(l(\widehat{\boldsymbol{b}})) \int \exp(\frac{1}{2}(\boldsymbol{b} - \widehat{\boldsymbol{b}})^T l''(\widehat{\boldsymbol{b}})(\boldsymbol{b} - \widehat{\boldsymbol{b}}))\,\mathrm{d}\,\boldsymbol{b}$.

This is a Gaussian integral and can be computed as $I \approx \exp(l(\widehat{\boldsymbol{b}}))(2\pi)^{r/2}|-l''(\widehat{\boldsymbol{b}})|^{-1/2}$. This approximation is available in most software dealing with non-linear mixed effects models.

SAS® (SAS Institute Inc., USA) software provides '*PROC NLMIXED*' for estimation of parameters in non-linear mixed effects models. It provides an option of integration of the likelihood function using adaptive Gaussian quadrature method presented by Pinheiro and Bates (1995). This method approximates the integral given in (5.23) by a weighted sum over predefined abscissae for the random effects $\boldsymbol{b}_i$. For a good approximation, this method centres the abscissae at the empirical Bayes estimate of $\boldsymbol{b}_i$, defined as the vector $\widehat{\boldsymbol{b}}_i$ that minimises $-\ln[p(\boldsymbol{y}_i|\boldsymbol{b}_i, \boldsymbol{X}_i; \boldsymbol{\Psi})p(\boldsymbol{b}_i; \boldsymbol{\Psi})]$. Standard optimisation algorithms can then be used to compute $\widehat{\boldsymbol{\Psi}}$.

Many statistical software have a dedicated submodule for estimation of parameters of non-linear mixed effects models. For example, *R* and S-Plus® both have packages called *nlme*. Specialist PK/PD software like NONMEM® and Monolix® present a library of PK/PD models and provide a choice of methods to estimate the parameters, of which the main ones were presented in this section.

The linear mixed effects (LME) estimation algorithm of Lindstrom and Bates (1990) is one of the most popular methods for estimation of parameters of non-linear mixed effects models and we use this method for our computations in this thesis. The MATLAB® package *nlmefit* implements this algorithm and returns estimates of the parameters and matrix of the estimated random effects. It also returns model fitting statistics such as the final maximised log-likelihood, the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). More information about the *nlmefit* package is available in MathWorks (2015).

5.3.3 Predictions in Non-linear Mixed Effects Models

One of the primary objectives of statistical modelling is to determine the relationship between the response and the explanatory variables. Existing values of the explanatory variables can be applied in the estimated model to obtain the fitted values of the response and new values of the variables can be applied in the estimated model to obtain the predicted response at those observations.

In the previous sections, we discussed some of the methods available in the literature for estimation of the parameters of non-linear mixed effects models. The estimated vector of fixed effects, $\widehat{\boldsymbol{\beta}}$ and the estimated vector of random effects, $\widehat{\boldsymbol{b}}_i$, can be applied back into the original model to get the fitted or predicted values. The fitted values are useful for checking the goodness of the fit as well as for drawing inferences

from the estimated model. In this section, we discuss the concept of predictions in the context of non-linear mixed effects models. We adapt the theory presented in the books Pinheiro and Bates (2000) and Davidian and Giltinan (1995). We continue using the notation already defined in Section 5.2.

Non-linear mixed effects models are useful in modelling the intra- as well as the inter-individual variability. They contain not only population mean effects but also random individual effects. This feature makes it possible to obtain predicted values of the response at not only the population level but also at the individual level. This is discussed below.

Population Level

Suppose the interest is in finding the predicted population response at the experimental settings contained in a vector denoted by $\widetilde{\boldsymbol{x}} = (\widetilde{\boldsymbol{x}}_1^T, ..., \widetilde{\boldsymbol{x}}_N^T)^T$, where $\widetilde{\boldsymbol{x}}_i^T$ are vectors of some new values of the explanatory variables. Let us consider the additive model expressed in Equation (5.5). Since the other models discussed can also be expressed in additive form, the theory presented here can be applied to these cases as well.

Population level predictions are obtained by estimating the expected response when the random effects, $\boldsymbol{b}$, are taken to be equal to their mean value, i.e., $\boldsymbol{0}$. This results in $\boldsymbol{\phi} = \boldsymbol{A\beta}$.

Taking expectation on both sides of Equation (5.5) we get $\mathrm{E}(\boldsymbol{y}) = \boldsymbol{\eta}(\widetilde{\boldsymbol{x}}, \boldsymbol{A\beta})$.

Estimated vector of the PK parameters, $\widehat{\boldsymbol{\beta}}$, can be used to obtain the predicted response at the population level, $\widetilde{\boldsymbol{y}} = (\widetilde{\boldsymbol{y}}_1^T, ..., \widetilde{\boldsymbol{y}}_N^T)^T$, at the experimental settings $\widetilde{\boldsymbol{x}}$ as

$$\widetilde{\boldsymbol{y}} = \boldsymbol{\eta}(\widetilde{\boldsymbol{x}}, \boldsymbol{A}\widehat{\boldsymbol{\beta}}).$$

If in a group of N patients the observations for all the individuals are collected at the same experimental settings, that is, $\widetilde{\boldsymbol{x}}_i = \widetilde{\boldsymbol{x}}_s$ (say) for $i = 1, ..., N$, then the predicted response at the population level, $\widetilde{\boldsymbol{y}}_s$, for all the N individuals will be the same. That is,

$$\widetilde{\boldsymbol{y}}_i = \widetilde{\boldsymbol{y}}_s = \boldsymbol{\eta}(\widetilde{\boldsymbol{x}}_s, \boldsymbol{A}\widehat{\boldsymbol{\beta}}), \quad i = 1, ..., N.$$

Individual Level

Here, we are interested in prediction of the expected response of the i^{th} individual, $i = 1, ..., N$, at the experimental settings $\widetilde{\boldsymbol{x}}_i = (\widetilde{x}_{i1}, ..., \widetilde{x}_{im})^T$.

In Section 5.3, we discussed some of the methods available in literature for estima-

tion of parameters of non-linear mixed effects models. For example, the linear mixed effects (LME) method outlined in Section 5.3.2, enables estimation of the vector of population parameters, $\boldsymbol{\beta}$, and the vectors of individuals' random effects, $\boldsymbol{b}_i$.

The estimates $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{b}}_i$ can then be used to obtain the predicted response for the i$^{\text{th}}$ subject. Firstly, estimated individual parameters for the i^{th} subject can be obtained from Equation (5.4) as

$$\widehat{\boldsymbol{\beta}}_i = \boldsymbol{A}_i\widehat{\boldsymbol{\beta}} + \boldsymbol{B}_i\widehat{\boldsymbol{b}}_i. \tag{5.27}$$

Now, taking expectation given the random effects $\boldsymbol{b}_i$ on both sides of Equation (5.3) we obtain the conditional expected response for the i^{th} subject at the setting $\widetilde{\boldsymbol{x}}_i$ as

$$\mathrm{E}(\boldsymbol{y}_i|\boldsymbol{b}_i) = \boldsymbol{\eta}(\widetilde{\boldsymbol{x}}_i, \boldsymbol{A}_i\boldsymbol{\beta} + \boldsymbol{B}_i\boldsymbol{b}_i), \qquad i = 1, ..., N.$$

The predicted response for the i^{th} subject, $\widetilde{\boldsymbol{y}}_i$ can be obtained by estimating the conditional expectation above by using $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{b}}_i$, i.e.,

$$\widetilde{\boldsymbol{y}}_i = \boldsymbol{\eta}(\widetilde{\boldsymbol{x}}_i, \boldsymbol{A}_i\widehat{\boldsymbol{\beta}} + \boldsymbol{B}_i\widehat{\boldsymbol{b}}_i).$$

Example

Let us consider the one-compartment model with zero order absorption described in Chapter 2, that is:

$$C(t) = \frac{de^{-K_e t}}{V_1}. \tag{5.28}$$

The parameters of the i^{th} subject are $\boldsymbol{\beta}_i = (K_{ei}, V_{1i})^T = (\beta_{1i}, \beta_{2i})^T$ and the population parameters are $\boldsymbol{\beta} = (K_e, V_1)^T = (\beta_1, \beta_2)^T$. Let $m = 3$ blood samples be collected from each subject at the same times. For the i^{th} subject, the experimental settings are $\boldsymbol{x}_i = (t_1, t_2, t_3)^T$, $i = 1, ..., N$.

Assuming an exponential error structure, the Stage 1 model can be represented as:

$$y_{ij} = \frac{de^{-\beta_{1i}t_j}}{\beta_{2i}}\exp(\epsilon_{ij}),$$

where $\epsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$, $i = 1, ..., N$ and $j = 1, 2, 3$.

Assuming exponential random effects, the Stage 2 model can be expressed as:

$$\begin{pmatrix} \beta_{1i} \\ \beta_{2i} \end{pmatrix} = \begin{pmatrix} \beta_1 \exp(b_{1i}) \\ \beta_2 \exp(b_{2i}) \end{pmatrix}.$$

Now, for the above model, the estimated vector of parameters $\widehat{\boldsymbol{\beta}}$ can be obtained by using any of the estimation methods described earlier. Then the predicted population response $\widetilde{y}_s$ at a new experimental setting $\widetilde{t}_s$ is

$$\widetilde{y}_s = \frac{de^{-\widehat{\beta}_1 \widetilde{t}_s}}{\widehat{\beta}_2}.$$

If estimates of the random effects for the i^{th} subject, $\widehat{\boldsymbol{b}}_i$, are available, then the predicted response $\widetilde{y}_i$ for this subject at the experimental setting $\widetilde{t}_s$ is

$$\widetilde{y}_i = \frac{de^{-\widehat{\beta}_{1i} \widetilde{t}_s}}{\widehat{\beta}_{2i}},$$

where

$$\begin{pmatrix} \widehat{\beta}_{1i} \\ \widehat{\beta}_{2i} \end{pmatrix} = \begin{pmatrix} \widehat{\beta}_1 \exp(\widehat{b}_{1i}) \\ \widehat{\beta}_2 \exp(\widehat{b}_{2i}) \end{pmatrix},$$

is the vector of estimated parameters of the i^{th} subject.

On account of different estimates of the random effects for different subjects, the corresponding values of the predicted response, even at the same experimental settings, are different.

Some authors such as Pinheiro and Bates (2000) argue that since the random effects $\boldsymbol{b}_i$ are sources of random variation in the model and not model parameters as such, it is more appropriate to say that they are 'predicted' rather than 'estimated' on the basis of the observed data.

Predictions are particularly useful in scientific studies involving multi-level or hierarchical models such as the ones described in this chapter as they enable model fitting for individual units in a study. For example, Stirnemann et al. (2012) predict the individual growth of foetuses in twin pregnancies and Brabec et al. (2008) present predictions of natural gas consumption by individual customers using non-linear mixed effects models.

In this thesis, we compute predicted responses in Section 7.1 to fit individual concentration profiles to each of the subjects in the cohort by applying the estimated population parameters and the estimated random effects in the originally assumed PK model. This enables estimation of the individual concentration profiles of the subjects which then facilitates individualisation of the dose regimen to each subject.

The predicted values described above are point estimates and their values depend

on the observed responses which are themselves subject to random error. It will be useful, therefore, to estimate an interval in which the predicted value will lie with a specified probability. Such an interval is called a prediction interval and it is analogous to the concept of a confidence interval which contains an unobservable population parameter with a specified probability. For the case of non-linear mixed effects models, the derivation of prediction intervals is often complex and heuristic methods have to be used for their computation as done in, for example, Stirnemann et al. (2012). Since we do not require prediction intervals for our work in this thesis we do not go into further details here.

5.4 Optimal Design of the Study

In the previous section, estimation of the vector of parameters $\boldsymbol{\Psi}$ was discussed. The samples of data are designated to be collected at the points T_j, $j = 1, ..., m$. Let the sampling time points be denoted in ξ as: $\xi = \{T_1, ..., T_m\}$ where $T_1 < ... < T_m$. Generally, the samples can only be collected within specific time intervals. The sample space from which the experimental settings can be chosen is called the design region.

In this section, we discuss how the choice of the experimental settings ξ, affects the precision of the parameters in $\boldsymbol{\Psi}$. We then briefly discuss the methods available in the software PopED to design an optimal experiment for non-linear mixed effects models. Firstly, we recall the notion of the Fisher information matrix and present some of the commonly used criteria of optimality.

5.4.1 Fisher Information Matrix and Optimality Criteria

Most of the criteria of optimality for design of experiments are functions of the Fisher Information Matrix (FIM).

Let $\boldsymbol{\Psi} = (\psi_1, ..., \psi_k)^T$ be the vector of parameters which are to be estimated. Given a random vector of observations $\boldsymbol{y}$, which depends on the vector of parameters $\boldsymbol{\Psi}$ having the likelihood function $l(\boldsymbol{y}; \boldsymbol{\Psi})$.

The score function, defined as the gradient of the log-likelihood with respect to $\boldsymbol{\Psi}$, is given as:

$$U(\boldsymbol{y}; \boldsymbol{\Psi}) = \frac{\partial \ \ln l(\boldsymbol{y}; \boldsymbol{\Psi})}{\partial \boldsymbol{\Psi}^T} = \left(\frac{\partial \ \ln l(\boldsymbol{y}; \boldsymbol{\Psi})}{\partial \psi_1}, \ \ldots \ , \frac{\partial \ \ln l(\boldsymbol{y}; \boldsymbol{\Psi})}{\partial \psi_k} \right)^T .$$

Fisher Information Matrix (FIM) is defined as the covariance of this score function

i.e.,

$$M(\boldsymbol{\Psi}) = E(U(\boldsymbol{y};\boldsymbol{\Psi})U^T(\boldsymbol{y};\boldsymbol{\Psi})). \tag{5.29}$$

Let us first consider the simple linear models:

$$\boldsymbol{y} = \boldsymbol{F}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{5.30}$$

where

$\boldsymbol{y}$ is the vector of N observations;

$\boldsymbol{\epsilon} = (\epsilon_1, ..., \epsilon_N)^T$ is the vector of random errors;

$\boldsymbol{\beta}$ is the p-dimensional vector of parameters;

$$\boldsymbol{F} = \begin{pmatrix} \boldsymbol{f}^T(x_1) \\ . \\ . \\ . \\ \boldsymbol{f}^T(x_N) \end{pmatrix} = \begin{pmatrix} f_1(x_1) & f_2(x_1) & ... & f_p(x_1) \\ f_1(x_2) & f_2(x_2) & ... & f_p(x_2) \\ . & . & . & . \\ . & . & . & . \\ f_1(x_N) & f_2(x_N) & ... & f_p(x_N) \end{pmatrix}$$

$\boldsymbol{f}^T(x) = (f_1(x), ..., f_p(x))$;

f_i is a known real-valued regression function, $i = 1, ..., p$;

x_j is the j^{th} experimental setting, $j = 1, ..., N$.

$\boldsymbol{F}$ is called the Design Matrix. For the above model, $E(\boldsymbol{y}) = \boldsymbol{F}\boldsymbol{\beta}$ and $\text{cov}(\boldsymbol{y}) = \text{cov}(\boldsymbol{\epsilon}) = \sigma^2\boldsymbol{I_N}$. The distribution of $\boldsymbol{y}$ is given as $\mathcal{N}_N(\boldsymbol{F}\boldsymbol{\beta}, \sigma^2\boldsymbol{I})$.

The log-likelihood function is given as:

$$\ln l(\boldsymbol{y};\boldsymbol{\beta}) = \frac{-N\ln(2\pi\sigma^2)}{2} - \frac{1}{2\sigma^2}(\boldsymbol{y} - \boldsymbol{F}\boldsymbol{\beta})^T(\boldsymbol{y} - \boldsymbol{F}\boldsymbol{\beta}).$$

Now, if a $n \times 1$ vector $\boldsymbol{X}$ is a function of the vector $\boldsymbol{\alpha}$, we have

$$\frac{\partial\, \boldsymbol{X}^T\boldsymbol{X}}{\partial\boldsymbol{\alpha}} = 2\boldsymbol{X}^T\frac{\partial\, \boldsymbol{X}}{\partial\boldsymbol{\alpha}}.$$

Thus, differentiating the log-likelihood function with respect to $\boldsymbol{\beta}$ we get:

$$U(\boldsymbol{y};\boldsymbol{\beta}) = \frac{\partial\ \ln l(\boldsymbol{y};\boldsymbol{\beta})}{\partial\boldsymbol{\beta}^T} = \frac{\boldsymbol{F}^T(\boldsymbol{y} - \boldsymbol{F}\boldsymbol{\beta})}{\sigma^2} = \frac{\boldsymbol{F}^T\boldsymbol{\epsilon}}{\sigma^2}.$$

Consequently, the FIM, $\boldsymbol{M}(\boldsymbol{\beta})$, is given as:

$$\boldsymbol{M}(\boldsymbol{\beta}) = \mathrm{E}\left(\frac{\boldsymbol{F}^T\boldsymbol{\epsilon}}{\sigma^2} \times \frac{\boldsymbol{\epsilon}^T\boldsymbol{F}}{\sigma^2}\right) = \frac{\boldsymbol{F}^T\boldsymbol{F}}{\sigma^2}.$$

It can be seen in the above expression that the expression for the FIM for a linear model does not depend on the model parameters. However, it does depend on the choice of the design matrix.

The design region, $\mathcal{X}$, is the set which contains all the possible experimental settings. For example, in a PK study, the design region could be $(0, 12]$ h, which means that a blood sample can be drawn at any time in this interval.

If m observations are to be made from a subject, they can all be taken from m distinct points in $\mathcal{X}$ or from $n(< m)$ distinct points by having replications at one or more points.

Let us consider n distinct points in $\mathcal{X}$. Then an approximate design can be represented by the measure ξ over $\mathcal{X}$ as follows:

$$\xi = \left\{ \begin{array}{cccc} x_1 & x_2 & ... & x_n \\ w_1 & w_2 & ... & w_n \end{array} \right\}, \; x_i \in \mathcal{X}, \quad w_i \in [0, 1] \quad \text{and} \quad \sum_{i=1}^{n} w_i = 1,$$

where x_is represent the design points and w_is the associated design weights, $i = 1, ..., n$. A design weight gives the number of replications at that design point, expressed as a fraction of m. If $n = m$, $w_i = 1/m \; \forall \; i$.

By dropping the constraint that mw_i is an integer, approximate designs make the process of design optimisation easier. Approximate designs were conceptualised by Kiefer and Wolfowitz (1959).

We will see later that the FIM for non-linear models is a function of the model parameters $\boldsymbol{\Psi}$, as well as the experimental settings ξ. Hereafter, the FIM will be represented as $\boldsymbol{M}(\boldsymbol{\Psi}, \xi)$.

We can now recall the following optimality criteria:

- A design ξ^* is called **D-optimal** if

  $$\Phi_D(\boldsymbol{M}(\boldsymbol{\Psi}, \xi)) = -\ln|\boldsymbol{M}(\boldsymbol{\Psi}, \xi)|$$

  is minimised by ξ^* or equivalently,

  $$\ln|\boldsymbol{M}(\boldsymbol{\Psi}, \xi^*)| \geq \ln|\boldsymbol{M}(\boldsymbol{\Psi}, \xi)|$$

for all $\xi \in \Xi$, where Ξ is a set of feasible designs.

D-optimality is the most popular criterion and it was introduced by Wald (1943).

- A design ξ^* is called **G-optimal** if

$$\Phi_G(\boldsymbol{M}) = \max_{x \in \mathcal{X}} \boldsymbol{f}^T(x) \boldsymbol{M}^{-1}(\boldsymbol{\Psi}, \xi) \boldsymbol{f}(x)$$

is minimised by ξ^* or equivalently,

$$\max_{x \in \mathcal{X}} \boldsymbol{f}^T(x) \boldsymbol{M}^{-1}(\boldsymbol{\Psi}, \xi^*) \boldsymbol{f}(x) \leq \max_{x \in \mathcal{X}} \boldsymbol{f}^T(x) \boldsymbol{M}^{-1}(\boldsymbol{\Psi}, \xi) \boldsymbol{f}(x)$$

for all $\xi \in \Xi$.

One of the earliest papers on the theory of optimum designs, Smith (1918), proposed the G-optimal designs. Kiefer and Wolfowitz (1960) later proved the equivalence of G- and D-optimal designs.

- A design ξ^* is called **A-optimal** if

$$\Phi_A(\boldsymbol{M}(\boldsymbol{\Psi}, \xi)) = \text{tr}(\boldsymbol{M}^{-1}(\boldsymbol{\Psi}, \xi))$$

is minimised by ξ^* or equivalently,

$$\text{tr}(\boldsymbol{M}^{-1}(\boldsymbol{\Psi}, \xi^*)) \leq \text{tr}(\boldsymbol{M}^{-1}(\boldsymbol{\Psi}, \xi))$$

for all $\xi \in \Xi$.

The A-optimality criterion is due to Elfving (1952).

- If the interest in estimation of a linear function $\boldsymbol{c}^T\boldsymbol{\beta}$, where $\boldsymbol{c}^T = (c_1, ..., c_p)$ is a vector of constants, the c-criterion is the most appropriate. Let the generalized inverse of $\boldsymbol{M}$ be denoted by $\boldsymbol{M}^-$.

  A design ξ^* is called **c-optimal** if

$$\Phi_c(\boldsymbol{M}) = \boldsymbol{c}^T \boldsymbol{M}^-(\boldsymbol{\Psi}, \xi) \boldsymbol{c}$$

is minimised by ξ^* or equivalently,

$$\boldsymbol{c}^T \boldsymbol{M}^-(\boldsymbol{\Psi}, \xi^*) \boldsymbol{c} \leq \boldsymbol{c}^T \boldsymbol{M}^-(\boldsymbol{\Psi}, \xi) \boldsymbol{c}$$

for all $\xi \in \Xi$.

The theory of optimal experimental design for linear models has been extensively

presented in several books: Fedorov (1972), Fedorov and Hackl (1997), Atkinson and Donev (1992), Atkinson et al. (2007), Pukelsheim (1993), Pázman (1986), Shah and Sinha (1989) are the major references.

Optimal Design for Non-linear Mixed Effects Models

The theory of optimum design for non-linear models is not as widely described as in the linear models case. Atkinson et al. (2007) and Fedorov and Hackl (1997) include some chapters on the non-linear models case. The recent book by Fedorov and Leonov (2014) covers numerical procedures for both parameter estimation and construction of optimal designs for the case of non-linear models with an explanation on using these techniques for dose-finding, PK, PD and other biopharmaceutical applications.

In the last section, we discussed the individual design ξ, which gives the sampling times for a single subject. However, in a clinical trial, treatments are often tested on several groups of subjects. In such situations, a *population design* is used which consists of individual designs ξ_1, ..., ξ_g, along with the proportions p_1, ..., p_g of the subjects of the sample population that are to be observed using these designs, where g is the number of groups.

A population design is then given as

$$\Xi = \left\{ \begin{array}{cccc} \xi_1 & \xi_2 & \dots & \xi_g \\ p_1 & p_2 & \dots & p_g \end{array} \right\}, \; p_i \in [0,1] \;\; \text{and} \;\; \sum_{i=1}^{g} p_i = 1.$$

For finding optimal population designs, one can use the same criteria as above, but the FIM will be more complicated. If $g = 1$, that is, the subjects are studied under the same design, the FIM for the population design will be the sum of individual FIMs.

As discussed in Section 5.3.2, the approximate distribution of $\boldsymbol{y}_i$ is $\mathcal{N}_m(\boldsymbol{E}_i, \boldsymbol{V}_i)$. The approximate FIM for the non-linear models can be derived from this distribution and the definition of the FIM. The parameter vector is $\boldsymbol{\Psi} = (\boldsymbol{\beta}^T, \boldsymbol{\lambda}^T)^T$ where $\boldsymbol{\lambda} = (\boldsymbol{\omega}^T, \sigma^2)^T$.

The log-likelihood of the vector of observations $\boldsymbol{y}_i$ is given as:

$$\ln l_i(\boldsymbol{\Psi}|\boldsymbol{y}_i) \approx -\frac{m}{2}\ln(2\pi) - \frac{1}{2}\ln|\boldsymbol{V}_i| - \frac{1}{2}(\boldsymbol{y}_i - \boldsymbol{E}_i)^T \boldsymbol{V}_i^{-1}(\boldsymbol{y}_i - \boldsymbol{E}_i). \tag{5.31}$$

Differentiating both sides of (5.31) with respect to β_g, $g = 1, ..., p$, we get:

$$\frac{\partial \ln l_i(\boldsymbol{\Psi}|\boldsymbol{y}_i)}{\partial \beta_g} \approx (\boldsymbol{y}_i - \boldsymbol{E}_i)^T \boldsymbol{V}_i^{-1} \left(\frac{\partial \boldsymbol{E}_i}{\partial \beta_g} \right).$$

Consequently, for $g, h = 1, ..., p$, we have

$$\mathrm{E}\left[\left(\frac{\partial \ln l_i(\boldsymbol{\Psi}|\boldsymbol{y}_i)}{\partial \beta_h} \right)^T \frac{\partial \ln l_i(\boldsymbol{\Psi}|\boldsymbol{y}_i)}{\partial \beta_g} \right] \approx \left(\frac{\partial \boldsymbol{E}_i}{\partial \beta_h} \right)^T \boldsymbol{V}_i^{-1} \left(\frac{\partial \boldsymbol{E}_i}{\partial \beta_g} \right),$$

since $\mathrm{E}[(\boldsymbol{y}_i - \boldsymbol{E}_i)(\boldsymbol{y}_i - \boldsymbol{E}_i)^T] = \boldsymbol{V}_i$.

Next, let us differentiate (5.31) with respect to λ_j, $j = 1, ..., p+1$:

$$\frac{\partial \ln l_i(\boldsymbol{\Psi}|\boldsymbol{y}_i)}{\partial \lambda_j} \approx -\frac{1}{2}\mathrm{tr}\left(\boldsymbol{V}_i^{-1} \frac{\partial \boldsymbol{V}_i}{\partial \lambda_j} \right) + \frac{1}{2}(\boldsymbol{y}_i - \boldsymbol{E}_i)^T \left(\boldsymbol{V}_i^{-1} \frac{\partial \boldsymbol{V}_i}{\partial \lambda_j} \boldsymbol{V}_i^{-1} \right) (\boldsymbol{y}_i - \boldsymbol{E}_i),$$

since,

$$\frac{\partial \ln |\boldsymbol{V}_i|}{\partial \lambda_j} = \mathrm{tr}\left(\boldsymbol{V}_i^{-1} \frac{\partial \boldsymbol{V}_i}{\partial \lambda_j} \right) \quad \text{and} \quad \frac{\partial \boldsymbol{V}_i^{-1}}{\partial \lambda_j} = -\boldsymbol{V}_i^{-1} \frac{\partial \boldsymbol{V}_i}{\partial \lambda_j} \boldsymbol{V}_i^{-1}.$$

Consequently,

$$\begin{aligned}
\mathrm{E}\left[\left(\frac{\partial \ln l_i(\boldsymbol{\Psi}|\boldsymbol{y}_i)}{\partial \lambda_k} \right)^T \frac{\partial \ln l_i(\boldsymbol{\Psi}|\boldsymbol{y}_i)}{\partial \lambda_j} \right] \approx{}& \frac{1}{4}\mathrm{tr}\left(\boldsymbol{V}_i^{-1} \frac{\partial \boldsymbol{V}_i}{\partial \lambda_k} \right) \mathrm{tr}\left(\boldsymbol{V}_i^{-1} \frac{\partial \boldsymbol{V}_i}{\partial \lambda_j} \right) - \\
& \frac{1}{4}\mathrm{tr}\left(\boldsymbol{V}_i^{-1} \frac{\partial \boldsymbol{V}_i}{\partial \lambda_j} \right) \mathrm{E}\left[(\boldsymbol{y}_i - \boldsymbol{E}_i)^T \left(\boldsymbol{V}_i^{-1} \frac{\partial \boldsymbol{V}_i}{\partial \lambda_k} \boldsymbol{V}_i^{-1} \right) (\boldsymbol{y}_i - \boldsymbol{E}_i) \right] - \\
& \frac{1}{4}\mathrm{tr}\left(\boldsymbol{V}_i^{-1} \frac{\partial \boldsymbol{V}_i}{\partial \lambda_k} \right) \mathrm{E}\left[(\boldsymbol{y}_i - \boldsymbol{E}_i)^T \left(\boldsymbol{V}_i^{-1} \frac{\partial \boldsymbol{V}_i}{\partial \lambda_j} \boldsymbol{V}_i^{-1} \right) (\boldsymbol{y}_i - \boldsymbol{E}_i) \right] + \\
& \frac{1}{4}\mathrm{E}\left[(\boldsymbol{y}_i - \boldsymbol{E}_i)^T \left(\boldsymbol{V}_i^{-1} \frac{\partial \boldsymbol{V}_i}{\partial \lambda_j} \boldsymbol{V}_i^{-1} \right) (\boldsymbol{y}_i - \boldsymbol{E}_i)(\boldsymbol{y}_i - \boldsymbol{E}_i)^T \left(\boldsymbol{V}_i^{-1} \frac{\partial \boldsymbol{V}_i}{\partial \lambda_k} \boldsymbol{V}_i^{-1} \right) (\boldsymbol{y}_i - \boldsymbol{E}_i) \right].
\end{aligned} \tag{5.32}$$

To evaluate the expectations of the above quadratic forms, we apply the following results given in Magnus (1978):

For a symmetric matrix $\boldsymbol{A}$ and for $\boldsymbol{\epsilon}_i \sim \mathcal{N}_m(\boldsymbol{0}, \boldsymbol{V}_i)$, where $\boldsymbol{V}_i$ is positive definite, we have:

- $\mathrm{E}[\boldsymbol{\epsilon}_i^T \boldsymbol{A} \boldsymbol{\epsilon}_i] = \mathrm{tr}(\boldsymbol{A}\boldsymbol{V}_i)$ and

- If $\boldsymbol{B}$ is another symmetric matrix, then $\mathrm{E}[\boldsymbol{\epsilon}_i^T \boldsymbol{A}\boldsymbol{\epsilon}_i \boldsymbol{\epsilon}_i^T \boldsymbol{B}\boldsymbol{\epsilon}_i] = \mathrm{tr}(\boldsymbol{A}\boldsymbol{V}_i)\ \mathrm{tr}(\boldsymbol{B}\boldsymbol{V}_i) + 2\ \mathrm{tr}(\boldsymbol{A}\boldsymbol{V}_i\boldsymbol{B}\boldsymbol{V}_i)$.

Using these results, the last term in (5.32) becomes:

$$\frac{1}{4}\mathrm{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \lambda_j}\right)\mathrm{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \lambda_k}\right) + \frac{1}{2}\mathrm{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \lambda_j}\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \lambda_k}\right).$$

Thus,

$$\mathrm{E}\left[\left(\frac{\partial \ln l_i(\boldsymbol{\Psi}|\boldsymbol{y}_i)}{\partial \lambda_k}\right)^T \frac{\partial \ln l_i(\boldsymbol{\Psi}|\boldsymbol{y}_i)}{\partial \lambda_j}\right] \approx \frac{1}{2}\mathrm{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \lambda_j}\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \lambda_k}\right).$$

Finally, we have:

$$\mathrm{E}\left[\left(\frac{\partial \ln l_i(\boldsymbol{\Psi}|\boldsymbol{y}_i)}{\partial \beta_g}\right)^T \frac{\partial \ln l_i(\boldsymbol{\Psi}|\boldsymbol{y}_i)}{\partial \lambda_j}\right] \approx \mathrm{E}\left[\left(\frac{\partial \boldsymbol{E}_i}{\partial \beta_h}\right)^T \boldsymbol{V}_i^{-1}(\boldsymbol{y}_i - \boldsymbol{E}_i)\ \times \right.$$
$$\left.\left\{-\frac{1}{2}\mathrm{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \lambda_j}\right) + \frac{1}{2}(\boldsymbol{y}_i - \boldsymbol{E}_i)^T\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \lambda_j}\boldsymbol{V}_i^{-1}\right)(\boldsymbol{y}_i - \boldsymbol{E}_i)\right\}\right]$$

$$= \frac{1}{2}\mathrm{E}\left[\left(\frac{\partial \boldsymbol{E}_i}{\partial \beta_h}\right)^T \boldsymbol{V}_i^{-1}(\boldsymbol{y}_i - \boldsymbol{E}_i)(\boldsymbol{y}_i - \boldsymbol{E}_i)^T\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \lambda_j}\boldsymbol{V}_i^{-1}\right)(\boldsymbol{y}_i - \boldsymbol{E}_i)\right].$$

The RHS of the above equation is a sum of terms of the type $\mathrm{E}[c(y_{ij} - E_{ij})(y_{ik} - E_{ik})(y_{il} - E_{il})]$ for $j, k, l = 1, ..., m$ where c is a term-specific constant.

Since $(\boldsymbol{y}_i - \boldsymbol{E}_i) \sim \mathcal{N}_m(\boldsymbol{0}, \boldsymbol{V}_i)$, according to Holmquist (1988):

$$\mathrm{E}[(y_{i1} - E_{i1})^{r_1}...(y_{im} - E_{im})^{r_m}] = 0, \quad \text{if } r_1 + ... + r_m \text{ is odd.}$$

Therefore,

$$\mathrm{E}\left[\left(\frac{\partial \ln l_i(\boldsymbol{\Psi}|\boldsymbol{y}_i)}{\partial \beta_g}\right)^T \frac{\partial \ln l_i(\boldsymbol{\Psi}|\boldsymbol{y}_i)}{\partial \lambda_j}\right] \approx 0, \text{ for } g = 1, ..., p \text{ and } j = 1, ..., p+1.$$

Thus, the approximated FIM for the non-linear mixed effects models can be expressed as:

$$\boldsymbol{M}_i(\boldsymbol{\Psi}, \xi_i) \approx \begin{pmatrix} \boldsymbol{P}_i & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Q}_i \end{pmatrix}, \tag{5.33}$$

where the matrices $\boldsymbol{P}_i$ $(p \times p)$ and $\boldsymbol{Q}_i$ $(p+1, p+1)$ are given as follows:

$$(\boldsymbol{P}_i)_{g,h} = \left(\frac{\partial \boldsymbol{E}_i}{\partial \beta_h}\right)^T \boldsymbol{V}_i^{-1} \left(\frac{\partial \boldsymbol{E}_i}{\partial \beta_g}\right) \quad \text{for } g,\, h = 1, ..., p$$

and

$$(\boldsymbol{Q}_i)_{g,h} = \frac{1}{2}\text{tr}\left(\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \lambda_j}\boldsymbol{V}_i^{-1}\frac{\partial \boldsymbol{V}_i}{\partial \lambda_k}\right) \quad \text{for } j,\, k = 1, ..., p+1.$$

The population FIM for the N individuals is given as the sum of individual FIMs:

$$M_F(\boldsymbol{\Psi}) = \sum_{i=1}^{N} \boldsymbol{M}_i(\boldsymbol{\Psi}, \xi_i). \tag{5.34}$$

For a single group of N individuals with identical designs ξ, the population FIM becomes:

$$M_F(\boldsymbol{\Psi}) = N\boldsymbol{M}(\boldsymbol{\Psi}, \xi). \tag{5.35}$$

Unlike linear models, the FIM for non-linear models depends on the initial values of the model parameters. As can be observed from Equation (5.33), the matrices $\boldsymbol{P}_i$ and $\boldsymbol{Q}_i$ depend on the values of the parameters contained in $\boldsymbol{\Psi}$. We show this with the help of an example.

Example

To show how the FIM can be computed for non-linear mixed effects models we consider the model described in Section 5.3.2, that is:

$$C(t) = \frac{de^{-K_e t}}{V_1}. \tag{5.36}$$

The parameters of the i^{th} subject are $\boldsymbol{\beta}_i = (K_{ei}, V_{1i})^T = (\beta_{1i}, \beta_{2i})^T$ and the population parameters are $\boldsymbol{\beta} = (K_e, V_1)^T = (\beta_1, \beta_2)^T$. Let $m = 3$ blood samples be collected per subject. Then, for the i^{th} subject, the sampling design is $\xi_i = \{T_{i1}, T_{i2}, T_{i3}\}$.

Assuming an exponential error structure, the Stage 1 model can be represented as:

$$y_{ij} = \frac{de^{-\beta_{1i}T_{ij}}}{\beta_{2i}} \exp(\epsilon_{ij}),$$

where $\epsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$, $i = 1, ..., M$ and $j = 1, 2, 3$.

Taking natural logarithm on both sides, the model becomes of the form of additive

errors model:

$$\ln(y_{ij}) = \ln\left(\frac{de^{-\beta_{1i}T_{ij}}}{\beta_{2i}}\right) + \epsilon_{ij} = \ln d - \beta_{1i}T_{ij} - \ln\beta_{2i} + \epsilon_{ij}.$$

Assuming exponential random effects, the Stage 2 model $\boldsymbol{g}(\boldsymbol{\beta}, \boldsymbol{b}_i)$ can be expressed as:

$$\begin{pmatrix}\beta_{1i}\\ \beta_{2i}\end{pmatrix} = \begin{pmatrix}\beta_1\exp(b_{1i})\\ \beta_2\exp(b_{2i})\end{pmatrix},$$

where $\boldsymbol{b}_i \sim \mathcal{N}_2(\boldsymbol{0}, \sigma^2\boldsymbol{\Omega})$ and $\boldsymbol{\Omega} = \text{diag}(\omega_1, \omega_2)$.

The regression function is:

$$\eta(T_{ij}; \boldsymbol{g}(\boldsymbol{\beta}, \boldsymbol{b}_i)) = \ln\left(\frac{de^{-\beta_1\exp(b_{1i})T_{ij}}}{\beta_2\exp(b_{2i})}\right) = \ln d - \beta_1 e^{b_{1i}}T_{ij} - \ln\beta_2 - b_{2i}. \tag{5.37}$$

Using the notation defined above, at the initial values of $\boldsymbol{\beta}$ and $\boldsymbol{b}$ equal to $\boldsymbol{\beta}^o = (\beta_1^o, \beta_2^o)^T$ and $\boldsymbol{b}^o = (0,0)^T$, we obtain the matrices $\boldsymbol{G}_i$ and $\boldsymbol{H}_i$ of the model approximation (5.24),

$$\boldsymbol{G}_i = \begin{pmatrix} -T_{i1} & \frac{-1}{\beta_2^o}\\ -T_{i2} & \frac{-1}{\beta_2^o}\\ -T_{i3} & \frac{-1}{\beta_2^o}\end{pmatrix} \text{ and } \boldsymbol{H}_i = \begin{pmatrix} -\beta_1^o T_{i1} & -1\\ -\beta_1^o T_{i2} & -1\\ -\beta_1^o T_{i3} & -1\end{pmatrix}.$$

Also, the dispersion matrix $\boldsymbol{V}_i$ of the linearised response (5.24) is

$$\boldsymbol{V}_i = \sigma^2\left[\begin{pmatrix}1&0&0\\0&1&0\\0&0&1\end{pmatrix} + \begin{pmatrix} -\beta_1^o T_{i1} & -1\\ -\beta_1^o T_{i2} & -1\\ -\beta_1^o T_{i3} & -1\end{pmatrix}\begin{pmatrix}\omega_1&0\\0&\omega_2\end{pmatrix}\begin{pmatrix} -\beta_1^o T_{i1} & -\beta_1^o T_{i2} & -\beta_1^o T_{i3}\\ -1 & -1 & -1\end{pmatrix}\right].$$

$$\Rightarrow \boldsymbol{V}_i = \sigma^2\begin{pmatrix} (\beta_1^o)^2T_{i1}^2\omega_1 + \omega_2 + 1 & (\beta_1^o)^2T_{i1}T_{i2}\omega_1 + \omega_2 & (\beta_1^o)^2T_{i1}T_{i3}\omega_1 + \omega_2\\ (\beta_1^o)^2T_{i2}T_{i1}\omega_1 + \omega_2 & (\beta_1^o)^2T_{i2}^2\omega_1 + \omega_2 + 1 & (\beta_1^o)^2T_{i2}T_{i3}\omega_1 + \omega_2\\ (\beta_1^o)^2T_{i3}T_{i1}\omega_1 + \omega_2 & (\beta_1^o)^2T_{i3}T_{i2}\omega_1 + \omega_2 & (\beta_1^o)^2T_{i3}^2\omega_1 + \omega_2 + 1\end{pmatrix}.$$

$$\Rightarrow \boldsymbol{V}_i = \sigma^2(\boldsymbol{I}_3 + \omega_2\boldsymbol{1}_3\boldsymbol{1}_3^T + (\beta_1^o)^2\omega_1\boldsymbol{T}_i\boldsymbol{T}_i^T),$$

where $\boldsymbol{1}_3 = (1,1,1)^T$ and $\boldsymbol{T}_i = (T_{i1}, T_{i2}, T_{i3})^T$ is the vector which contains the sampling points for the i^{th} subject.

Then, the FIM for the i^{th} subject for $\boldsymbol{\Psi} = (\boldsymbol{\beta}^T\ \boldsymbol{\lambda}^T)^T$ where $\boldsymbol{\lambda} = (\omega_1, \omega_2, \sigma^2)^T$ is given by Equation (5.33).

The 2×2 matrix $\boldsymbol{P}_i$ can be computed using the expressions for $\boldsymbol{G}_i$ and $\boldsymbol{V}_i$ above.

To compute the matrix $\boldsymbol{Q}_i$, we need $\dfrac{\partial \boldsymbol{V}_i}{\partial \lambda_g}$, $g = 1, 2, 3$, as shown below:

$$\frac{\partial \boldsymbol{V}_i}{\partial \omega_1} = \sigma^2 (\beta_1^o)^2 \boldsymbol{T}_i \boldsymbol{T}_i^T,$$
$$\frac{\partial \boldsymbol{V}_i}{\partial \omega_2} = \sigma^2 \mathbf{1}_3 \mathbf{1}_3^T \quad \text{and}$$
$$\frac{\partial \boldsymbol{V}_i}{\partial \sigma^2} = \sigma^{-2} \boldsymbol{V}_i.$$

It can be seen from the above expressions that the FIM for the parameters $\boldsymbol{\Psi}$ is a function of the sampling times, the variances of the random effects and the initial values of the parameters.

Now we consider the case of multiple doses, where n doses of the drug are administered. Using Equations (5.1) and (5.37), the j^{th} observation from the i^{th} subject can be expressed as:

$$\ln(y_{ij}) = \ln\left(\sum_{k=1}^{n} I_{\{T_{ij} \geq t_k\}} \frac{d_k e^{-\beta_1 \exp(b_{1i})(T_{ij} - t_k)}}{\beta_2 \exp(b_{2i})}\right) + \epsilon_{ij}. \tag{5.38}$$

For the regression function in this model, the expression for the matrices $\boldsymbol{P}_i$ and $\boldsymbol{Q}_i$ will be complicated. We give below the expressions for the matrices $\boldsymbol{G}_i$ and $\boldsymbol{H}_i$ of (5.24) from which the matrices $\boldsymbol{E}_i$ and $\boldsymbol{V}_i$ and subsequently the matrices $\boldsymbol{P}_i$ and $\boldsymbol{Q}_i$ can be computed.

$$\boldsymbol{G}_i = \begin{pmatrix} e^{-\eta(T_{i1}; \boldsymbol{g}(\boldsymbol{\beta}^o, \mathbf{0}))} \sum_{k=1}^{n} I_{\{T_{i1} \geq t_k\}} d_k e^{-\beta_1^o (T_{i1} - t_k)} (\beta_2^o)^{-1} (t_k - T_{i1}) & -1/\beta_2^o \\ e^{-\eta(T_{i2}; \boldsymbol{g}(\boldsymbol{\beta}^o, \mathbf{0}))} \sum_{k=1}^{n} I_{\{T_{i2} \geq t_k\}} d_k e^{-\beta_1^o (T_{i2} - t_k)} (\beta_2^o)^{-1} (t_k - T_{i2}) & -1/\beta_2^o \\ e^{-\eta(T_{i3}; \boldsymbol{g}(\boldsymbol{\beta}^o, \mathbf{0}))} \sum_{k=1}^{n} I_{\{T_{i3} \geq t_k\}} d_k e^{-\beta_1^o (T_{i3} - t_k)} (\beta_2^o)^{-1} (t_k - T_{i3}) & -1/\beta_2^o \end{pmatrix},$$

$$\boldsymbol{H}_i = \begin{pmatrix} e^{-\eta(T_{i1}; \boldsymbol{g}(\boldsymbol{\beta}^o, \mathbf{0}))} \sum_{k=1}^{n} I_{\{T_{i1} \geq t_k\}} d_k e^{-\beta_1^o (T_{i1} - t_k)} (\beta_1^o / \beta_2^o) (t_k - T_{i1}) & -1 \\ e^{-\eta(T_{i2}; \boldsymbol{g}(\boldsymbol{\beta}^o, \mathbf{0}))} \sum_{k=1}^{n} I_{\{T_{i2} \geq t_k\}} d_k e^{-\beta_1^o (T_{i2} - t_k)} (\beta_1^o / \beta_2^o) (t_k - T_{i2}) & -1 \\ e^{-\eta(T_{i3}; \boldsymbol{g}(\boldsymbol{\beta}^o, \mathbf{0}))} \sum_{k=1}^{n} I_{\{T_{i3} \geq t_k\}} d_k e^{-\beta_1^o (T_{i3} - t_k)} (\beta_1^o / \beta_2^o) (t_k - T_{i3}) & -1 \end{pmatrix}.$$

As can be seen from this example, the computation of the FIM even for a simple compartmental model is a non-trivial problem and computationally expensive. The optimisation of a function of the FIM is an even more difficult problem. However, a variety of software are available for computation of the optimum design. The general procedure in these software is to numerically approximate the FIM in each iteration and then use special algorithms to find the design which will give a better value of the objective function (the desired function of the FIM which is to be optimised) in the next iteration. Such iterations continue until a user-specified convergence criterion has been met. At the termination of the iterations, the software will report the optimum design along with the optimised value of the objective function.

The two software which are the most commonly used are *PopED* (Foracchia et al. (2004)), a MATLAB® based software and *PFIM* (Bazzoli et al. (2009)), an *R* based package. Nyberg et al. (2015) compared these two software along with three others and concluded that they give similar results. In this thesis, we use PopED for computation of the optimal blood sampling points. For us, the choice of PopED was natural as the ED algorithm had already been programmed in MATLAB® and so it was easy to embed the submodules of PopED in our master code. We briefly explain below some nuances of PopED which will be useful for our work in this thesis.

5.5 PopED for Design Optimisation

In this section, we explain some of the features in PopED which are relevant to our work in this thesis along with the underlying statistical theory. The material of this section is adapted from Foracchia et al. (2004), the scientific paper associated with PopED and PopED Manual (2012).

The FIM is computed by using Equation (5.29) and approximating the likelihood function FO or FOCE methods. PopED provides options for optimising over several variables, for example the number of samples per subject, sampling schedule or the covariates. For our work, we only optimise the FIM with respect to the sampling schedule, i.e., the sampling time points.

The derivatives are computed by PopED numerically. Two methods are available for this: complex difference and central difference.

The complex difference method is the default method for finding derivatives in PopED and is robust and faster than the central difference method. A prerequisite for this method is that the model must be able to handle complex numbers. If h is taken to be the step size, the complex derivative of a function $F(x)$ at x_o is defined as:

$$F'(x_o) = \frac{\text{Im}\left(F(x_o + ih)\right)}{h},$$

where Im(.) gives the imaginary part of the argument. The step h is generally taken to be 10^{-8}.

If the function is not amenable to complex numbers, then the central difference method is used to find derivatives. This method is from the calculus of finite differences. If h is taken to be the uniform width of the abscissae, then the first derivative of a function $F(x)$ at x_o is defined as:

$$F'(x_o) = \frac{F(x_o + h) - F(x_o - h)}{2h}.$$

As mentioned earlier, for our work we are interested in finding D-optimal time points to collect the blood samples. This consists in maximising the determinant or the natural logarithm of the determinant of the FIM. The algorithms available in PopED for searching for the best design, i.e. the ξ which maximises the objective function are:

- **Random Search** : This method begins with a random search over the entire search space but as soon as it finds a good objective function's value, it collapses to an Adaptive Narrowed Random Search. It will then search close to the previous best objective function value. Thus, while the Random Search is global, the Adaptive Narrowed Random Search is local.

- **Stochastic Gradient**: This is a local search method. For a D-optimal design, this algorithm is simply the steepest descent algorithm.

- **Line Search**: Line search does a grid search in one dimension, i.e. over one parameter at a time. It is a global search method but it alone does not converge to the global optimum even if infinitely many grid points are taken. Therefore, this method must be used in conjunction with a local search method.

- **Modified Fedorov Exchange Algorithm**: This method consists in searching for that sample which changes the objective function's value the most. That sample is then exchanged into the current optimal design and the next iteration is performed with the current optimal design. It terminates when the exchange results in an increase in the objective function's value which is less than a threshold value. Ogungbenro et al. (2005) describe this method in detail.

- **BFGS Minimizer (Broyden-Fletcher-Goldfarb-Shannon)**: This method uses a quasi-Newton based method to do a local search. It is described in detail in Battiti and Masulli (1990). It is an alternative to the stochastic gradient method and so either of them can be used at a time.

The authors of PopED recommend that all the search types should be used together to obtain the best result with the exception of the modified Fedorov exchange algorithm which does not work with other methods and has to be used alone. For our work, we use the random search, the stochastic gradient and the line search methods simultaneously.

For optimisation of the sampling time points, the following inputs are required to be entered or selected by the user:

- The PK model and the variability models. The variability models consist of the intra-individual variability and the inter-individual variability (the random

effects). To specify the PK model, PopED comes with a library from which a model can be selected. It also provides a feature for the user to specify a custom model. In Chapter 6, we shall use this feature to define our own multiple dose model described in Section 5.1. Similarly, the user can choose the error structure from the options available or specify a custom model. The error models defined in Section 5.2 are all available in PopED.

- As can be seen from Equation (5.34), the FIM depends on the parameter vector $\mathbf{\Psi}$. Therefore, initial values of the parameters contained in the vector $\mathbf{\Psi}$ need to entered. The computation of the objective function's value will be at these parameter values.

- The optimality criterion (D-optimality for our work), the log-likelihood approximation type and the search algorithm.

- PopED provides the option to optimise the FIM with respect to other design variables e.g., the number of samples per subject, covariates and the dose sizes. We do not use these features for our work.

- The design region for the m samples needs to be specified. Along with the minimum and maximum values of time at which samples are permissible to be collected, initial values for each of the samples needs to be specified to enable PopED to compute the optimal values.

- Convergence settings of the search algorithms and FIM approximation can also be changed by the user, if required.

Having discussed the methods and the software available for estimation of the parameters and for optimal design of the experiment for the case of non-linear mixed effects models, we are now in a position to describe another feature of the ED algorithm introduced in Chapter 4. In the next chapter, we show the applicability of the ED algorithm when the parameters of the mechanistic model describing the drug concentrations are unknown. This will be done in an adaptive setting where the knowledge about the model is updated in each cohort, using the techniques of estimation and optimal design. Based on this knowledge, the most efficient dose regimen is administered to the next cohort with the help of the ED algorithm. PK samples are then collected from this cohort and this data, along with the data collected from previous cohorts are used to find the updated estimates of the parameters. This procedure is continued until a stopping rule is applied.

Chapter 6

The Efficient Dosing Algorithm for the Case of Unknown Parameters

The ED algorithm introduced in Chapter 4 to solve the problems described in Section 3.3 requires knowledge of the PK models and estimates of the model parameters. In Chapter 4, we assumed that the estimated mean parameters are available before the computation of the optimal dose regimen. In this chapter, we show that the ED algorithm can be used in an adaptive setting to simultaneously optimise the dose regimen for the population and estimate the unknown model parameters and their variances. This chapter describes the methodology and an associated simulation study for the problem of Type 1 for a single drug described in Section 3.3. The proposed method can also be adapted for optimising Type 1 problem for combination therapies and for Type 2 problems.

In adaptive clinical trials, data are analysed at interim stages and changes are made in the trial so that it is more efficient (in terms of, e.g., shorter duration or lesser number of volunteers enrolled) or is more informative (in terms of, e.g., exposing broader dose-response relationship or demonstrating the efficacy of the test drug, if it exists), FDA (2010). For these reasons, such trial designs are also referred to as flexible. The advantages of adaptive designs include optimum utilisation of resources which generally translates to improved economics and enhanced ethical standards through quicker elimination of substandard treatments or dosages and reduction in the sample size. A disadvantage of such designs is that the added complexity in the trial protocol may be difficult to follow and may cause operational challenges. Chow and Chang (2008) present a review of the types of adaptive designs, the practical issues that are encountered in their application and some practical examples. The setting in which we will study the ED algorithm in this chapter is adaptive in nature since it will involve interim analyses on cohorts of patients and the results will be used to decide the further course of the trial.

In this chapter, we use the statistical models discussed in Chapter 5. The variability in the PK parameters and the errors in the measured concentrations are also included in the model. This differentiates such statistical models from the deterministic models discussed in Chapter 4.

6.1 Dose Regimen Optimisation in an Adaptive Trial

An adaptive approach is planned to ascertain the average PK parameters of the drug in a population while minimising the desired objective function $\vartheta(\mathcal{R})$ defined in Section 3.3. We start with initial values of the model parameters for the first cohort and administer the optimised dose regimen to it. Based on the initial values of the parameters, optimal time points are computed for collection of blood samples from the subjects contained in the first cohort. PK sampling is then performed and the collected samples are used to estimate the model parameters. This procedure is repeated for the next cohort, using the parameter estimates obtained from the first cohort. For each cohort, the estimates are based on the information contained in the current as well as previous cohorts. To stop the adaptive procedure, different stopping rules will be considered. The methodology is formally introduced below.

6.1.1 Notation

We consider the notation introduced in Chapter 5. The parameters of interest can be put in vector form as: $\boldsymbol{\Psi} = (\boldsymbol{\beta}^T, \boldsymbol{\omega}^T, \sigma^2)^T$. Furthermore, we introduce the following symbols:

C: maximum number of cohorts in the adaptive procedure. Depending on a stopping rule used the number of cohorts can be smaller than C;

c: number of subjects in each cohort;

m: number of PK samples per subject;

$\widehat{\boldsymbol{\Psi}}_i$: vector of estimates of the parameters updated at the i^{th} cohort, $i = 1, ..., C$;

$\boldsymbol{\Psi}_{true}$: the vector of true values of the parameters for the population (used for simulation studies);

$\boldsymbol{\Psi}_o$: the vector of values of the parameters which is available *a priori* (initial values);

ξ_i^*: set of optimal sampling time points for the i^{th} cohort. ξ_1^* will be computed using the initial values $\boldsymbol{\Psi}_o$. Subsequent optimal sampling times will be computed using the updated vector of parameters $\widehat{\boldsymbol{\Psi}}_i$ for each cohort i, where $i = 2, ..., C$;

$\boldsymbol{D}_i^*$: the vector of doses optimising the objective function ϑ; the doses administered to the i^{th} cohort, where $i = 1, ..., C$.

6.1.2 The Adaptive Procedure

We now explain the steps of the methodology for optimisation of $\vartheta(\mathcal{R}_n(\boldsymbol{D}|n, \boldsymbol{t}, T, \theta)$, defined in Section 3.3, for the objective of minimising the area around a pre-chosen C_{tgt}. The methodology for optimisation of $\vartheta(\mathcal{R}_n(\boldsymbol{D}|n, \boldsymbol{t}, T, \theta)$ for maintenance of the concentration between a therapeutic range, say, (C_{tgt}^-, C_{tgt}^+) will be the same with the exception of the ED algorithm's version that is used to find the efficient dose regimen for each cohort.

The PK model followed by the drug's concentrations is assumed to be known up to the model parameters. The prefixed (given) number of doses and the dosing time points are n and $\boldsymbol{t}$ respectively. Let the current cohort under study be number i, $i = 1, ..., C$. The adaptive procedure is presented in the flowchart shown in Figure 6.1. The steps are further explained below.

Step 1. Optimisation of the dose regimen for cohort 1.

Initial values of the model parameters are available at this step. For cohort 1, these will be contained in the vector of initial values $\boldsymbol{\Psi}_o$. Using these values, the optimal dose regimen $\boldsymbol{D}^*$ is obtained by the ED algorithm and administered to each of the c subjects in cohort 1 at the dosing time points $\boldsymbol{t}$ (pre-fixed).

Step 2. Computation of optimal PK sampling time points and collection of samples.

Using $\widehat{\boldsymbol{\Psi}}_{i-1}$ and $\boldsymbol{D}_i^*$, ξ_i^* is computed. ξ_i^* contains the m population D-optimal blood sampling time points for cohort i. From every subject in cohort i, m samples are collected at the times contained in ξ_i^*. Thus, we have $c \times m$ blood samples from this cohort. From these samples, the drug's concentrations at the time points contained in ξ_i^* are determined for each subject in the cohort.

Step 3. Computation of $\widehat{\boldsymbol{\Psi}}_i$.

The concentration values obtained from the previous step, together with the val-

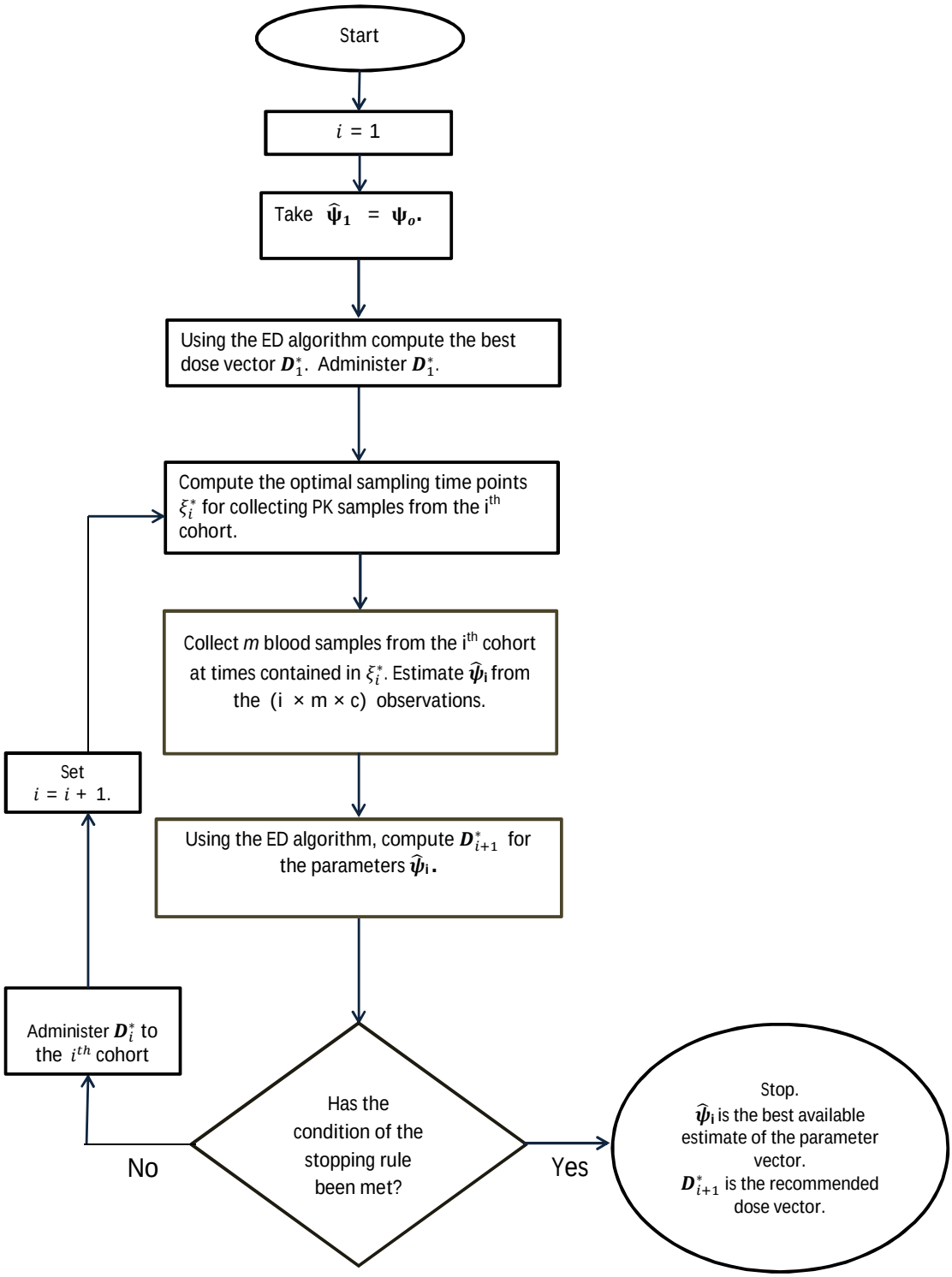


Figure 6.1: Flowchart explaing the methodology of using the ED algorithm in an adaptive setting, serving the dual purpose - parameter estimation and dose regimen optimisation to adhere to a desired target concentration or range of concentrations.

ues obtained from all the previous cohorts are now used to compute $\widehat{\boldsymbol{\Psi}}_i$, the updated estimate of the model parameters. For estimation, any appropriate method, including the ones mentioned in Section 5.3 can be used. For our work, we apply the LME method which was discussed in Section 5.3.

Step 4. Computation of $\boldsymbol{D}^*_{i+1}$ and stopping rule test

The ED algorithm is run to find $\boldsymbol{D}^*_{i+1}$ corresponding to the estimates contained in $\widehat{\boldsymbol{\Psi}}_i$. At this step, it is examined if the condition of a stopping rule is met. If a stopping rule applies at this cohort, the procedure is terminated with $\widehat{\boldsymbol{\Psi}}_i$ being the best available estimate of the parameters and $\boldsymbol{D}^*_{i+1}$ being the recommended optimal dose vector.

If no stopping rule is applied at this step, $\boldsymbol{D}^*_{i+1}$ is administered to the next cohort, i.e., cohort $i+1$. We then go back to step 2 with i replaced with $i+1$.

These steps are repeated until a stopping rule terminates the procedure or the maximum number of cohorts (C) have been analysed. The parameter estimates derived from the cohort at which the stopping rule applies are denoted by $\widehat{\boldsymbol{\Psi}}$. In this chapter, unless explicitly mentioned otherwise, $\widehat{\boldsymbol{\Psi}}$ will refer to the estimates obtained by the application of SR1, i.e., when all the C cohorts available are analysed.

Thus, during the course of the iterations, the estimates are expected to improve because of the accumulating data while minimising the objective function $\vartheta(\mathcal{R}_n(\boldsymbol{D}|n, \boldsymbol{t}, T, \theta))$ for the best available estimate of the vector of parameters.

As discussed previously, a stopping rule is required to terminate the adaptive procedure. We consider three stopping rules defined below.

6.1.3 Stopping Rules

We consider three stopping rules as discussed below:

SR1. The procedure is terminated when all the C cohorts have been analysed. This is the most trivial rule.

SR2. The procedure is terminated at the i^{th} iteration if,

$$\frac{|\mathbf{1}_n^T \boldsymbol{D}^*_{i+1} - \mathbf{1}_n^T \boldsymbol{D}^*_i|}{\mathbf{1}_n^T \boldsymbol{D}^*_i} \leq .05,$$

where $\mathbf{1}_n$ is a column vector of order n. In other words, the procedure is terminated when the cumulative dose for the current cohort differs from the cumulative dose of the previous cohort by less than 5% of the latter. This stopping rule should perform well when the elements of $\boldsymbol{D}$ are almost of the same magnitude. However, if e.g., the loading dose is quite large as compared to the maintenance doses, SR2 will be dominated by the larger dose and therefore, a stricter stopping rule should be enforced, such as the one given below.

SR3. Terminate the procedure at the i^{th} iteration if

$$\bigcap_{j=1}^{n} \left\{ \frac{|d^*_{(i+1)j} - d^*_{ij}|}{d^k_{ij}} \leq .05 \right\} \text{ is a true event,}$$

where d^*_{ij} is the j$^{\text{th}}$ dose in the dose vector $\boldsymbol{D}^*_i$, $i \in \{1, ..., C\}$ and $j = 1, ..., n$. That is, the procedure is terminated when every element of the optimal dose vector $\boldsymbol{D}^*_{i+1}$ for the $(i+1)^{th}$ cohort differs from the corresponding elements of $\boldsymbol{D}^*_i$ by not more than 5% of the latter. SR3 is a stricter rule than SR2, as when SR3 gets triggered, that will imply that the condition of SR2 has also been met. It may be noted here that since the stopping rule can become effective before the last cohort is analysed, the actual number of cohorts which are involved in the adaptive procedure can be less than C.

6.2 Simulation Studies

To explain the above methodology, we use an example. The example is for a hypothetical drug, for which the mechanistic model which describes the concentration-time relationship is known, but the model parameters $\boldsymbol{\Psi}$ are unknown. However, initial values of the parameters, $\boldsymbol{\Psi}_o$, are available.

The code for this simulation study is given in Appendix D.4 and was implemented in MATLAB®. The studies were run on a PC with Intel® Core i5-4210U CPU @ 1.70 GHz. The *poped* submodule which computes the D-optimal sampling time points for given values of $\boldsymbol{D}$ and $\boldsymbol{\Psi}$ has been adapted from PopED software, Foracchia et al. (2004). The estimation of parameters was carried out using the in-built libraries of MATLAB® software. The ED algorithm was used for dose regimen optimisation.

The objective is to determine and estimate $\boldsymbol{\Psi}$ and to find a dose vector $\boldsymbol{D}^*$ such that $\vartheta(\mathcal{R}_n|n, \boldsymbol{t}, T, \theta)$ is minimised. In the problem considered in this chapter, ϑ will correspond to the case of maintenance of target concentration. That is, to minimise the deviations about a pre-chosen C_{tgt}. This will be solved using the methodology described earlier in the chapter. We firstly define the PK model for our study.

6.2.1 Example and the Simulation Study Set-up

The concentration-time curve is given by the one compartment first order absorption given in Equation (2.3) and recalled here:

$$C(t) = \frac{dK_{a_i}}{V(K_{a_i} - K_{e_i})}(e^{-K_{e_i}t} - e^{-K_{a_i}t}). \tag{6.1}$$

To handle the population PK sampling described in the above methodology, we use the theory of non-linear mixed effects models described in Chapter 5. Also, since multiple doses of the drug are administered, we use Equation (5.1) to express the drug's concentration in a sample collected at time T_j.

Let $\boldsymbol{\beta_i} = (K_{ai}, K_{ei}, V_i)^T$ be the vector of PK parameters of the i^{th} individual and $\boldsymbol{\beta} = (K_a, K_e, V)^T$ be the vector of mean PK parameters for the population. The model for the random effects is:

$$\begin{pmatrix} K_{a_i} \\ K_{e_i} \\ V_i \end{pmatrix} = \begin{pmatrix} K_a \exp(b_{1i}) \\ K_e \exp(b_{2i}) \\ V \exp(b_{3i}) \end{pmatrix}, \tag{6.2}$$

where $\boldsymbol{b}_i = (b_{1i}, b_{2i}, b_{3i})^T$ is a vector of random effects such that $\boldsymbol{b}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega})$. In this thesis, we assume $\boldsymbol{\Omega}$ to be a diagonal matrix. Let $\boldsymbol{\omega} = (\omega_1, \omega_2, \omega_3)^T$ represent the three diagonal elements of $\boldsymbol{\Omega}$.

Then the multiple dose model of observations can be expressed as a non-linear mixed effects model shown below:

$$y_{ij} = \sum_{k=1}^{n} I_{\{T_j \geq t_k\}} \frac{d_k K_a e^{b_{1i}}}{V e^{b_{3i}}(K_a e^{b_{1i}} - K_e e^{b_{2i}})} \left(e^{-K_e e^{b_{2i}}(T_j - t_k)} - e^{-K_a e^{b_{1i}}(T_j - t_k)} \right) e^{\epsilon_{ij}}, \tag{6.3}$$

where y_{ij} is the j^{th} sample from the i^{th} subject at time T_j, where $i = 1, ..., c$ and $j = 1, ..., m$.

The random errors, ϵ_{ij}, are normally distributed, that is $\epsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$ and they are assumed to be independent of the elements in $\boldsymbol{b}_i$'s. The vector of parameters for this model is given by: $\boldsymbol{\Psi} = (\boldsymbol{\beta}^T, \boldsymbol{\omega}^T, \sigma^2)^T = (K_a,\ K_e,\ V,\ \omega_1,\ \omega_2,\ \omega_3,\ \sigma^2)^T$.

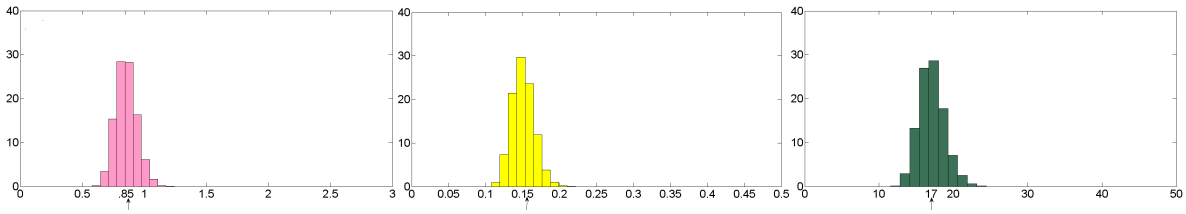
For simulation purposes, we assume that the true values of the parameters are:

$$\boldsymbol{\Psi}_{true} = (.85, .15, 17, .1, .1, .1, .1)^T. \tag{6.4}$$

The variance 0.1 (i.e. standard deviation of about 0.32) of the random effects of $\boldsymbol{b}_i$ amounts to approximately 32% coefficient of variation (CV) in the PK parameters in $\boldsymbol{\beta}_i$ in log scale. In Appendix C this relationship between the normal and the

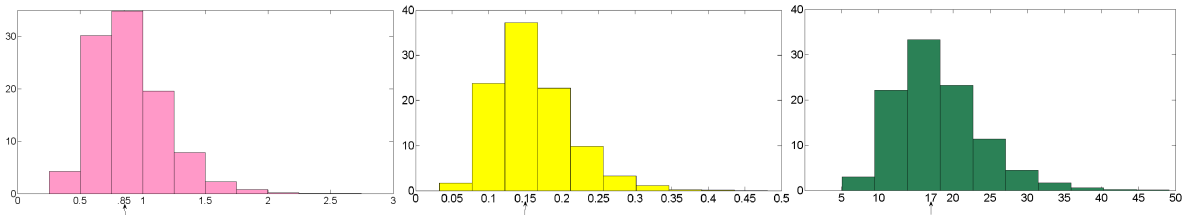
lognormal distributions is explained. We shall see later in Figure 6.5 that this choice of variance produces a reasonable degree of inter-individual variability in the simulated population.

Figures 6.2 and 6.3 show how the distributions of subjects' PK parameters depend on the variance parameters $\boldsymbol{\omega}$. As expected, the spread in the simulated distributions increases with the increase in variance of the random effects. Also, if $X \sim \log \mathcal{N}(\mu, \sigma^2)$, the skewness in X is given as $\text{skew}(X) = (e^{\sigma^2} + 2)\sqrt{e^{\sigma^2} - 1}$, Weisstein (2016). Thus, a higher variance results in greater skewness, as can be seen from these figures. For our study, we choose the following values:



(a) K_{ai} (b) K_{ei} (c) V_i

Figure 6.2: Simulated distributions of elements of $\boldsymbol{\beta}_i$ using the Stage II model expressed in Equation (6.2). The mean PK parameters are $\boldsymbol{\beta}_{true} = (.85, .15, 17)^T$ and $\boldsymbol{\omega} = (.01, .01, .01)^T$.



(a) K_{ai} (b) K_{ei} (c) V_i

Figure 6.3: Simulated distributions of elements of $\boldsymbol{\beta}_i$ using the Stage II model expressed in Equation (6.2). The mean PK parameters are $\boldsymbol{\beta}_{true} = (.85, .15, 17)^T$ and $\boldsymbol{\omega} = (.1, .1, .1)^T$.

n: the number of doses, is taken to be 5.
$\boldsymbol{t}$: the dosing time points are $\boldsymbol{t} = (0, 8, 16, 24, 32)^T$ h.
The target concentration $C_{tgt} = 5$ mg/L is to be maintained for $T = 40$ hours.
c: the cohort size is 10 subjects. Although we consider cohorts of equal sizes, the algorithm permits the use of unequal cohort sizes.
m: the number of PK samples per patient is taken to be 3. As there are three fixed effects in the model, for estimation of $\boldsymbol{\Psi}$, at least three samples are required to be collected from each subject. Although we consider equal number of blood samples per patient for all the cohorts, the algorithm permits the use of unequal number of samples.
C: the number of cohorts is 10. We consider some other values of C in the section on

sensitivity analysis.
κ: the discretised version of the ED algorithm is used with $\kappa = 10$. This means that the dose sizes can only be in multiples of 10. The resolution can also be chosen as δ if the doses are permitted to be real numbers. We consider discretised doses to fasten the convergence of the algorithm.
$d_{max} = 200$ mg is the maximum dose which can be administered.
The initial dose sets for the ED algorithm are taken as $L_i^k = \{20, 100, 200\}$.
The design region for the purpose of PK sampling is taken to be [.1, 48] h. For reasons of practicality, a minimum gap of 0.25 h is enforced between any two successive sampling time points.

The initial values of the parameters available before a simulation is commenced are: $\boldsymbol{\Psi}_o = (1, .1, 20, .05, .15, .05, .15)^T$. Each of the N_{sim} simulations starts with these initial values.

Based on $\boldsymbol{\beta}_{true} = (.85, .15, 17)^T$ and $\kappa = 10$, the φ_A-efficient dose vector is $\boldsymbol{D}^* = (140, 90, 90, 100, 90)^T$ and the associated efficiency measure is $\varphi_A(\boldsymbol{\Delta}^*|\kappa = 10) = 7.1561$. Since we retain the same resolution in this chapter, $\varphi_A(\boldsymbol{\Delta}^*|\kappa = 10)$ is written as just φ_A for convenience. This is the ideal dose regimen which should be administered to the population. It will be unknown in practice but in a simulation study, it can be used as a benchmark to evaluate the performance of our methodology.

For $\boldsymbol{\Psi}_{true}$ given in (6.4) and the chosen design region [.1, 48], the three D-optimal points are: $\xi^* = \{.1, 6.60, 48\}$.

Figure 6.4 shows the concentration profile generated by $\boldsymbol{\beta}_{true}$ and $\boldsymbol{D}^*$. Also plotted on the same graph are the D-optimal time points ξ^* for the vector of parameters $\boldsymbol{\Psi}_{true}$ and the model (6.3).

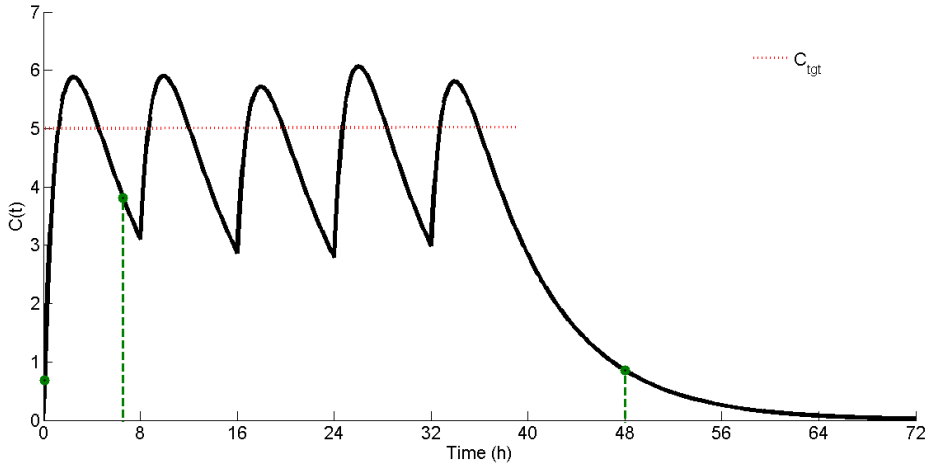
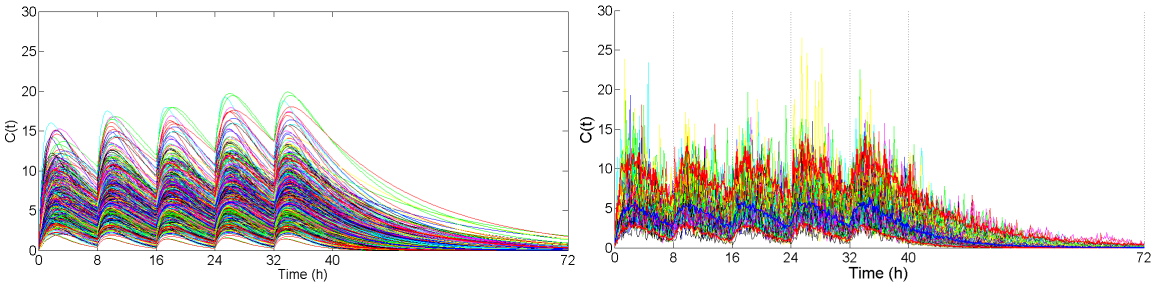


Figure 6.4: Concentration profile for the population parameters $\boldsymbol{\beta}_{true}$ and dose regimen $\boldsymbol{D}^*$, along with the three D-optimal PK sampling time points.

To give an idea of the inter- and intra-individual variability generated in the concentration-profiles by $\mathbf{\Psi}_{true}$, we simulated some profiles as shown in Figure 6.5. The range of the simulated profiles is quite wide. For example, the lowest C_{max} value is around 2 mg/L while the highest is around 20 mg/L in Figure 6.5a. Note that the simulated profiles in Figure 6.5a are smooth as they are obtained by setting the residual error variance to 0, i.e., $\sigma_{true} = 0$. On the other hand, the profiles in Figure 6.5b are obtained by setting $\sigma_{true} = .1$. The inclusion of this intra-individual variability makes the simulated profiles fuzzy.



(a) $\sigma_{true} = 0$ in $\mathbf{\Psi}_{true}$ (b) $\sigma_{true} = 0.1$ in $\mathbf{\Psi}_{true}$

Figure 6.5: Simulated PK profiles for model (6.3) for parameters $\mathbf{\Psi}_{true}$. (a): 1000 profiles were generated. (b): 50 profiles were generated (to keep the figure comprehensible).

It is important at this stage to differentiate between 'an iteration', 'a simulation' and 'a simulation study' in the context of our work. An iteration refers to a single cohort and is defined as the process of collection of blood samples at optimal time points, estimation of the model parameters (from all available data) and optimisation of the dose regimen to be applied in the next iteration.

On the other hand, by a simulation we mean the complete adaptive procedure starting from cohort 1 to the last cohort used in the trial. Therefore, a simulation will consist of maximum C iterations.

A number of simulations should be performed so that the aggregated data can offer more insights into the performance of the algorithm by averaging out random variations. A simulation study consists of N_{sim} simulations. To determine an appropriate value of N_{sim} for our work, we ran many test simulation studies. We observed that while there was variability in the parameter estimates in the beginning of each study, they stabilised around the 600^{th} simulation. We therefore decided to take $N_{sim} = 1000$.

In step 3 of the methodology described in Section 6.1.2, we mentioned about analysing the PK samples collected from the cohort in step 2. For the simulation study we use the assumed true values of the model parameters and simulate the samples at the required optimal time points. Using *normrnd* function in MATLAB®

the random vector $(r_1, r_2, r_3)^T$ is generated from $\mathcal{N}(\mathbf{0}, \mathbf{\Omega})$, where $\mathbf{\Omega} = \text{diag}(.1, .1, .1)$ and r_4 is independently generated from $\mathcal{N}(0, \sigma^2)$, where $\sigma^2 = 0.1$.

Then the j^{th} observed concentration for subject i at time T_j is simulated as:

$$y_{ij} = \sum_{k=1}^{n} I_{\{T_j \geq t_k\}} \frac{d_k K_a e^{r_1}}{V e^{r_3}(K_a e^{r_1} - K_e e^{r_2})} \left(e^{-K_e e^{r_2}(T_j - t_k)} - e^{-K_a e^{r_1}(T_j - t_k)}\right) e^{r_4}, \qquad (6.5)$$

where $\boldsymbol{D} = (d_1, ..., d_5)^T = (140, 90, 90, 100, 90)^T$, $\boldsymbol{t} = (t_1, ..., t_5)^T = (0, 12, 24, 36, 48)^T$, $K_a = 0.85$, $K_e = 0.15$, $V = 17$, $j = 1, 2, 3$, $i = 1, ..., 10$ and $n = 5$.

These concentration values are simulated in every iteration of all the simulations in the study.

In every simulation, $\boldsymbol{D}_1^*$ and ξ_1^* are computed using the initial values $\mathbf{\Psi}_o$. We use the same initial values in every simulation.

At the end of the simulation study, the data generated from the N_{sim} simulations need to be analysed to test the validity of our methodology. This is done by examining the convergence of the estimated vector of parameters $\mathbf{\Psi}$ and the performance of the stopping rules.

For the simulation study, we define the following metrics to accomplish the above:

- We define the average stopping cohort number (ACN) for the j^{th} stopping rule as:

$$\bar{C}_j = \frac{1}{N_{sim}} \sum_{k=1}^{N_{sim}} C_j^{(k)},$$

  where $C_j^{(k)}$ denotes the cohort number at which SRj is applied, $j = 2, 3$ and $k = 1, ..., N_{sim}$.

  If in a simulation the respective conditions of SRj are not met even after the last, i.e., the C^{th} cohort has been analysed, we let $C_j^{(k)} = C$ for $j = 2, 3$. For SR1, by definition, the stopping rule is applied at cohort C.

  By definition of SR1, $\bar{C}_1 = C$ and $\bar{C}_j \leq \bar{C}_1$, for $j = 2, 3$. Also, as indicated previously, $\bar{C}_2 \leq \bar{C}_3$.

- The final estimated vector of parameters is obtained by averaging $\widehat{\mathbf{\Psi}}^{(k)}$ over the N_{sim} simulations:

$$\bar{\widehat{\mathbf{\Psi}}} = \frac{1}{N_{sim}} \sum_{k=1}^{N_{sim}} \widehat{\mathbf{\Psi}}^{(k)},$$

  where $\widehat{\mathbf{\Psi}}^{(k)}$ is defined as the estimate of $\mathbf{\Psi}$ obtained in the k^{th} simulation. As mentioned earlier, $\widehat{\mathbf{\Psi}}^{(k)}$ corresponds to the case when SR1 is used to terminate

the simulation, unless mentioned otherwise.

- The mean administered dose regimen is obtained by averaging $\boldsymbol{D}^{*^{(k)}}$ over the N_{sim} simulations:

$$\boldsymbol{D}^{*} = \frac{1}{N_{sim}} \sum_{k=1}^{N_{sim}} \boldsymbol{D}^{*^{(k)}},$$

  where $\boldsymbol{D}^{*^{(k)}}$ is defined as the dose regimen computed by the ED algorithm for the vector of parameters $\widehat{\boldsymbol{\Psi}}^{(k)}$. Like $\widehat{\boldsymbol{\Psi}}^{(k)}$, $\boldsymbol{D}^{*^{(k)}}$ refers to the case when SR1 is used to terminate the simulation, unless mentioned otherwise.

- The average φ_A-efficiency measure is obtained by averaging $\varphi_A^{(k)}$ over the N_{sim} simulations:

$$\bar{\varphi}_A = \frac{1}{N_{sim}} \sum_{k=1}^{N_{sim}} \varphi_A^{(k)},$$

  where $\varphi_A^{(k)}$ is the φ_A-efficiency of $\boldsymbol{D}^{*^{(k)}}$.

An estimate of the bias of estimator $\widehat{\boldsymbol{\Psi}}$ is given as:

$$\widehat{\text{bias}}\left(\widehat{\boldsymbol{\Psi}}\right) = \widehat{\text{E}}\left(\widehat{\boldsymbol{\Psi}}\right) - \boldsymbol{\Psi}_{true} = \frac{1}{N_{sim}} \sum_{k=1}^{N_{sim}} \widehat{\boldsymbol{\Psi}}^{(k)} - \boldsymbol{\Psi}_{true} = \bar{\widehat{\boldsymbol{\Psi}}} - \boldsymbol{\Psi}_{true}.$$

Since the parameters differ substantially in their magnitude, the bias for a parameter can also be reported in a more useful way by expressing it as a percentage of the corresponding true parameter value. For example, the estimated percentage bias in $\bar{\widehat{K}}_a$ is:

$$\frac{\bar{\widehat{K}}_a - K_{a_{true}}}{K_{a_{true}}} \times 100.$$

An estimate of the mean squared error (mse) of the estimator $\widehat{\boldsymbol{\Psi}}$ is given as:

$$\widehat{\text{mse}}\left(\widehat{\boldsymbol{\Psi}}\right) = \text{diag}\left(\widehat{\text{E}}\left[(\widehat{\boldsymbol{\Psi}} - \boldsymbol{\Psi}_{true})(\widehat{\boldsymbol{\Psi}} - \boldsymbol{\Psi}_{true})^T\right]\right)$$

$$= \text{diag}\left(\frac{1}{N_{sim}} \sum_{k=1}^{N_{sim}} (\widehat{\boldsymbol{\Psi}}^{(k)} - \boldsymbol{\Psi}_{true})(\widehat{\boldsymbol{\Psi}}^{(k)} - \boldsymbol{\Psi}_{true})^T\right).$$

$\text{var}(\widehat{\boldsymbol{\Psi}})$ can be estimated by replacing $\boldsymbol{\Psi}_{true}$ with $\bar{\widehat{\boldsymbol{\Psi}}}$ in the above formula.

Furthermore, $\bar{\widehat{\boldsymbol{\Psi}}}$ and $\widehat{\text{var}}(\widehat{\boldsymbol{\Psi}})$ can be used to find the coefficients of variations (CV) of the parameters as:

$$\text{CV}(\widehat{K}_a) = \left(\sqrt{\widehat{\text{var}}(\widehat{K_a})} \times 100\right) \div \bar{\widehat{K}}_a.$$

6.2.2 Results of the Simulation Studies

The study described above was carried out for $N_{sim} = 1000$ simulations and the results are summarised below. We examine the performances of the three stopping rules, the distributions of the estimated parameters, the distributions of the D-optimal sampling time points and the efficiencies of the administered dose regimens.

Stopping Rules

The average cohort number (ACN) for SR2 was computed as $\bar{C}_2 = 3.81$ and for SR3, it was $\bar{C}_3 = 7.52$. Table 6.1 shows the percentage distribution of $C_j^{(k)}$, $j = 2, 3$, for the two stopping rules. As expected, SR2 tends to get applied earlier than SR3 because of the way they have been defined. It can be inferred from the table that the probabilities that the adaptive trial is terminated at the 5^{th} cohort or earlier according to the two stopping rules are about 0.86 and 0.28. The high percentage for cohort 10 for the case of SR3 is due to the fact that it includes both - simulations in which SR3 was actually applied and the simulations in which SR3 did not apply but the maximum number of cohorts ($C = 10$) was reached.

Cohort	$C_2^{(k)}$	$C_3^{(k)}$
1	0.0	0.0
2	19.5	0.0
3	36.3	6.8
4	20.0	11.0
5	10.5	10.2
6	5.5	8.7
7	3.1	8.0
8	1.7	7.8
9	1.9	9.3
10	1.5	38.2

Table 6.1: Percentage distributions of the cohort number at which stopping rules SR2 and SR3 were applied. For example, at cohort number 3, SR2 was applied 36.3% of the times while SR3 was applied just 6.8% of the times.

As mentioned before, if the condition of a stopping rule is not met even at the last available cohort, then the trial stops. Therefore, the numbers shown in the table for the 10^{th} cohort are the percentage of the simulations for which the conditions of SR2 and SR3 were not met at the preceding nine cohorts.

Estimates of the Model Parameters

The parameters' estimates were computed for each cohort for $N_{sim} = 1000$ simulations. Table 6.2 presents the estimates at each cohort, averaged over the N_{sim} simulations and Figure 6.6 shows the distributions of the estimated parameters for the three stopping rules.

Cohort	$\bar{\hat{K}}_a$	$\bar{\hat{K}}_e$	$\bar{\hat{V}}$	$\bar{\hat{\omega}}_1$	$\bar{\hat{\omega}}_2$	$\bar{\hat{\omega}}_3$	$\bar{\hat{\sigma}}^2$
1	0.8479	0.1551	16.8318	0.1013	0.0884	0.0939	0.0876
2	0.8029	0.1657	15.6445	0.0932	0.0861	0.1082	0.0990
3	0.8053	0.1643	15.3679	0.0884	0.0887	0.1124	0.1021
4	0.8101	0.1619	15.4250	0.0856	0.0903	0.1122	0.1038
5	0.8145	0.1604	15.6578	0.0874	0.0918	0.1119	0.1042
6	0.8167	0.1595	15.9801	0.0868	0.0928	0.1124	0.1044
7	0.8180	0.1588	16.3427	0.0865	0.0933	0.1136	0.1049
8	0.8173	0.1587	16.6338	0.0871	0.0938	0.1149	0.1048
9	0.8185	0.1583	16.8809	0.0866	0.0937	0.1158	0.1049
10	0.8192	0.1580	17.0449	0.0870	0.0940	0.1165	0.1050

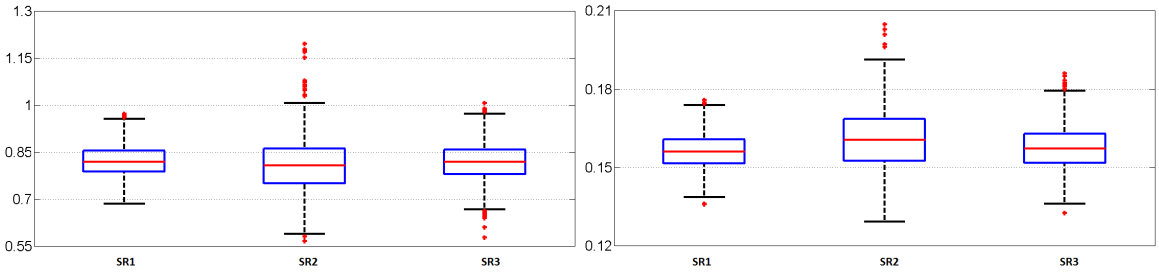
Table 6.2: Average estimates of the model parameters over the course of the adaptive procedure, derived at each of the C cohorts. The assumed true values are $\boldsymbol{\Psi}_{true} = (.85, .15, 17, .1, .1, .1, .1)^T$ and the initial values are $\boldsymbol{\Psi}_o = (1, .1, 20, .05, .15, .05, .15)^T$.

The boxplots that we present in this thesis have the box made up of the 3^{rd} and 1^{st} quartile, i.e., $Q3$ and $Q1$ with the median as the red line in the middle of the box. The length of each of the two whiskers is $1.5(Q3 - Q1)$.

It seems from Figure 6.6 that bias is present in most of the estimates, since most of the distributions are not centred around the true values. For the stopping rule SR1, which entails evaluation of all the cohorts, the biases in general are smaller as compared to scenarios in which the trial is terminated early due to application of stopping rules SR2 and SR3.

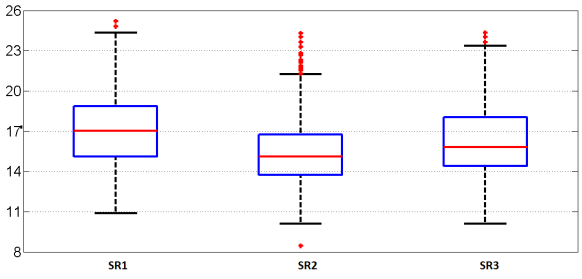
As expected, the spread for SR1 is the smallest for all the seven parameters, since maximum information is utilised. By the same logic, the spread in the distributions when SR2 is used is largest since, on an average, lesser number of cohorts are analysed. Further, as can be seen from Table 6.3, the precision of the estimated parameters is highest for SR1 followed by SR3 and then SR2.

Figure 6.7 shows the distributions for the parameter estimates obtained in the simulations. It can be observed from the figure that bias is present in most of the estimates, since the distributions are generally not centred around the true values.

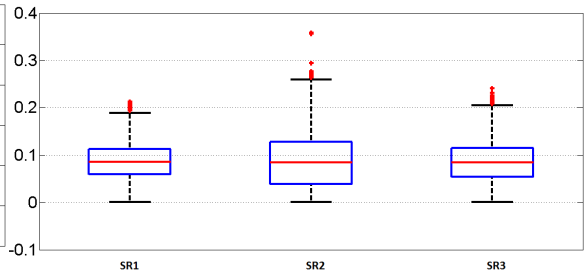


(a) $\widehat{K}_a$

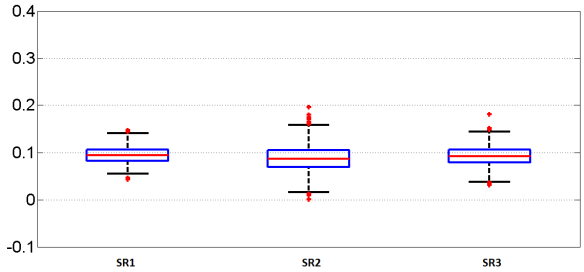
(b) $\widehat{K}_e$



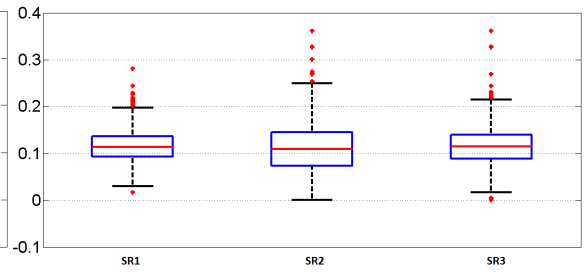
(c) $\widehat{V}$



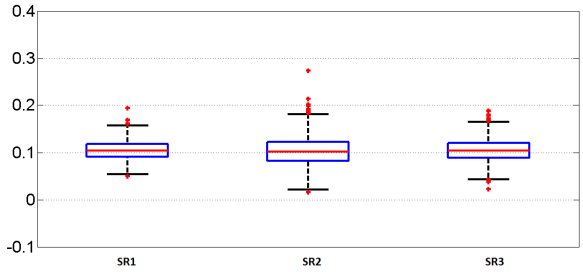
(d) $\widehat{\omega}_1$



(e) $\widehat{\omega}_2$



(f) $\widehat{\omega}_3$



(g) $\widehat{\sigma}^2$

Figure 6.6: Simulated distribution of the estimated parameters for the three stopping rules, averaged over the N_{sim} simulations. For each of the seven parameters, the spread is least when SR1 is used.

It seems from Tables 6.2 and 6.3 that an increase in the number of cohorts does not necessarily lead to a proportionate improvement in the quality of the parameters' estimates (as measured by the associated bias and the mse). Indeed, considering the factors of cost and patient inconvenience, it might be sensible to work with a fewer number of cohorts. Or it might be more informative to have fewer cohorts but

		K_a	K_e	V	ω_1	ω_2	ω_3	σ^2
$\bar{\hat{\boldsymbol{\Psi}}}$	SR1	0.8192	0.1580	17.0449	0.0870	0.0940	0.1165	0.1050
	SR2	0.8062	0.1628	15.3183	0.0892	0.0878	0.1117	0.1029
	SR3	0.8162	0.1593	16.2209	0.0863	0.0927	0.1159	0.1049
$\widehat{\text{Bias}}(\hat{\boldsymbol{\Psi}})$ p.c.	SR1	-3.6	5.3	0.3	-13.0	-6.0	16.5	5.0
	SR2	-5.1	8.5	-9.9	-10.8	-12.2	11.7	2.9
	SR3	-4.0	6.2	-4.6	-13.7	-7.3	15.9	4.9
$\widehat{\text{mse}}(\hat{\boldsymbol{\Psi}})$	SR1	0.0047	0.0013	7.7326	0.0018	0.0003	0.0014	0.0004
	SR2	0.0107	0.0017	8.9386	0.0049	0.0009	0.0031	0.0010
	SR3	0.0060	0.0014	8.1502	0.0025	0.0005	0.0019	0.0006

Table 6.3: Average estimated model parameters for the three stopping rules and their properties.

larger in sizes. We explore such ideas later, in the section on sensitivity analysis. The quality of the estimates is better for SR3 as compared to SR2. The reason for this is the stricter convergence criteria for SR3 which results in higher average cohort number (ACN) of 7.52 as compared to 3.81 for SR2. The extra information utilised in SR3 improves the statistical properties of the estimates. Given the fact that, on average, SR2 requires about 38% and SR3 75% of the C cohorts, the two stopping rules certainly improve the cost effectiveness of the trial.

D-optimal time points

Table 6.4 shows the D-optimal sampling time points for the different cohorts, averaged over the N_{sim} simulations. The D-optimal points corresponding to $\boldsymbol{\Psi}_{true}$ are $\xi^*_{true} = \{.10, 6.60, 48.00\}$ and the optimal points corresponding to $\boldsymbol{\Psi}_o$ are $\xi^*_o = \{.10, 5.42, 48.00\}$. The optimal sampling points for each cohort are derived from the estimated $\boldsymbol{\Psi}$ using the information collected from all the preceding cohorts.

There are no significant changes in the optimal sampling times from the third cohort. Since the optimal sampling times are a function of the parameters' values, this can be attributed to the relative stability in the parameter estimates from that cohort onward. The two points on the boundary of the design region i.e., [.1, 48] are stable throughout.

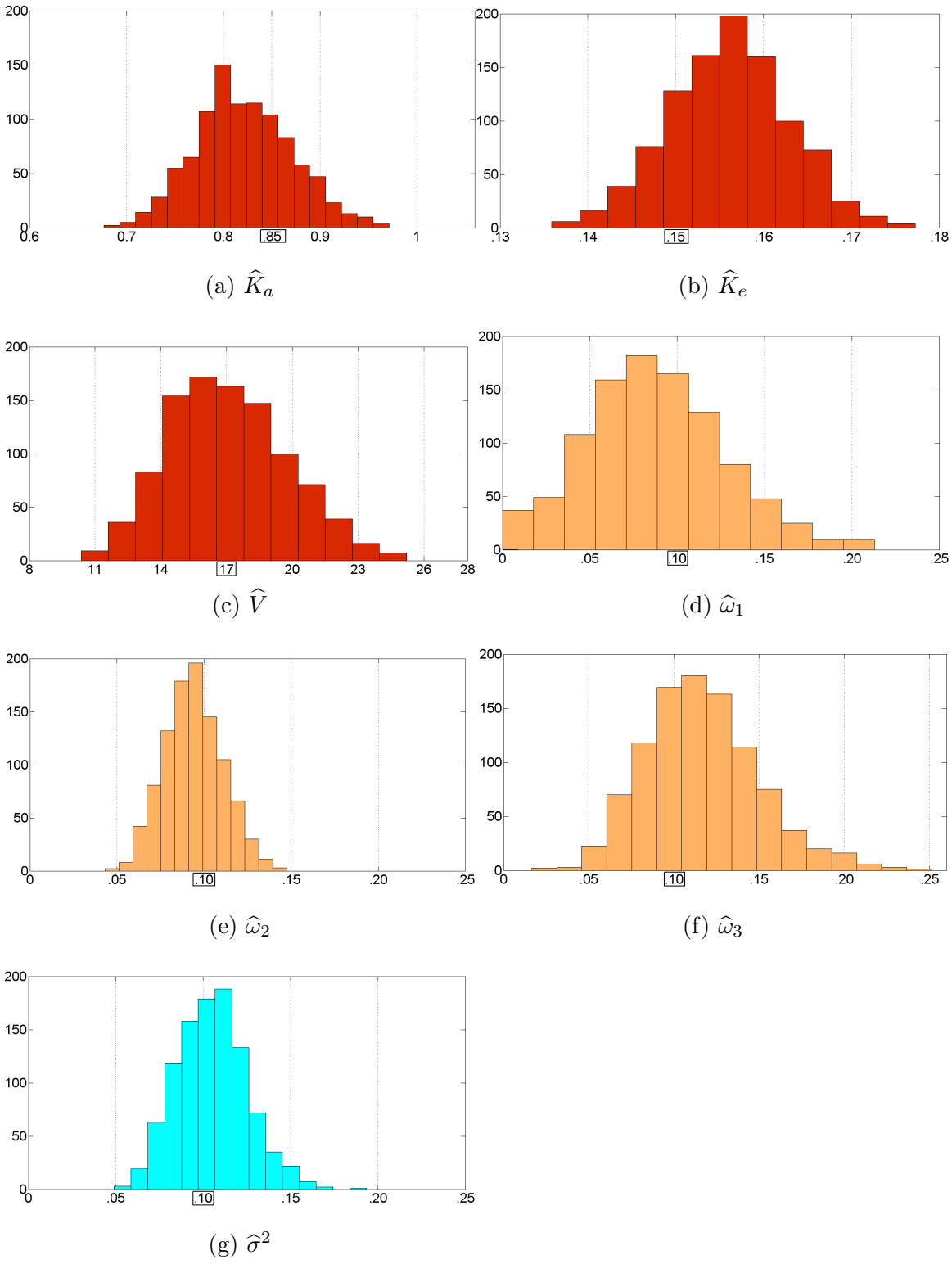


Figure 6.7: Simulated distribution of the mean parameters computed at the last, i.e., the C^{th} cohort, averaged over the N_{sim} simulations. The distributions seem to be symmetric, although not centered at the true values of the parameters owing to the presence of bias.

Dose Regimen

Using the vector of parameter estimates $\bar{\bar{\boldsymbol{\beta}}}$ derived from the C cohorts, the optimal φ_A-efficient dose vector is computed in each simulation using the ED algorithm.

Cohort	T_1	T_2	T_3
1	0.10	5.42	48.00
2	0.38	7.40	47.47
3	0.10	6.72	47.99
4	0.10	6.79	47.99
5	0.10	6.76	48.00
6	0.10	6.72	48.00
7	0.10	6.74	48.00
8	0.10	6.74	48.00
9	0.10	6.75	48.00
10	0.10	6.75	48.00

Table 6.4: D-optimal sampling points for the $C = 10$ cohorts, averaged over the $N_{sim} = 1000$ simulations. The D-optimal sampling times corresponding to $\boldsymbol{\Psi}_{true}$ are $\xi^*_{true} = \{.10, 6.60, 48.00\}$.

Averaging over the N_{sim} simulations, we obtained the average dose regimen administered to the last cohort as $\boldsymbol{D}^* = (145.58, 92.79, 97.02, 96.67, 97.02)^T$ with $\bar{\varphi}_A = 8.5577$. The optimal dose regimen corresponding to the true PK parameters $\boldsymbol{\Psi}_{true}$ is $\boldsymbol{D}^* = (140, 90, 90, 100, 90)^T$ and the associated efficiency measure is $\varphi_A = 7.1561$. Table 6.5 presents the average φ_A-efficient dose regimen for each cohort in the adaptive procedure. As far as $\bar{\varphi}_A$ values are concerned, there is no clear advantage of having a larger number of cohorts. We look into this aspect later in the chapter. Since in each simulation the first cohort receives the dose regimen which is based on the fixed initial values $\boldsymbol{\Psi}_o$, cohort 1 receives the same dose regimen in all the N_{sim} simulations. Hence, the average dose regimen for cohort 1 is equal to $\boldsymbol{D}^*_1$.

Figure 6.8 presents the variability in $\boldsymbol{D}^*$ across the $N_{sim} = 1000$ simulations. The medians of the doses d^*_1, d^*_2 and d^*_4 coincide with their true values of 140, 90 and 100 respectively. The distributions for d^*_3, d^*_4 and d^*_5 are quite similar to each other. The coefficients of variation for all the five doses were found to be approximately 16%, whereas the assumed inter-individual variability in the true PK parameters is around 32%.

We also study the distributions of φ_A for the three stopping rules to compare their performances. Figure 6.9 shows the distribution of φ_A for the three stopping rules. The median value of φ_A is lowest for SR1, which is expected since more information is utilised. The spread of φ_A is smallest for SR1 for the same reason. However, no significant difference is observed between the boxplots for SR2 and SR3. Furthermore, $\bar{\varphi}_A$ for SR2 is 8.8241 and for SR3 it is 8.9036. Thus, SR2 and SR3 have similar performances in determining the φ_A-efficient dose regimens.

Cohort	d_1^*	d_2^*	d_3^*	d_4^*	d_5^*	$\bar{\varphi}_A$
1	150.00	70.00	80.00	70.00	80.00	9.0129
2	143.80	90.43	94.79	94.12	94.82	8.4296
3	137.76	89.08	93.58	92.86	93.59	8.7375
4	135.00	87.02	91.32	90.56	91.32	8.5854
5	134.59	86.20	90.52	89.91	90.52	8.3465
6	135.97	86.63	91.15	90.44	91.16	8.2156
7	138.53	88.17	92.44	91.95	92.44	8.2484
8	141.27	89.86	94.39	93.54	94.39	8.3576
9	143.64	91.55	95.86	95.37	95.86	8.4805
10	145.58	92.79	97.02	96.67	97.02	8.5577

Table 6.5: φ_A-efficient dose vector for the $C = 10$ cohorts, averaged over the N_{sim} simulations along with the corresponding average φ_A values. $\boldsymbol{D}^*$ corresponding to $\boldsymbol{\beta}_{true}$ is $(140, 90, 90, 100, 90)^T$ with $\varphi_A = 7.1561$.

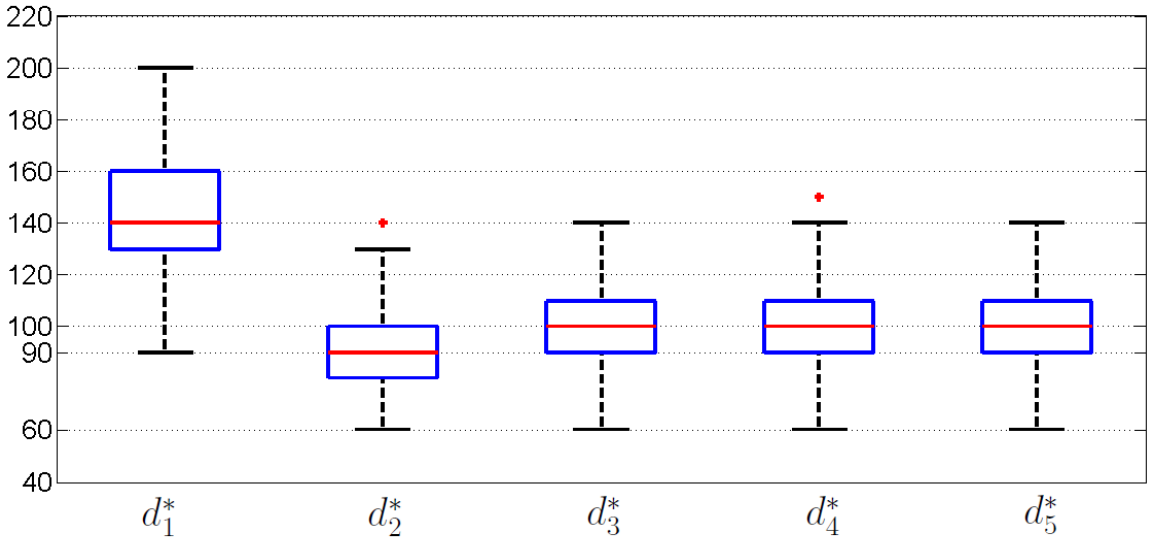


Figure 6.8: Distribution of $\boldsymbol{D}^*$ across the $N_{sim} = 1000$ simulations. The individual doses are more or less symmetrically distibuted around their corresponding true values.

6.2.3 Distributions of the Simulations

To facilitate a better understanding of the simulation study, it is important to evaluate the extent of variability that is induced in the simulated data. This is important not just for assessing the overall performance of the methodology but also for conducting sensitivity analysis in Section 6.3.

In this simulation study, two kinds of data were simulated. The first is the vector $\boldsymbol{\beta}_i$ of PK parameters of a patient i, as described in Equation (6.2) and the second is

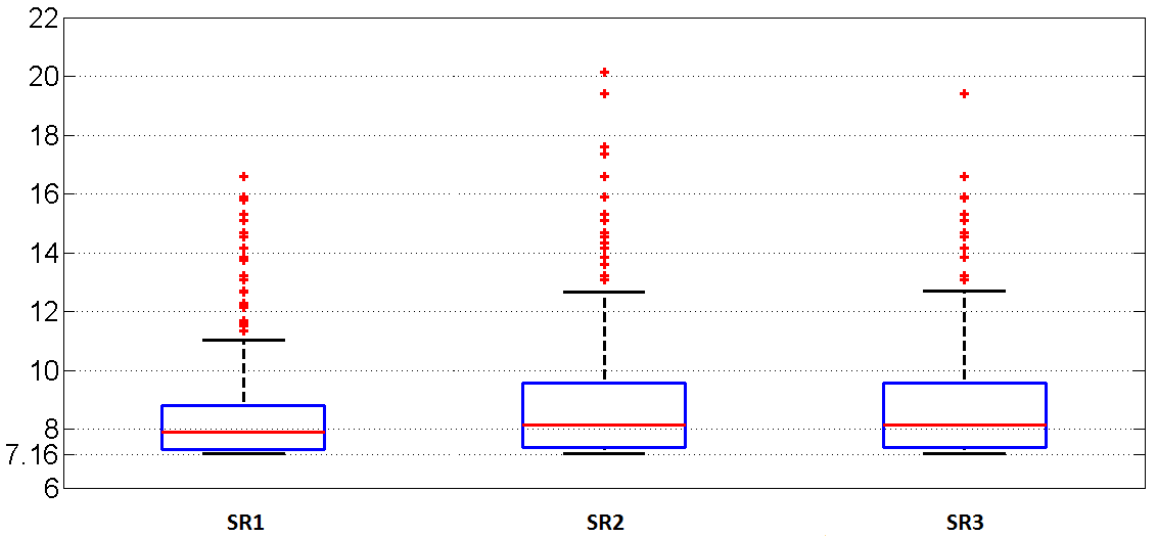


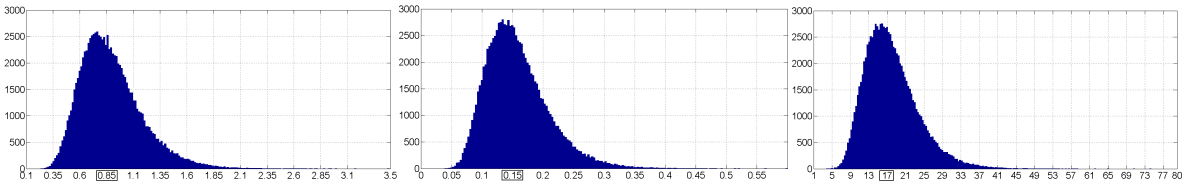
Figure 6.9: Distribution of the values of φ_A efficiency measure for the dose regimens according to the three stopping rules SR1, SR2 and SR3.

the response y_{ij}, as described in Equation (6.3).

Simulations of Parameters

Each study consists of $N_{sim} = 1000$ simulations of $C = 10$ cohorts of size $c = 10$ patients. Therefore, in one simulation study, we simulate 100,000 vectors of $\boldsymbol{\beta}_i$, as explained in Equation (6.5).

As shown in Section 5.2, the distributions of the PK parameters for the model defined in this example are lognormal. Here we assumed that $K_{a_i} \sim \ln\mathcal{N}(\ln(.85), 0.1)$, $K_{e_i} \sim \ln\mathcal{N}(\ln(.15), 0.1)$ and $V_i \sim \ln\mathcal{N}(\ln(17), 0.1)$. The distributions of the simulated profiles are shown in Figure 6.10.



(a) K_{a_i} (b) K_{e_i} (c) V_i

Figure 6.10: Simulated distributions of the parameters contained in $\boldsymbol{\beta}_i$.

The histograms indicate that the distributions are positively skewed, which is consistent with the fact that they are lognormally distributed.

Table 6.19 compares the statistics derived from the simulated distributions of the three parameters with their theoretical values. They are observed to be in good agreement. This supports our previous assessment that $N_{sim} = 1000$ is a sufficient number of simulations which should be run in a study.

Statistic	Distribution	K_a	K_e	V
Mean	Theoretical	0.8936	0.1577	17.8716
	Simulated	0.8931	0.1577	17.8947
Median	Theoretical	0.8500	0.1500	17.0000
	Simulated	0.8500	0.1500	17.0313
IQR	Theoretical	0.3654	0.0645	7.3071
	Simulated	0.3659	0.0651	7.2641

Table 6.6: Comparison of statistics related to the simulated distribution of the three PK parameters with the values obtained from their theoretical distributions.

Simulation of Responses

In the context of this study, a response refers to the concentration of the drug measured in the blood sample collected from a subject. The subjects' responses are central to the adaptive procedure as they give insights into the physiological profile of the subject. In the adaptive procedure, at the time points contained in ξ_i, $m = 3$ concentration values (response) were simulated from each patient in the i^{th} cohort, as explained in Equation (6.5). Since in the adaptive procedure the sampling time points can be different for each of the C cohorts, in order to correctly study the distribution of the responses, it is imperative that the responses for the cohorts are simulated at the same time points.

Therefore, for the parameters simulated in the study, the responses at the time points $\xi^* = \{.10, 6.60, 48\}$ are recorded for each of the 100,000 subjects. Figure 6.11 presents the distributions of the response at the three sampling time points and Table 6.7 presents the statistics related to these distributions.

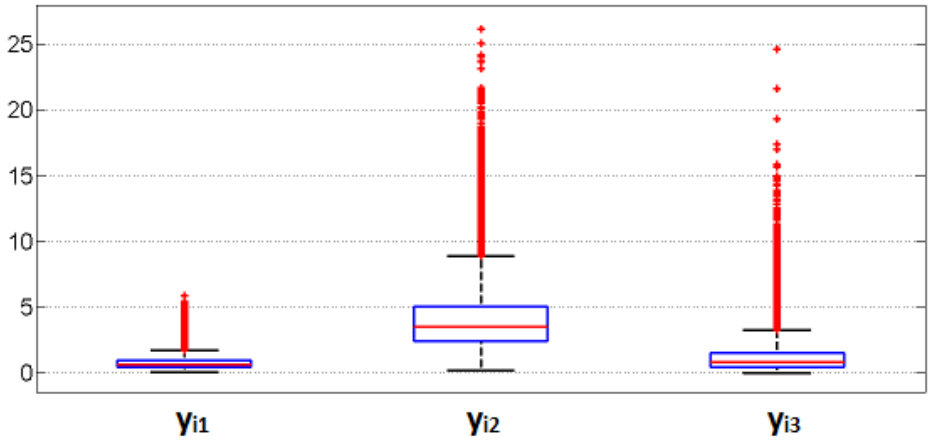


Figure 6.11: Distributions of the response at the sampling points $\xi^* = \{.10, 6.60, 48\}$.

From the figure, it seems that the variability in the response depends on its magni-

tude, which is expected since the error structure in the model used in Equation (6.3) is proportional to the response.

Statistic	y_{i1}	y_{i2}	y_{i3}
Mean	0.7697	4.0412	1.1962
Median	0.6601	3.5484	0.8289
IQR	0.5069	2.5786	1.1277

Table 6.7: Mean, median and inter-quartile range of the simulated responses at the three sampling points.

The data from the table also confirm this observation. A large mean or median concentration is associated with a large value of the inter-quartile range. Overall, the variability in the response has been assumed to be moderately large in the simulation study.

In Section 6.3, where we analyse the sensitivity of the algorithm to changes in the inputs, we re-examine the distributions of the response whenever there is a change in the assumed model or the model parameters. For example, in Section 6.3.5, where we study the effect of misspecification of the PK model on the performance of the adaptive procedure, we compare the distributions of the simulated responses and the simulated parameters for the assumed and the true models. Similarly, when we study the effect of the error variance on the quality of the parameters' estimates in Section 6.3.7, the distributions of the response shed more light on the scale of variability induced by the magnitude of error variance. As mentioned before, studying the distribution of the response is important for gaining better insights into the performance of the methodology.

6.2.4 Distinction Between Population and Individual Dosing Risk

In Section 4.4.1, we explored how misspecification of the PK model affects the efficiency of the administered dose regimen. The dose regimen computed by the ED algorithm is a function of the vector of model parameters, $\boldsymbol{\beta}$. In most practical situations, this vector is unknown and has to be estimated. Any deviation of the estimated parameters' values from the true values may, therefore, result in a suboptimal dose regimen.

The extent of this suboptimality can be quantitatively assessed at two levels. Firstly, at the population level, where the estimated vector of mean PK parameters, $\widehat{\boldsymbol{\beta}}$, is used to compute the optimal dose regimen for the entire population instead

of the true PK parameters $\boldsymbol{\beta}$. Secondly, at the individual level, such that the i^{th} individual in the cohort having PK parameters $\boldsymbol{\beta}_i$, $i = 1, ..., c$, is administered a dose regimen based on $\widehat{\boldsymbol{\beta}}$.

In this section, we highlight the difference between error in dosing at the population level and at an individual level. At the population level, the error is measured around the optimal dose regimen based on the mean PK parameters. At the individual level, the error is measured around the optimal dose regimen based on the subject's own PK parameters.

Population Level

Consider a population having vector of mean PK parameters $\boldsymbol{\beta}$. The ED algorithm can be used to find the optimal dose regimen $\boldsymbol{D}^*(\boldsymbol{\beta}) = (d_1(\boldsymbol{\beta}), ..., d_n(\boldsymbol{\beta}))^T$. This notation emphasizes the dependence of the optimal dose regimen on $\boldsymbol{\beta}$. As mentioned above, usually an estimate $\widehat{\boldsymbol{\beta}}$ is used to compute an estimate of the optimal dose regimen, $\boldsymbol{D}^*(\widehat{\boldsymbol{\beta}}) = (d_1(\widehat{\boldsymbol{\beta}}), ..., d_n(\widehat{\boldsymbol{\beta}}))^T$.

We define the *Dosing Risk* at the population level as the root mean squared error:

$$\mathrm{R}\left(\boldsymbol{\beta}\right) = \left(\sqrt{\mathrm{E}\left(d_1(\widehat{\boldsymbol{\beta}}) - d_1(\boldsymbol{\beta})\right)^2}, ..., \sqrt{\mathrm{E}\left(d_n(\widehat{\boldsymbol{\beta}}) - d_n(\boldsymbol{\beta})\right)^2}\right)^T.$$

An estimate of the dosing risk, using estimated values from the simulation study, $\widehat{\boldsymbol{\beta}}^{(k)}$, $k = 1, ..., N_{sim}$, can be obtained as:

$$\widehat{\mathrm{R}}\left(\boldsymbol{\beta}\right) = \left(\sqrt{\frac{1}{N_{sim}}\sum_{k=1}^{N_{sim}}\left(d_1(\widehat{\boldsymbol{\beta}}^{(k)}) - d_1(\boldsymbol{\beta})\right)^2}, ..., \sqrt{\frac{1}{N_{sim}}\sum_{k=1}^{N_{sim}}\left(d_n(\widehat{\boldsymbol{\beta}}^{(k)}) - d_n(\boldsymbol{\beta})\right)^2}\right)^T.$$

Since the square root is computed over addition of squared numbers, the dosing risk is always real. It follows from the definition that a large spread in the distribution of $\widehat{\boldsymbol{\beta}}$ will generally result in increased dosing risk.

For the simulation study described in Section 6.2.1, $\boldsymbol{\beta} = (.85, .15, 17)^T$ and the optimal dose regimen based on $\boldsymbol{\beta}$ is $\boldsymbol{D}^*(\boldsymbol{\beta}) = (140, 90, 90, 100, 90)^T$.

Using the estimated values $\widehat{\boldsymbol{\beta}}^{(k)}$ from each of the $N_{sim} = 1000$ simulations in Section 6.2.1, we ran the ED algorithm to find the optimal dose regimen $D^*(\widehat{\boldsymbol{\beta}}^{(k)})$. Figure 6.12 presents the distribution of $d_i(\widehat{\boldsymbol{\beta}}^{(k)}) - d_i(\boldsymbol{\beta})$, $i = 1, ..., n$ and $k = 1, ..., N_{sim}$. The deviations are centred around zero, apart from the cases of doses 3 and 5.

It can be seen from the figure that for doses 1, 2 and 4, the chance of over-exposure

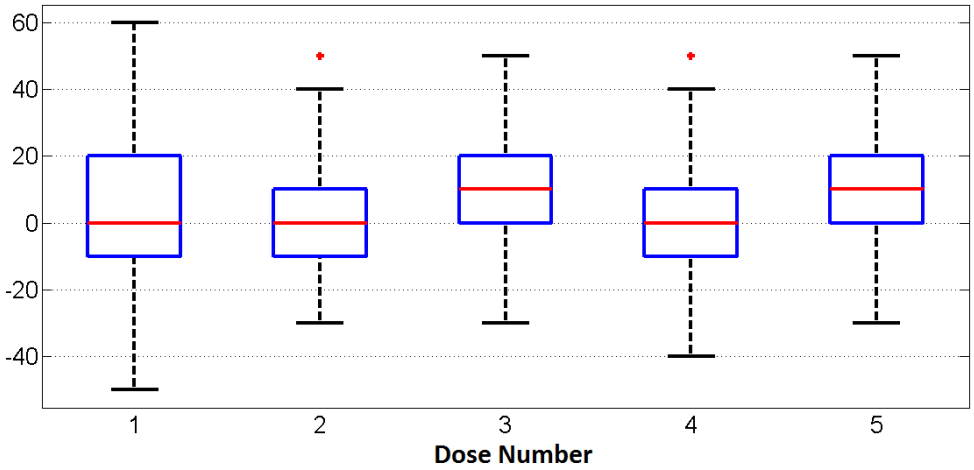


Figure 6.12: Distributions of $d_i(\widehat{\boldsymbol{\beta}}^{(k)}) - d_i(\boldsymbol{\beta})$, $i = 1, ..., 5$ and $k = 1, ..., 1000$.

is the same as of under-exposure since the median of the deviations is zero. For doses 3 and 5, the chances of over- and under-exposure are 75% and 25% respectively. The estimated dosing risk at the population level was found in the simulation study to be $\widehat{\mathrm{R}}\,(\boldsymbol{\beta}) = (23.20, 15.08, 16.76, 15.70, 16.76)^T$. The larger value of the dosing risk for the first dose is consistent with the larger spread inherent in its deviations with $d_1(\boldsymbol{\beta})$, as can be observed in Figure 6.12. The first dose, being the loading dose is relatively more affected by the variability in the parameters' estimates, as was also seen in Figure 6.8.

The analysis presented above is at the population level, that is, the deviations were measured from the dose regimen which was optimised for the population parameters. Next, we assess the dosing risk at the individual level. That is, the deviation between the optimal dose regimen a subject should be administered and the dose regimen that is actually administered.

Individual Level

In the adaptive procedure described in this chapter, each of the c subjects in a cohort are administered the dose regimen $\boldsymbol{D}^*(\widehat{\boldsymbol{\beta}})$, where the estimate $\widehat{\boldsymbol{\beta}}$ is derived from the data collected from all the previous cohorts.

However, for the subjects individually, the optimal dose regimen is the one based on their own PK parameters, i.e., $\boldsymbol{\beta}_i$ and not $\widehat{\boldsymbol{\beta}}$, where $\boldsymbol{\beta}_i$ is the vector of PK parameters for the i^{th} subject in a cohort of the adaptive trial, $i = 1, ..., c$.

As a measure of this discrepancy, we define the dosing risk for the i^{th} subject as

the root mean squared error:

$$\mathrm{R}\left(\boldsymbol{\beta}_i\right) = \left(\sqrt{\mathrm{E}\left(d_1(\widehat{\boldsymbol{\beta}}) - d_1(\boldsymbol{\beta}_i)\right)^2}, ..., \sqrt{\mathrm{E}\left(d_n(\widehat{\boldsymbol{\beta}}) - d_n(\boldsymbol{\beta}_i)\right)^2}\right)^T,$$

for $i = 1, ..., c$.

The dosing risk for a subject will increase with the deviation between the administered dose regimen $\boldsymbol{D}^*(\widehat{\boldsymbol{\beta}})$ from the optimal dose regimen $\boldsymbol{D}^*(\boldsymbol{\beta}_i)$. If the magnitude of the random effects is large, the spread in the distribution of $\boldsymbol{\beta}_i$ will be large and consequently, the dosing risk for the subjects in the trial will be high.

The risk can be evaluated by a simulation study, where we obtain an estimate of the dosing risk for the i^{th} subject using estimates $\widehat{\boldsymbol{\beta}}_i^{(k)}$. That is,

$$\widehat{\mathrm{R}}\left(\boldsymbol{\beta}_i\right) = \left(\sqrt{\frac{1}{N_{sim}}\sum_{k=1}^{N_{sim}}\left(d_1\left(\widehat{\boldsymbol{\beta}}^{(k)}\right) - d_1\left(\widehat{\boldsymbol{\beta}}_i^{(k)}\right)\right)^2}, ..., \sqrt{\frac{1}{N_{sim}}\sum_{k=1}^{N_{sim}}\left(d_n\left(\widehat{\boldsymbol{\beta}}^{(k)}\right) - d_n\left(\widehat{\boldsymbol{\beta}}_i^{(k)}\right)\right)^2}\right)^T.$$

As mentioned in Section 5.3.2, values $\widehat{\boldsymbol{\beta}}_i^{(k)}$ can be computed by applying the estimated fixed effects and the estimated random effects in the original Stage II model. That is, $\widehat{\boldsymbol{\beta}}^{(k)}$ and random effects $\widehat{\boldsymbol{b}}_i^{(k)}$ are applied back into Equation (6.2) to get vectors of individual parameters for each patient:

$$\begin{pmatrix} \widehat{K}_{a_i}^{(k)} \\ \widehat{K}_{e_i}^{(k)} \\ \widehat{V}_i^{(k)} \end{pmatrix} = \begin{pmatrix} \widehat{K}_a^{(k)} \exp(\widehat{b}_{1i}^{(k)}) \\ \widehat{K}_e^{(k)} \exp(\widehat{b}_{2i}^{(k)}) \\ \widehat{V}^{(k)} \exp(\widehat{b}_{3i}^{(k)}) \end{pmatrix}, \tag{6.6}$$

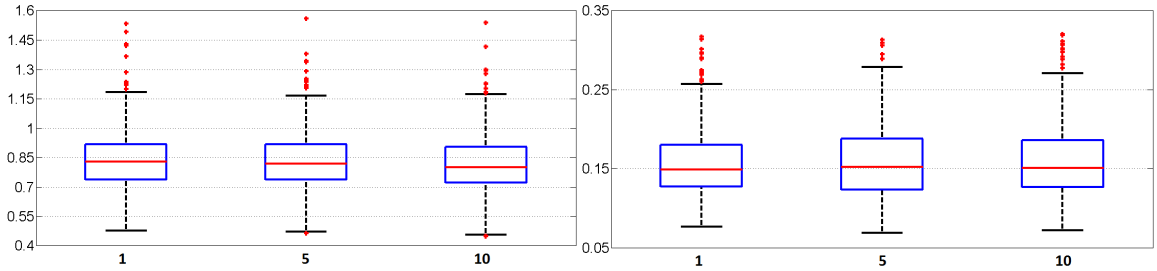
for $i = 1, ..., c$ and $k = 1, ..., N_{sim}$.

For example, let us consider the last, that is, the C^{th} cohort of the trial. The distributions of the individual parameters for the $c = 10$ subjects contained in this cohort were found to be broadly similar. For example, Figure 6.13 presents the distributions of $\widehat{K}_{ai}^{(k)}$, $\widehat{K}_{ei}^{(k)}$ and $\widehat{V}_i^{(k)}$ for subjects 1, 5 and 10 only as they well represent all the subjects.

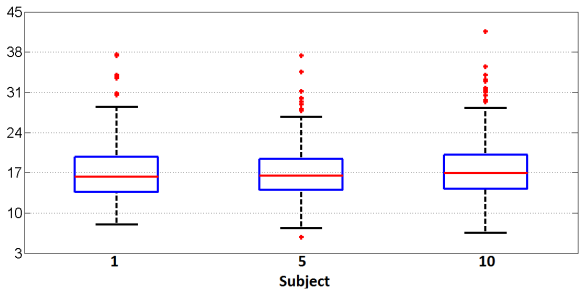
The respective medians of the individual parameters are approximately equal to the respective population values which, as discussed in Section 6.2.3, is expected.

Table 6.8 presents the estimated dosing risk $\widehat{R}(\boldsymbol{\beta}_i)$ for subjects 1, 5 and 10.

In comparison to the estimated dosing risk at the population level $\widehat{R}(\boldsymbol{\beta})$, the estimated dosing risk for a subject is much higher. This is on account of the fact that the



(a) $\widehat{K}_{ai}^{(k)} = \widehat{K}_{a}^{(k)} \exp(\widehat{b}_{1i}^{(k)})$ (b) $\widehat{K}_{ei}^{(k)} = \widehat{K}_{e}^{(k)} \exp(\widehat{b}_{2i}^{(k)})$



(c) $\widehat{V}_{i}^{(k)} = \widehat{V}^{(k)} \exp(\widehat{b}_{3i}^{(k)})$

Figure 6.13: Distributions of $\widehat{K}_{ai}^{(k)}$, $\widehat{K}_{ei}^{(k)}$ and $\widehat{V}_{i}^{(k)}$ for subjects 1, 5 and 10, $k = 1, ..., N_{sim}$.

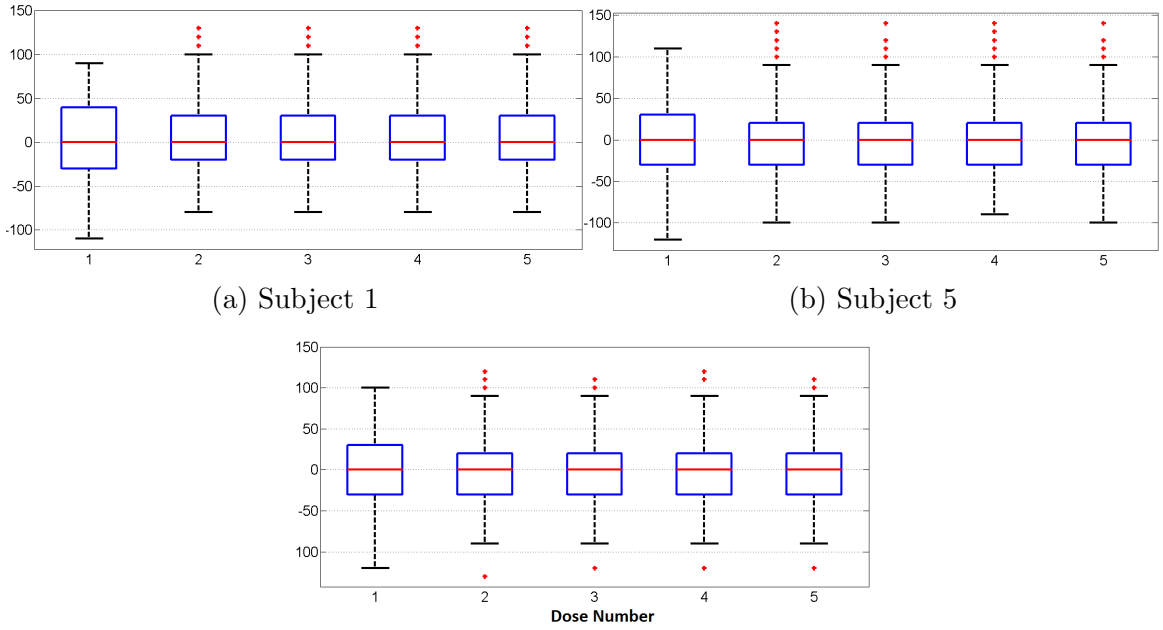
Subject	d_1	d_2	d_3	d_4	d_5
1	39.51	41.90	40.16	40.31	40.12
5	40.34	41.03	39.54	39.47	39.51
10	37.91	38.18	36.58	36.40	36.65

Table 6.8: Estimated dosing risk, $\widehat{R}(\boldsymbol{\beta}_i)$, for the three subjects in the terminal cohort.

random effects that influence the PK parameters at the individual level are absent at the population level. This additional variability induced by the random effects predisposes the subjects to a higher chance of being administered a substandard dose regimen. The dosing risk for a subject is expected to increase with an increase in inter-individual variability, although, as discussed in Section 4.4.1, the net effect of variation in the PK parameters on the efficiency of the administered dose regimen can sometimes balance out.

Since the distributions of the parameters for the three subjects are mostly similar, the distributions of the deviations $d_j\left(\widehat{\boldsymbol{\beta}}^{(k)}\right) - d_j\left(\widehat{\boldsymbol{\beta}}_i^{(k)}\right)$, $j = 1, ..., 5$ are also similar. Figure 6.14 presents the distribution of the deviations for subjects 1, 5 and 10 in the cohort. The distributions are centred around zero.

From the figure, it can be seen that the chance of over- and under-exposure for each of the five doses is 50%. Furthermore, the spread in the distribution for the first (loading) dose is the largest. This was also observed for the case of dosing risk at the



(a) Subject 1

(b) Subject 5

(c) Subject 10

Figure 6.14: Distributions of $d_j\left(\widehat{\boldsymbol{\beta}}^{(k)}\right) - d_j\left(\widehat{\boldsymbol{\beta}}_i^{(k)}\right)$, $k = 1, ..., N_{sim}$ and $j = 1, 5$ and 10. The distribution of the deviations for each dose is centred at 0.

population level. Similar conclusions can be made for all other seven patients in the cohort.

One way to reduce the dosing risk of a subject is to administer an individualised dose regimen. This can be achieved when the optimisation of the dose regimen is performed by using the individual PK parameters of the subject as inputs to the ED algorithm. We present a methodology to accomplish this in Section 7.1.

If individualisation of the dose regimen is not an option, another way of reducing the dosing risk for a subject could be to include relevant covariates in the PK model. Covariates could explain some of the inter-individual variability in the subjects which may result in a reduction of an individual's dosing risk. This is discussed briefly in Section 7.1.4.

The ability of the adaptive methodology outlined in this chapter to ascertain the population PK parameters while administering the efficient dose regimen to patients in each cohort is established. Optimisation of the number of cohorts, cohort sizes and other design variables could improve the quality of the inferences drawn from the trial, as well as it could be useful in enhancing the ethical standards and the cost effectiveness. Although we do not attempt formal optimisation of these variables in this thesis, in the next section we discuss some scenarios in which the effect of the variables' choices is studied. We also study the performance of the methodology in the presence of some practical issues such as model misspecification, missing data and

non-compliance with the recommended dose regimen.

6.3 Sensitivity Analysis

Sensitivity analysis, in the context of a simulation study, entails measuring the uncertainty in the output of a model that can be attributed to the uncertainty in the input of the model, Saltelli et al. (2008). As mathematical modelling is an attempt to imitate the real world, there will be departures from the assumptions made during the modelling process, and the ability of the model to withstand them is what is called its robustness. In the context of our work, it is the uncertainty in the initial values of the PK parameters which is of concern since most estimation methods for non-linear mixed effects models are sensitive to the initial values. Cohort size, number of cohorts and number of PK sampling times are design variables and can be optimised. We plan to extend the algorithm for this optimisation. For now, however, the procedure does not include this option and so we examine the effect of changing these numbers on the outcome in the simulation studies.

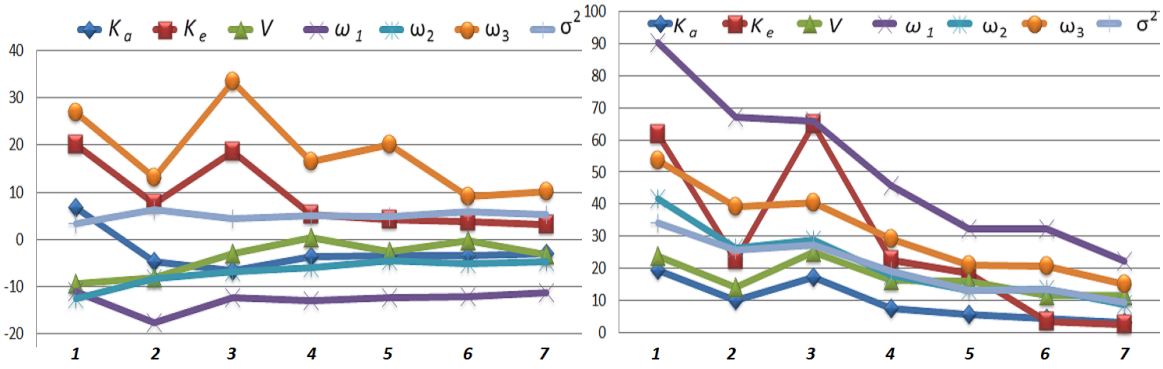
While conducting sensitivity analysis, the researcher needs to be sensible in varying the input variables as a large change might introduce so much variability in the output that it ceases to be discernible while a small change may not produce any noticeable difference in the output. The econometrician Edward E. Leamer has well articulated this balance in Leamer (1983):

> I have proposed a form of organized sensitivity analysis that I call 'global sensitivity analysis' in which a neighbourhood of alternative assumptions is selected and the corresponding interval of inferences is identified. Conclusions are judged to be sturdy only if the neighbourhood of assumptions is wide enough to be credible and the corresponding interval of inferences is narrow enough to be useful.

In this section, we shall explore the effects of varying some inputs on the results produced by our methodology, while keeping the other inputs unchanged. We firstly study the effect of using different cohort sizes and the number of cohorts on the performance of the algorithm.

6.3.1 Dependence on the Cohort Size and the Number of Cohorts

Cohort size, c, is a design variable which is entangled with the number of cohorts, C. In the original simulation study, we took $(c, C) = (10, 10)$. Table 6.9 presents the data obtained by re-running the simulation study $N_{sim} = 1000$ times for different pairs of (c, C) for the stopping rule SR1. The data corresponding to the original run with (10, 10) have been inserted to facilitate comparisons. The number in bold in the percentage bias row is the average absolute percentage bias, computed by taking average of the absolute values of the bias for the seven parameters estimates. Similarly, the bold number in the CV row is the average CV per parameter estimate. Figure 6.15 plots the percentage bias and CV of the seven parameters. In general, the bias and the



(a) Percentage bias (b) CV

Figure 6.15: Variation of the percentage bias and CV of each of the seven parameters for the seven scenarios of the choice of pair (c, C): 1. (5, 5), 2. (10, 5), 3. (5, 10), 4. (10, 10), 5. (10, 20), 6. (20, 10), 7. (20, 20).

variability in the estimated parameters decrease with increase in the cohort size and the number of cohorts. However, the percentage bias is more effectively reduced by taking a larger cohort size rather than having more cohorts. This can be inferred by noticing in Table 6.9 that the magnitude of the decrease in the absolute bias is larger when c is increased keeping C fixed than the other way around. For example, keeping C fixed at 10 cohorts, the average percentage biases for $c = 5$, 10 and 20 subjects are 12.1, 7.1 and 5.7. However, keeping c fixed at 10 subjects, the average percentage biases for $C = 5$, 10 and 20 subjects are 9.4, 7.1 and 7.4.

The average CV also appears to follow a similar relationship with (c, C). The dose vector administered to the last cohort, $\boldsymbol{D}^*$, tends to get closer to the true optimal dose vector as the bias decreases, which is also supported by the simulated data.

$\bar{\varphi}_A$ values for the seven scenarios also support the idea of preferring larger cohort sizes to having more cohorts. For fixed C, $\bar{\varphi}_A$ decreases with higher c but this is not always true when C is increased keeping c fixed.

Further, Figure 6.16 shows that the spread of the distribution of φ_A for the seven scenarios also follows the pattern exhibited by $\bar{\varphi}_A$. Comparing scenarios 2 and 3, 5 and 6, it can be seen that the spread in φ_A is curbed by having larger cohort sizes rather than more cohorts. Furthermore, the similarity between the distributions for scenarios 2 and 4, 4 and 5 and 6 and 7 shows that no significant benefit is accrued by having a larger number of cohorts without increasing the cohort size.

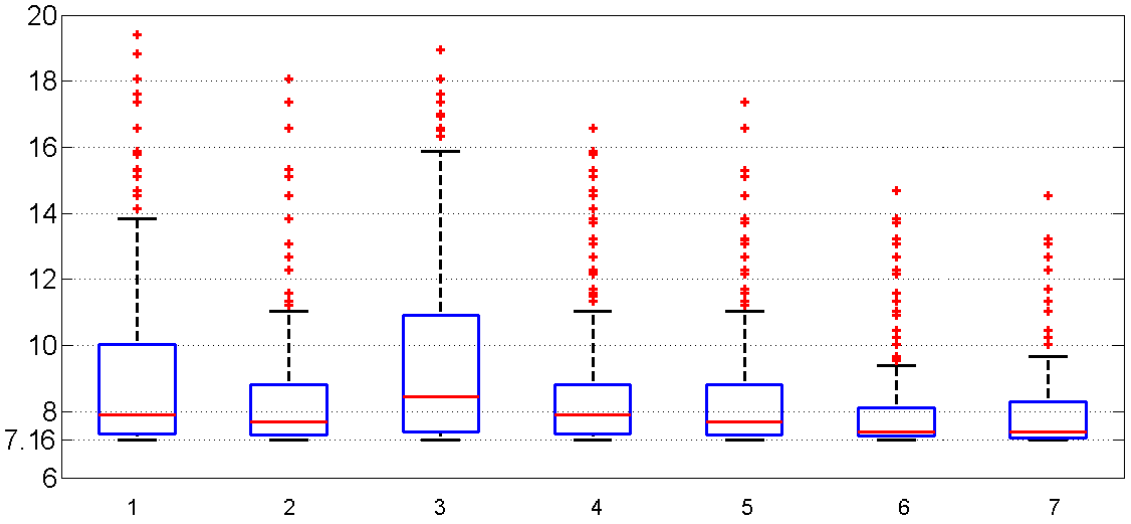


Figure 6.16: Distibution of φ_A in the $N_{sim} = 1000$ simulations for each of the seven scenarios of the choice of pair (c, C): 1. (5, 5), 2. (10, 5), 3. (5, 10), 4. (10, 10), 5. (10, 20), 6. (20, 10), 7. (20, 20).

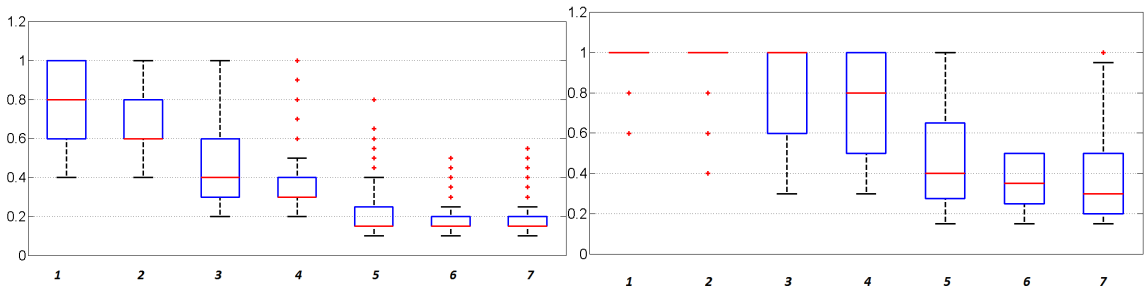
Thus, the data suggest that for a given number of subjects available for the adaptive trial, it may be better to divide them into fewer but larger cohorts to get the maximum information. However, a potential disadvantage of having large cohorts is that more subjects will get under- or overexposed to the drug before more credible parameters are derived from the subsequent cohorts. One possible remedy could be to start with a small cohort size and increase it during the trial.

	c	C	K_a	K_e	V	ω_1	ω_2	ω_3	σ^2	
	5	5	0.7920	0.1799	15.3946	0.0889	0.0875	0.1269	0.1033	
	10	5	0.8098	0.1614	15.6219	0.0824	0.0918	0.1130	0.1063	
	5	10	0.7945	0.1778	16.4839	0.0877	0.0932	0.1333	0.1044	
$\bar{\hat{\boldsymbol{\Psi}}}$	10	10	0.8192	0.1580	17.0449	0.0870	0.0940	0.1165	0.1050	
	10	20	0.8201	0.1565	16.5505	0.0876	0.0955	0.1200	0.1048	
	20	10	0.8203	0.1556	16.9598	0.0879	0.0949	0.1090	0.1058	
	20	20	0.8237	0.1548	16.4561	0.0888	0.0953	0.1102	0.1052	
	5	5	-6.8	20.00	-9.4	-11.1	-12.5	26.9	3.3	[**10.3**]
	10	5	-4.7	7.6	-8.1	-17.6	-8.2	13.0	6.3	[**9.4**]
	5	10	-6.5	18.5	-3.0	-12.3	-6.8	33.3	4.4	[**12.1**]
$\widehat{\text{Bias}}(\hat{\boldsymbol{\Psi}})$	10	10	-3.6	5.3	0.3	-13.0	-6.0	16.5	5.0	[**7.1**]
p.c.	10	20	-3.5	4.3	-2.6	-12.4	-4.5	20.0	4.8	[**7.4**]
	20	10	-3.5	3.8	-0.2	-12.1	-5.1	9.0	5.8	[**5.7**]
	20	20	-3.1	3.2	-3.2	-11.2	-4.7	10.2	5.2	[**5.8**]
	5	5	19.5	61.8	23.7	90.2	41.7	53.7	34.2	[**46.4**]
	10	5	9.9	22.7	14.0	67.0	26.4	39.1	25.7	[**29.2**]
	5	10	17.3	65.0	25.2	65.7	28.8	40.5	27.2	[**38.6**]
CV	10	10	7.5	22.5	16.3	45.8	17.90	29.0	19.1	[**22.6**]
	10	20	5.6	18.6	16.0	32.3	13.1	21.0	13.2	[**17.1**]
	20	10	4.3	3.3	11.6	32.3	13.4	20.7	13.5	[**14.1**]
	20	20	3.26	2.48	11.65	22.32	8.83	15.18	9.42	[**10.4**]
			d_1^*	d_2^*	d_3^*	d_4^*	d_5^*		$\bar{\varphi}_A$	
	5	5	135.37	86.69	90.94	90.33	90.94		9.0909	
	10	5	135.16	86.85	91.28	90.53	91.29		8.2798	
	5	10	144.16	92.14	96.22	95.90	96.22		9.5193	
$\boldsymbol{D}^*$	10	10	145.58	92.79	97.02	96.67	97.02		8.5577	
	10	20	143.66	90.47	95.43	94.64	95.43		8.4487	
	20	10	144.16	92.14	96.22	95.90	96.22		7.8798	
	20	20	142.81	89.55	94.17	93.16	94.17		7.8876	
				$\bar{C}_2$		$\bar{C}_3$				
	5	5		3.70		4.83				
	10	5		3.42		4.72				
	5	10		4.47		8.15				
ACN	10	10		3.81		7.52				
	10	20		3.88		9.50				
	20	10		3.43		6.90				
	20	20		3.35		7.49				

Table 6.9: Summary of the statistics related to the simulation studies for different values of the cohort size (c) and the number of cohorts (C). For each case, the numbers in the bold are the average absolute percentage bias and the average CV of the parameter estimates.

By definition, the average cohort number (ACN) is the average cohort at which the stopping rule applies. Higher variability in the simulated distributions of the PK parameters would lead to increased variability in the efficient dose regimens computed at each iteration. This would lead to a delay in the application of SR2 or SR3, and consequently would lead to an increased ACN. Therefore, as the average CV of the simulated parameters decreases, generally ACNs $\bar{C}_2$ and $\bar{C}_3$ for the two stopping rules also decrease. However, in Table 6.9, the low ACN values for the two cases of $C = 5$ are due to the fact that if in a simulation the condition of SRj ($j = 2, 3$) was not met, C_j was assigned the maximum possible value of 5. For this reason, the ACN for small values of C will be smaller in spite of the large underlying variability. In general, the ACNs of any two or more scenarios should only be compared if they have the same cohort sizes.

Figure 6.17 presents the distributions of the cohorts at which SR2 and SR3 are applied. Since the cohort sizes used in the seven scenarios are different, we need to



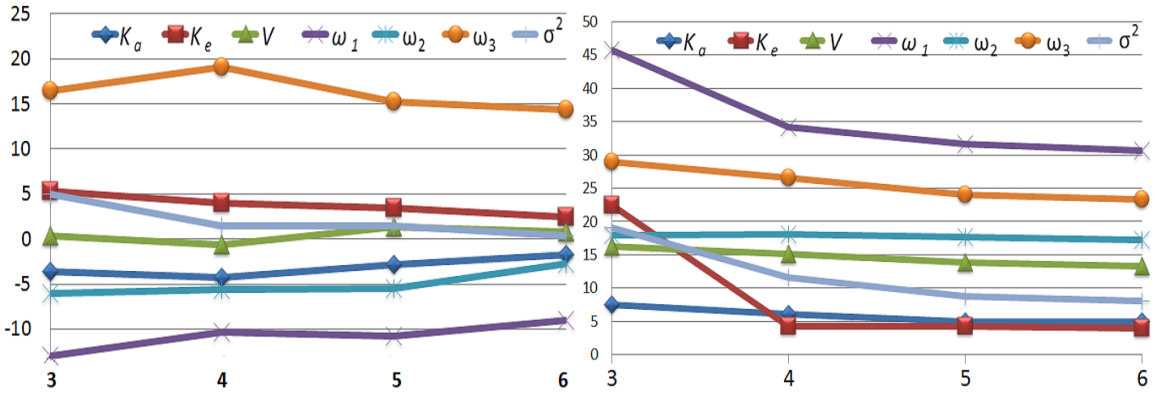
(a) SR2 (b) SR3

Figure 6.17: Distributions of the ratio $C_2^{(k)}/C$ for SR2 and $C_3^{(k)}/C$ for SR3, for the seven scenarios of the choice of pair (c, C): 1. (5, 5), 2. (10, 5), 3. (5, 10), 4. (10, 10), 5. (10, 20), 6. (20, 10), 7. (20, 20).

standardise the cohort sizes to make comparisons between the scenarios. In each of the 1000 simulations, we divide the cohort number at which SR2 was applied by the maximum number of cohorts (C) in that scenario. For example, for scenarios 1 and 2, the cohort numbers were divided by 5 while for 3, 4 and 6, they were divided by 10. The distribution of this statistic is shown in Figure 6.17a. Similar distributions are plotted for SR3 in Figure 6.17b. We can see from the figures that, in general, higher values of c result in lower ACN and lesser spread of the distribution. For SR3, for the first 2 scenarios, the distributions show that in most of the 1000 simulations, all 5 cohorts had to be analysed. The ACN for SR3 tends to be greater than SR2. This is ascribable to the more stringent criterion for application of SR3. Also, the distributions for SR3, in general, have greater spread than that of SR2. This can also be attributed to the same reason as it makes the stoppage of the trial more dependent on the underlying variability in the estimated parameters.

6.3.2 Dependence on the Number of PK Samples per Subject

Collecting more blood samples per subject (m) should give more information about the model parameters. We explore this conjecture by repeating the original simulation study which had $m = 3$ with three other values of m: $\{4, 5, 6\}$. The same design region of [.10, 48] h is used for finding the D-optimal sampling time points. For reasons of practicality, a minimum gap of 0.25 h is ensured between any two successive sampling time points. $N_{sim} = 1000$ simulations are performed for each of these three values and the performance metrics for the stopping rule SR1 are summarised in Table 6.10. ξ_C^* is defined as the optimal sampling times computed using the parameter estimates from the last cohort, i.e., $\widehat{\mathbf{\Psi}}$. The results pertaining to the original run with $m = 3$ observations per subject are added in the table for the purpose of making comparisons. Figure 6.18 plots the average percentage bias and CV of the parameters for the four values of m.



(a) Percentage bias (b) CV

Figure 6.18: Variation of the percentage bias and CV of the seven parameters for the four different values of m: 3, 4, 5 and 6.

The results suggest that in general, collection of more blood samples per subject leads to a reduction in the average absolute bias and the coefficient of variation of the parameter estimates. However, the gain in precision has to be weighed against the loss in economy with respect to the added costs and patient inconvenience. $m = 4$ seems to be a good trade-off between precision and economy.

A noteworthy point in the presented data is that the magnitude of decrease, in the average CV per parameter when $m = 4$ samples are taken instead of $m = 3$, is considerably higher than the pairs of $m = 5$ and $m = 4$ and also when comparing the outputs for $m = 6$ and $m = 5$. The explanation for this can be traced to the theory of optimal designs of experiments. We compare the values of the optimised determinants of the Fisher information matrices (i.e. the objective function values) corresponding to the D-optimal designs for the four values of m and the parameters

	m	K_a	K_e	V	ω_1	ω_2	ω_3	σ^2	
	3	0.8192	0.1580	17.0449	0.0870	0.0940	0.1165	0.1050	
$\bar{\hat{\boldsymbol{\Psi}}}$	4	0.8131	0.1560	16.9007	0.0897	0.0944	0.1191	0.1014	
	5	0.8254	0.1550	17.2243	0.0892	0.0945	0.1152	0.1015	
	6	0.8352	0.1536	17.1299	0.0909	0.0973	0.1144	0.1004	
	3	-3.6	5.3	.3	-13.0	-6.0	16.5	5.0	[**7.1**]
$\widehat{\text{Bias}}(\hat{\boldsymbol{\Psi}})$	4	-4.3	4.0	-0.6	-10.3	-5.6	19.1	1.4	[**6.5**]
p.c.	5	-2.9	3.4	1.3	-10.8	-5.5	15.2	1.5	[**5.8**]
	6	-1.7	2.4	0.8	-9.0	-2.7	14.4	0.4	[**4.5**]
	3	7.5	22.5	16.3	45.8	17.90	29.0	19.1	[**22.6**]
CV	4	6.1	4.3	15.1	34.2	18.1	26.6	11.6	[**16.6**]
	5	4.9	4.2	13.9	31.7	17.7	24.0	8.8	[**15.0**]
	6	5.0	4.0	13.3	30.6	17.3	23.3	8.0	[**14.5**]
		d_1^*	d_2^*	d_3^*	d_4^*	d_5^*		$\bar{\varphi}_A$	
	3	145.58	92.79	97.02	96.67	97.02		8.5577	
$\boldsymbol{D}^*$	4	144.55	91.56	96.24	95.58	96.24		8.4742	
	5	146.30	92.63	97.22	96.75	97.23		8.3155	
	6	145.67	92.00	96.43	95.96	96.43		8.1781	
		T_1^*	T_2^*	T_3^*	T_4^*	T_5^*	T_6^*		
	3	.10	6.75	48.00					
ξ_C^*	4	.11	0.36	6.81	48.00				
	5	.10	.35	6.28	7.35	48.00			
	6	.10	.35	5.83	6.04	47.71	48.00		
			$\bar{C}_2$		$\bar{C}_3$				
	3		3.81		7.52				
ACN	4		3.69		7.44				
	5		3.64		7.35				
	6		3.59		7.23				

Table 6.10: Comparison of statistics related to the simulation studies for different values of m.

$\boldsymbol{\Psi}_{true}$, as shown in Table 6.11.

From the table, it can be seen that the percentage increase in the optimum determinant value is much higher when 4 samples are collected per subject as compared to the other two cases. Also, from Table 6.10, the reduction in the CV from taking 6 instead of 5 samples is considerably smaller, just like the comparatively smaller

m	ξ^*	$\lvert\boldsymbol{M}(\boldsymbol{\Psi}_{true}, \xi^*)\rvert$	%-age increase in $\lvert\boldsymbol{M}(\boldsymbol{\Psi}_{true}, \xi^*)\rvert$
3	{.1, 6.60, 48}	2.96×10^6	-
4	{.1, .35, 6.67, 48}	2.45×10^7	727.7
5	{.1, .35, 6.51, 6.95, 48}	9.88×10^7	303.3
6	{.1, .35, 5.63, 6.95, 47.75, 48}	2.62×10^8	165.2

Table 6.11: The optimal sampling time points and the optimal objective function value for the different values of m. The values in the last column are the percentage increases in $\lvert\boldsymbol{M}(\boldsymbol{\Psi}_{true}, \xi^*)\rvert$ when m samples are collected instead of $m-1$, $m = 4, 5, 6$.

increase in the optimum value of the determinant of the corresponding FIMs.

The pattern exhibited by the average CVs is also reflected in the two average cohort numbers for the different values of m as presented in Table 6.10. This seems reasonable, in the light of the argument made in the previous section that larger variability in the simulated parameters leads to higher ACNs for stopping rules SR2 and SR3. A decrease in variability in the estimated parameters results in earlier application of the two stopping rules which decreases their ACNs. Therefore, the ACNs for the two stopping rules decrease as m increases, although the reduction is very small.

Regarding the effect on the average efficiency of the dose regimens administered to the final cohort, $\bar{\varphi}_A$ decreases as m increases. Figure 6.19 presents the distributions of φ_A values for the four values of m. The distributions are similar, although the median of φ_A slightly improves (i.e. decreases) as m increases.

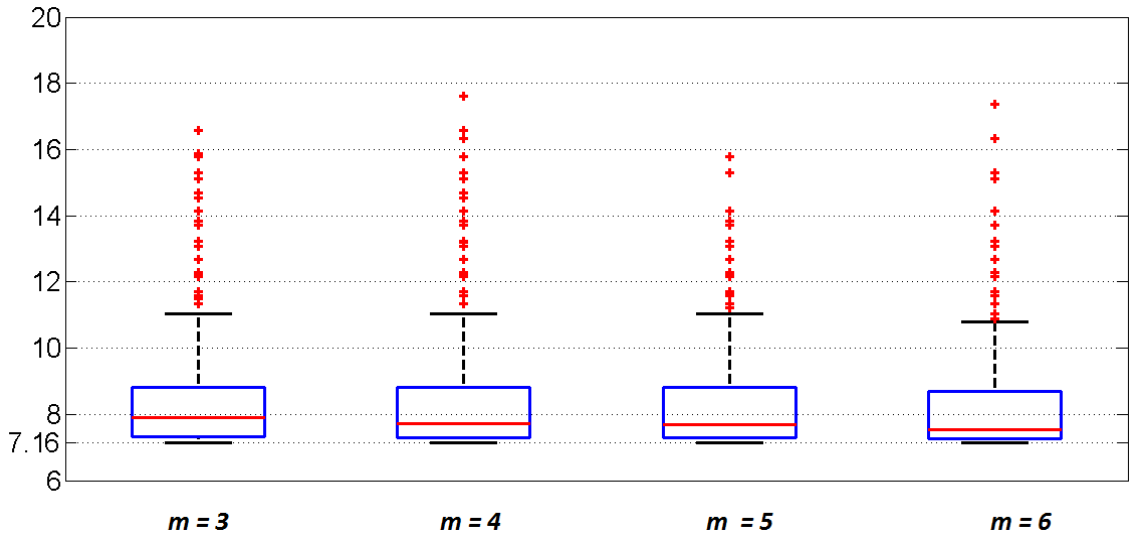


Figure 6.19: Distribution of φ_A values associated with $\boldsymbol{D}^*$ in the $N_{sim} = 1000$ simulations for each of the four values of m.

An interesting observation can be made regarding the D-optimal sampling times ξ_C^*. As m increases, the optimal points tend to be around the points contained in ξ_C^*

for $m = 3$. It seems that there are only three informative sampling time points and any new points would be close to one of these three points.

We conclude from our analyses that although taking $m = 4$ might improve the bias and the CV of the estimated parameters to some extent, there is no strong justification for selecting higher values of m than 4. However, from practical point of view, $m = 3$ could be more suitable as it may be difficult to collect blood samples within 15 minutes.

6.3.3 Sensitivity to the Initial Values

In the original simulation study, the true values of the parameters were taken to be $\mathbf{\Psi}_{true} = (.85, .15, 17, .1, .1, .1, .1)^T$ and the initial values were taken as $\mathbf{\Psi}_o = (1, .1, 20, .05, .15, .05, .15)^T$. We now examine if the choice of the initial values has any bearing on the performance of the methodology.

We showed in Section 5.2 that for the model given in Equation (6.2), the standard deviation of K_{ai} is given by $K_a\sqrt{e^{\omega_1}(e^{\omega_1}-1)}$. For $\mathbf{\Psi}_{true}$, this is computed as 0.29. Similarly, the standard deviations of K_{ei} and V_i are computed as 0.05 and 5.80. If we consider intervals around the true values of K_a, K_e and V, of width equal to two standard deviations of the corresponding random effect, then the intervals around K_a, K_e and V are given as: (.56, 1.14), (.1, .2) and (11.2, 22.8). To see the effect of initial values on our algorithm, we consider six new scenarios as described below.

In the first, we take the upper limits of these intervals as the initial values of the parameters and keep the initial values of the variance parameters unchanged, that is, as in $\mathbf{\Psi}_o$. In the second, we choose the lower limits of these intervals as the initial values of the parameters and keep the initial values of the variance parameters unchanged.

For the variance parameters $(\omega_1, \omega_2, \omega_3, \sigma^2)$, since true value of each is 0.1, we consider intervals of (.01, .5) for all the four parameters. For the third scenario, we keep the three PK parameters as they were in the original simulation study and assign the value 0.01 to all the variance parameters. In the fourth scenario, we keep the three PK parameters as they were in the original simulation study and assign the value 0.5 to all the variance parameters. As discussed in Section 6.2.1, the variances of .01 and .5 are approximately equivalent to CVs of 10% and 70%.

We also explore the performance of the method when the initial values are very close or coincide with the true parameter values contained in $\mathbf{\Psi}_{true}$. In all, we run six

scenarios which have the following vectors as the initial values of the parameters:

$$\boldsymbol{\Psi}_{o1} = (1.14, .2, 22.8, .05, .15, .05, .15)^T,$$

$$\boldsymbol{\Psi}_{o2} = (.56, .1, 11.2, .05, .15, .05, .15)^T,$$

$$\boldsymbol{\Psi}_{o3} = (1, .1, 20, .01, .01, .01, .01)^T,$$

$$\boldsymbol{\Psi}_{o4} = (1, .1, 20, .50, .50, .50, .50)^T,$$

$$\boldsymbol{\Psi}_{o5} = (.9, .13, 18, .12, .12, .12, .12)^T,$$

$$\boldsymbol{\Psi}_{o6} = \boldsymbol{\Psi}_{true}.$$

Through these six vectors, we want to see the effect of using vague initial values on the performance of our algorithm. In fact, the above vectors of initial values roughly cover the boundaries of the simulated parameters' distributions depicted in Figure 6.6.

The results from the simulation studies for these six scenarios are presented in Tables 6.12 and 6.13. The data corresponding to the initial values in the original study ($\boldsymbol{\Psi}_o$) are re-presented for making comparisons.

The values in Table 6.12 suggest that the choice of the initial values does not significantly affect the variability in the simulated distributions, as evidenced by the nearly same CVs of the parameters for the different initial values.

The biases in the estimated parameters also seem to be invariant to the choice of the initial values of the variance parameters. This can be observed from the data corresponding to the initial values $\boldsymbol{\Psi}_{o3}$ and $\boldsymbol{\Psi}_{o4}$. However, for $\boldsymbol{\Psi}_{o1}$ and $\boldsymbol{\Psi}_{o2}$, the biases seem to be correlated with the corresponding initial values, especially for the parameters V and ω_3. Selection of initial values close to the true values do not seem to yield any additional significant benefit in reducing the bias and the variability. This is evident from the insignificant difference between the data corresponding to $\boldsymbol{\Psi}_o$, $\boldsymbol{\Psi}_{o5}$ and $\boldsymbol{\Psi}_{o6}$.

The D-optimal sampling times ξ^*_{true} corresponding to the true parameters $\boldsymbol{\Psi}_{true}$ are {.10, 6.60, 48.00}. The mean D-optimal sampling times ξ^* over the C cohorts for the considered seven scenarios are presented in Figure 6.20.

The first and third sampling times converge after cohort 4 to 0.10 and 48. This behaviour of the border points was observed in the original simulation study as well. The second sampling times, also stabilise from cohort 4 to values around 6.60. Since no variability in the first and third sampling times was observed in the simulations at cohort C, we explore the distributions of the second sampling time for the five

		K_a	K_e	V	ω_1	ω_2	ω_3	σ^2	
$\hat{\hat{\boldsymbol{\Psi}}}$	$\boldsymbol{\Psi}_o$	0.8192	0.1580	17.0449	0.0870	0.0940	0.1165	0.1050	
	$\boldsymbol{\Psi}_{o1}$	0.8273	0.1542	18.9483	0.0856	0.0977	0.1441	0.1039	
	$\boldsymbol{\Psi}_{o2}$	0.8177	0.1572	15.1813	0.0851	0.0949	0.1276	0.1045	
	$\boldsymbol{\Psi}_{o3}$	0.8175	0.1580	16.9072	0.0835	0.0945	0.1182	0.1059	
	$\boldsymbol{\Psi}_{o4}$	0.8117	0.1612	17.1278	0.0835	0.0934	0.1193	0.1066	
	$\boldsymbol{\Psi}_{o5}$	0.8248	0.1557	16.5119	0.0872	0.0948	0.1173	0.1050	
	$\boldsymbol{\Psi}_{o6}$	0.8232	0.1552	16.3992	0.0851	0.0964	0.1161	0.1052	
$\widehat{\text{Bias}}(\hat{\boldsymbol{\Psi}})$ p.c.	$\boldsymbol{\Psi}_o$	-3.6	5.3	.3	-13.0	-6.0	16.5	5.0	[**7.1**]
	$\boldsymbol{\Psi}_{o1}$	-2.7	2.8	11.5	-14.4	-2.3	44.1	3.9	[**11.7**]
	$\boldsymbol{\Psi}_{o2}$	-3.8	4.8	-10.7	-14.9	-5.1	27.6	4.5	[**10.2**]
	$\boldsymbol{\Psi}_{o3}$	-3.8	5.4	-.5	-16.5	-5.5	18.2	6.0	[**8.0**]
	$\boldsymbol{\Psi}_{o4}$	-4.5	7.5	.75	-16.5	-6.6	19.3	6.6	[**8.8**]
	$\boldsymbol{\Psi}_{o5}$	-3.0	3.8	-2.9	-12.8	-5.2	17.3	5.0	[**7.1**]
	$\boldsymbol{\Psi}_{o6}$	-3.1	3.5	-3.5	-14.9	-3.6	16.1	5.2	[**7.1**]
CV	$\boldsymbol{\Psi}_o$	7.5	22.5	16.3	45.8	17.90	29.0	19.1	[**22.6**]
	$\boldsymbol{\Psi}_{o1}$	6.5	13.5	14.8	46.2	18.4	29.2	19.4	[**21.2**]
	$\boldsymbol{\Psi}_{o2}$	7.1	19.3	16.0	47.0	18.6	28.4	19.3	[**22.2**]
	$\boldsymbol{\Psi}_{o3}$	7.5	24.1	15.9	47.7	18.9	30.9	18.1	[**23.3**]
	$\boldsymbol{\Psi}_{o4}$	9.0	35.4	17.6	50.1	19.1	29.8	18.5	[**25.6**]
	$\boldsymbol{\Psi}_{o5}$	7.2	18.0	15.9	46.4	18.8	28.1	18.5	[**21.9**]
	$\boldsymbol{\Psi}_{o6}$	6.8	12.2	15.3	46.6	17.9	28.3	18.5	[**20.8**]

Table 6.12: Comparison of statistics related to different vectors of initial values. For each case, the numbers in the bold are the average absolute percentage bias and the average CV of the parameter estimates.

scenarios.

In Figure 6.21, we present the distribution of T_2^*, the second element of ξ_C^*, for the seven scenarios.

The spreads of the distributions are not very different but the median does seem to be influenced by the choice of the initial values of the PK parameters in the second and third scenarios. Overall, we do not see a large effect of the choice of initial values on the D-optimal sampling points, however, the second sampling points for the cases of $\boldsymbol{\Psi}_{o1}$ and $\boldsymbol{\Psi}_{o2}$ are slightly different from each other.

An apparently counter-intuitive observation from Table 6.13 is the relatively higher values of the ACNs for $\boldsymbol{\Psi}_{o1}$ and $\boldsymbol{\Psi}_{o2}$ despite no significantly larger variability in

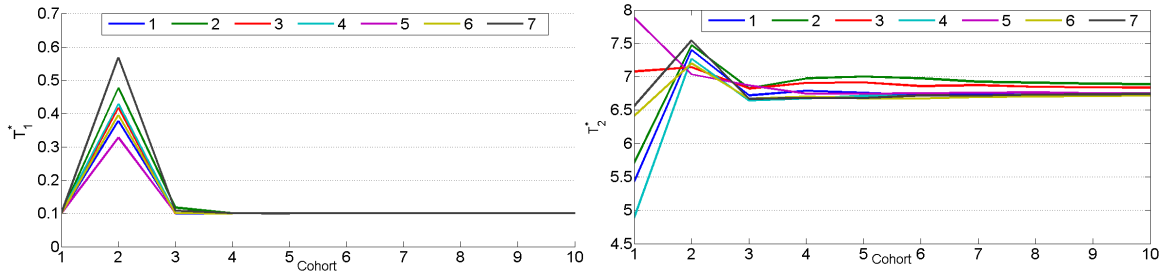
		d_1^*	d_2^*	d_3^*	d_4^*	d_5^*	$\bar{\varphi}_A$
$\boldsymbol{D}^*$	$\boldsymbol{\Psi}_o$	145.58	92.79	97.02	96.67	97.02	8.5577
	$\boldsymbol{\Psi}_{o1}$	156.28	98.88	103.43	102.75	103.43	8.7446
	$\boldsymbol{\Psi}_{o2}$	134.18	84.56	89.41	88.45	89.42	8.9229
	$\boldsymbol{\Psi}_{o3}$	144.73	92.02	96.29	95.74	96.30	8.5012
	$\boldsymbol{\Psi}_{o4}$	146.34	93.36	97.51	97.01	97.53	8.6045
	$\boldsymbol{\Psi}_{o5}$	141.42	89.43	93.67	93.22	93.69	8.5684
	$\boldsymbol{\Psi}_{o6}$	140.61	88.80	93.21	92.80	93.22	8.4978
		T_1^*	T_2^*	T_3^*			
ξ_C^*	$\boldsymbol{\Psi}_o$	.10	6.75	48			
	$\boldsymbol{\Psi}_{o1}$	.10	6.89	48			
	$\boldsymbol{\Psi}_{o2}$	.10	6.84	48			
	$\boldsymbol{\Psi}_{o3}$	.10	6.74	48			
	$\boldsymbol{\Psi}_{o4}$	.10	6.75	48			
	$\boldsymbol{\Psi}_{o5}$	.10	6.72	48			
	$\boldsymbol{\Psi}_{o6}$	.10	6.74	48			
			$\bar{C}_2$		$\bar{C}_3$		
ACN	$\boldsymbol{\Psi}_o$		3.81		7.52		
	$\boldsymbol{\Psi}_{o1}$		5.38		7.90		
	$\boldsymbol{\Psi}_{o2}$		4.97		8.31		
	$\boldsymbol{\Psi}_{o3}$		3.73		7.60		
	$\boldsymbol{\Psi}_{o4}$		3.83		7.65		
	$\boldsymbol{\Psi}_{o5}$		3.74		7.44		
	$\boldsymbol{\Psi}_{o6}$		3.56		7.35		

Table 6.13: Comparison of statistics related to different vectors of initial values.

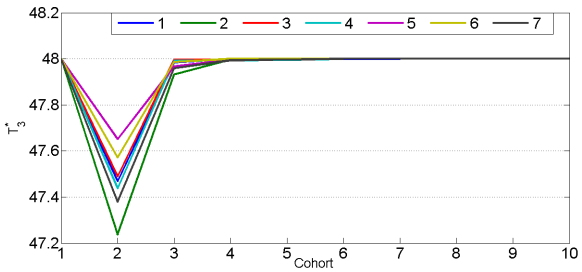
the estimated parameters (as measured by the CV). To understand this anomaly, we examined the distributions for the two stopping rules for $\boldsymbol{\Psi}_{oi}$, $i = 1, ..., 6$ and compared them with the original run with $\boldsymbol{\Psi}_o$. Figure 6.22 presents the percentage distribution of the cohort number at which the two stopping rules SR2 and SR3 apply.

For SR2, the distributions for $\boldsymbol{\Psi}_{o1}$ and $\boldsymbol{\Psi}_{o2}$ seem to be slightly shifted to the right, as compared with the other five vectors of initial values.

The explanation for this behaviour can be traced to the starting dose regimen $\boldsymbol{D}_1^*$ which is administered to cohort 1 using the given initial values. For $\boldsymbol{\Psi}_o$, $\boldsymbol{D}_1^*$ for cohort 1 was found to be $(150, 70, 80, 70, 80)^T$. This is the same regimen which



(a) T_1^* (b) T_2^*



(c) T_3^*

Figure 6.20: Average D-optimal sampling times over the course of the trial for the seven scenarios consisting of choice of the initial values: $\boldsymbol{\Psi}_o$ and $\boldsymbol{\Psi}_{oi}$, $i = 1, ..., 6$ over the $N_{sim} = 1000$ simulations. The first and the third sampling times stabilise to the values of .10 and 48 from cohort 4, which are at the boundary of the design region while the second sampling times stabilise to around 6.60, which is the second sampling time in ξ^*_{true}.

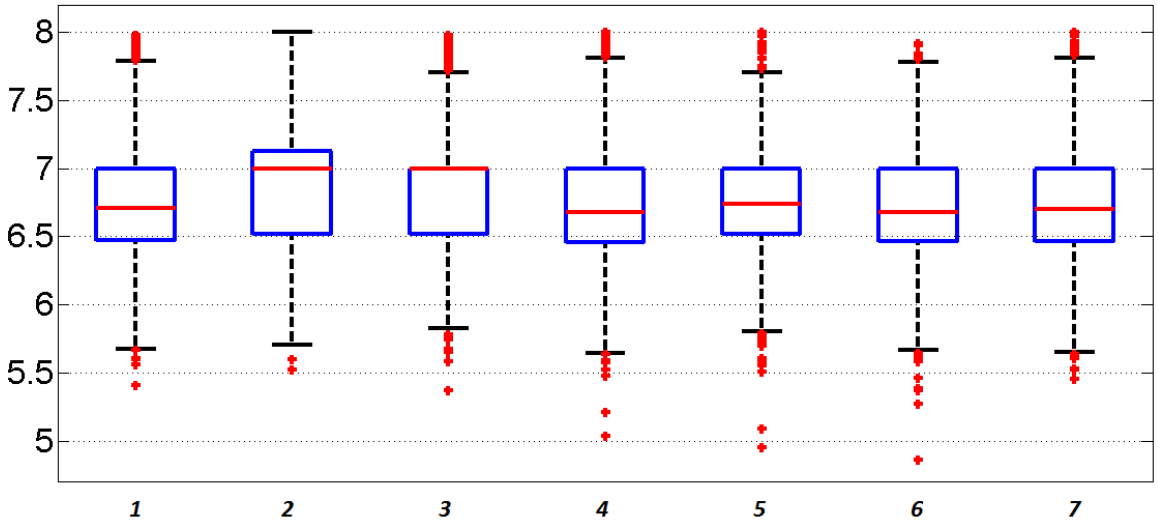
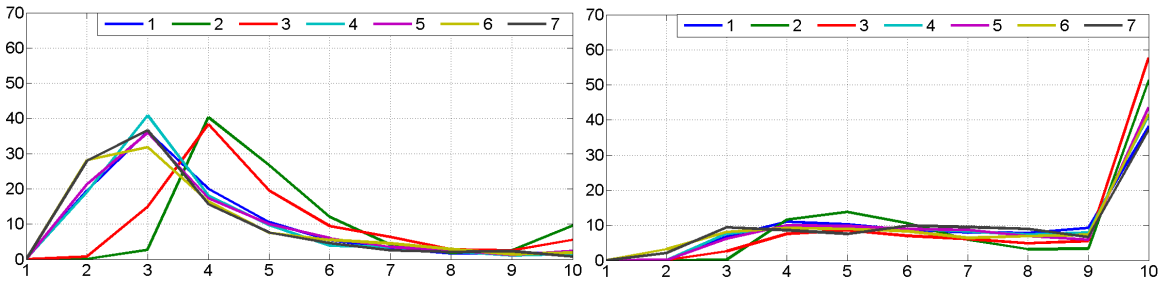


Figure 6.21: Distribution of T_2^*, the second element of ξ_C^* for the seven scenarios consisting of choice of the initial values: $\boldsymbol{\Psi}_o$ and $\boldsymbol{\Psi}_{oi}$, $i = 1, ..., 6$ over the $N_{sim} = 1000$ simulations.

is administered to cohort 1 when $\boldsymbol{\Psi}_{o3}$ and $\boldsymbol{\Psi}_{o4}$ are used as initial values since the ED algorithm is not dependent on the variance parameters. However, for $\boldsymbol{\Psi}_{o1}$ and $\boldsymbol{\Psi}_{o2}$ the corresponding $\boldsymbol{D}_1^*$ are $(200, 160, 160, 160, 160)^T$ and $(110, 50, 60, 60, 60)^T$. Now, for the vector of true parameters $\boldsymbol{\Psi}_{true}$, the true dose regimen $\boldsymbol{D}_1^*$ is given as



(a) SR2 (b) SR3

Figure 6.22: Percentage distribution for the two stopping rules SR2 and SR3 for the seven scenarios consisting of choice of the initial values: $\boldsymbol{\Psi}_o$ and $\boldsymbol{\Psi}_{oi}$, $i = 1, ..., 6$.

$(140, 90, 90, 100, 90)^T$. For $\boldsymbol{\Psi}_o$ (as well as for $\boldsymbol{\Psi}_{o3}$ and $\boldsymbol{\Psi}_{o4}$), since the starting dose regimen is closer to the true dose regimen, the conditions of the stopping rules are met sooner. For example, for SR2, it can be seen from the figure that in about 60% of the simulations the trial is terminated at cohort 2 or 3 for scenarios 1, 4 and 5. However, for scenarios 2 and 3, the corresponding percentages are approximately 3 and 15. Because of this, the ACNs for scenarios 2 and 3 get shifted towards the latter cohorts.

For SR3, the distributions for $\boldsymbol{\Psi}_{o1}$ and $\boldsymbol{\Psi}_{o2}$ are mostly similar except that in a higher percentage of the simulations, the trial is terminated at cohort 10. For the same reason as above, scenarios 2 and 3 have higher percentages of simulations in which the trial is stopped at the last cohort that is cohort 10.

The same argument can be used to explain the low ACN values for scenario 7, i.e., for the case of $\boldsymbol{\Psi}_{o6}$. Since the initial values are equal to the true parameter values, the starting dose regimen $\boldsymbol{D}_1^*$ coincides with the φ_A-efficient dose regimen corresponding to $\boldsymbol{\Psi}_{true}$. The proximity of $\boldsymbol{D}_1^*$ to the true dose regimen results in earlier application of the stopping rules SR2 and SR3. In fact, the ACN values for this scenario are the lowest. Furthermore, since $\boldsymbol{\Psi}_{o5}$ and $\boldsymbol{\Psi}_{o6}$ are closest to $\boldsymbol{\Psi}_{true}$, it can be seen from Figure 6.22 that in around 30% of the simulations, SR2 is applied at cohort 2 for these two scenarios compared to around 20% for scenarios 1, 4 and 5 and almost nil for scenarios 2 and 3. This implies that the closer the initial values are to the true values, the trial will, on average, terminate earlier.

As can be seen in Table 6.13, $\bar{\varphi}_A$ values are relatively larger for scenarios 2 and 3. This is on account of the deviation of the final administered dose regimens in these scenarios from the true optimal dose regimen, $\boldsymbol{D}^* = (140, 90, 90, 100, 90)^T$. Figure 6.23 presents the distribution of φ_A corresponding to $\boldsymbol{D}^*$ at cohort C over the $N_{sim} = 1000$ simulations. The distributions are mostly similar, however, the spread is relatively larger for the second and the third scenario. Also, $\bar{\varphi}_A$ and the median of φ_A values are larger for these two scenarios. A higher spread of φ_A values means that a greater

number of subjects in a simulation study are administered sub-optimal dose regimens. No significant differences are observed in the other five scenarios, though the $\bar{\varphi}_A$ value for the case of $\boldsymbol{\Psi}_{o6}$ is the lowest.

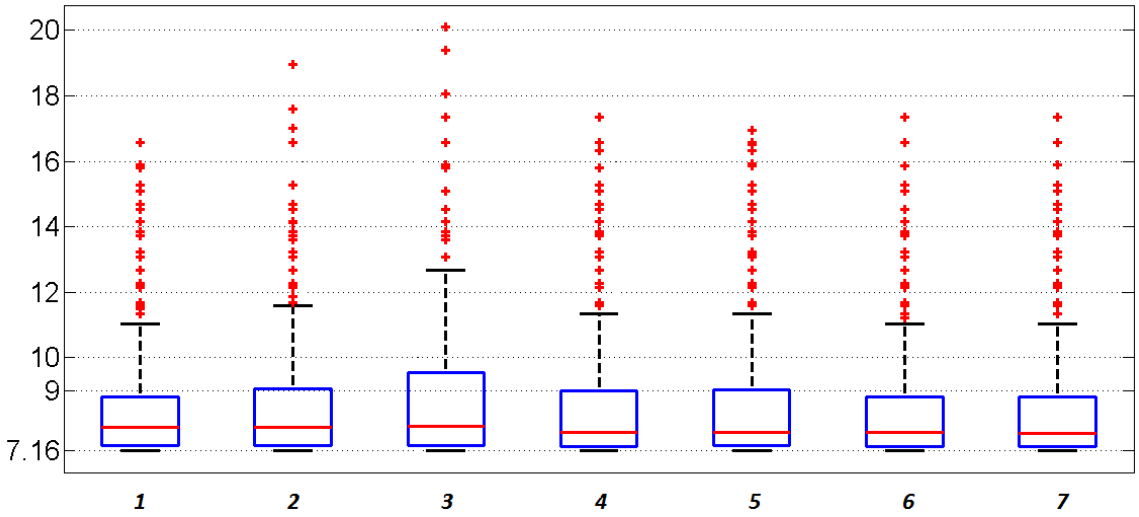


Figure 6.23: Distribution of φ_A for the seven scenarios consisting of choice of the initial values: $\boldsymbol{\Psi}_o$ and $\boldsymbol{\Psi}_{oi}$, $i = 1, ..., 6$.

We conclude from our analyses that while vague initial values for the variance parameters do not significantly affect the outcome, vague initial values of the PK parameters can amplify the bias and variability in the parameter estimates, lead to inefficient dose regimens (evident from the higher values of $\bar{\varphi}_A$) and can delay the termination of the trial (evident from the higher ACNs). Initial values of the mean PK parameters should therefore be closer to the true values to minimise these departures. In the next section, we examine the performance when the initial values have a stronger deviation from the true values.

6.3.4 Stronger Deviation of the Initial Values

In the original simulation study in Section 6.2.1, the true values of the parameters were taken to be $\boldsymbol{\Psi}_{true} = (.85, .15, 17, .10, .10, .10, .10)^T$ and the initial values were taken as $\boldsymbol{\Psi}_o = (1, .10, 20, .05, .15, .05, .15)^T$. In the previous section, we studied the effect of using different vectors of initial values on the performance of our methodology. The initial values we assumed previously were within the distance of two standard deviations from the mean taken to be the true parameter value.

In this section, we evaluate the case when the initial values are substantially different from the true values. The difference is expressed in two contexts. The first is the difference between the true values and the initial values. The second is the deviation of the ratios of the true parameters from the ratios of the initial values. This is discussed below.

Difference between parameters' values

Using a trial and error approach, the algorithm was run for several vectors of initial values which were quite distant from the true values. In many cases, for large deviations of the initial values from the true values, the algorithm terminated abruptly due to computational problems. The maximum workable deviation for a parameter was determined to be about 3 times of the respective true value on the higher side and about (1/3) times on the lower side. Let us consider the following two vectors as the initial values of the parameters:

$$\boldsymbol{\Psi}_{o-} = (.30, .05, 6, .02, .02, .02, .03)^T,$$

$$\boldsymbol{\Psi}_{o+} = (2.60, .50, 51, .50, .30, .30, .30)^T.$$

In terms of standard deviations, the initial values in $\boldsymbol{\Psi}_{o+}$ are about 5 standard deviations more than the corresponding true values of the PK parameters whereas the initial values in $\boldsymbol{\Psi}_{o-}$ are about 2 standard deviations less than the corresponding true values. Taking initial values of these parameters outside the interval $(\boldsymbol{\Psi}_{o-}, \boldsymbol{\Psi}_{o+})$ generally causes the algorithm to terminate unsuccessfully. The simulation study described in Section 6.2.1 is now run for these two vectors of initial values. The results from the simulation studies for these two vectors are presented in Table 6.14. The data corresponding to the initial values in the original study ($\boldsymbol{\Psi}_o$) are re-presented for making comparisons.

The biases in the estimates for the initial values $\boldsymbol{\Psi}_{o-}$ and $\boldsymbol{\Psi}_{o+}$ are larger only for some parameters as compared to when the initial values are $\boldsymbol{\Psi}_o$. Parameters like V, ω_1 and ω_3 are more severely affected than the other parameters. The increase in the variability of the simulated parameters, as measured by their CV, is also significant for some of the parameters. Overall, quality of the estimates of most parameters deteriorates when the initial values are strongly deviated from the true values.

As can be seen in Table 6.14, the average dose regimens administered to the last cohort for the initial values $\boldsymbol{\Psi}_{o-}$ and $\boldsymbol{\Psi}_{o+}$ are quite different from the optimal $\boldsymbol{D}^*_{true}$. Consequently, the associated $\bar{\varphi}_A$ values are quite large for $\boldsymbol{\Psi}_{o-}$ and $\boldsymbol{\Psi}_{o+}$ as compared to when $\boldsymbol{\Psi}_o$ is used as initial values. The reason that the average recommended doses for the initial values $\boldsymbol{\Psi}_{o-}$ and $\boldsymbol{\Psi}_{o+}$ are large is the overestimation of the parameter V in these two scenarios. Since V appears in the denominator of the assumed compartmental model, the ED algorithm compensates for the large values of this parameter by increasing the magnitude of the administered doses.

Figure 6.24 presents the distribution of φ_A over the course of $C = 10$ cohorts for the three vectors of initial values.

		K_a	K_e	V	ω_1	ω_2	ω_3	σ^2	
	$\boldsymbol{\Psi}_{o-}$	0.8578	0.1475	20.7526	0.0716	0.1096	0.4990	0.1102	
$\bar{\widehat{\boldsymbol{\Psi}}}$	$\boldsymbol{\Psi}_o$	0.8192	0.1580	17.0449	0.0870	0.0940	0.1165	0.1050	
	$\boldsymbol{\Psi}_{o+}$	0.8235	0.1522	19.3055	0.0843	0.0986	0.1563	0.1054	
	$\boldsymbol{\Psi}_{o-}$	0.9	-1.7	22.1	-28.4	9.6	398.9	10.2	[**67.4**]
$\widehat{\text{Bias}}(\widehat{\boldsymbol{\Psi}})$	$\boldsymbol{\Psi}_o$	-3.6	5.3	0.3	-13.0	-6.0	16.5	5.0	[**7.1**]
p.c.	$\boldsymbol{\Psi}_{o+}$	-3.1	1.5	13.6	-15.7	-1.4	56.3	5.4	[**13.9**]
	$\boldsymbol{\Psi}_{o-}$	6.8	5.2	20.2	67.8	17.8	31.9	21.2	[**24.4**]
CV	$\boldsymbol{\Psi}_o$	7.5	22.5	16.3	45.8	17.9	29.0	19.1	[**22.6**]
	$\boldsymbol{\Psi}_{o+}$	6.1	5.1	21.2	84.0	18.6	34.4	20.8	[**27.2**]
		d_1^*	d_2^*	d_3^*	d_4^*	d_5^*		$\bar{\varphi}_A$	
$\boldsymbol{D}^*_{true}$		140.00	90.00	90.00	100.00	90.00		7.1561	
	$\boldsymbol{\Psi}_{o-}$	30.00	10.00	10.00	10.00	10.00		33.6442	
$\boldsymbol{D}^*$	$\boldsymbol{\Psi}_o$	150.00	70.00	80.00	70.00	80.00		9.0129	
Cohort 1	$\boldsymbol{\Psi}_{o+}$	200.00	200.00	200.00	200.00	200.00		31.0280	
	$\boldsymbol{\Psi}_{o-}$	177.85	119.40	121.23	121.00	121.28		11.7678	
$\boldsymbol{D}^*$	$\boldsymbol{\Psi}_o$	145.58	92.79	97.02	96.67	97.02		8.5577	
Last cohort	$\boldsymbol{\Psi}_{o+}$	157.84	99.75	103.84	103.81	103.83		12.2722	
	$\boldsymbol{\Psi}_{o-}$	170.00	110.00	110.00	110.00	110.00		8.7036	
$\boldsymbol{D}^*(\widehat{\boldsymbol{\Psi}})$	$\boldsymbol{\Psi}_o$	150.00	90.00	100.00	100.00	100.00		7.2585	
Recommended	$\boldsymbol{\Psi}_{o+}$	170.00	100.00	110.00	110.00	110.00		8.3987	
		T_1^*	T_2^*	T_3^*					
	$\boldsymbol{\Psi}_{o-}$	.10	6.94	48.00					
ξ_C^*	$\boldsymbol{\Psi}_o$	.10	6.75	48.00					
	$\boldsymbol{\Psi}_{o+}$	.10	6.98	48.00					
			$\bar{C}_2$		$\bar{C}_3$				
	$\boldsymbol{\Psi}_{o-}$		4.76		5.38				
ACN	$\boldsymbol{\Psi}_o$		3.81		7.52				
	$\boldsymbol{\Psi}_{o+}$		5.64		8.10				

Table 6.14: Comparison of the data related to the three vectors of initial values: $\boldsymbol{\Psi}_{o-}$, $\boldsymbol{\Psi}_o$ and $\boldsymbol{\Psi}_{o+}$.

The ranges of the distributions for $\boldsymbol{\Psi}_{o-}$ and $\boldsymbol{\Psi}_{o+}$ are quite large as compared to the case of $\boldsymbol{\Psi}_o$. This implies that substantially suboptimal dose regimens are administered in a number of simulations, with some values of φ_A being about four times of the optimal value of $\varphi_A^* = 7.16$. Therefore, large deviations of the initial values from the true values can result in significant levels of under- and over-exposure experienced by

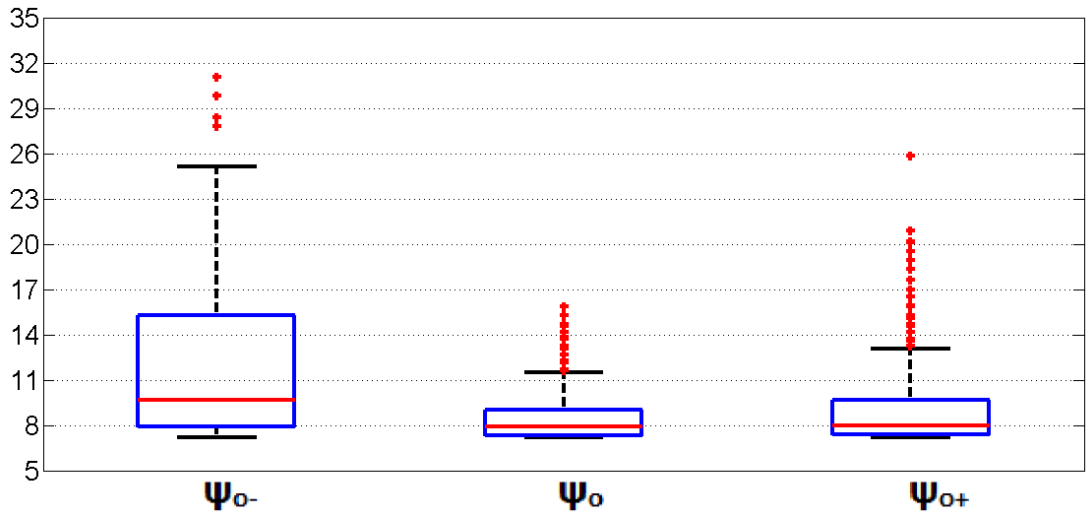
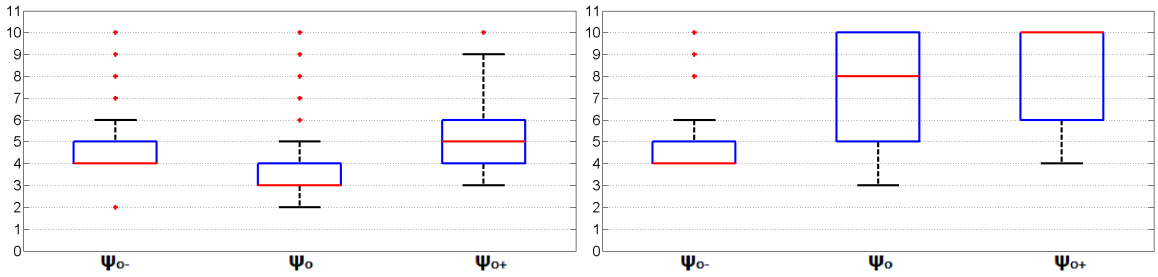


Figure 6.24: Distribution of φ_A for the three vectors of initial values of the parameters.

the subjects over the course of the trial.

A few more observations can be made from Table 6.14. Firstly, the dose regimens that are administered to the first cohort for the vectors $\boldsymbol{\Psi}_{o-}$ and $\boldsymbol{\Psi}_{o+}$ are significantly different from the optimal dose regimen $\boldsymbol{D}^*_{true} = (140, 90, 90, 100, 90)^T$. As a result, the subjects in the first cohort in both the cases are administered unethical dose levels. However, towards the end of the trial, the administered dose regimens are closer to the optimal dose regimen. Moreover, the recommended dose regimens for $\boldsymbol{\Psi}_{o-}$ and $\boldsymbol{\Psi}_{o+}$, based on the final parameters' estimates obtained from the trial, are not very far from the one corresponding to $\boldsymbol{\Psi}_o$. This points to an advantage of using the adaptive procedure which is discussed in more detail in Section 6.4.

The ACN values are, in general, higher for vectors $\boldsymbol{\Psi}_{o-}$ and $\boldsymbol{\Psi}_{o+}$ as compared to $\boldsymbol{\Psi}_o$. Again, as explained in Section 6.3.3, this is attributable to the fact that the starting dose regimens for these vectors of initial vectors are significantly different from the optimal dose regimen which delays the application of the stopping rule. Figure 6.25 presents the distribution of the cohorts at which the two stopping rules, SR2 and SR3, were applied for the three vectors of initial values.



(a) SR2 (b) SR3

Figure 6.25: Percentage distribution for the two stopping rules SR2 and SR3 for the three scenarios.

An interesting observation can be made from Figure 6.25b. The value of $\bar{C}_3$ and the spread of C_3 for this vector of initial values is the lower than even the case of $\boldsymbol{\Psi}_o$ when the initial values are moderately close to the true values. This seems counter-intuitive since a larger deviation from the true values would have caused delay in the application of the stopping rules which would have resulted in a larger value of $\bar{C}_3$.

To explore this anomaly, we compare the modal dose regimens administered to each of the ten cohorts for the three vectors of initial values. These data are presented in Table 6.15 and Figure 6.26.

	$\boldsymbol{\Psi}_{o-}$					$\boldsymbol{\Psi}_o$					$\boldsymbol{\Psi}_{o+}$				
Cohort	d_1	d_2	d_3	d_4	d_5	d_1	d_2	d_3	d_4	d_5	d_1	d_2	d_3	d_4	d_5
1	30	10	10	10	10	150	70	80	70	80	200	200	200	200	200
2	140	90	90	90	90	130	90	90	90	90	140	90	90	90	90
3	**200**	**200**	**200**	**200**	**200**	130	80	90	90	90	120	70	80	70	70
4	**200**	**200**	**200**	**200**	**200**	130	80	90	90	90	110	70	70	70	70
5	**200**	**200**	**200**	**200**	**200**	130	80	90	90	90	110	70	70	70	70
6	**200**	**200**	**200**	**200**	**200**	130	90	90	90	90	120	70	80	80	80
7	200	200	140	140	140	130	90	90	90	90	130	80	80	80	80
8	200	120	130	130	130	130	90	90	90	90	140	80	90	90	90
9	200	120	120	120	120	140	90	90	90	90	140	90	90	90	90
10	200	110	110	110	110	130	90	100	90	100	150	100	100	100	100

Table 6.15: Modal dose regimens at each of the ten cohorts for the three sets of the initial values.

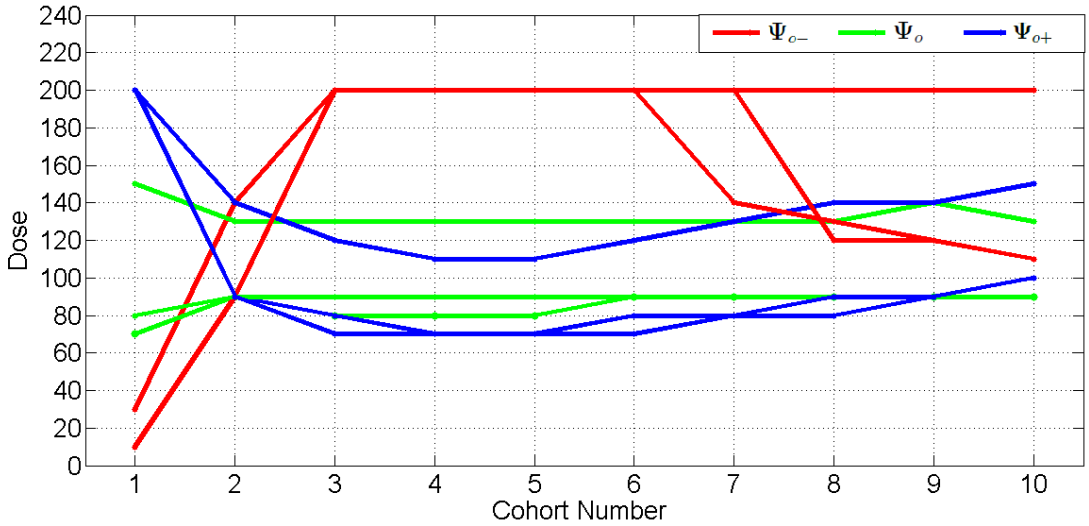


Figure 6.26: Plot of the modal dose regimens at each of the ten cohorts for the three sets of the initial values: $\boldsymbol{\Psi}_{o+}$, $\boldsymbol{\Psi}_{o+}$ and $\boldsymbol{\Psi}_{o+}$

From the data, it can be seen that for the vector $\boldsymbol{\Psi}_{o-}$, the modal dose regimens administered to cohorts 3, 4, 5 and 6 consist of the maximum permissible dose size, $d_{max} = 200$. This consistency in the modal dose regimens for these successive four cohorts results in early application of the stopping rule SR3 that is reflected in the lowest $\bar{C}_3$ value for $\boldsymbol{\Psi}_{o-}$ followed by the one corresponding to $\boldsymbol{\Psi}_o$ and then to $\boldsymbol{\Psi}_{o+}$.

This stability in the dose regimens of the successive cohorts triggers application of both stopping rules SR2 and SR3 which results in similar distributions of SR2 and SR3 for $\boldsymbol{\Psi}_{o-}$.

Furthermore, the large dose sizes make the dose regimens administered to these cohorts highly suboptimal that results in large values of φ_A for $\boldsymbol{\Psi}_{o-}$. On the other hand, the dose regimens administered to the cohorts for initial values $\boldsymbol{\Psi}_o$ and $\boldsymbol{\Psi}_{o+}$ are relatively closer to the true optimal dose regimen $\boldsymbol{D}^*$ resulting in relatively smaller spread in the distribution of φ_A.

To understand the reason behind the large dose sizes corresponding to $\boldsymbol{\Psi}_{o-}$, we also explored the distribution of the estimated parameters for each cohort. Table 6.16 presents the mean estimators at each cohort for the vectors $\boldsymbol{\Psi}_{o-}$, $\boldsymbol{\Psi}_o$ and $\boldsymbol{\Psi}_{o+}$.

	$\bar{\hat{K}}_a$			$\bar{\hat{K}}_e$			$\bar{\hat{V}}$		
Cohort	$\boldsymbol{\Psi}_{o-}$	$\boldsymbol{\Psi}_o$	$\boldsymbol{\Psi}_{o+}$	$\boldsymbol{\Psi}_{o-}$	$\boldsymbol{\Psi}_o$	$\boldsymbol{\Psi}_{o+}$	$\boldsymbol{\Psi}_{o-}$	$\boldsymbol{\Psi}_o$	$\boldsymbol{\Psi}_{o+}$
1	0.9557	0.8479	0.8365	0.1398	0.1551	0.1550	15.0157	16.8318	16.7674
2	0.8416	0.8029	0.8465	0.1683	0.1657	0.1414	**32.2667**	15.6445	14.3686
3	0.8097	0.8053	0.8448	0.1800	0.1643	0.1431	**36.5948**	15.3679	13.4200
4	0.7942	0.8101	0.8383	0.1797	0.1619	0.1460	**32.2973**	15.4250	13.5122
5	0.8171	0.8145	0.8304	0.1686	0.1604	0.1485	**28.6288**	15.6578	14.2044
6	0.8345	0.8167	0.8278	0.1601	0.1595	0.1499	26.4795	15.9801	15.2269
7	0.8440	0.8180	0.8266	0.1545	0.1588	0.1508	24.9094	16.3427	16.4234
8	0.8519	0.8173	0.8263	0.1507	0.1587	0.1512	23.5405	16.6338	17.6276
9	0.8577	0.8185	0.8250	0.1483	0.1583	0.1517	22.1858	16.8809	18.6269
10	0.8578	0.8192	0.8235	0.1475	0.1580	0.1522	20.7526	17.0449	19.3055

Table 6.16: Averaged estimated parameters at each cohort for the three vectors of initial values.

It can be seen from the table that for the initial values $\boldsymbol{\Psi}_{o-}$, the estimated values of the parameter V for cohorts 2 to 5 are very large. This leads to choice of the maximum permissible doses for cohorts 3 to 6. For $\boldsymbol{\Psi}_o$ and $\boldsymbol{\Psi}_{o+}$, the average estimated parameters for any cohort do not take such extreme values and so the administered dose regimens are closer to the optimal dose regimen.

In conclusion, large deviation of the initial values from the true values can result in significant increase in bias in some of the estimated parameters. This may lead to administration of inefficient dose regimens to some of the cohorts. Therefore, it may be sensible to impose an upper limit which is not too high on the dose sizes, as we did in this study by having $d_{max} = 200$. This protects the subjects from potentially experiencing significant levels of over-exposure to the drug.

Furthermore, for $\boldsymbol{\Psi}_{o-}$ and $\boldsymbol{\Psi}_{o+}$, the dose regimens administered to cohort number 10 are significantly more efficient than the ones administered to cohort number 1. Also, as can be observed from Table 6.16, the mean estimated parameters generally

improve over the course of the adaptive trial. Therefore, when the initial values are not known with reasonable confidence, it is better to evaluate all the cohorts in the trial rather than using a stopping rule to terminate the trial early. This, as discussed above, will lead to a good estimate of the optimal dose regimen towards the end of the trial.

Difference between parameters' ratios

In Section 4.4.1, we demonstrated that the efficiency of the administered dose regimens depends not only on the degree of misspecification but also on which parameters are misspecified. Since the dose regimen administered to the first cohort depends on the initial values of parameters, we further explore this idea here.

Let us consider the following two vectors as the initial values of the parameters:

$$\boldsymbol{\Psi}_{or-} = (1.5, .1, 6, .05, .15, .05, .15)^T,$$

$$\boldsymbol{\Psi}_{or+} = (.5, .1, 30, .05, .15, .05, .15)^T.$$

Let us consider the ratio $R_p = V/K_a$. As can be seen from the model defined in Equation (6.1) that keeping other parameters fixed, a large V is associated with smaller values of the concentration whereas increasing K_a results in larger values of the concentration. Therefore, other factors held fixed, the larger the ratio R_p is, the lower is the exposure to the drug.

Now, the true values of V and K_a are 17 and 0.85, that is, $R_p = 20$. Previously, for the initial values $\boldsymbol{\Psi}_{o-}$ and $\boldsymbol{\Psi}_{o+}$, the values of this ratio were 20 and 19.6 respectively which are almost the same as the ratio of the true values of the parameters, despite the initial values themselves being far-off from the true values.

For the vectors $\boldsymbol{\Psi}_{or-}$ and $\boldsymbol{\Psi}_{or+}$ we have defined above, the value of this ratio is $6/1.5 = 4$ and $30/.5 = 60$ respectively. These values have a large deviation from 20, the ratio's value for the true parameters. For the values contained in $\boldsymbol{\Psi}_{or-}$, the first cohort is expected to be severely over-exposed to the drug because of small value of R_p. Similarly, the first cohort is expected to be under-exposed to the drug for the case of $\boldsymbol{\Psi}_{or+}$. The performance of the adaptive procedure at these two vectors of initial values is studied in this section.

The results from the simulation studies for these two vectors are presented in Table 6.17. The data corresponding to the initial values in the original study ($\boldsymbol{\Psi}_o$) are re-presented for making comparisons.

		K_a	K_e	V	ω_1	ω_2	ω_3	σ^2	
	$\boldsymbol{\Psi}_{or-}$	0.8264	0.1550	19.2341	0.0833	0.0916	0.1508	0.1056	
$\bar{\widehat{\boldsymbol{\Psi}}}$	$\boldsymbol{\Psi}_o$	0.8192	0.1580	17.0449	0.0870	0.0940	0.1165	0.1050	
	$\boldsymbol{\Psi}_{or+}$	0.8484	0.1473	17.3552	0.0879	0.1036	0.3319	0.1040	
	$\boldsymbol{\Psi}_{or-}$	-2.8	3.4	13.1	-16.7	-8.4	50.8	5.6	[**14.4**]
$\widehat{\text{Bias}}(\widehat{\boldsymbol{\Psi}})$	$\boldsymbol{\Psi}_o$	-3.6	5.3	0.3	-13.0	-6.0	16.5	5.0	[**7.1**]
p.c.	$\boldsymbol{\Psi}_{or+}$	-0.2	-1.8	2.1	-12.1	3.6	231.9	4.0	[**36.5**]
	$\boldsymbol{\Psi}_{or-}$	6.1	4.5	15.1	48.8	17.5	46.8	19.0	[**22.5**]
CV	$\boldsymbol{\Psi}_o$	7.5	22.5	16.3	45.8	17.9	29.0	19.1	[**22.6**]
	$\boldsymbol{\Psi}_{or+}$	6.8	5.1	15.5	53.2	18.4	14.4	19.3	[**19.0**]
		d_1^*	d_2^*	d_3^*	d_4^*	d_5^*		$\bar{\varphi}_A$	
$\boldsymbol{D}^*_{true}$		140.00	90.00	90.00	100.00	90.00		7.1561	
	$\boldsymbol{\Psi}_{or-}$	40.00	20.00	30.00	20.00	20.00		30.6532	
$\boldsymbol{D}^*$	$\boldsymbol{\Psi}_o$	150.00	70.00	80.00	70.00	80.00		9.0129	
First cohort	$\boldsymbol{\Psi}_{or+}$	200.00	120.00	120.00	110.00	120.00		11.0700	
	$\boldsymbol{\Psi}_{or-}$	158.63	101.26	105.80	105.19	105.80		9.0313	
$\boldsymbol{D}^*$	$\boldsymbol{\Psi}_o$	145.58	92.79	97.02	96.67	97.02		8.5577	
Last cohort	$\boldsymbol{\Psi}_{or+}$	158.76	97.99	102.73	102.11	102.73		8.4055	
	$\boldsymbol{\Psi}_{or-}$	170.00	100.00	110.00	110.00	110.00		8.3987	
$\boldsymbol{D}^*(\widehat{\boldsymbol{\Psi}})$	$\boldsymbol{\Psi}_o$	150.00	90.00	100.00	100.00	100.00		7.2585	
Recommended	$\boldsymbol{\Psi}_{or+}$	150.00	90.00	90.00	100.00	90.00		7.1708	
		T_1^*	T_2^*	T_3^*					
	$\boldsymbol{\Psi}_{or-}$	.10	6.94	48					
ξ_C^*	$\boldsymbol{\Psi}_o$	.10	6.75	48					
	$\boldsymbol{\Psi}_{or+}$	.10	7.07	48					
			$\bar{C}_2$		$\bar{C}_3$				
	$\boldsymbol{\Psi}_{or-}$		5.13		8.04				
ACN	$\boldsymbol{\Psi}_o$		3.81		7.52				
	$\boldsymbol{\Psi}_{or+}$		6.27		8.43				

Table 6.17: Comparison of the data related to the three vectors of initial values: $\boldsymbol{\Psi}_{or-}$, $\boldsymbol{\Psi}_o$ and $\boldsymbol{\Psi}_{or+}$.

As far as bias is concerned, the parameter ω_3 is most adversely affected. Except ω_3, there is no large effect on quality of the parameters' estimates.

As can be seen in Table 6.17, $\bar{\varphi}_A$ values for the last cohort are not very large for $\boldsymbol{\Psi}_{or-}$ and $\boldsymbol{\Psi}_{or+}$ as compared to when $\boldsymbol{\Psi}_o$ is used as the vector of initial values. In this case, therefore, the effect on the terminal cohorts' subjects is less adverse as compared

to when $\mathbf{\Psi}_{o-}$ and $\mathbf{\Psi}_{o+}$ were used as the initial values (Table 6.14). Furthermore, the average recommended dose regimens corresponding to $\mathbf{\Psi}_{or-}$ and $\mathbf{\Psi}_{or+}$ are also closer to the true optimal dose regimen, $\boldsymbol{D}^*_{true}$. This shows that towards the end of the trial, the procedure is able to approximately determine the optimal dose regimen for the initial values $\mathbf{\Psi}_{or-}$ and $\mathbf{\Psi}_{or+}$.

Figure 6.27 presents the distribution of φ_A over the course of $C = 10$ cohorts for the three vectors of initial values.

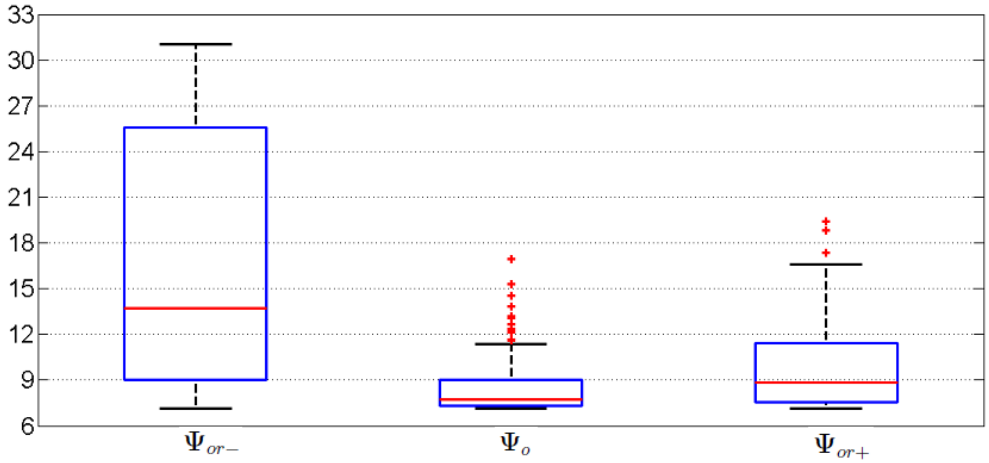


Figure 6.27: Distribution of φ_A over the course of the adaptive trial for the three vectors of initial values: $\mathbf{\Psi}_{or-}$, $\mathbf{\Psi}_o$ and $\mathbf{\Psi}_{or+}$.

The range of the distribution for $\mathbf{\Psi}_{or+}$ is larger than that for $\mathbf{\Psi}_o$, however, for $\mathbf{\Psi}_{or-}$ the spread is extraordinarily large. To explore this further, we compare the modal dose regimens administered to each of the ten cohorts for the three vectors of initial values. These data are presented in Table 6.18.

	$\mathbf{\Psi}_{or-}$					$\mathbf{\Psi}_o$					$\mathbf{\Psi}_{or+}$				
Cohort	d_1	d_2	d_3	d_4	d_5	d_1	d_2	d_3	d_4	d_5	d_1	d_2	d_3	d_4	d_5
1	40	20	30	20	20	150	70	80	70	80	200	120	120	110	120
2	140	90	90	90	90	130	90	90	90	90	150	90	90	100	90
3	**200**	**200**	**200**	**200**	**200**	130	80	90	90	90	120	80	80	80	80
4	200	200	200	190	200	130	80	90	90	90	110	70	80	70	80
5	200	160	160	160	160	130	80	90	90	90	110	70	80	70	80
6	200	130	140	140	140	130	90	90	90	90	120	80	80	80	80
7	200	120	130	130	130	130	90	90	90	90	120	80	80	80	80
8	200	120	120	120	120	130	90	90	90	90	140	90	90	90	90
9	170	110	110	110	110	140	90	90	90	90	150	90	100	100	100
10	150	100	100	100	100	130	90	100	90	100	150	100	100	100	100

Table 6.18: Modal dose regimens at each of the ten cohorts for the three sets of the initial values.

As mentioned previously, a small value of R_p is associated with over-exposure to the drug. As a result, to counterbalance this effect, the dose regimen computed by the ED algorithm for the first cohort consists of small dose sizes for the initial values

contained in $\boldsymbol{\Psi}_{or-}$. By the same argument, since value of R_p is large for the initial values contained in $\boldsymbol{\Psi}_{or+}$, the dose regimen computed by the ED algorithm for the first cohort consists of large doses to offset the potential under-exposure.

To determine the average under- and over-exposure experienced by the subjects in each of the $C = 10$ cohorts, Figure 6.28 presents the average values of φ_A for each of the ten cohorts corresponding to the three vectors of initial values.

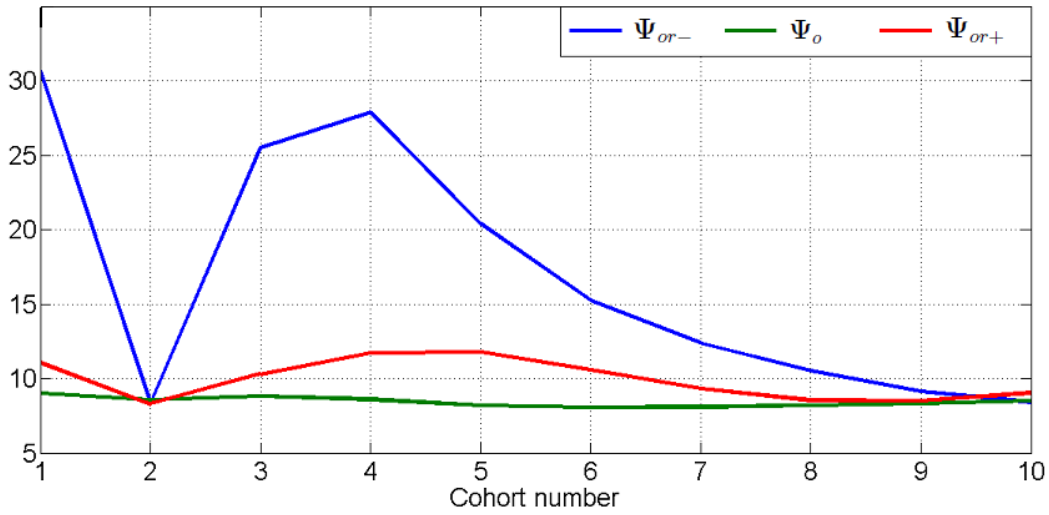


Figure 6.28: Average values of φ_A over the course of the trial for the three vectors of initial values of the parameters: $\boldsymbol{\Psi}_{or-}$, $\boldsymbol{\Psi}_o$ and $\boldsymbol{\Psi}_{or+}$.

From the figure, it can be observed that the average φ_A value for the first cohort for the set of initial values $\boldsymbol{\Psi}_{or-}$ is quite large. Also, there is large variation in the average φ_A value over the course of the trial for this vector of initial values. This explains the large spread in the distribution of φ_A observed for $\boldsymbol{\Psi}_{or-}$ in Figure 6.27.

The ACN values for the vectors $\boldsymbol{\Psi}_{or-}$ and $\boldsymbol{\Psi}_{or+}$ are larger as compared to those for the vector $\boldsymbol{\Psi}_o$. This was also observed previously when $\boldsymbol{\Psi}_{o-}$ and $\boldsymbol{\Psi}_{o+}$ were used as initial values. Imprecise initial values, therefore, delay the application of the two stopping rules thereby diminishing the chance of an early termination of the trial. In any case, as observed from Tables 6.15 and 6.18, the administered dose regimens get closer to the optimal dose regimen $\boldsymbol{D}^*_{true}$ towards the last cohort of the trial. Therefore, if the available information about the initial values is not reliable, early stoppage of the trial should be avoided and all the cohorts should be evaluated.

In conclusion, some of the parameters' estimates are significantly affected when the initial values are strongly deviated from the true values. Estimates of the variance parameters seem to be affected more than that of the PK parameters. However, since the ED algorithm is a function of only the PK parameters and not the variance parameters, the procedure is able to approximately determine the optimal dose regimen towards the end of the trial. Therefore, if reliable information about the initial values

is not available, it is advisable that all the available cohorts are evaluated and the trial is not terminated early.

The observations made in this section underpin the advantage of carrying out this procedure in an adaptive setting rather than having a single cohort. If there is a strong deviation from the initial values, the subjects in the first cohort may be exposed to highly unethical dose regimens. However, the accumulating data and the interim analyses provide an opportunity of course-correction following starting on initial values that have a large deviation from the true values. This point is expounded in Section 6.4 where we make a comparison between the adaptive and non-adaptive approach to achieving the objectives of PK estimation and dose regimen optimisation. Furthermore, if there is little information about the possible values of the parameters, it is prudent to evaluate all the available cohorts and not to terminate the trial earlier.

6.3.5 Misspecification of the PK Model

In the original simulation study, we considered the one-compartment PK model and assumed that it was the true underlying model. However, in reality, the assumed model may not be correct and the true model could be different. It is pertinent to study the performance of the methodology in such situations, that is when the assumed and the true underlying models can, possibly, differ.

Let us consider an example in which the assumed model for conducting the adaptive trial is the one-compartment model, as described in Equation (6.1) but the true, underlying model is the two-compartment model, as described in Equation (2.7). That is, the assumed model is given as

$$C(t) = \frac{dK_a}{V(K_a - K_e)}(e^{-K_e t} - e^{-K_a t}),$$

whereas the true model is

$$C(t) = Ae^{-\lambda t} + Be^{-\mu t} - (A + B)e^{-K_{at} t},$$

where A, B, λ and μ are functions of the PK parameters CL, V_1, K_{at}, Q and V_2, as shown in Equation (2.7). To distinguish from the absorption rate constant of the assumed one-compartment model, we use K_{at} for the true model.
Let the true values of these parameters be $\boldsymbol{\beta}_t' = (CL, V_1, K_{at}, Q, V_2)^T = (2, 20, 1, 1, 10)^T$.

For comparing the concentration profiles generated by the respective population parameters of the assumed and the true models, $\boldsymbol{\beta}_t$ and $\boldsymbol{\beta}_t'$, Figure 6.29 plots them for the dose regimen $\boldsymbol{D}^* = (140, 90, 90, 100, 90)^T$ previously determined to be optimal

for the assumed model with parameters $\boldsymbol{\beta}_t = (.85, .15, 17)^T$.

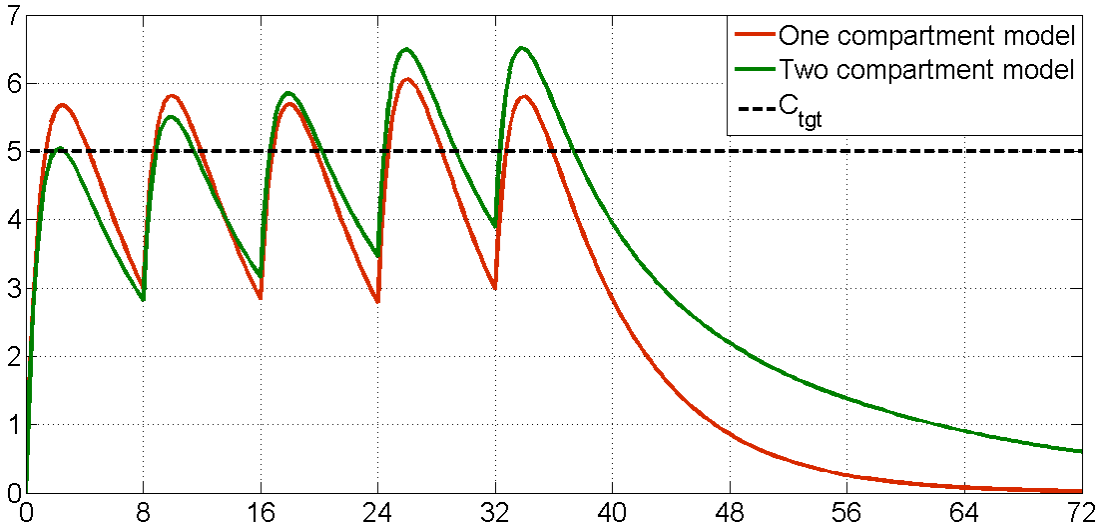


Figure 6.29: Comparison of the assumed model (in red) and the correct model (in green). It appears that the dose regimen $\boldsymbol{D}^*$, optimised for the assumed model, is suboptimal for the true model.

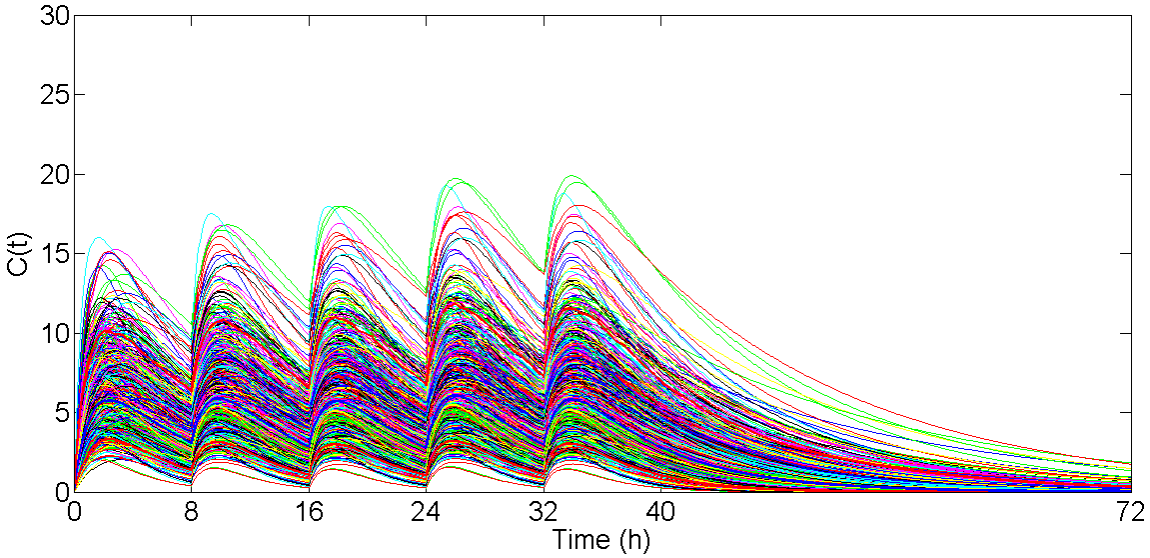
From the figure, it seems that $\boldsymbol{D}^*$ is not optimal for the true (two-compartment) model, which is not unexpected as $\boldsymbol{D}^*$ has been optimised by the ED algorithm for the assumed model. Later in this section, we will compute $\boldsymbol{D}'^*$, the φ_A-efficient dose regimen for the true model.

For the assumed model, the random effects for the parameters are as defined in Equation (6.2). For the two-compartment model (true model), let the random effects be given as:

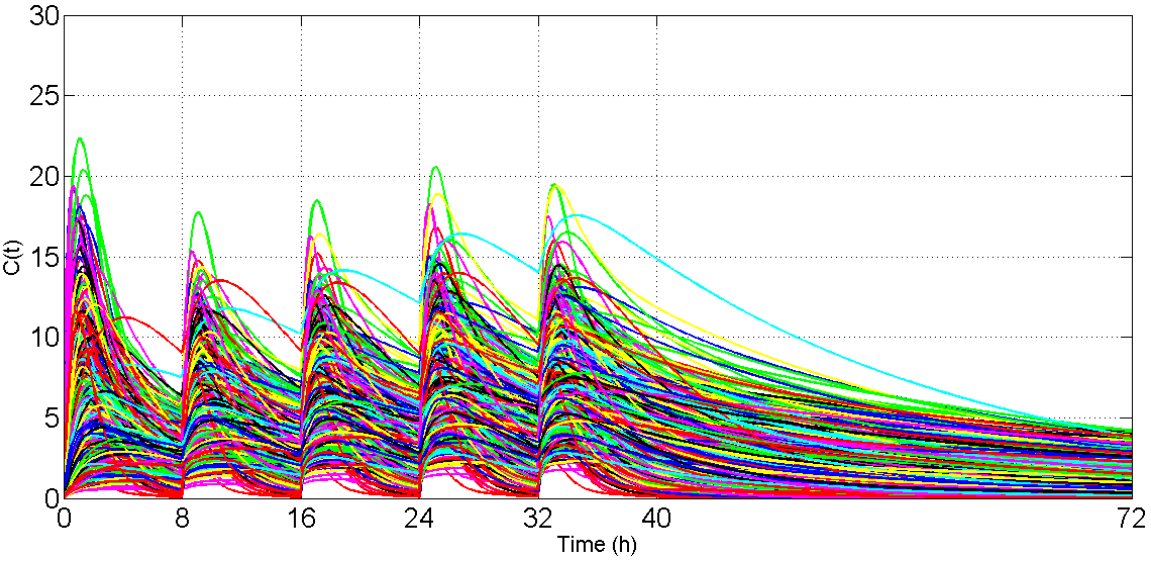
$$\begin{pmatrix} CL_i \\ V_{1_i} \\ K_{at_i} \\ Q_i \\ V_{2_i} \end{pmatrix} = \begin{pmatrix} CL\exp(b_{1i}) \\ V_1\exp(b_{2i}) \\ K_{at}\exp(b_{3i}) \\ Q\exp(b_{4i}) \\ V_2\exp(b_{5i}) \end{pmatrix}, \tag{6.7}$$

where $\boldsymbol{b}_i = (b_{1i}, b_{2i}, b_{3i}, b_{4i}, b_{5i})^T$ is a vector of random effects such that $\boldsymbol{b}_i \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Omega}')$. As before, we assume $\boldsymbol{\Omega}'$ is a diagonal matrix. Let $\boldsymbol{\omega}' = (\omega'_1, \omega'_2, \omega'_3, \omega'_4, \omega'_5)^T$ represent the five diagonal elements of this matrix. For the purpose of simulation, we choose the following values for the variance parameters: $\boldsymbol{\omega}' = (.2, .5, .2, .5, 1)^T$. The error structure for this model is assumed to be exponential, i.e., the same as it was chosen for the assumed (one-compartment) model. The value of the error variance is also taken to be the same as before, i.e., $\sigma^2 = .10$. Figure 6.30b presents 1000 simulated concentration profiles from the two-compartment model and compares them with profiles from the assumed one-compartment model in Figure 6.30a.

As discussed in Section 2.1.3, the two-compartment model displays biphasic behaviour. That is, the decline in concentration in the central department after reaching



(a) One-compartment model (assumed)



(b) Two-compartment model (true)

Figure 6.30: 1000 concentration profiles simulated from the assumed and the true PK models. The PK parameters of the assumed model are $\boldsymbol{\beta}_t = (K_a, K_e, V)^T = (.85, .15, 17)^T$ and the parameters of the true model are $\boldsymbol{\beta}'_t = (CL, V_1, K_{at}, Q, V_2)^T = (2, 20, 1, 1, 10)^T$.

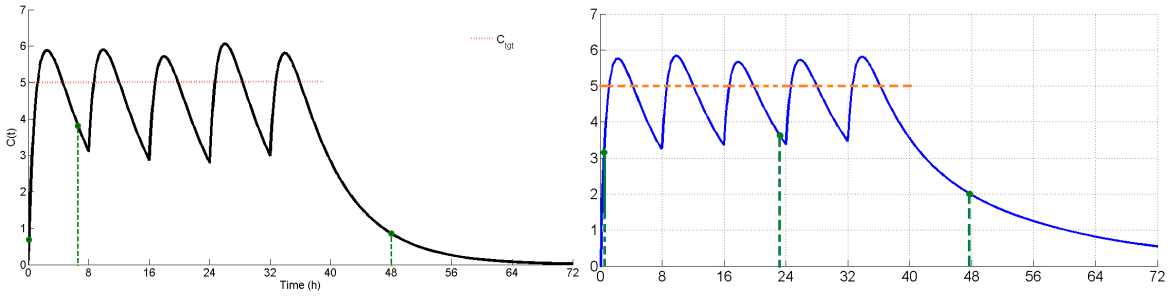
C_{max} is initially rapid but then reduces to zero gradually. This generally leads to a longer retention or accumulation of the drug in the body. This behaviour can be observed in Figure 6.30b. On the other hand, concentration profiles in Figure 6.30a decrease gradually after achieving their respective C_{max} values after a dose is administered.

As before, the goals of our adaptive procedure are to estimate the PK parameters of the underlying model and to maintain the concentration profile of the average subject in the cohort around some target value. Here we choose $C_{tgt} = 5$ mg/L to be

maintained for $T = 40$ h. However, as in this case the underlying PK model is the two-compartment model, the estimated parameters of the assumed one-compartment model will not correspond to the true ones.

Furthermore, the dose regimen optimisation and the computation of the D-optimal sampling points will be based on the assumed one-compartment model instead of the true model. It will be of interest to see the deviation of the optimal dose regimen and the D-optimal sampling points based on the assumed model (which are used in practice in the trial) from the corresponding values based on the true values (which should have been used in the trial). Based on the two-compartmental model, the D-optimal sampling times are $\xi^{*'} = \{0.53, 23.23, 47.76\}$. In comparison, the D-optimal time points for the assumed one-compartment model are $\xi^* = \{0.10, 6.60, 48.00\}$.

It may be recalled here that $\boldsymbol{D}^* = (140, 90, 90, 100, 90)^T$ was determined to be φ_A-efficient for the assumed (one-compartment) model in the original study with $\varphi_A = 7.1561$. As can be seen in Figure 6.29, $\boldsymbol{D}^*$ does not seem optimal for the true (two-compartment) model. The φ_A-efficient dose regimen for the two compartmental model at resolution $\kappa = 10$ was computed by the ED algorithm to be $\boldsymbol{D}'^* = (160, 90, 80, 80, 80)^T$ with $\varphi'_A = 5.8352$. Figure 6.31 compares the concentration profiles of the two models, based on their fixed parameters and the respective optimal dose regimens. The D-optimal sampling time points are also shown.



(a) Assumed: one-compartment model (b) True: two-compartment model

Figure 6.31: D-optimal sampling times ξ^* and ξ'^* and the concentration profiles generated from the respective optimal dose regimens, $\boldsymbol{D}^*$ and $\boldsymbol{D}'^*$, for the assumed and the true models.

It can be observed from the figure that the areas of under-exposure for the two-compartment model are smaller than the corresponding areas for the one-compartment model which is on the account of more gradual elimination of the two-compartment model. This explains why φ'_A is smaller than φ_A.

In Section 6.2.3, we discussed the distributions of the simulated parameters and response from the one-compartment model. We now present these distributions for the two-compartment model, taken as the true model in this study.

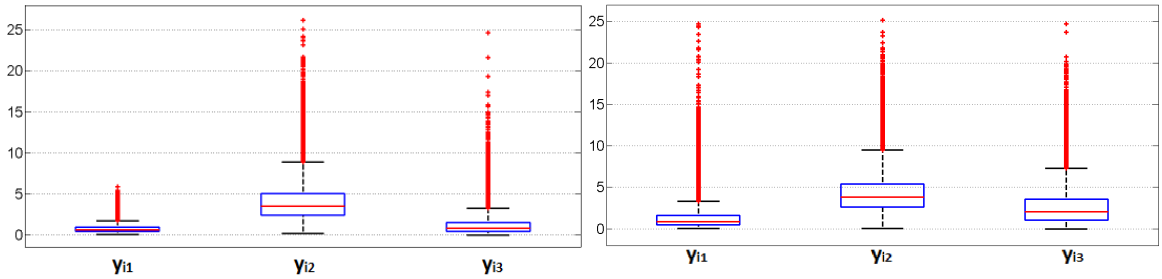
Each simulation study consists of $N_{sim} = 1000$ simulations and each simulation consists of $C = 10$ cohorts, each of size $c = 10$ patients. Therefore, in one simulation study, we simulate 100,000 vectors of $\boldsymbol{\beta}'$.

As shown in Section 5.2, the distribution of the PK parameters for the model defined in Equation (6.7) is lognormal. That is, $CL_i \sim \ln\mathcal{N}(\ln(2), .2)$, $V_{1_i} \sim \ln\mathcal{N}(\ln(20), .5)$, $K_{at_i} \sim \ln\mathcal{N}(\ln(1), .2)$, $Q_i \sim \ln\mathcal{N}(\ln(1), .5)$ and $V_{2_i} \sim \ln\mathcal{N}(\ln(10), 1)$. Table 6.19 compares the statistics derived from the simulated distributions of the five parameters with their theoretical values. They are observed to be in good agreement.

Statistic	Distribution	CL	V_1	K_{at}	Q_i	V_2
Mean	Theoretical	2.2103	25.6805	1.1052	1.2840	16.4872
	Simulated	2.2047	25.7813	1.1008	1.2795	16.5340
Median	Theoretical	2.0000	20.0000	1.0000	1.0000	10.0000
	Simulated	1.9838	19.8110	0.9970	0.9921	9.9193
IQR	Theoretical	1.2249	19.8090	0.6125	0.9904	14.5361
	Simulated	1.2211	20.2069	0.5955	0.9900	14.6514

Table 6.19: Comparison of statistics related to the simulated distribution of the five PK parameters with the values obtained from their theoretical distributions.

To compare the responses from the one-compartment model (assumed) and the two-compartment model (true), we simulate responses from both the models at the time points $\xi^* = \{.10, 6.60, 48\}$ and for dose regimen $\boldsymbol{D}^*$. For the purpose of comparison, the time points and the dose regimen have to be the same for both the models. Figure 6.32 presents the distributions of the response at the three sampling time points and Table 6.20 presents the statistics related to these distributions.



(a) Assumed: one-compartment model (b) True: two-compartment model

Figure 6.32: Distribution of the simulated responses for the assumed and the true models.

From the table and the figure, it can be seen that the inter-individual variability assumed in the two-compartment model is larger than that assumed in the one-compartment model. This is clear from the higher IQR values and a large number

	Assumed Model			True Model		
Statistic	y_{i1}	y_{i2}	y_{i3}	y_{i1}	y_{i2}	y_{i3}
Mean	0.7697	4.0412	1.1962	1.3165	4.3092	2.6002
Median	0.6601	3.5484	0.8289	0.8976	3.8113	2.0938
IQR	0.5069	2.5786	1.1277	1.1244	2.7464	2.5060

Table 6.20: Distribution of the simulated responses from the assumed and the true models.

of outliers in the case of the two-compartment model. The mean (or the median) simulated response at the third sampling point, i.e., y_{i3} for the the two-compartment model is more than twice of the corresponding value for the one-compartment model. As explained before, this is on account of the property of the two-compartment model which postulates slower elimination or longer retention of the drug in the body. The median response at the first time point, y_{i1}, for the two models is similar but the mean response and the IQR value for the two-compartment model are almost twice of the corresponding values for the one-compartment model. This is on account of the larger variability assumed in the two-compartment model which results in a large number of outliers at this point.

Let us now study the effect on the methodology's performance when a wrong model is assumed to be true. The original simulation study described in Section 6.2 is now run with the change that the responses are simulated from a two-compartment model but computation of the optimal dose regimen and D-optimal sampling points are based on the one-compartment model, which is assumed to be true.

Table 6.21 compares the following two scenarios: first, in which the assumed and the true models are same, i.e., the one-compartment model (original study). Second, the assumed model is a one-compartment model and the true model is a two-compartment model, as defined above.

The estimated parameters in the second scenario have a large deviation from the true parameters, $\mathbf{\Psi}_{true}$, of the one-compartment model (given in Equation (6.4)) since the underlying model is altogether different. Also, the variability in the estimated parameters is larger in the second scenario.

For the two-compartment model, the optimal D-sampling time points were determined to be $\xi^{'*} = \{0.53, 23.23, 47.76\}$, as mentioned before. For the scenario, where a wrong model is assumed to be true, the average sampling points for the last cohort were found to be $\{0.10, 3.52, 47.18\}$. Table 6.22 presents the average D-optimal sampling time points for the second scenario.

	True model	K_a	K_e	V	ω_1	ω_2	ω_3	σ^2	
$\bar{\hat{\boldsymbol{\Psi}}}$	One-comp.	0.8192	0.1580	17.0449	0.0870	0.0940	0.1165	0.1050	
	Two-comp.	1.1502	0.0952	21.0931	0.5903	0.1439	0.0469	0.0954	
CV	One-comp.	7.5	22.5	16.3	45.8	17.9	29.0	19.1	[**22.6**]
	Two-comp.	16.3	12.4	23.5	28.3	30.9	115.8	36.0	[**37.6**]
		d_1^*	d_2^*	d_3^*	d_4^*	d_5^*			
$\boldsymbol{D}^*$	One-comp.	145.58	92.79	97.02	96.67	97.02		$\bar{\varphi}_A = 8.5577$	
	Two-comp.	152.01	72.58	76.31	76.10	76.33		$\bar{\varphi}'_A = 8.6600$	
		T_1^*	T_2^*	T_3^*					
ξ_C^*	One-comp.	.10	6.75	48.00					
	Two-comp.	.10	3.52	47.18					
			$\bar{C}_2$		$\bar{C}_3$				
ACN	One-comp.		3.81		7.52				
	Two-comp.		4.04		8.71				

Table 6.21: Comparison of statistics related to the simulation studies for the two underlying models. The assumed model in both scenarios is the one-compartment model.

Cohort	T_1	T_2	T_3
1	0.10	5.42	48.00
2	0.35	6.47	45.08
3	0.13	5.10	46.45
4	0.11	4.67	46.97
5	0.10	4.44	46.97
6	0.10	3.89	46.99
7	0.10	3.83	47.15
8	0.10	3.60	47.27
9	0.10	3.66	47.33
10	0.10	3.52	47.18

Table 6.22: D-optimal sampling points for the $C = 10$ cohorts, averaged over the $N_{sim} = 1000$ simulations. The D-optimal sampling times corresponding to $\boldsymbol{\Psi}_{true}$ are $\xi^{'*} = \{0.53, 23.23, 47.76\}$.

The larger ACN values for the second scenario mean that assuming a wrong PK model could cause a delay in the termination of the trial when the stopping rules SR2 and SR3 are used instead of complete examination of C cohorts.

The average optimal dose regimen in the second scenario is computed as $(152.01, 72.58, 76.31, 76.10, 76.33)^T = \bar{\boldsymbol{D}}_{II}^*$ (say). Given the fact that this is computed on the fallacious assumption of the true model, this is quite close to the true optimal dose regimen for $\boldsymbol{D}^{'*} = (160, 90, 80, 80, 80)^T$. In this case, the adaptive procedure

shows a good degree of robustness being able to approximately determine the optimal dose regimen for the population.

To explore this further, we compare in Figure 6.33 the 'ideal' concentration profile generated by the true parameters of the two-compartment model, $\boldsymbol{\beta}'_t$, and the dose regimen $\boldsymbol{D}'^*$ with the 'in-practice' average concentration profile generated by the one-compartment model with dose regimen $\bar{\boldsymbol{D}}^*_{II}$ and parameters $\bar{\hat{\boldsymbol{\beta}}}_{II}$, where $\bar{\hat{\boldsymbol{\beta}}}_{II}$ represents the average estimated parameters from the last cohort in the second scenario. From Table 6.21, $\bar{\hat{\boldsymbol{\beta}}}_{II} = (1.1502, .0952, 21.0931)^T$.

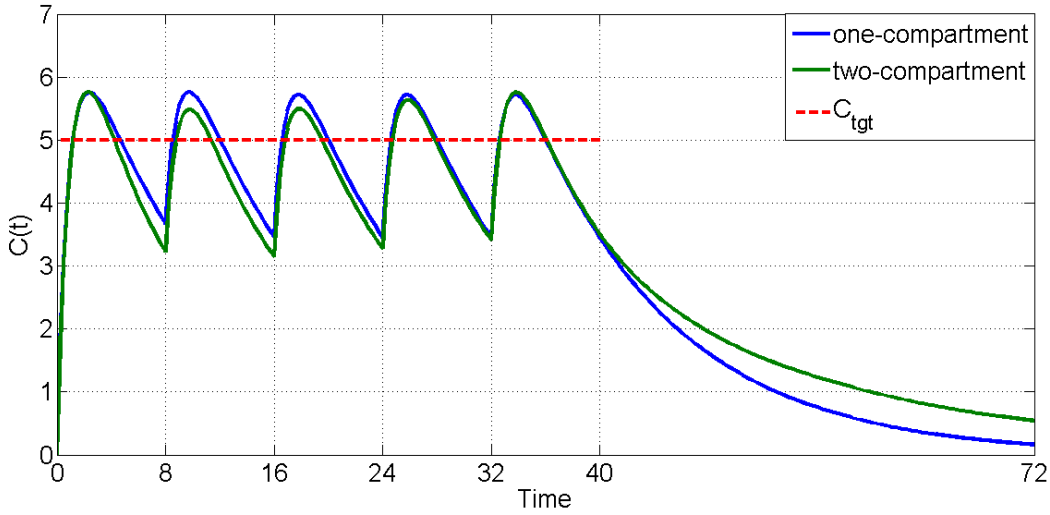


Figure 6.33: Comparison of the average concentration profiles: the ideal (in green) and the one actually achieved (in blue).

It can be seen from this figure that the two concentration profiles are quite close to each other, especially for the duration of 40 h for which it is desired to maintain C_{tgt}. This shows that even if the assumed model is different from the underlying model, the estimated parameters of the assumed model 'adapt' to the response in such a way that the concentration profile generated by them resembles the true concentration profile. Comparing Figures 6.29 and 6.33, it can be concluded that despite assuming a wrong PK model, the adaptive procedure is able to approximately determine the optimal dose regimen on the strength of the subjects' response data.

However, this inference is based only on the average concentration profiles. The robustness of the adaptive procedure can be established when the variability in the estimated parameters and its effect on the variability of the performance are also taken into account. For this, we use the measure defined in Section 4.4.1 where we introduced the notion of relative efficiency of a dose regimen which is actually administered to the dose regimen which should have been administered.

Specifically, for this problem, we use

$$\varphi_{ref} = \frac{\varphi_t(\boldsymbol{D}'^*)}{\varphi_t(\boldsymbol{D}^*_{II})} \times 100,$$

where the function φ_t measures the average over- and under-exposure of a dose regimen when the underlying model is the two compartment model. φ_{ref} represents the relative efficiency of $\boldsymbol{D}^*_{II}$, the dose regimen which is administered to the dose regimen which should have been administered, i.e., $\boldsymbol{D}'^*$. The range of φ_{ref} is $[0, 100]$. A value of φ_{ref} close to 100 signifies that the administered dose regimen was not very inferior to the optimal dose regimen. As mentioned before, $\varphi_t(\boldsymbol{D}'^*) = 5.8352$.

Figure 6.34 presents the distribution of φ_{ref} of the recommended dose regimens in the 1000 simulations.

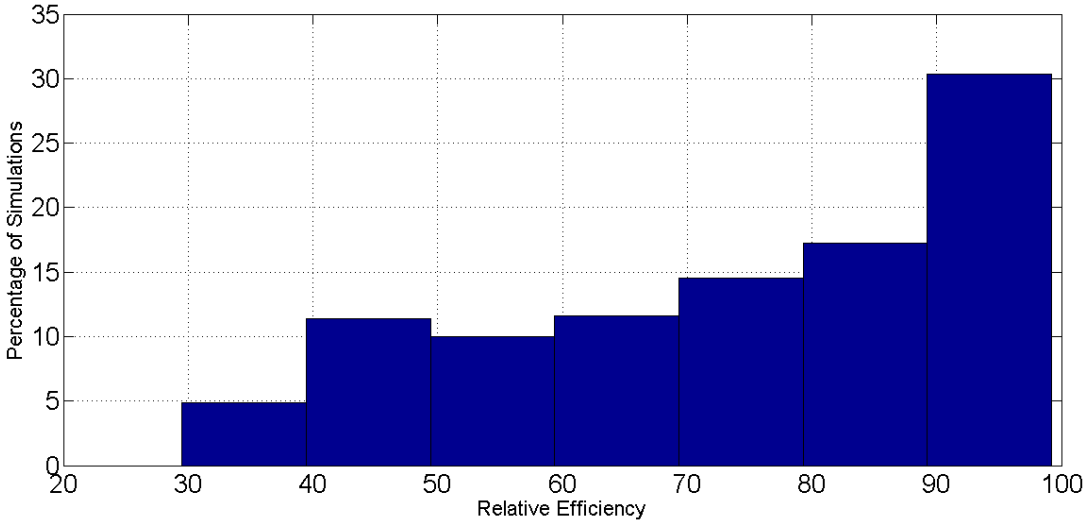


Figure 6.34: Distribution of φ_{ref} for the 1000 simulations.

From the figure, it can be seen that about 47% of the recommended dose regimens are at least 80% as efficient as the ideal dose regimen. About 16% of the recommended dose regimens are at most 50% as efficient. The remaining 37% of the recommended dose regimens are of intermediate efficiency - between 50% and 80%. The mean efficiency, $\overline{\varphi}_{ref}$ was computed as 73.9%, median efficiency, $\tilde{\varphi}_{ref}$, as 79% and the IQR as 35.2%. Given that the assumed and the true models in this study were two different compartmental models, these data show that methodology performed reasonably well in estimating the concentration-time relationship which resulted in the recommended dose regimens to be of a high degree of efficiency, despite being based on a wrong PK model.

In conclusion, the adaptive procedure described in this chapter is reasonably robust against misspecification of the underlying PK model. In the next section, we explore how assuming a wrong error structure affects the methodology's performance.

6.3.6 Deviation from the Assumed Error Structure

In the original simulation study, the model assumed in Equation (6.3) consisted of an exponential error structure. Since the concentration values are always non-negative, the exponential error structure is suitable for such data. We defined two other error structures in Section 5.2, additive and proportional. For these error structures, the response can take negative values as well.

In this section, we explore the performance of the methodology when an incorrect error structure is assumed to be the true one. Suppose the true underlying model is the proportional error model, but the exponential error model is believed to be true. That is, the assumed model is as given in Equation (6.3) but the true model is

$$y_{ij} = \sum_{k=1}^{n} I_{\{T_j \geq t_k\}} \frac{d_k K_a e^{b_{1i}}}{V e^{b_{3i}} (K_a e^{b_{1i}} - K_e e^{b_{2i}})} \left(e^{-K_e e^{b_{2i}} (T_j - t_k)} - e^{-K_a e^{b_{1i}} (T_j - t_k)} \right) (1 + \epsilon'_{ij}), \tag{6.8}$$

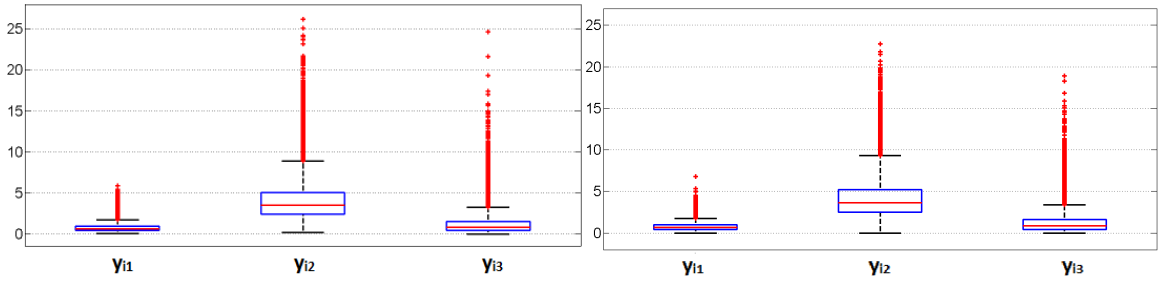
where y_{ij} is the j^{th} sample from the i^{th} subject at time T_j, where $i = 1, ..., c$ and $j = 1, ..., m$. The random errors, ϵ'_{ij}, are normally distributed, that is $\epsilon'_{ij} \sim \mathcal{N}(0, \sigma'^2)$ and they are assumed to be independent of the elements in $\boldsymbol{b}_i$'s.

The problem is to see the effect of this wrong assumption on the performance of the methodology. To examine in a simulation study the effect of assuming a wrong error structure, it is important to keep similar the magnitude of variability in the response induced by the assumed and the true error models. This is to avoid confounding the effects of a different error structure with the effects of difference in variability. We separately explore how the magnitude of the error variance affects the performance of the methodology in Section 6.3.7. The effect of assuming a different magnitude of the variability, i.e., taking initial values of the error variance significantly different from the true value, was discussed in Sections 6.3.3 and 6.3.4. It was observed that, in general, if the difference between the initial value and the true value of the variance parameter is not very large, there is no significant effect on the performance of the methodology. However, significant deviation may result in large bias in some parameters.

For the assumed model, that is the exponential error model given in Equation (6.3), the vector of initial values of the parameters $\boldsymbol{\Psi} = (K_a, K_e, V, \omega_1, \omega_2, \omega_3)^T$ was chosen to be $\boldsymbol{\Psi}_o = (1, .1, 20, .05, .15, .05, .15)^T$. The vector of parameter values for the true model, that is the proportional error model given in Equation (6.8), is chosen to be $\boldsymbol{\Psi}_{true} = (.85, .15, 17, .1, .1, .1, .05)^T$, i.e., $\sigma'^2 = .05$.

For comparing the variability generated by the assumed and the true models, Figure 6.35 presents the distributions of simulated response at three points {.10, 6.60, 48.00}. Table 6.23 presents the data related to these responses. As explained before,

the variability induced by the two error models has been deliberately kept similar. For the chosen value of σ'^2, no negative value of the response were observed in the $N_{sim} = 1000$ simulations.



(a) Assumed error structure (Exponential) (b) True error structure (Proportional)

Figure 6.35: Distribution of the simulated responses for the two error structures.

	Assumed	Error	Model	True	Error	Model
Statistic	y_{i1}	y_{i2}	y_{i3}	y_{i1}	y_{i2}	y_{i3}
Mean	0.7697	4.0412	1.1962	0.7922	4.1419	1.2502
Median	0.6601	3.5484	0.8289	0.6833	3.6732	0.8554
IQR	0.5069	2.5786	1.1277	0.5599	2.8723	1.2318

Table 6.23: Mean, Median and IQR of the 1000 simulated responses from the assumed and the true error models.

Now, assuming the exponential model to be true when, in fact, the true model is the proportional model, a simulation study with N_{sim} simulations was run. The data from the study are presented in Table 6.24. Data related to the original simulation study, when the assumed and the true error models are the same, i.e., the exponential error model, have been re-presented to compare with the effect of the erroneous assumption.

From the table, it can be seen that barring the optimal sampling time points ξ_C^*, there is no significant effect on other aspects of the adaptive methodology. The explanation for this is as follows. Since the variability induced in the response by the two models is similar and the two models differ only with respect to their error structure, excepting the estimate of the error variance, other parameters' estimates do not differ significantly in the two situations. The error variance for the exponential model (assumed error model) is $\sigma^2 = .10$ whereas it is $\sigma'^2 = .05$ for the proportional model (true error model). The estimate of σ'^2 is found to be .0624 which is reasonably close to the true value of .05, given that a different error model was assumed to be true.

Regarding the optimal dose regimens, it may be recalled here that the ED algorithm uses only the PK parameters as inputs and does not involve any of the variance

	True model	K_a	K_e	V	ω_1	ω_2	ω_3	σ^2	
$\bar{\hat{\boldsymbol{\Psi}}}$	Exponential	0.8192	0.1580	17.0449	0.0870	0.0940	0.1165	0.1050	
	Proportional	0.8295	0.1541	17.5501	0.0857	0.0940	0.1154	0.0624	
$\widehat{\text{Bias}}(\hat{\boldsymbol{\Psi}})$	Exponential	-3.6	5.3	.3	-13.0	-6.0	16.5	5.0	[**7.1**]
p.c.	Proportional	-2.4	2.7	3.2	-14.3	-6.0	15.4	24.7	[**9.8**]
CV	Exponential	7.5	22.5	16.3	45.8	17.9	29.0	19.1	[**22.6**]
	Proportional	5.3	3.7	13.7	39.4	16.4	26.5	23.4	[**18.3**]
		d_1^*	d_2^*	d_3^*	d_4^*	d_5^*		$\bar{\varphi}_A$	
$\boldsymbol{D}^*$	Exponential	145.58	92.79	97.02	96.67	97.02		8.5577	
	Proportional	149.08	94.25	98.80	98.00	98.80		8.5367	
		T_1^*	T_2^*	T_3^*					
ξ_C^*	Exponential	.10	6.75	48.00					
	Proportional	.10	7.38	48.00					
			$\bar{C}_2$		$\bar{C}_3$				
ACN	Exponential		3.81		7.52				
	Proportional		3.77		7.45				

Table 6.24: Comparison of statistics related to the simulation studies for the two underlying models. The assumed model in both scenarios is the exponential error model.

parameters in computation of the optimal dose regimen. As a result, the optimal dose regimens are similar as in the original case even though the assumed error model is different from the true error model. By the same argument, $\bar{\varphi}_A$ values and the ACN values are similar to each other in both cases. Figure 6.36 presents the distribution of φ_A for the two scenarios.

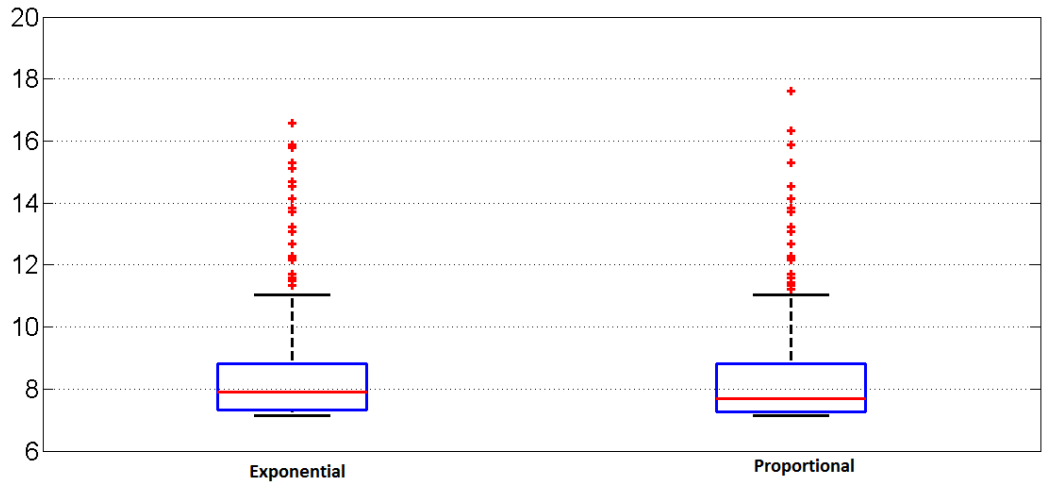


Figure 6.36: Distribution of φ_A for the two underlying error models. The assumed model in both cases is the exponential error model.

The distributions are similar which shows that the degree of over- and under-

exposure experienced by the subjects will not be very different even if an incorrect error model is used, provided that variability induced in the response is similar by both.

Next, let us consider the D-optimal sampling time points. Based on $\mathbf{\Psi}_{true}$, the D-optimal sampling time points for the proportional error model are $\xi^* = \{.10, 7.90, 48.00\}$. In comparison, the D-optimal points for the exponential error model in the original simulation study were $\xi^* = \{0.10, 6.60, 48\}$. In the original simulation study, ξ^*_C, the D-optimal dosing time points of the terminal, i.e., the C^{th} cohort were found to be $\xi^*_C = \{.10, 6.75, 48\}$. In the current study, when the underlying error model is proportional, $\xi^*_C = \{0.1, 7.38, 48\}$ which is reasonably close to $\xi^* = \{.10, 7.90, 48\}$.

For the current simulation study, Table 6.25 shows the D-optimal sampling time points for the different cohorts, averaged over the $N_{sim} = 1000$ simulations.

Cohort	T_1	T_2	T_3
1	0.10	5.42	48.00
2	0.85	8.76	47.04
3	0.11	7.62	47.99
4	0.10	7.54	47.99
5	0.10	7.43	47.99
6	0.10	7.47	47.99
7	0.10	7.36	47.99
8	0.10	7.35	48.00
9	0.10	7.34	48.00
10	0.10	7.38	48.00

Table 6.25: D-optimal sampling points for the $C = 10$ cohorts, averaged over the $N_{sim} = 1000$ simulations. The D-optimal sampling times corresponding to $\mathbf{\Psi}_{true}$ are $\xi^*_{true} = \{.10, 7.90, 48.00\}$.

Since the D-optimal sampling points for the first cohort are based on the initial values, $\mathbf{\Psi}_o = (1, .1, 20, .05, .15, .05, .15)^T$, these sampling points are the same as that for the first cohort when both - the assumed and the true models - are same, as can be observed in Table 6.4. However, from the second cohort onwards, the sampling time points, especially the second point, take different values. This difference is attributable to the fact that in the current study, exponential error model with error variance $\sigma^2 = .10$ is assumed whereas the true model is the proportional error model with error variance $\sigma^{'2} = .05$. Because of this difference, the estimates of σ^2 at each cohort in the current study are different from the estimates in the original study.

Therefore, the difference in the assumed and the true error structures can make the supposedly optimal sampling points suboptimal for collection of the blood samples. Furthermore, the bias in the estimate of the error variance can be significantly larger since the underlying error structure is different from the assumed error structure.

As the ED algorithm does not require the error variance as an input, the optimal dose regimens and the associated φ_A values are not directly affected by assuming an incorrect error model. Therefore, for similar levels of variability, the adaptive procedure is observed to be reasonably robust against deviation from an assumed error structure.

6.3.7 Effect of Error Variance on the Parameters' Estimates

In the original simulation study, the true value of the error variance was taken to be $\sigma^2_{true} = .10$ and the initial value was taken as $\sigma^2_o = .15$.

In this section, we study how the magnitude of the error variance influences the distribution of the estimated parameters. We assume five different true values of the error variance, much smaller than 0.10 and much larger than this value. That is, we assume : $\sigma^2_{true} = .02, .05, .25, .40$ and $.60$.

We consider two scenarios for this problem: (a) The initial values for the error variance are in the same proportion to σ^2_{true} as they were in the original study, that is, the corresponding initial values are $\sigma^2_{oa} = .03, .075, .38, .60$ and $.90$ and (b) the initial values have the same difference $(\sigma^2_o - \sigma^2_{true})$ as in the original study. That is, the corresponding initial values are $\sigma^2_{ob} = .07, .10, .30, .45$ and $.65$. The other parameters in $\mathbf{\Psi}_{true}$ and $\mathbf{\Psi}_o$ remain unchanged in both scenarios.

Before studying the effect of the error variance on the bias of the estimators, it will be useful to analyse the distributions of the simulated responses for different values of the error variance. In Section 6.2.3, we presented the distribution of the responses at the time points $\xi^* = \{.10, 6.60, 48\}$ for the case when σ^2_{true} was taken to be 0.1.

Figures 6.37, 6.38 and 6.39 present the distribution of the simulated responses for different values of the true error variance, σ^2_{true} at each of the three time points contained in ξ^*. The previously considered case of $\sigma^2_{true} = .10$ is added for comparison. Table 6.26 presents the statistics related to the simulated responses.

The mean response increases with increase in the error variance while the median response decreases, at the three time points. As the distribution of the response is approximately lognormal (as shown in Section 5.3.2) and since the skewness of lognormal distribution is a function of the error variance (discussed in Section 6.2.1), this is related to increase in the skewness of the distribution. The IQR and the CV of the distributions increase with an increase in the magnitude of the error variance, which is expected since a higher variance induces more variability in the response.

The range of variabilities induced by the different values of σ^2_{true} is quite large. For

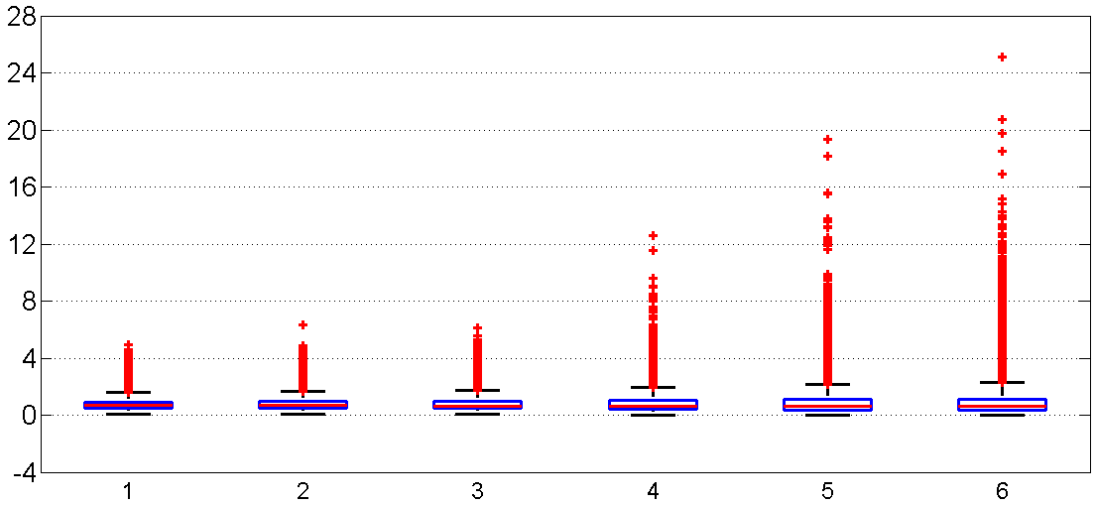


Figure 6.37: Box-plots of the simulated response at the first sampling time point, y_{i1}, for the six values of the true error variance, σ^2_{true}, labelled as 1, ..., 6 for σ^2_{true} = .02, .05, .10, .25, .40 and .60 respectively.

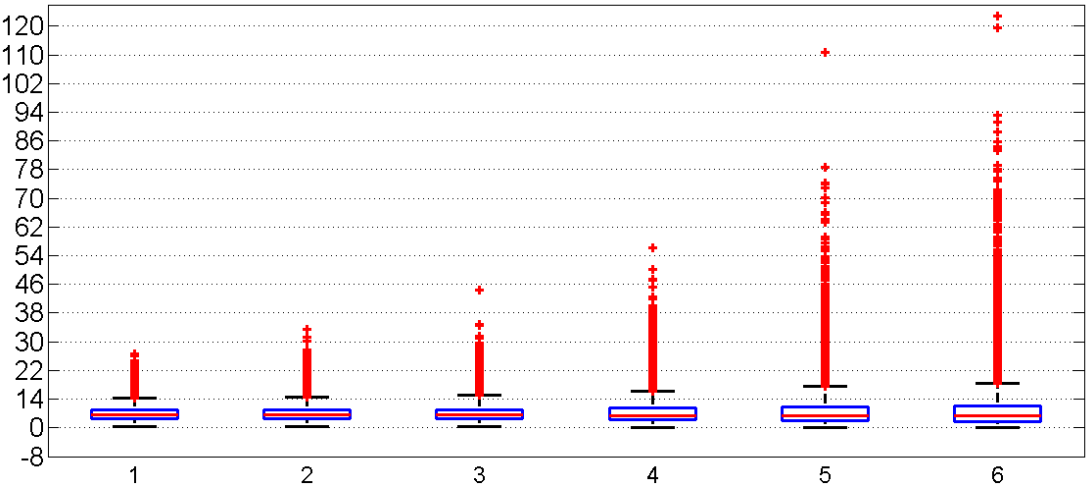


Figure 6.38: Box-plots of the simulated response at the second sampling time point, y_{i2}, for the six values of the true error variance, σ^2_{true}, labelled as 1, ..., 6 for σ^2_{true} = .02, .05, .10, .25, .40 and .60 respectively.

example, the maximum simulated response at y_{i2} for $\sigma^2_{true} = 0.60$ is nearly 5 times of the maximum response for $\sigma^2_{true} = .02$. Furthermore, the number of outliers (in this case, excessively large values) in the responses increase with increasing σ^2_{true}. The values of σ^2_{true} considered here, therefore, cover a large range of values of the response.

The bias and the CV of the estimated parameters for different values of σ^2_{true} and Scenarios (a) and (b) are presented in Tables 6.27 and 6.28. The data corresponding to $\sigma^2_{true} = 0.10$ have been inserted to facilitate comparisons.

From the tables, it can be seen that while bias for most parameters in general increases with an increase in the magnitude of the error variance, the effect is more pronounced for some parameters. The parameter most adversely affected by the magnitude of the error variance is the variance of random effect acting on the parameter

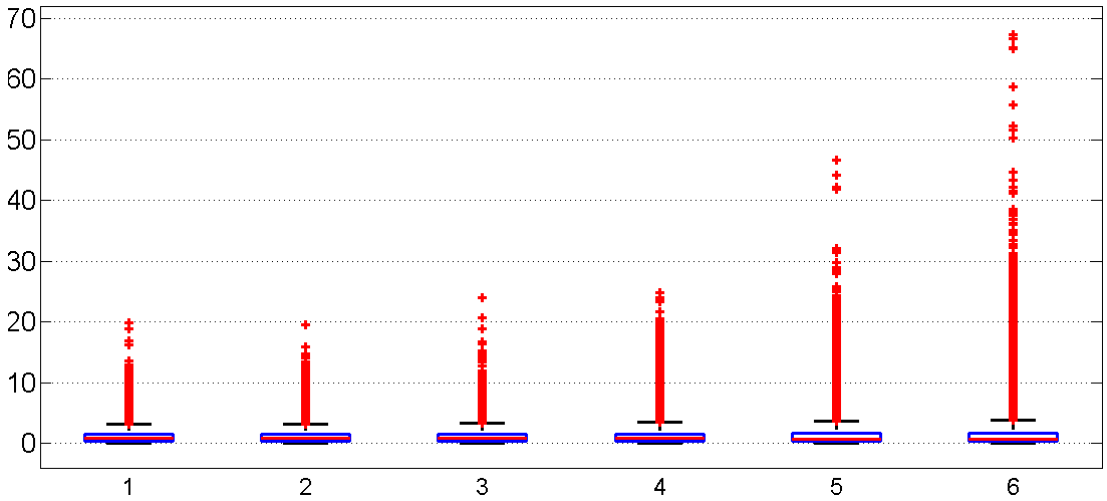


Figure 6.39: Box-plots of the simulated response at the third sampling time point, y_{i3}, for the six values of the true error variance, σ^2_{true}, labelled as 1, ..., 6 for σ^2_{true} = .02, .05, .10, .25, .40 and .60 respectively.

Statistic	σ^2_{true}	y_{i1}	y_{i2}	y_{i3}
Mean (Median)	.02	0.7542 (0.6702)	3.9774 (3.6200)	1.1794 (0.8438)
	.05	0.7610 (0.6692)	3.9981 (3.6010)	1.1902 (0.8469)
	.10	0.7697 (0.6601)	4.0412 (3.5484)	1.1962 (0.8289)
	.25	0.8158 (0.6432)	4.3005 (3.4468)	1.2867 (0.8090)
	.40	0.8624 (0.6253)	4.5338 (3.3556)	1.3963 (0.7875)
	.60	0.9296 (0.6140)	4.9376 (3.2539)	1.5139 (0.7892)
IQR (CV)	.02	0.4502 (51.89)	2.2755 (47.70)	1.1099 (95.74)
	.05	0.4645 (54.11)	2.3740 (50.08)	1.1134 (97.67)
	.10	0.5069 (60.06)	2.5786 (56.60)	1.1277 (101.84)
	.25	0.6217 (78.61)	3.2485 (75.22)	1.2196 (118.10)
	.40	0.7241 (96.52)	3.8298 (92.55)	1.3242 (138.12)
	.60	0.7950 (114.50)	4.2117 (113.72)	1.4051 (165.01)

Table 6.26: Four statistics, mean, median in the brackets, interquartile range and coefficient of variation in the brackets of the simulated responses for the six values of the true error variance, σ^2_{true}.

V, i.e, ω_3. Large σ^2_{true} seriously increases bias in ω_3 and it is worse with less accurate σ^2_o. As seen in Table 6.26, the variability in the simulated responses increases with the error variance. Since the parameter V acts as a scaling parameter in the compartmental model, a possible reason for the large positive bias in ω_3 could be the overestimation of the inter-individual variability in the parameter V. It may be

	σ^2_{true}	K_a	K_e	V	ω_1	ω_2	ω_3	σ^2	
	.02	0.8297	0.1533	17.3693	0.0879	0.0931	0.1137	0.0236	
	.05	0.8293	0.1539	17.1974	0.0873	0.0933	0.1178	0.0543	
	.10	0.8192	0.1580	17.0449	0.0870	0.0940	0.1165	0.1050	
$\bar{\hat{\mathbf{\Psi}}}$	.25	0.7997	0.1607	17.0928	0.0802	0.0897	0.1427	0.2594	
	.40	0.8092	0.1644	21.3341	0.0625	0.0810	0.5984	0.4149	
	.60	0.8094	0.1670	20.6649	0.0565	0.0765	0.6313	0.6031	
	.02	-2.4	2.2	2.2	-12.1	-6.9	13.7	18.0	[**8.2**]
	.05	-2.4	2.6	1.2	-12.7	-6.7	17.8	8.6	[**7.4**]
$\widehat{\text{Bias}}(\hat{\mathbf{\Psi}})$	.10	-3.6	5.3	0.3	-13.0	-6.0	16.5	5.0	[**7.1**]
p.c.	.25	-5.9	7.1	0.5	-19.8	-10.3	42.7	3.7	[**12.9**]
	.40	-4.8	9.6	25.5	-37.5	-19.0	498.4	3.7	[**85.5**]
	.60	-4.8	11.3	21.6	-43.5	-23.5	531.3	0.5	[**90.9**]
	.02	4.9	4.0	13.4	27.4	15.3	24.1	29.2	[**16.9**]
	.05	5.0	4.1	14.5	33.8	16.6	24.4	21.2	[**17.1**]
	.10	7.5	22.5	16.3	45.8	17.9	29.0	19.1	[**22.6**]
CV	.25	8.7	5.6	22.9	81.1	21.3	84.1	16.0	[**34.3**]
	.40	13.8	11.0	69.9	121.6	50.0	69.7	14.8	[**50.1**]
	.60	15.8	11.1	75.8	150.1	61.2	66.7	14.2	[**56.4**]

Table 6.27: Comparison of statistics related to different values of the error variance with corresponding initial values, σ^2_{oa} = .03, .075, .15, .38, .60 and .90 (Scenario I). The vector of true parameters is $\mathbf{\Psi}_{true} = (.85, .15, 17, .1, .1, .1, .1)^T$.

useful, therefore, to include appropriate covariates in the model which could perhaps reduce the variability in the parameters' estimates to some extent. The parameter σ^2 seems well estimated in all cases.

Interestingly, the percentage bias in the estimates of σ^2 decreases with an increase in the magnitude of the error variance. The reduction in the percentage bias is due to the larger value of σ^2_{true} which is in the denominator of the percentage bias. The absolute bias in σ^2 is similar across the different values considered.

The variability in the estimates generally increases with an increase in the magnitude of the error variance. This is shown by the CVs of the estimated parameters. The CVs of the variance parameters i.e., ω_1, ω_2 and ω_3 are relatively more affected by the magnitude of the error variance than other parameters in $\mathbf{\Psi}$.

Comparing Tables 6.27 and 6.28, it can be seen that, in general, bias is larger in Scenario I than in Scenario II. The difference in the estimates between the two set

	σ^2_{true}	K_a	K_e	V	ω_1	ω_2	ω_3	σ^2	
	.02	0.8302	0.1532	17.2464	0.0895	0.0940	0.1121	0.0238	
	.05	0.8284	0.1539	17.2801	0.0866	0.0929	0.1140	0.0543	
	.10	0.8192	0.1580	17.0449	0.0870	0.0940	0.1165	0.1050	
$\bar{\hat{\boldsymbol{\Psi}}}$	.25	0.8037	0.1602	17.1423	0.0837	0.0890	0.1364	0.2564	
	.40	0.7833	0.1648	17.3461	0.0875	0.0841	0.1847	0.4063	
	.60	0.7716	0.1693	18.0542	0.0870	0.0842	0.2454	0.5959	
	.02	-2.3	2.1	1.4	-10.5	-6.0	12.1	19.2	[**7.7**]
	.05	-2.5	2.6	1.2	-13.4	-7.1	14.0	8.7	[**7.1**]
$\widehat{\text{Bias}}(\hat{\boldsymbol{\Psi}})$	.10	-3.6	5.3	0.3	-13.0	-6.0	16.5	5.0	[**7.1**]
p.c.	.25	-5.4	6.8	0.8	-16.3	-11.0	36.4	2.5	[**11.3**]
	.40	-7.8	9.8	2.0	-12.5	-15.9	84.7	1.6	[**19.2**]
	.60	-9.2	12.9	6.2	-13.0	-15.8	145.4	-.70	[**29.0**]
	.02	4.9	3.9	13.6	27.6	15.3	22.3	28.0	[**16.5**]
	.05	5.0	4.1	14.5	33.8	16.6	24.4	21.2	[**17.1**]
	.10	7.5	22.5	16.3	45.8	17.9	29.0	19.1	[**22.6**]
CV	.25	9.0	5.6	21.2	74.5	20.4	70.9	15.9	[**31.0**]
	.40	11.4	6.9	34.2	94.0	25.9	99.2	15.3	[**41.0**]
	.60	13.6	7.8	42.1	112.0	32.0	102.0	13.9	[**46.2**]

Table 6.28: Comparison of statistics related to different values of the error variance and the corresponding initial values, $\sigma^2_{ob} = .07, .10, .15, .30, .45$ and $.65$ (Scenario II). The vector of true parameters is $\boldsymbol{\Psi}_{true} = (.85, .15, 17, .1, .1, .1, .1)^T$.

of initial values is particularly prominent for larger values of σ^2_{true}, i.e., .40 and .60. The reason for this is that the initial values in σ^2_{oa} are farther away from the true values than the initial values in σ^2_{ob}, particularly for larger values of σ^2_{true}. This is consistent with the observations made in Section 6.3.4 that when the initial values are significantly deviated from the true values, the bias in the estimates tends to be large.

In conclusion, large values of the error variance can have an adverse effect on the quality of estimates of some parameters. As the parameters' estimates drive the ED algorithm in computation of the optimal dose regimen, it is important to control the magnitude of the error variance as much as possible. Inclusion of relevant covariates in the model, ensuring that the assumed model is a good fit to the response data and reducing observational error are some of the steps that can be taken to reduce the intra-individual variability in the response. Although we do not work with covariates

in this thesis, a brief discussion is given in Section 7.1.4.

6.3.8 Robustness to Design Implementations

The adaptive method described in this chapter may suffer from some procedural imperfections when it is implemented in practice. These imperfections may arise primarily from two sources: missing data and subjects' non-compliance with the recommended dose regimen. In this section, we explore how prevalence of such factors influence the performance of the adaptive procedure.

Missing Data

The problem of missing data is widespread in clinical research and can compromise the conclusions drawn from the trial, as discussed in Dziura et al. (2013). Missing data are defined by Little et al. (2012) as values that are not available and that would have been meaningful for analysis if they could have been observed.

Missing data have been broadly classified by Little and Rubin (2002) into three categories on the basis of how they are generated. One of the most basic missing data mechanisms is 'missing completely at random' (MCAR), Little (2013). Under this mechanism, the probability of an observation being missing is the same across the data irrespective of factors such as patients' covariates. The data points could be missing due to factors such as equipment failure or subjects dropping out at random during the study.

The second mechanism is called 'missing at random' (MAR). This is when the probability of missing data are related to observed variables but not to the unobserved ones. For example, in a clinical trial study, males may have a higher dropout rate as compared to females but all males have an equal chance of dropping out and all females have an equal chance of dropping out, Dziura et al. (2013). That is, the chance of a male dropping out from the study does not depend on the variable of interest in the clinical trial. In such a case, the missing data are referred to as MAR.

The third mechanism is termed as 'missing not at random' (MNAR). In this case, the chance of an observation being missing depends on the unobserved variables in the study. For example, in a weight-loss clinical trial, morbidly obese patients may have a higher dropout rate as compared to the other participants in the trial. In such cases, the mere presence of missing data points may indicate information about response.

In the context of the current simulation study, since the only data which are collected from the subjects are the response data about the drug's concentrations, we

assume that the missing data are generated by the MCAR mechanism.

There are various methods available for handling missing values in a study. The appropriate method to be used also depends on the underlying mechanism which generates the missing data, if known.

The most trivial method is the 'complete-case analysis' which simply ignores all the subjects for which one or more values are missing and analyses data from only those subjects for whom complete information is available. For example, in the adaptive procedure, if m_s (≥ 1) of the m concentration values were not collected from a subject, the $(m - m_s)$ non-missing values from this subject are also ignored from consideration in the analysis. This procedure is straightforward but may lead to significant underutilisation of the collected data since some observations are excluded from the analysis.

Missing values are frequently handled by using one of the 'imputation' techniques, which involve replacement of the missing values with imputed values derived from the available data. One of the methods that could be considered is 'last observation carried forward' (LOCF). In this method, a missing value is replaced with the subject's last recorded observation. For example, suppose a subject drops out after the second observation y_{i2} has been collected. That is, the last observation, y_{i3}, is missing. In this case, the LOCF method sets $y_{i3} = y_{i2}$. This method can introduce serious bias in the parameters' estimates and is not suitable for the MCAR mechanism.

One of the most commonly used imputation techniques is the mean imputation. It consists of replacing the missing values of the variable with the simple mean of the available values for that variable. Although it ignores information from other variables in the study, the mean imputation method generally provides valid estimates for the MCAR mechanism, Dziura et al. (2013). Since we have not considered covariates or other response variables in our example, let us explore the performance of the procedure when the missing values of the response are handles using the mean imputation technique.

Let the blood samples $\{y_{i1j}, y_{i2j}, y_{i3j}\}$ from the i^{th} subject in the j^{th} cohort be collected at time points $\xi = \{T_{1j}, T_{2j}, T_{3j}\}$, $i = 1, ..., c$ and $j = 1, ..., C$.

Then, the imputed values to replace the missing observations in the data can be computed in the following way. Suppose in the j^{th} cohort, only one observation of one subject, say, y_{i1j} is missing. Then, the mean of the available observations at T_{1j} for the other 9 subjects in the cohort can be used to impute for the missing value. If two subjects have missing values at T_{1j}, then the mean of the observations collected from the other 8 subjects can be imputed for the two missing values. Similarly, missing values for the variables y_{i2j} and y_{i3j} can be imputed by the respective means based

on the available observations.

Since our methodology is adaptive in nature, it is pertinent to consider whether data collected from the previous cohorts can be utilised for computing the imputed values. That is, whether the observations collected from the subjects contained in the previous $(j-1)$ cohorts can be used for computing the mean for imputation for the missing values in the j^{th} cohort, $j = 2, ..., C$. That way, the mean is expected to be a better representative of the underlying process. However, the information from previous cohorts can not be used because of two reasons: the dose regimens administered to the cohorts are different as the estimated parameters from each cohort are different. Secondly, the D-optimal sampling times, which are also functions of the estimated parameters, are different for every cohort. Because of these two reasons, the values of the response in the previous cohorts, being collected at different experimental settings, can not be directly used for imputation in the current cohort. Therefore, for our work, we compute the mean based only on the observations collected from the current cohort. We describe the procedure below.

In the adaptive methodology described in Section 6.2.1, $m = 3$ blood samples are collected from each of the $c \times C = 100$ subjects for measuring the concentration of the drug. Thus, a total of $m \times c \times C = 300$ blood samples are collected in the trial. So far, in every scenario considered by us, all of these 300 response points were assumed to be available for analysis.

Let q be the probability with which an observation may be missing. This probability is assumed to be independent of factors such as the magnitude of the response, subject's ID and the cohort number. Therefore, in a simulation, the expected number of missing values of the response is $q \times c \times C \times m$. Under this set-up, a subject can have more than one missing value and more than one subject in the trial can have missing values. Furthermore, the probability of each of the $m = 3$ observations to be missing is the same, that is, q.

To implement this in the computer code, we replace the simulated responses in the current cohort with 'NaN' (not-a-number) having probability q, where NaN in MATLAB® denotes a missing or an unavailable data point. That is, from a cohort of $c = 10$ subjects each contributing $m = 3$ concentration values, the expected number of NaN values is $30q$. Thereafter, each of these NaN values are replaced by the corresponding mean values, which are computed as explained above. After this replacement, the model parameters for this cohort are estimated using the completed dataset of 30 observations. The other steps of the adaptive procedure remain the same.

We consider three values of probability q, that is $q = .02, .10, .25$. Simulation

studies were run according to these values and the results are presented in Table 6.29. The data related to the original simulation study when there are no missing observations, that is when $q = 0$, have been re-presented to facilitate comparisons.

	q	K_a	K_e	V	ω_1	ω_2	ω_3	σ^2	
	.00	0.8192	0.1580	17.0449	0.0870	0.0940	0.1165	0.1050	
$\bar{\hat{\boldsymbol{\Psi}}}$	.02	0.8241	0.1556	17.0919	0.0820	0.0898	0.1134	0.1098	
	.10	0.8276	0.1541	17.3150	0.0598	0.0842	0.1007	0.1247	
	.25	0.8456	0.1518	17.2170	0.0426	0.0739	0.0737	0.1372	
	.00	-3.6	5.3	0.3	-13.0	-6.0	16.5	5.0	[**7.1**]
$\widehat{\text{Bias}}(\hat{\boldsymbol{\Psi}})$	.02	-3.0	3.7	0.5	-18.0	-10.2	13.4	9.8	[**8.4**]
p.c.	.10	-2.6	2.7	1.9	-40.2	-15.8	0.7	24.7	[**12.6**]
	.25	-0.5	1.2	1.3	-57.4	-26.1	-26.3	37.2	[**21.4**]
	.00	7.5	22.5	16.3	45.8	17.90	29.0	19.1	[**22.6**]
CV	.02	6.3	4.5	14.8	50.0	18.1	33.7	18.7	[**20.8**]
	.10	7.2	4.9	16.2	67.3	19.2	50.2	19.2	[**26.3**]
	.25	8.2	5.8	18.4	94.8	22.6	59.3	19.2	[**32.6**]
		d_1^*	d_2^*	d_3^*	d_4^*	d_5^*		$\bar{\varphi}_A$	
	.00	145.58	92.79	97.02	96.67	97.02		8.5577	
$\boldsymbol{D}^*$	.02	145.75	92.70	97.00	96.81	97.01		8.2282	
	.10	146.65	92.70	97.33	96.70	97.34		8.7061	
	.25	144.33	90.80	95.45	94.38	95.45		9.3376	
		T_1^*	T_2^*	T_3^*					
	.00	.10	6.75	48.00					
ξ_C^*	.02	.10	6.68	48.00					
	.10	.10	6.45	48.00					
	.25	.10	6.12	48.00					
		$\bar{C}_2$	$\bar{\varphi_A}(C_2)$				$\bar{C}_3$	$\bar{\varphi_A}(C_3)$	
	.00	3.81	8.8241				7.52	8.9036	
ACN	.02	3.80	9.2114				7.67	9.4253	
	.10	3.93	9.5173				7.63	9.7214	
	.25	4.05	9.7459				7.82	10.1241	

Table 6.29: Comparison of data related to the simulation studies when some observations are missing.

From the table, it can be observed that the adaptive procedure is quite robust in situations in which the percentage of missing values is mild, i.e., up to 2%. The bias and the CV of the estimated parameters are similar to the case when no missing

values are present. Furthermore, the φ_A values, average D-optimal sampling times ξ_C^* and the ACNs for $q = .00$ and $q = .02$ are very similar. The distributions of φ_A for the four values of q are presented in Figure 6.40.

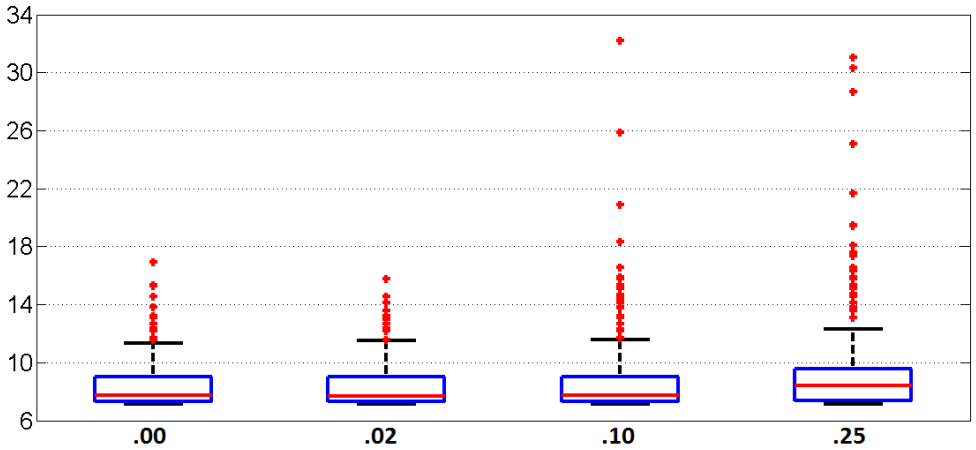


Figure 6.40: Distribution of φ_A for the four values of q: .00, .02, .10 and .25.

The distributions of φ_A for $q = 0$ and $q = .02$ are also quite similar. Therefore, no significant effect is observed on the performance of the adaptive procedure even when about 2% of the responses are missing.

For $q = .10$, the adverse effects of missing observations begin showing-up in the simulated data. The bias and the CV of the estimated parameters increase, although the increase is proportionally not as large as the increase in the value of q. The $\bar{\varphi}_A$ value increases which implies that the dose regimen administered to the last cohort is, on average, inferior to the case when missing values are not present. The increase in the variability of the estimated parameters also results in larger ACN values which means that, on average, there is a delay in the termination of the trial according to the two stopping rules. Also, as can be seen in Figure 6.40, for some cohorts, φ_A is significantly larger for $q = .10$ as compared to $q = .00$ and $q = .02$.

Finally, for $q = .25$ when, on average, one out of every four observations is missing, the bias and the variability in some of the parameters' estimates increase significantly. Also, $\bar{\varphi}_A$ and the spread in the distribution of φ_A are larger. This shows that in the presence of a large number of missing values, some cohorts may be administered dose regimens which are significantly inefficient. Furthermore, the ACN values in this case are larger which means that presence of missing values may delay the termination of the trial according to the defined stopping rules.

A noteworthy observation that can be made from the simulated data is the underestimation of the variance parameters, ω_1, ω_2 and ω_3 for relatively larger values of q. This indicates a potential disadvantage of the imputation technique used in this example. Single imputation methods such as the mean imputation technique treat the imputed values as if they were observed and do not account for the uncertainty

associated with the missing response values, Ravina et al. (2012). Since the ED algorithm does not require the variance parameters as inputs, this disadvantage of the mean imputation procedure does not affect the performance of the methodology significantly. Although mean imputation works reasonably well for this example, other methods could be used to improve the estimates of the variance parameters.

For example, the method 'multiple imputation' described in Rubin (1987) is able to overcome to some extent the problem of under-representation of uncertainty associated with the single imputation methods. The multiple imputation procedure consists of three steps: Firstly, the missing values in the dataset are replaced with imputed values not just once but a number of times, say, $M(\geq 2)$ to create M completed datasets. These imputed values are drawn from specified distributions which can be different for each of the missing value. In the second step, each of these M datasets are analysed which results in M analyses. Finally, these analyses are pooled together to compute statistics such as the mean and the variance for the variable of interest.

In conclusion, the adaptive procedure outlined in this chapter is reasonably robust when a small percentage of the response values are missing. Increase in the proportion of missing values can result in administration of inefficient dose regimens and deterioration in the quality of some of the parameters' estimates. Inclusion of covariates in the PK model could lead to further explorations into the robustness of the procedure under the MAR and MNAR mechanisms.

Non-compliance with the Recommended Dose Regimen

Another possible obstacle in proper implementation of the adaptive procedure is non-compliance of the subjects to the recommended dose regimen. Non-compliance is a significant medical challenge with rates of non-compliance being as much as 25% for some short-term therapies, Jin et al. (2008). The most common actions that constitute non-compliance are: not following the prescribed interval between two successive doses and omission of the recommended doses. Some drugs are 'forgiving', that is, their duration of therapeutic action is more than double of the time interval between two successive doses, which gives them a certain degree of robustness against non-compliance of the subjects towards the prescribed dose regimen, Urquhart (1996). But even for such drugs, a large degree of non-compliance may lead to therapeutic failure. Below we explore the effect of non-compliance on the adaptive procedure described in this chapter.

In the original simulation study, the dosing time points were $\boldsymbol{t} = (t_1, ..., t_5)^T = (0, 8, 16, 24, 32)^T$. That is, there was a uniform time interval of 8 h between any two successive doses. Let the i^{th} dosing interval be $\tau_i = t_{i+1} - t_i$, $i = 1, ..., 4$. Then, the time

intervals can be expressed in the form of a vector as $\boldsymbol{\tau} = (\tau_1, ..., \tau_4)^T = (8, 8, 8, 8)^T$. When the method is implemented in practice, it is unreasonable to expect that all $c \times C$ subjects participating in the study will follow the prescribed dosing time points perfectly. As a result, the time interval between two successive doses will generally be a random variable centred around 8 h.

Regarding skipping of dose, there are a total of $n \times c \times C = 5 \times 10 \times 10 = 500$ doses administered in a single trial. Let a dose be skipped by a subject with the probability p_d. Then, the expected number of skipped doses during the trial is $500p_d$. Under this set-up, the subjects are assumed to be equally non-compliant, that is, the chance of skipping a dose is the same irrespective of the subject and is equally probable for any of the $n = 5$ doses administered to a subject. The probability of a subject missing at least one dose out of the administered dose regimen of $n = 5$ doses can be obtained by subtracting from 1 the probability of not missing any of the n doses. That is, the probability of a subject following an incomplete dose regimen is $1 - (1 - p_d)^n$.

To evaluate the effect of non-compliance on the adaptive method, we define three scenarios:

(I) The vector of time intervals for a subject is defined as $\boldsymbol{\tau}_u = (u_1, u_2, u_3, u_4)^T$, where u_is, $i = 1, ..., 4$, are drawn independently from uniform distribution on interval [7,9] for the subject. The mean of U[7,9] distribution is 8, so these random intervals are centred around the original time interval of 8 h. Consequently, the corresponding dosing time points are $\boldsymbol{t} = \left(0, u_1, \sum_{i=1}^{2} u_i, \sum_{i=1}^{3} u_i, \sum_{i=1}^{4} u_i\right)^T$. As a result, the dosing schedules for the subjects are slightly different from each other. To implement this in our MATLAB® code, the random dosing time points as defined above are applied in Equation (6.5) to simulate y_{ij}, the j^{th} response from the i^{th} subject at time T_j, $j = 1, 2, 3$ and $i = 1, ..., c$. This is done for each of the C cohorts.

(II) The vector of time intervals for a subject is defined as $\boldsymbol{\tau}_v = (v_1, v_2, v_3, v_4)^T$, where v_is, $i = 1, ..., 4$ are drawn independently from N(8,1) distribution for the subject. As compared to Scenario I, this scenario assumes a higher degree of non-compliance of the subjects to the recommended dose regimen. For U[7,9] distribution, 100% of the dosing intervals lie within the interval [7, 9]. However, in this scenario, by definition of the normal distribution, 32% of the dosing intervals lie outside the interval [7,9] since for a $\mathrm{N}(\mu, \sigma^2)$ distribution, 68% of the area lies between $[\mu - \sigma, \mu + \sigma]$ and the rest 32% outside this interval. In this scenario, therefore, the dose schedules followed by some subjects will be significantly non-compliant to the recommended dose schedules.

(III) In this scenario, we consider the case of dose omission, i.e., inadvertent skipping of one or more doses by the subjects. Let p_d, the probability of a subject missing a

dose, be 0.5%. The expected number of doses missed randomly in a trial consisting of $c \times C = 100$ subjects is 2.5. For an individual, the probability of not ingesting all the prescribed n doses is $1 - (1 - p_d)^n = 2.47\%$. To implement this in MATLAB® , we equate the value of the k^{th} dose, d_k, to zero in Equation 6.5 with probability p_d, $k = 1, ..., n$.

For each of these these three scenarios, the simulation studies were run and the data are presented in Table 6.30. Data pertaining to the original run, denoted by Scenario 'O', when perfect compliance was assumed, have been re-presented to facilitate comparison. Figure 6.41 presents the distribution of φ_A for the three scenarios described in this section and the original run.

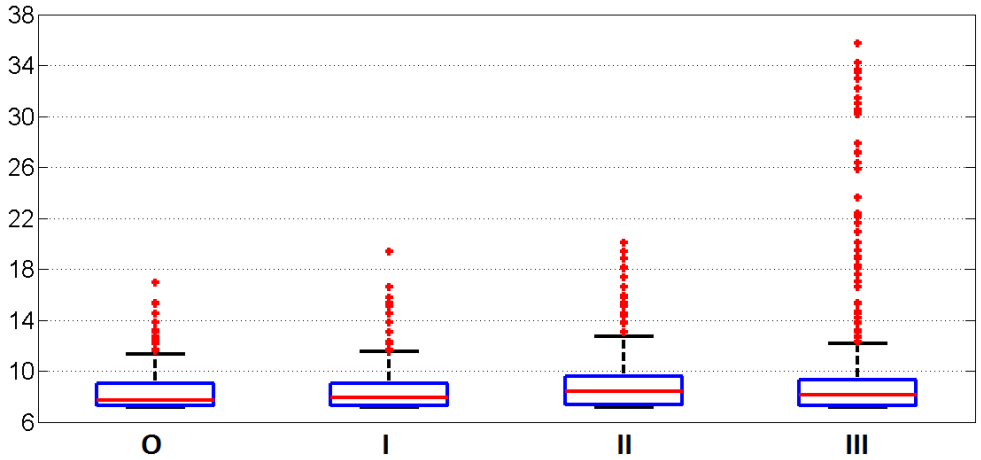


Figure 6.41: Distribution of φ_A for the three scenarios considered in this section and the original study.

It can be seen from the table that in Scenario I the parameter estimates are only mildly affected when the dosing intervals are drawn from U[7,9] distribution. Other performance indicators of the procedure such as the average optimal dose regimen, ACN, $\bar{\varphi}_A$ and the spread in φ_A are also not much different from the case when each subject in the trial takes the drug strictly every 8 h. This shows that the adaptive procedure is robust when the deviation from the prescribed dosing interval is not more than 1 hour on either side.

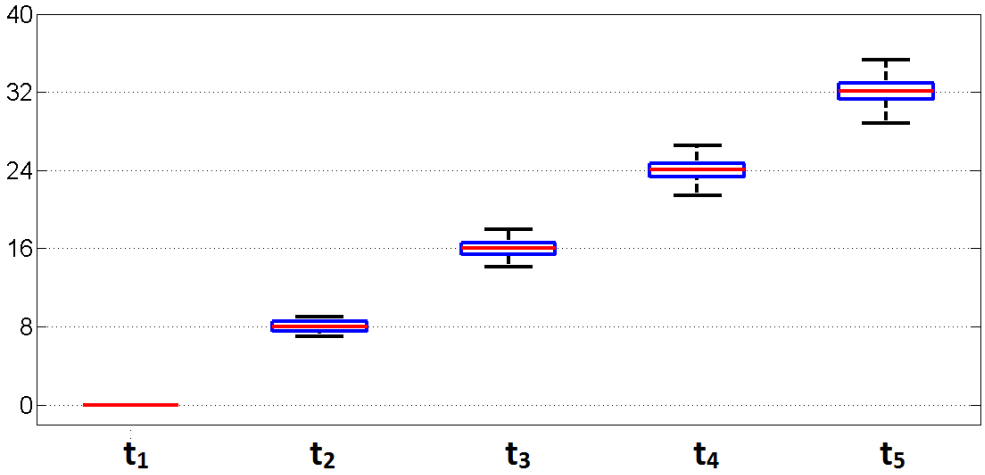
The adaptive procedure does not seem quite robust for Scenario II, when about 32% of the dosing intervals have a deviation of more than 1 hour from the prescribed interval of 8 h. The parameters have, in general, larger bias and variability. The parameters which are most adversely affected by the variability in the dosing time points are ω_1, ω_3 and σ^2. The value of $\bar{\varphi_A}$ is larger which means that on average, the subjects in the last cohort experience a higher degree of under- and over-exposure. The values of $\bar{\varphi_A}(C_2)$ and $\bar{\varphi_A}(C_3)$, i.e., the average values of under- and exposure experienced by the subjects when the trial is stopped according to SR2 and SR3 are also large which shows that it will be imprudent to terminate the trial early in such a

		K_a	K_e	V	ω_1	ω_2	ω_3	σ^2	
	O	0.8192	0.1580	17.0449	0.0870	0.0940	0.1165	0.1050	
	I	0.8059	0.1558	16.7899	0.0710	0.0966	0.1192	0.1206	
$\bar{\widehat{\boldsymbol{\Psi}}}$	II	0.7727	0.1572	15.5259	0.0446	0.1050	0.1112	0.1519	
	III	0.7814	0.1598	17.7924	0.2882	0.0901	0.1967	0.1985	
	O	-3.6	5.3	.3	-13.0	-6.0	16.5	5.0	[**7.1**]
	I	-5.1	3.8	-1.2	-30.0	-3.4	19.2	20.6	[**11.8**]
$\widehat{\text{Bias}}(\widehat{\boldsymbol{\Psi}})$	II	-9.1	4.8	-8.7	-55.4	5.0	11.2	51.9	[**20.9**]
p.c.	III	-8.1	6.6	4.7	188.2	-9.9	96.7	98.5	[**58.9**]
	O	7.5	22.5	16.3	45.8	17.9	29.0	19.1	[**22.6**]
	I	8.8	5.6	17.6	59.8	17.8	40.0	18.8	[**24.1**]
CV	II	11.2	7.1	19.4	89.5	16.9	43.3	20.1	[**29.6**]
	III	10.4	6.0	26.3	140.1	27.4	141.8	70.5	[**60.4**]
		d_1^*	d_2^*	d_3^*	d_4^*	d_5^*		$\bar{\varphi}_A$	
	O	145.58	92.79	97.02	96.67	97.02		8.5577	
	I	142.29	90.09	94.57	93.98	94.57		8.3045	
$\boldsymbol{D}^*$	II	132.84	83.58	88.25	87.56	88.26		9.0544	
	III	149.54	99.69	103.20	102.83	103.23		10.3046	
		T_1^*	T_2^*	T_3^*					
	O	.10	6.75	48.00					
	I	.10	6.79	48.00					
ξ_C^*	II	.10	6.90	48.00					
	III	.10	6.11	48.00					
		$\bar{C}_2$	$\bar{\varphi}_A(C_2)$				$\bar{C}_3$	$\bar{\varphi}_A(C_3)$	
	O	3.81	8.8241				7.52	8.9036	
	I	3.88	9.1568				7.59	9.2746	
ACN	II	3.98	9.8442				7.56	10.4362	
	III	3.97	10.1534				7.61	10.5003	

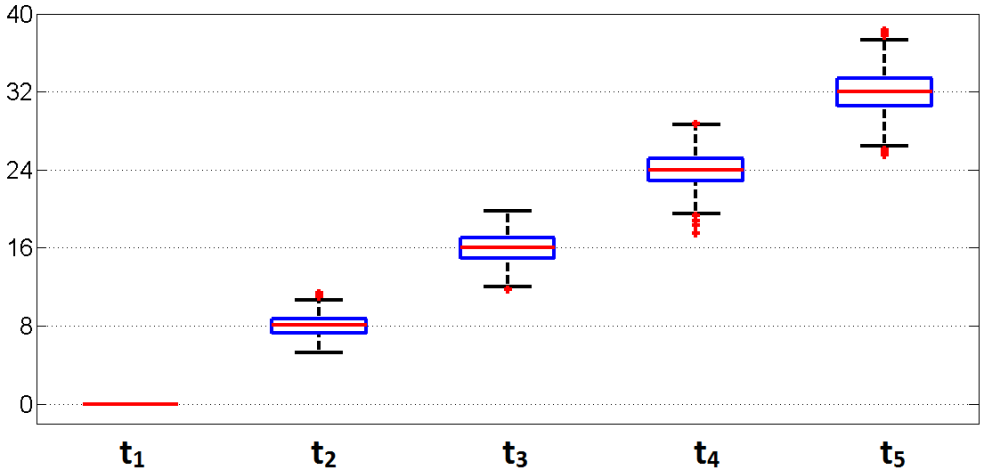
Table 6.30: Comparison of the data from simulation studies when some subjects are non-compliant to the treatment regimen.

scenario. Also, as can be seen in Figure 6.41, the spread of φ_A is larger for the cohorts in this scenario as compared to Scenarios I and to the case of perfect compliance. As a result, when implemented in practice, it should be endeavoured that the subjects comply with the prescribed dosing intervals as far as possible. The mean D-optimal sampling time points do not seem to be significantly affected in any of the scenarios.

For Scenarios I and II, the distributions of the simulated dosing time points are presented in Figure 6.42. The simulated dosing time points are centred around the planned time points $\boldsymbol{t} = (0, 8, 16, 24, 32)^T$. As discussed before, the spread for case of the normal distribution is larger than that for the uniform distribution. Also, the variability increases with the index of the dosing time point t_i since the variances associated with the previous dosing time points add up in the variance of the current time point.



(a) Scenario I: time intervals are drawn from uniform distribution.

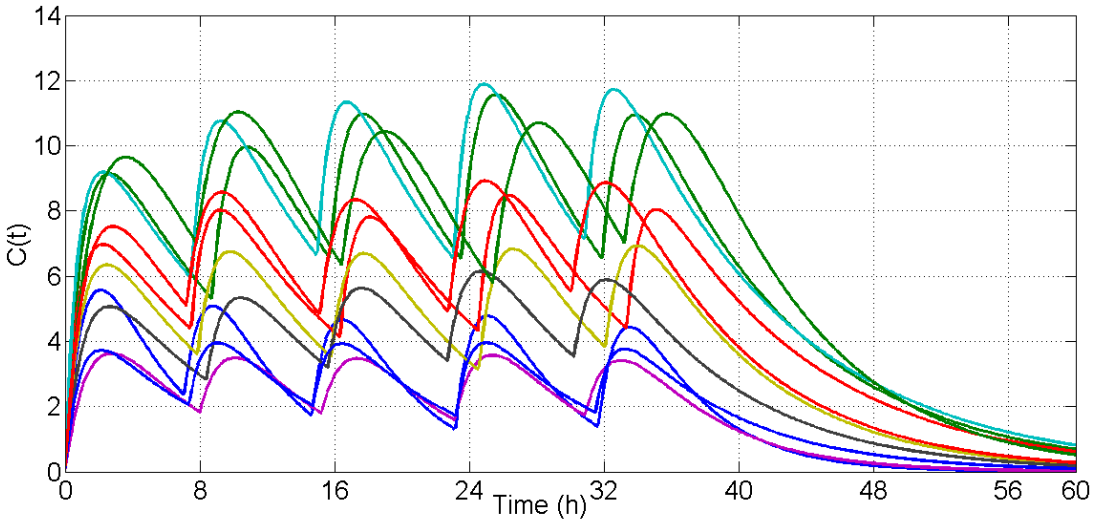


(b) Scenario II: time intervals are drawn from normal distribution.

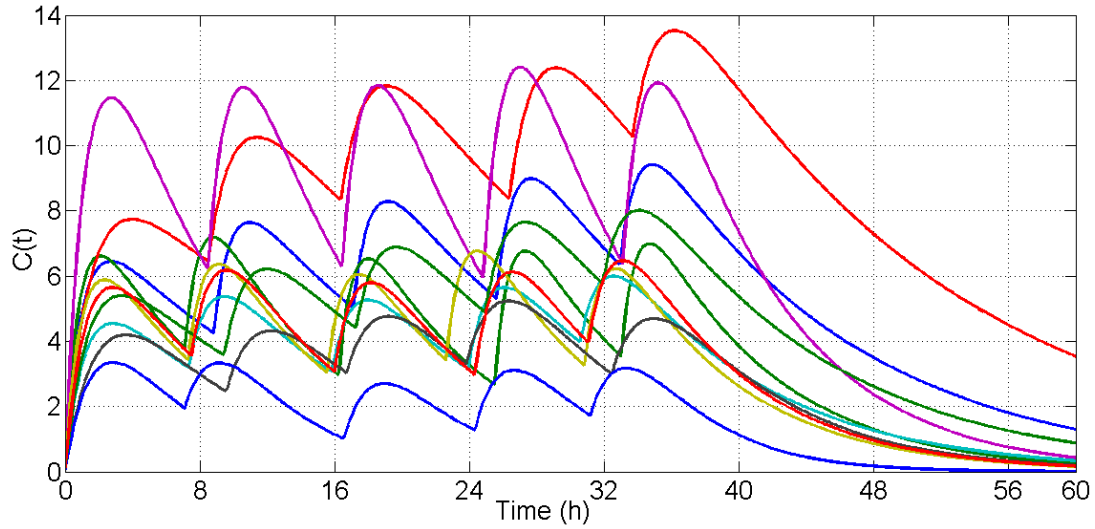
Figure 6.42: Distribution of the simulated dosing time points $\boldsymbol{t}$ for Scenarios I and II. The prescibed dosing time points (perfect compliance) are $\boldsymbol{t} = (0, 8, 16, 24, 32)^T$.

Figure 6.43 presents concentration profiles corresponding to ten vectors of the simulated dosing time points $\boldsymbol{t}$, as an example. The departure from the prescribed dosing time points is larger in Figure 6.43b since the assumed variability in the dosing time points is larger in this case, as discussed above.

In Scenario III, when the dose is missed by a subject with probability $p_d = .005$, the effect on the performance of the methodology is severe. The bias and the variability in the estimates, especially for the variance parameters, is significantly large. The value



(a) Scenario I: time intervals are drawn from uniform distribution.



(b) Scenario II: time intervals are drawn from normal distribution.

Figure 6.43: Concentration profiles corresponding to the simulated dosing time points $\boldsymbol{t}$ for Scenarios I and II. The prescibed dosing time points (perfect compliance) are $\boldsymbol{t} = (0, 8, 16, 24, 32)^T$.

of $\bar{\varphi}_A$ is quite large which means that even in the terminal cohort, the subjects, on average, experience excessive over- and under-exposure to the treatment drug. This shows that the adaptive procedure is quite sensitive to omission of the prescribed doses by the subjects, even when only about 2 to 3 doses, on average, are skipped in a trial which involves $c \times C = 100$ subjects. For values of p_d greater than .005, estimation problems arose and the algorithm terminated abruptly. As a result, when carried out in practice, it should be ensured as much as possible that the subjects do not skip the prescribed doses. If any concentration value is measured to be too low, it should be checked, if possible, from the concerned subject that whether a dose was missed. If so, the affected response value could be treated as a missing value and imputed with the series mean for which, as shown previously, the algorithm is

reasonably robust.

In conclusion, the adaptive procedure is found to be reasonably robust for mild to medium deviations from the planned design, except that care must be exercised that the prescribed doses are not missed by the subjects. The procedure is quite sensitive to the instances of dose omission. Therefore, if a measured concentration value is observed to be suspicious, or is out-of-sync from the other values collected from that subject, it might be sensible to investigate, if possible, whether a dose was missed. Inclusion of covariates could further increase the robustness of the procedure as that information could be utilised for computing the imputed value in case of missing values. Covariates' data could also be useful for validating the measured responses and for detecting if doses have been missed during the trial.

6.4 Comparison with a Non-adaptive Approach

As explained at the beginning of this chapter, our methodology for ascertaining the PK parameters and optimisation of the dose regimen is adaptive in nature. The expected benefits of the adaptive design are: fewer patients to be administered sub-optimal dose regimen as the parameters' estimates and the dose regimens should improve over the course of the adaptive trial on account of accumulating data and possibility of an early termination of the trial. That is, what we expect from adaptive designs are higher economy and ethical standards with at least as good outcomes as from the traditional approach.

However, the logistics of such trials can be complicated and so clinicians may be discouraged from applying them. A question then arises whether we can achieve similar properties from non-adaptive designs, which by definition are simpler. The objectives, as before, are PK estimation and dose regimen optimisation. We compare the performance of non-adaptive approach with the adaptive one with the help of simulation studies.

To make a fair comparison between the two approaches, it is important to keep the total number of subjects equal in both. We consider two cases: when the total number of subjects is 50 and when it is 200. For the adaptive case, we consider two pairs of values of (c, C): (5, 10) and (20, 10) such that the number of cohorts in both pairs is equal, i.e., $C = 10$. For the non-adaptive approach, which involves only one cohort, the two scenarios can be expressed as $(c, C) = (50, 1)$ and $(c, C) = (200, 1)$.

Furthermore, as seen in the previous sections, the choice of initial values has important implications on the quality of the parameters' estimates and the dose regimen.

We therefore consider three vectors of initial values: initial values very close, moderately close and far-off from the true values. These cases are described below.

Initial Values Very Close to the True Values

The vector of initial values is taken to be $\mathbf{\Psi}_{oa} = (.9, .13, 18, .12, .12, .12, .12)^T$.
The true values used in the simulations are $\mathbf{\Psi}_{true} = (.85, .15, 17, .1, .1, .1, .1)^T$. That is, here we consider a very good guess. We repeat the original study using these initial values and taking four pairs of the cohort size and number of cohorts.

Table 6.31 presents the data related to the four scenarios of (c, C): *1.* (5, 10), *2.* (50,1), *3.* (20, 10) and *4.* (200,1).

As far as quality of the estimates is concerned, it is clear that most of the estimates in Scenarios *2* and *4*, i.e., for the case of non-adaptive approach, have lower bias and variability. However, the overall bias in the estimates for a cohort of size 50 is almost as much as that for a cohort of size 200, although, a reduction in variability of the estimates is observed in the latter case.

As explained in Section 6.3.3, the dose regimen administered to the first cohort in a simulation study is the same across all N_{sim} simulations since the inputs to the ED algorithm are the initial values of the parameters. Since the assumed initial values are very close to the true values, the dose regimen administered to the first cohort is quite close to the optimal dose regimen, as can be seen in the table and, accordingly, the corresponding φ_A value is also close to the optimal value φ_A^*.

Since there is only one cohort in the non-adaptive method, the φ_A value for this cohort is the same in all the N_{sim} simulations, irrespective of the cohort size. Therefore, the φ_A values for scenarios *2* and *4* are equal and there is no variability in their distributions. This can also be observed in Figure 6.44, which plots the distribution of φ_A for these four and other scenarios which we discuss later.

In general, the non-adaptive approach performs better than the adaptive approach on this count as well, since $\bar{\varphi}_A$ values for the adaptive approach are larger than the φ_A values of the non-adaptive approach. $\bar{\varphi}_A$ is the average value of φ_A for the subjects in the last, i.e., the C^{th} cohort of the trial. As φ_A measures average over- and under-exposure to the drug, its low, constant value in the non-adaptive case means that all the patients are treated with good doses during the trial thereby enhancing the ethical standards. Furthermore, as can be seen in Figure 6.44, the large spread in the distribution of φ_A for Scenarios *1* and *3* shows that a large number of patients are exposed to suboptimal dose regimens over the course of the trial, although it is better for Scenario *3* which consists of cohorts of larger sizes. Overall, the non-adaptive method has superior ethical standards when the initial values are very close to the

	c	C	K_a	K_e	V	ω_1	ω_2	ω_3	σ^2	
	5	10	0.8244	0.1548	16.9025	0.0825	0.0917	0.1352	0.1058	
$\bar{\widehat{\boldsymbol{\Psi}}}$	50	1	0.8319	0.1543	16.5268	0.0894	0.0922	0.0988	0.1022	
	20	10	0.8245	0.1542	16.4614	0.0883	0.0923	0.1100	0.1042	
	200	1	0.8264	0.1542	16.4937	0.0900	0.0931	0.0991	0.1051	
	5	10	-3.0	3.2	-0.6	-17.5	-8.3	35.3	5.8	[**10.5**]
$\widehat{\text{Bias}}(\widehat{\boldsymbol{\Psi}})$	50	1	-2.1	2.9	-2.8	-10.6	-7.8	-1.3	2.2	[**4.2**]
p.c.	20	10	-3.0	2.8	-3.2	-11.7	-7.7	10.0	4.2	[**6.1**]
	200	1	-2.8	2.8	-3.0	-10.0	-6.9	-0.9	5.1	[**4.5**]
	5	10	8.8	6.3	21.4	67.7	26.9	55.8	25.4	[**30.3**]
CV	50	1	8.8	5.5	7.4	60.0	24.0	38.8	25.6	[**24.2**]
	20	10	4.4	3.4	11.3	32.0	12.9	21.2	13.2	[**14.0**]
	200	1	3.9	2.8	3.6	31.5	12.0	18.8	13.6	[**12.3**]
			d_1^*	d_2^*	d_3^*	d_4^*	d_5^*		$\bar{\varphi}_A$	
$\boldsymbol{D}^*_{true}$			140.00	90.00	90.00	100.00	90.00		7.1561	
$\boldsymbol{D}^*$ (cohort 1)			140.00	80.00	90.00	90.00	90.00		7.2888	
	5	10	144.40	91.91	96.49	95.85	96.51		9.6096	
$\boldsymbol{D}^*$	50	1	140.00	80.00	90.00	90.00	90.00		7.2888	
Last cohort	20	10	141.15	89.04	93.71	92.75	93.71		7.8883	
	200	1	140.00	80.00	90.00	90.00	90.00		7.2888	
	5	10	150.00	90.00	100.00	90.00	100.00		7.1916	
$\boldsymbol{D}^*(\widehat{\boldsymbol{\Psi}})$	50	1	140.00	90.00	90.00	100.00	90.00		7.1561	
Recommended	20	10	140.00	90.00	90.00	100.00	90.00		7.1561	
	200	1	140.00	90.00	90.00	100.00	90.00		7.1561	
			$\bar{C}_2$		$\bar{\varphi}_A(C_2)$		$\bar{C}_3$		$\bar{\varphi}_A(C_3)$	
	5	10	4.40 (22)		10.1464		8.04 (40.2)		10.0027	
ACN (AN)	50	1	1 (50)		7.2888		1 (50)		7.2888	
	20	10	3.11 (62.2)		7.9284		6.63 (132.6)		7.9034	
	200	1	1 (200)		7.2888		1 (200)		7.2888	

Table 6.31: Data related to the simulation studies for the four scenarios: *1.* (5, 10), *2.* (50,1), *3.* (20, 10) and *4.* (200, 1), when the initial values of the parameters, $\boldsymbol{\Psi}_{oa}$, are very close to the true values.

true values.

One of the objectives of conducting the trial is to recommend a dose regimen for the future based on the best available estimates of the model parameters, $\bar{\widehat{\boldsymbol{\Psi}}}$. Using $\bar{\widehat{\boldsymbol{\Psi}}}$, the recommended dose regimen for the population is computed in each simulation and the averages for each of the four scenarios are presented in the table. The average recommended dose regimens for both adaptive and non-adaptive are reasonably close to the true optimal dose regimen. Except Scenario *1* for which the recommended dose regimen is quite close to the optimal dose regimen, the recommended dose regimens for the other three scenarios are exactly the same as the optimal regimen. Therefore, when initial values are very close to the true values, both adaptive and non-adaptive

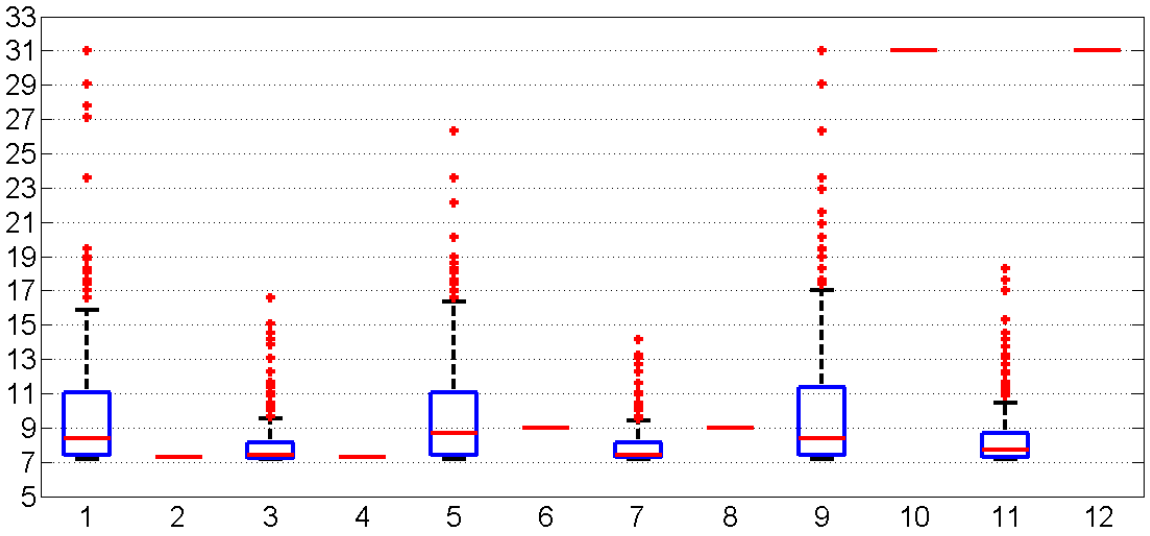


Figure 6.44: Distribution of φ_A over the course of the trial for the twelve scenarios considered in this section: (*1 - 4*) $\mathbf{\Psi}_{oa}$ very close to $\mathbf{\Psi}_{true}$, (*5 - 8*) $\mathbf{\Psi}_{ob}$ moderately close to $\mathbf{\Psi}_{true}$ and (*9 - 12*) $\mathbf{\Psi}_{oc}$ far from $\mathbf{\Psi}_{true}$.

procedures are successful in determination of the appropriate dose regimen for the population.

The average stopping cohort number (ACN) for a non-adaptive approach is always 1 as there is only one cohort in the trial. The average number of subjects utilised in a trial, denoted by AN, has been presented for each scenario in the table where $AN = ACN \times c$. Let us consider scenarios *3* and *4*. The ACN value according to the second stopping rule, SR2, for scenario *3* is 3.11. Since there are $c = 20$ subjects in a cohort, this means that on average, $3.11 \times 20 \approx 63$ subjects are examined before the trial is terminated. The average φ_A value at the cohort number at which SR2 applies, i.e., $\bar{\varphi_A}(C_2)$ is about 7.93. In scenario *4*, the φ_A value for the only cohort of 200 subjects is about 7.29. This shows that for similar values of φ_A, it might be possible to stop the trial early, resulting in significant savings in cost. This feature of the adaptive approach is absent in the non-adaptive method.

It can be concluded that when the initial values of the parameters are very close to the true values, the non-adaptive approach is better than the adaptive approach, although, it may be possible to accrue some savings by adaptation when the cohort size is reasonably large, as we see here for $c = 20$.

Next, we make this comparison between the two approaches when the initial values of the parameters are moderately close to the true values.

Initial Values Moderately Close to the True Values

The vector of initial values is now taken to be $\mathbf{\Psi}_{ob} = (1, .1, 20, .05, .15, .05, .15)^T$. We repeat the original study using these initial values and taking four pairs of the cohort size and number of cohorts, (c, C): *5.* (5, 10), *6.* (50,1), *7.* (20, 10) and *8.* (200,1). Table 6.32 presents the data related to these four scenarios, obtained for $\mathbf{\Psi}_{ob}$.

	c	C	K_a	K_e	V	ω_1	ω_2	ω_3	σ^2	
$\hat{\bar{\mathbf{\Psi}}}$	5	10	0.8293	0.1562	17.1534	0.0859	0.0907	0.1355	0.1062	
	50	1	0.8343	0.1532	16.6336	0.0896	0.0911	0.0987	0.1030	
	20	10	0.8201	0.1553	17.0027	0.0866	0.0908	0.1105	0.1058	
	200	1	0.8348	0.1530	16.6458	0.0912	0.0930	0.1014	0.1026	
$\widehat{\text{Bias}}(\hat{\mathbf{\Psi}})$	5	10	-2.4	4.1	0.9	-14.1	-9.3	35.5	6.2	[**10.4**]
	50	1	-1.9	2.1	-2.2	-10.4	-8.9	-1.3	3.0	[**4.2**]
p.c.	20	10	-3.5	3.6	0.1	-13.4	-9.2	10.5	5.8	[**6.6**]
	200	1	-1.8	2.0	-2.1	-8.8	-7.0	1.4	2.6	[**3.7**]
CV	5	10	8.8	6.1	21.2	67.8	25.0	49.7	26.6	[**29.3**]
	50	1	8.7	5.6	7.1	63.5	23.3	38.8	26.8	[**24.8**]
	20	10	4.3	3.3	11.7	32.1	12.5	21.3	12.8	[**14.0**]
	200	1	4.0	2.7	3.5	30.9	11.9	18.3	13.4	[**12.1**]
			d_1^*	d_2^*	d_3^*	d_4^*	d_5^*		$\bar{\varphi}_A$	
$\boldsymbol{D}^*_{true}$			140.00	90.00	90.00	100.00	90.00		7.1561	
$\boldsymbol{D}^*$ (cohort 1)			150.00	70.00	80.00	70.00	80.00		9.0129	
$\boldsymbol{D}^*$	5	10	145.63	93.50	97.59	96.90	97.59		9.6302	
	50	1	150.00	70.00	80.00	70.00	80.00		9.0129	
Last cohort	20	10	144.46	91.64	95.99	95.49	95.99		7.9568	
	200	1	150.00	70.00	80.00	70.00	80.00		9.0129	
$\boldsymbol{D}^*(\hat{\mathbf{\Psi}})$	5	10	150.00	90.00	100.00	100.00	100.00		7.2585	
	50	1	140.00	90.00	90.00	100.00	90.00		7.1561	
Recommended	20	10	150.00	90.00	100.00	90.00	100.00		7.1916	
	200	1	140.00	90.00	90.00	100.00	90.00		7.1561	
			$\bar{C}_2$		$\bar{\varphi}_A(C_2)$		$\bar{C}_3$		$\bar{\varphi}_A(C_3)$	
ACN (*AN*)	5	10	4.51 (22.6)		9.7706		8.18 (40.9)		9.9511	
	50	1	1 (50)		9.0129		1 (50)		9.0129	
	20	10	3.40 (68)		7.8890		6.83 (136.6)		7.8701	
	200	1	1 (200)		9.0129		1 (200)		9.0129	

Table 6.32: Data related to the simulation studies for the four scenarios: *5.* (5, 10), *6.* (50,1), *7.* (20, 10) and *8.* (200, 1), when the initial values of the parameters, $\mathbf{\Psi}_{ob}$, are moderately close to the true values.

As far as quality of the estimates is concerned, the results are quite similar to the previous case, when the initial values were taken to be close to the true values. That is, the quality of the parameters' estimates in the non-adaptive approach is overall better than in the adaptive approach.

As compared to the previous case, the dose regimen administered to the first cohort is farther from the optimal dose regimen. This is expected since a larger deviation from the true values will generally increase the extent of sub-optimality of the dose regimen computed by the ED algorithm resulting in larger φ_A values. Only in some special cases when some misspecified parameter values cancel the effects of other misspecified parameter values the resulting exposure to the drug may not change much. Such issues were discussed in Section 4.4.1.

The $\bar{\varphi}_A$ value associated with this dose regimen is about 9.01 whereas the optimal value is 7.16. All the patients in Scenarios *6* and *8* (non-adaptive approach) are administered this dose regimen. $\bar{\varphi}_A$ in case of the non-adaptive approach is, therefore, higher than when the initial values were closer to the true values.

A distinct advantage of the adaptive approach becomes apparent here. While in Scenarios *6* and *8*, all the 50 and 200 subjects are administered suboptimal dose regimens resulting in $\bar{\varphi}_A$ value of about 9.01, in scenario *7*, which represents the adaptive approach, the $\bar{\varphi}_A$ for the last, i.e., the 10^{th} cohort is about 7.96. By utilising the accumulating data, improved dose regimens are administered to successive cohorts in an adaptive trial which results in lower values of φ_A.

Furthermore, in Scenario *7*, the average stopping cohort number (ACN) according to SR2 is 3.40 and the average value of φ_A when the trial is stopped is about 7.89. This shows that the adaptive approach also enables possible early termination of the trial without a significant decrease in efficiency, as compared to the non-adaptive approach in which the entire cohort of 200 patients is used.

The above two features contribute in improving the ethical standards of the adaptive approach as compared to the non-adaptive method and lead, on average, to a better recommendation of dose regimen. This is more evident in the case of larger cohort size. Figure 6.44 shows that for Scenario 7 even the upper quartile of φ_A is below the value of φ_A achieved in Scenario 8.

Although the ethical standards of the adaptive procedure are superior in this case, the non-adaptive approach, on the strength of superior parameters' estimates, has an edge in determination of the recommended dose regimen. As can be seen from the table, the non-adaptive procedure is able to exactly determine the optimal dose regimen, $\boldsymbol{D}^*_{true}$. Even the adaptive approach is fairly successful in recommending a dose regimen which is reasonably close to $\boldsymbol{D}^*_{true}$.

Next, we take initial values significantly farther from the true values.

Initial Values Far-Off from the True Values

The vector of initial values is now taken to be $\mathbf{\Psi}_{oc} = (2.60, .50, 51, .3, .3, .3, .3)^T$. These initial values have a stronger deviation from the true values of the parameters. As before, we repeat the original study using these initial values and taking four pairs of the cohort size and number of cohorts, (c, C): *9.* (5, 10), *10.* (50,1), *11.* (20, 10) and *12.* (200,1). Table 6.33 presents the data related to these four scenarios and Figure 6.44 plots the distribution of φ_A.

The parameter estimates' in Scenarios *10* and *12*, i.e., for non-adaptive approach, do not seem to be affected by the stronger deviation of the initial values from the true values. In fact, the bias and CV for the estimated parameters, for the non-adaptive approach in Scenarios *2*, *6*, *10* and Scenarios *4*, *8*, *12* are quite similar. This shows that from the parameter estimates point of view, the non-adaptive approach has a certain degree of robustness against the choice of initial values.

The dose regimen administered to the first cohort is highly suboptimal with $\varphi_A = 31.03$. This means that the first cohort in the adaptive method and all the patients in the non-adaptive scheme are exposed to such dose regimens.

However, the non-adaptive approach is not without a major drawback. As can be seen in the table, the φ_A value for Scenarios *10* and *12* is about 31.03, which is quite large. Also, based on the initial values, the recommended dose regimen for these two scenarios consists of the maximum permissible dose. And since the procedure is non-adaptive, all the subjects in the cohort are administered a highly inefficient dose regimen.

On the other hand, in Scenarios *9* and *11*, the $\bar{\varphi}_A$ values for the last cohort are not much larger than when the initial values are closer to the true values. As compared to the adaptive procedure in Scenario *11* for which $\bar{\varphi}_A$ for the last cohort is 8.21, the non-adaptive procedure administers a dose regimen to the entire cohort which is almost four times inefficient than the one administered by the adaptive procedure. As mentioned before, the adaptive procedure, by utilising accumulating data at interim stages is able to arrive closer to the optimum dose regimen already during the trial. Figure 6.44 which presents the distribution of φ_A over the course of the trial for these four scenarios shows that for scenarios *9* and *10* related to adaptive procedure protect patients from wrong dose regimens much more effectively than in Scenarios *10* and *12*.

Therefore, the adaptive procedure in this case is decidedly more ethical than the non-adaptive one. However, as mentioned previously, the parameter estimates are of a higher quality from the non-adaptive procedure. The recommended dose regimen,

	c	C	K_a	K_e	V	ω_1	ω_2	ω_3	σ^2	
$\bar{\widehat{\boldsymbol{\Psi}}}$	5	10	0.8291	0.1531	19.1731	0.0819	0.0968	0.1811	0.1046	
	50	1	0.8257	0.1543	16.5247	0.0899	0.0898	0.1026	0.1010	
	20	10	0.8233	0.1522	19.0746	0.0866	0.0991	0.1452	0.1040	
	200	1	0.8224	0.1545	16.4081	0.0897	0.0937	0.1006	0.1032	
$\widehat{\text{Bias}}(\widehat{\boldsymbol{\Psi}})$	5	10	-2.5	2.0	12.8	-18.1	-3.2	81.1	4.6	[**17.8**]
	50	1	-2.9	2.9	-2.8	-10.1	-10.2	2.6	1.0	[**4.6**]
p.c.	20	10	-3.1	1.4	12.2	-13.4	-0.9	45.2	4.0	[**11.5**]
	200	1	-3.3	3.0	-3.5	-10.3	-6.3	0.6	3.2	[**4.3**]
CV	5	10	9.1	6.3	20.2	70.3	25.1	79.0	26.7	[**33.8**]
	50	1	8.7	5.6	7.1	62.1	27.1	35.8	25.6	[**24.6**]
	20	10	4.4	3.2	12.3	34.2	12.1	25.2	13.2	[**15.0**]
	200	1	4.4	2.8	3.9	30.9	12.4	19.5	13.7	[**12.5**]
			d_1^*	d_2^*	d_3^*	d_4^*	d_5^*		$\bar{\varphi}_A$	
$\boldsymbol{D}^*_{true}$			140.00	90.00	90.00	100.00	90.00		7.1561	
$\boldsymbol{D}^*$ (cohort 1)			200.00	200.00	200.00	200.00	200.00		31.0280	
$\boldsymbol{D}^*$	5	10	156.90	101.18	104.90	104.48	104.90		9.7594	
	50	1	200.00	200.00	200.00	200.00	200.00		31.0280	
Last cohort	20	10	156.55	98.40	102.89	102.40	102.89		8.2128	
	200	1	200.00	200.00	200.00	200.00	200.00		31.0280	
$\boldsymbol{D}^*(\widehat{\boldsymbol{\Psi}})$	5	10	160.00	100.00	110.00	110.00	110.00		8.1225	
	50	1	140.00	90.00	90.00	100.00	90.00		7.1561	
Recommended	20	10	160.00	100.00	110.00	100.00	110.00		7.8896	
	200	1	140.00	90.00	90.00	90.00	90.00		7.1744	
			$\bar{C}_2$		$\bar{\varphi}_A(C_2)$		$\bar{C}_3$		$\bar{\varphi}_A(C_3)$	
ACN (AN)	5	10	5.95 (29.8)		11.5632		8.40 (42)		10.4853	
	50	1	1 (50)		31.0280		1 (50)		31.0280	
	20	10	5.24 (104.8)		10.1026		7.58 (151.6)		9.1337	
	200	1	1 (200)		31.0280		1 (200)		31.0280	

Table 6.33: Data related to the simulation studies for the four scenarios: *9.* (5, 10), *10.* (50,1), *11.* (20, 10) and *12.* (200, 1), when the initial values of the parameters, $\boldsymbol{\Psi}_{oc}$, are far-off from the true values.

which is a function of the final parameters' estimates is, therefore, closer to the true optimal dose regimen, $\boldsymbol{D}^*_{true}$. As can be seen from the table, the φ_A values for the recommended dose regimens corresponding to the adaptive procedure are higher than those associated with the non-adaptive procedure. Since the parameters' estimates and the recommended dose regimen for non-adaptive approach are observed to be of good quality in all the three cases of different initial values, it can be said that the non-adaptive method is much more robust against the choice of initial values than the adaptive one.

While the adaptive procedure is found to be more ethical, the non-adaptive one is more effective in achieving the trial's objectives of determination of the correct dose

regimen. However, the large φ_A value of 31.02 for the non-adaptive procedure, which is more than 4 times the optimal value, could be unacceptable to be experienced by a single large cohort of patients. On the other hand, efficient determination of the best dose regimen for the population, could be of paramount importance. A middle path between the adaptive and the non-adaptive procedure could be to follow a two-stage design which incorporates the ethics of the adaptive procedure and the precision of the non-adaptive procedure. This is explored later.

To summarise, the non-adaptive approach is superior to the adaptive procedure in estimation of the parameters and seems to be robust against deviation of the initial values from the true values. Good parameters' estimates also lead to recommendation of the correct dose regimen for the population. However, since the first dose regimen is administered to the subjects using the initial values of the parameters as inputs to the ED algorithm, the dose regimen administered to the cohort in a non-adaptive setting can be significantly unethical resulting in excessive over- and under-exposure to the subjects of the trial. Thus, the non-adaptive procedure can be significantly unethical. On the other hand, the parameters' estimates in the adaptive approach, are relatively more affected by the choice of initial values. However, even for large deviation of the initial values, the nature of the adaptive approach makes it possible to apply the optimal dose regimen during the course of the trial. The advantage of this is at least double-fold: first of all, patients taking part in the trial are more likely to be treated with the right dose regimen; second, some information on the response at these doses can already be gathered at the trial. This can be very useful for further investigations of the drug's efficacy and toxicity.

Therefore, if initial values are known to be very close to the true values, non-adaptive approach is superior. In case there is a wide difference between them, the non-adaptive approach, although still effective, may be significantly unethical as compared to the adaptive method.

The difference between the initial values and the true values will generally be not known in advance. A practical approach could be, therefore, to divide a given number of subjects into a small number of large cohorts in the trial. This is consistent with the conclusion we made in Section 6.3.1. Motivated by this observation, we next evaluate the performance of a two-stage design, i.e., an adaptive design in which all subjects are divided between only two cohorts.

A Two-Stage Design

For a two-stage design, $C = 2$, we denote the cohort sizes by c_1 and c_2 so that $c_1 + c_2 = c$. For the total number of available subjects equal to 50 and 200, we

consider the following six scenarios of pairs of (c_1, c_2): *1.* (5, 45), *2.* (10, 40), *3.* (25, 25), *4.* (10, 190), *5.* (40, 160) and *6.* (100, 100). We expect that such designs should benefit from both approaches: the adaptive and non-adaptive. The first cohort is a learning one and we would like it to be small. The second cohort will use the information gathered in the first stage and so the study for this one should be better informed than in a non-adaptive case. Setting up these scenarios we investigate what division of the cohort sizes is most beneficial.

The original simulation study described in Section 6.2.1 is run for these six pairs of cohort sizes and the results are presented in Table 6.34. The initial values for these scenarios were taken to be $\mathbf{\Psi}_{oc} = (2.60, .50, 51, .3, .3, .3, .3)^T$, i.e., the initial values are assumed to be far-off from the true values and the true parameter values were $\mathbf{\Psi}_{true} = (.85, .15, 17, .1, .1, .1, .1)^T$. The maximum dose level was assumed to be $d_{max} = 200$.

As was seen previously, since the initial values are far from the true values, the dose regimen administered to the first cohort, that is, at stage 1 is highly suboptimal with a large value of φ_A. In all the six scenarios, the patients in the first cohort are exposed to this dose regimen.

From the data presented in the table, it can be seen that, indeed, the two-stage design combines to a certain degree the advantages of both: the non-adaptive approach and the adaptive one. Let us consider, for example, the pairs $(c, C) = (200, 1)$ and $(c, C) = (20, 10)$ in Table 6.33 and the pair $(c_1, c_2) = (10, 190)$ in Table 6.34. The quality of the estimates is best for the pair (200, 1) and it is marginally lower for the pair (10, 190). However, for the pair (200, 1), the average value of φ_A for the 200 subjects in the cohort is 31.03 which is quite high. For the pair (10, 190), the average φ_A value for the first cohort of 10 subjects is the same at 31.03, but for the 190 subjects in the second cohort it is 8.68. Thus, for a marginal loss in the quality of the estimates, substantial gains can be made in enhancing the ethical standards of the trial. Similarly, the quality of the estimates for the pair (20, 10) is slightly lower than that for the pair $(c_1, c_2) = (200, 1)$ but only the first cohort of 20 subjects has the average φ_A value as high as 31.03. As compared to (20, 10), the variability in the estimates is higher for the pair (10, 190), although the bias is lower for the latter case.

This shows that a two-stage design, by exposing only a small number of subjects in the first cohort to suboptimal dose regimen, is able to efficiently determine parameters' estimates and the recommended dose regimen on the strength of the large size of the second cohort. This feature makes the two-stage design more ethical than the non-adaptive approach and more efficient from the estimation point of view as well as simpler to implement than the adaptive approach which has a large number of

	c_1	c_2	K_a	K_e	V	ω_1	ω_2	ω_3	σ^2	
	5	45	0.8688	0.1536	15.6565	0.0897	0.0996	0.1315	0.1034	
	10	40	0.8406	0.1480	15.5650	0.0881	0.1052	0.1370	0.1043	
$\bar{\widehat{\boldsymbol{\Psi}}}$	25	25	0.8507	0.1402	14.2366	0.0834	0.1169	0.1431	0.1036	
	10	190	0.8538	0.1537	16.0727	0.0933	0.0956	0.1120	0.1038	
	40	160	0.8327	0.1484	15.4728	0.0841	0.1060	0.1275	0.1047	
	100	100	0.8470	0.1408	14.1543	0.0845	0.1163	0.1392	0.1053	
	5	45	2.2	2.4	-7.9	-10.3	-0.4	31.5	3.4	[**8.3**]
	10	40	-1.1	-1.3	-8.4	-11.9	5.2	37.0	4.3	[**9.9**]
$\widehat{\text{Bias}}(\widehat{\boldsymbol{\Psi}})$	25	25	0.1	-6.5	-16.3	-16.6	16.9	43.1	3.6	[**14.7**]
p.c.	10	190	0.4	2.4	-5.5	-6.7	-4.4	12.0	3.8	[**5.0**]
	40	160	-2.0	-1.1	-9.0	-15.9	6.0	27.5	4.7	[**9.5**]
	100	100	-0.4	-6.1	-16.7	-15.5	16.3	39.2	5.3	[**14.2**]
	5	45	88.2	11.9	38.5	76.6	33.5	72.6	25.6	[**49.5**]
	10	40	11.1	6.9	10.6	120.4	27.3	75.5	28.3	[**40.0**]
CV	25	25	11.6	7.6	13.2	70.2	27.1	42.6	26.4	[**28.4**]
	10	190	61.5	8.8	13.7	73.9	16.3	30.0	13.9	[**31.1**]
	40	160	5.1	3.4	6.2	35.4	13.7	22.8	13.9	[**14.4**]
	100	100	8.6	5.0	9.5	36.9	13.5	24.4	14.6	[**16.1**]
			d_1^*	d_2^*	d_3^*	d_4^*	d_5^*		$\bar{\varphi}_A$	
$\boldsymbol{D}^*_{true}$			140.00	90.00	90.00	100.00	90.00		7.1561	
$\boldsymbol{D}^*$ (cohort 1)			200.00	200.00	200.00	200.00	200.00		31.0280	
	5	45	141.05	89.45	93.71	92.90	93.70		10.1707	
	10	40	142.52	89.14	93.67	92.96	93.65		8.5599	
$\boldsymbol{D}^*$	25	25	141.38	89.23	93.30	93.20	93.30		7.7924	
cohort 2	10	190	142.48	89.40	94.10	93.30	94.11		8.6754	
	40	160	140.85	88.98	92.99	92.54	92.99		7.4371	
	100	100	141.10	88.84	93.33	92.94	93.33		7.3418	
	5	45	130.00	90.00	90.00	90.00	90.00		7.3132	
	10	40	130.00	80.00	90.00	80.00	90.00		7.6863	
$\boldsymbol{D}^*(\widehat{\boldsymbol{\Psi}})$	25	25	120.00	70.00	70.00	80.00	70.00		10.8063	
Recommended	10	190	140.00	90.00	90.00	90.00	90.00		7.1744	
	40	160	130.00	80.00	90.00	80.00	90.00		7.6863	
	100	100	120.00	70.00	70.00	80.00	70.00		10.8063	

Table 6.34: Data related to the simulation studies for the six pairs of values of (c_1, c_2): *1.* (5, 45), *2.* (10, 40), *3.* (25, 25), *4.* (10, 190), *5.* (40, 160) and *6.* (100, 100), when the initial values of the parameters, $\boldsymbol{\Psi}_{oc}$, are far-off from the true values.

cohorts.

An important question that arises in planning the two-stage design is how should the total number of subjects be divided between the first and the second cohort. This

can be answered by examining the data in Table 6.34 along with Figure 6.45 which presents the distribution of φ_A for the second cohort in the six scenarios.

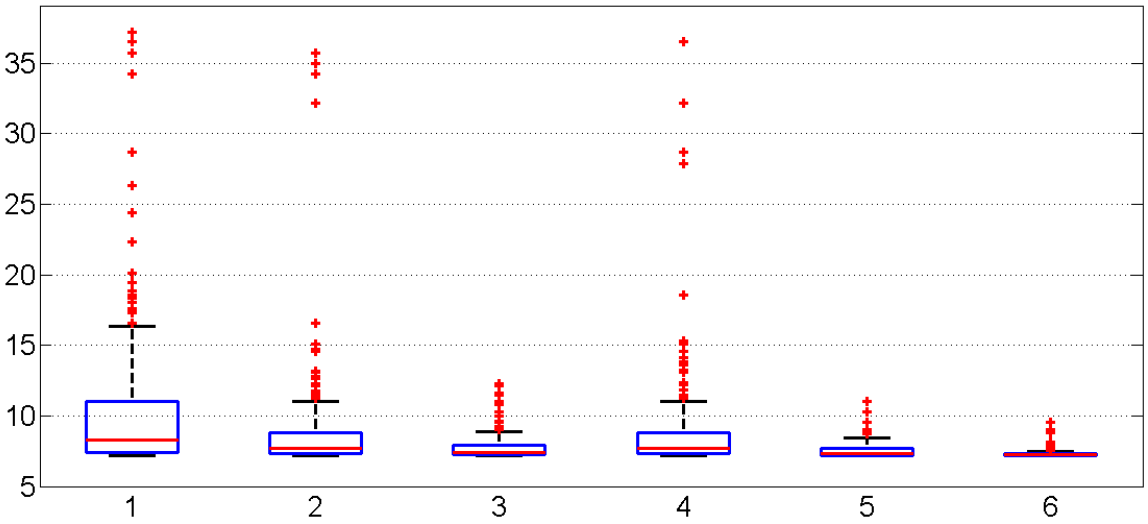


Figure 6.45: Distribution of φ_A for the second cohort in the six scenarios consisting of different pairs of (c_1, c_2): *1.* (5, 45), *2.* (10, 40), *3.* (25, 25), *4.* (10, 190), *5.* (40, 160) and *6.* (100, 100).

The advantages of having the first cohort of a smaller size are that fewer subjects are administered suboptimal dose regimens and the large size of the second cohort reduces the bias in the estimates. Good quality of the parameters' estimates translates to recommendation of a dose regimens which is closer to the optimal dose regimen. As can be seen from the table, the recommended dose regimens and the associated φ_A values are closest to their optimal values for Scenarios *1* and *4* in which the sizes of the first cohort have been taken to be small.

On the other hand, a larger first cohort enables better estimation of the parameters in the first stage and thus, the average φ_A values for the second cohort are significantly lower in Scenarios *3* and *6*. Also, it can be seen in the figure that when the two cohorts are of same sizes, the spread in the distribution of φ_A for the second cohort is much lower. However, this is at the cost of exposing an equally large number of subjects to suboptimal dose regimen in the first stage. Furthermore, the variability in the estimates derived from the second cohort is lower when the two cohorts are of identical sizes.

However, in recommendation of a dose regimen for the population, the dose regimen determined by the pair $(10, 190)$ is closest to the optimal dose regimen. Therefore, considering the higher ethical standards and correct recommendation of the dose regimen associated with this pair, it is better to have a small first cohort and a larger second cohort.

In conclusion, the non-adaptive procedure works best when the initial values of the

parameters are close to their respective true values. However, for initial values far-off from the true values, the ethical standards of the non-adaptive procedure can be significantly inferior to the adaptive method. The ethical standards associated with a two-stage design may be superior than that of a non-adaptive design. Also, the two-stage design could be easier to implement than an adaptive design which has a large number of cohorts, without compromising on the precision of the estimates and the recommended dose regimen. Choice of a small first cohort and a much larger second cohort was found to be ethical as well as effective in achieving the trial's objectives.

6.5 Conclusions

In this chapter we were able to demonstrate that the ED algorithm can find the optimal dose regimens without assuming that the PK parameters are precisely known, although, initial values are needed to initiate the procedure. The features of the proposed method to analyse interim data and decide about the further course of the trial make it an adaptive design.

In Section 6.2, we presented a simulation study to evaluate the adaptive procedure. The stopping rules make it possible to terminate the trial without making use of the maximum number of cohorts (C). SR2 can be used for drugs which have a broader therapeutic index while SR3 could be more suited to drugs for which it is critical to maintain the concentration in a narrow range. This can lead to more savings in cost and time and make the trial more ethical.

The example we discussed in this chapter had equal cohort sizes (c) for every cohort and the same number of blood samples per subject (m). However, it is not necessary to impose these constraints. Indeed, as mentioned before, these variables, along with the number of cohorts (C) are design variables which can be optimised for every cohort in the trial. The optimisation would further improve the trial's cost effectiveness.

We have used the φ_A criterion to find the efficient dose regimens for each cohort. Other efficiency criteria introduced in Chapter 4 such as φ_G or φ_C could be used if the loading and the maintenance doses are to be assigned different importance in the optimisation process. Optimality criteria for finding the blood sampling times other than the D-optimality criterion could also be explored to find the sampling times for each cohort. In particular, c-optimality for minimising the variance of the estimate of AUC would also be an appropriate criterion.

The adaptive procedure is successful in estimation of the PK parameters and determination of efficient dose regimens. Section 6.3 revealed some interesting relation-

ships between some of the inputs and the outcomes from the algorithm. Given a fixed number of volunteers who can be enrolled in the trial, we found that it is statistically better to form larger cohorts than have a greater number of cohorts. However, for drugs whose safety profile has not been fully established, having larger cohorts exposes more volunteers to the risk of toxicity, especially at the start of the trial. A possible remedy could be to start with small cohort sizes and make them larger during the course of the trial. Collection of more blood samples per subject improves the bias and the variability in the parameters' distributions to some extent. However, the gains are capped and we found that taking excessive number of blood samples may be an unnecessary cost and inconvenience to the volunteers.

We have used the LME algorithm of Lindstrom and Bates (1990) for estimation of parameters in this chapter. Other methods of estimation can be explored to see the effect on estimated parameters' bias and sensitivity to initial values. As pointed out by Mentré and Lavielle (2008), some of the other methods of estimation such as FO, FOCE and Laplace approximation (discussed in Section 5.3) also suffer from the problem of sensitivity to the initial values. Since the efficient dose regimen computed by the ED algorithm is a function of the estimated PK parameters, any gains in bias reduction are accrued in enhancement of the efficiency of the recommended dose regimen. Nevertheless, this simulation study establishes the ability of the proposed method for simultaneous estimation of the parameters and optimisation of dose regimen.

The PK model we used in the example did not include patient covariates. As discussed in Section 2.3, inclusion of covariates in the PK model can mitigate residual variability and improve the model fitness. Although in this thesis we have not carried out a simulation study with a PK model that contains covariates, we plan to take it up in the future. This is discussed in some detail in Section 7.1.4.

The example in this chapter specifically dealt with the case of maintenance of serum concentration around a pre-determined target level. An extension of the ED algorithm - optimisation for a therapeutic range - can also be accomplished in this adaptive setting.

For optimisation of combination ratios and fixed dose combinations, the problem can be solved using this adaptive procedure, although there are some challenges which need to be dealt with. Since two PK models will be involved, the computation of design points need to be informative for both the models. For example, Almohaimeed and Donev (2014) present methods for design optimisation in combination studies for a variety of statistical models and error distributions. Also, estimation of the parameters of the two PK models has to be done simultaneously. There is always a possibility of two drugs interacting with each other and this interaction can significantly affect

their PK and PD. The interaction needs to be considered carefully as it might lead to revision of the respective target concentrations of the two drugs and some additional constraints on the dose regimens. This is briefly discussed in Section 8.2.

In the next chapter, we discuss some other potential applications of the ED algorithm.

Chapter 7

Potential Applications of the ED Algorithm

In this chapter, we present two other potential applications of the ED algorithm. The first application illustrates the possible use of the algorithm for dose individualisation.

In the second application, we use a pharmacodynamic (PD) model to optimise the dose regimen for anti-infective drugs, where instead of aiming for a target concentration, we optimise the dose regimen for a target reduction in the microbial load. This illustrates the use of the algorithm for achieving a PD target.

7.1 The ED Algorithm as a Dose Individualisation Tool

In Section 3.1.2, we introduced the concept of dose individualisation and briefly discussed the capabilities of dose individualisation software for estimation of a patient's parameters and dose adjustment. In this section, we show that the ED algorithm can also be considered as one of the tools for individualisation of the dose regimen. We take a step further to not just optimise the dose regimen but also quantify the compliance with the target concentration or a range of concentrations for every individual. We explain this in the procedure outlined below.

7.1.1 Procedure

We assume that the mechanistic model followed by the drug's concentration is known along with an estimate of the population PK parameters. Let the parameters be

contained in the vector $\boldsymbol{\Psi} = (\boldsymbol{\beta}^T, \boldsymbol{\omega}^T, \sigma^2)^T$. These parameters are for the general population and are *a priori* available from previous trials. In standard dose individualisation software, this information is available for a number of drugs. The problem is to find the best dose regimen for a patient which launches the drug quickly into its steady state and maintains it in the desired target range in that subject. Let c be the number of subjects for whom the dose regimen is to be individualised. We assume that $c > 1$. These c subjects form a group in the context of this problem.

We use the known population parameter estimates of $\boldsymbol{\Psi}$ as initial values in our procedure and we denote them by $\boldsymbol{\Psi}_o = (\boldsymbol{\beta}_o^T, \boldsymbol{\omega}_o^T, \sigma_o^2)^T$. $\widehat{\boldsymbol{\Psi}} = (\widehat{\boldsymbol{\beta}}^T, \widehat{\boldsymbol{\omega}}^T, \widehat{\sigma}^2)^T$ denotes the vector of estimated mean parameters for the specific group of c subjects under study. $\widehat{\boldsymbol{\Psi}}$ has to be computed from these subjects.

The ED algorithm can be used here through the following steps:

- Based on $\boldsymbol{\beta}_o$, the optimum dose vector $\boldsymbol{D}_o = (d_{o1}, ..., d_{on})^T$ is computed using the ED algorithm and everyone in the group of c subjects is administered the first $n_o (< n)$ doses from $\boldsymbol{D}_o$, i.e. the doses $d_{o1}, ..., d_{on_o}$.

- Based on the initial population parameters' values $\boldsymbol{\Psi}_o$, we compute D-optimal sampling time points for collection of m blood samples from the subjects in this group. As in Chapter 6, we use PopED for this.

- Based on the collected $c \times m$ blood samples, we estimate the parameters $\boldsymbol{\beta}$ of the cohort. We accomplish this with the help of the LME procedure of Lindstrom and Bates (1990) which also enables estimation of the individual vector of the random effects $\boldsymbol{b}_i$. The procedure is available in the MATLAB® module *nlmefit*.

- The estimated vectors of fixed effects $\widehat{\boldsymbol{\beta}}$ and random effects $\widehat{\boldsymbol{b}}_i$ are applied back into the original stage II model to get vectors of individual parameters for each patient. For example, if the stage II model was defined as in Equation (6.2), the estimated individual parameters would be :

$$\begin{pmatrix} \widehat{K}_{a_i} \\ \widehat{K}_{e_i} \\ \widehat{V}_i \end{pmatrix} = \begin{pmatrix} \widehat{K}_a \exp(\widehat{b}_{1i}) \\ \widehat{K}_e \exp(\widehat{b}_{2i}) \\ \widehat{V} \exp(\widehat{b}_{3i}) \end{pmatrix} \tag{7.1}$$

- Using the estimated individual PK parameters, the ED algorithm is re-run to obtain the updated dose vectors D_i^* for each of the c patients. These c dose vectors will have the initial n_o doses common, but the $n - n_o$ subsequent doses will be individualised to the patients. For the ED algorithm to compute these $n - n_o$ doses, the first n_o dose levels are restricted to the administered doses so that the algorithm optimises the remaining $n - n_o$ doses. Ideally, n_o should be

as small as possible so that the doses are individualised to the subjects as soon as possible.

7.1.2 Example

We consider a group of size $c = 10$ subjects. For every subject in this group, individualised dose regimens are to be determined for a drug which is known to follow a one compartmental model described in Equations (6.1), (6.2) and (7.2). The target concentration is $C_{tgt} = 5$ mg/L which is desired to be maintained in every patient in the group for the treatment duration $T = 60$ h. We take $n = 5$ and the dosing time points as: $\boldsymbol{t} = (0, 12, 24, 36, 48)^T$. The number of PK blood samples per subject is taken to be $m = 3$. The maximum dose which can be administered at an occasion is $d_{max} = 300$ mg.

We describe the simulation study through the following steps:

Step 1 - Definition of the model and the model parameters $\boldsymbol{\Psi}$

As mentioned above, the PK model followed by the drug is assumed to be known along with an estimate of the population model parameters $\boldsymbol{\Psi}$. This is a standard assumption in all dose individualisation software.

For this example, we assume that the one compartment first order absorption model explains the concentration-time relationship of the drug. As before, let y_{ij} be the j^{th} sample from the i^{th} subject at time T_j, where $i = 1, ..., c$ and $j = 1, ..., m$.

We use the model given in Equation (6.3):

$$y_{ij} = \sum_{k=1}^{n_o} I_{\{T_j \geq t_k\}} \frac{d_k K_a e^{b_{1i}}}{V e^{b_{3i}}(K_a e^{b_{1i}} - K_e e^{b_{2i}})} \left(e^{-K_e e^{b_{2i}}(T_j - t_k)} - e^{-K_a e^{b_{1i}}(T_j - t_k)} \right) e^{\epsilon_{ij}}, \quad (7.2)$$

where $\epsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$ and $\boldsymbol{b} \sim \mathcal{N}_3(\boldsymbol{0}, \boldsymbol{\Omega})$. The matrix $\boldsymbol{\Omega}$ is assumed to be diagonal.

The vector of parameters for this model is

$$\boldsymbol{\Psi} = (\boldsymbol{\beta}^T, \boldsymbol{\omega}^T, \sigma^2)^T = (K_a,\ K_e,\ V,\ \omega_1,\ \omega_2,\ \omega_3,\ \sigma^2)^T.$$

The values of the vector $\boldsymbol{\Psi}_o$ for the general population (known before the individualisation is started) are: $\boldsymbol{\Psi}_o = (.85, .15, 17, .1, .1, .1, .1)^T$. These values are the same as those used in Chapter 6.

Step 2 - Computation of $\boldsymbol{D}_o$ and optimal sampling points ξ_o based on $\boldsymbol{\Psi}_o$

The ED algorithm is used for the vector of parameters $\boldsymbol{\Psi}_o$ to find the optimal dose vector $\boldsymbol{D}_o$. We use the discretised version of the ED algorithm with $\kappa = 2$ to get the optimal dose vector as: $\boldsymbol{D}_o = (158,\ 126,\ 128,\ 126,\ 128)^T$ with $\varphi_A(\boldsymbol{\Delta}_o|\kappa = 2) = 17.81$.

We take $n_o = 1$, i.e., the first dose from $\boldsymbol{D}_o$ is administered at $t = 0$ to every subject in the group. The remaining $n - n_o = 4$ doses are to be individualised to the subjects, after estimating their PK parameters. So, each of the 10 subjects will be administered the dose $d_{o1} = 158$ mg at time $t_1 = 0$.

Using $\boldsymbol{\Psi}_o$ and d_{o1} and the model in Equation (7.2), D-optimal design ξ_o is obtained for collection of blood samples using PopED as it was done in Chapter 6. The design region for the PK sampling is chosen to be [.1, 10] h. The optimal design was found to be $\xi_o = \{.1, 3.66, 10\}$.

Figure 7.1 plots the D-optimal points ξ_o and the concentration-time curve generated by $\boldsymbol{\beta}_o$, $d_{o1} = 158$ in the interval [0, 12] h.

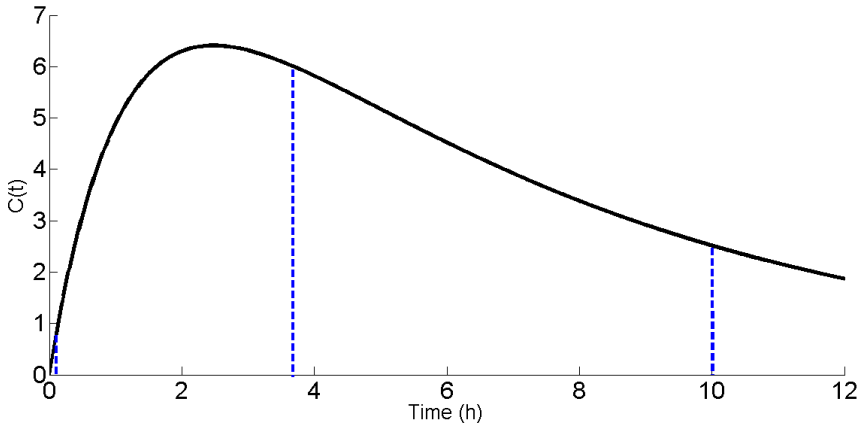


Figure 7.1: The concentration-time curve generated by the dose d_{o1} for the model defined in (7.2) for $\boldsymbol{b} = \boldsymbol{0}$, along with the $m = 3$ D-optimal sampling times ξ_o as computed by PopED.

Step 3 - Simulation of $c \times m$ blood concentrations at ξ_o

If this procedure was carried out in real life, $m = 3$ blood samples would have been collected at the time points ξ_o found in the previous step and the drug's concentrations would have been determined by assay.

For our study, we simulate these $c \times m = 30$ observations using the vector of parameters $\boldsymbol{\Psi}_o$, Equation (7.2) and the model assumptions. Taking $n_o = 1$ in Equation

(7.2), we simulate the observation y_{ij} at a given time $T_j \in \xi_o$ as follows:

$$y_{ij} = \frac{d_k K_a e^{r_1}}{V e^{r_3}(K_a e^{r_1} - K_e e^{r_2})} \left(e^{-K_e e^{r_2} T_j} - e^{-K_a e^{r_1} T_j}\right) \exp(r_4),$$

where $(r_1, r_2, r_3)^T$ and r_4 are random numbers drawn from $\mathcal{N}_3(\mathbf{0}, \mathbf{\Omega})$ and $\mathcal{N}(0, \sigma^2)$ distributions respectively using the *normrnd* function in MATLAB® . Thus, at each of the m sampling times contained in ξ_o, c observations of the concentration values are simulated.

A data file is created at this step whose first column represents the subject number, the second column represents the times of sampling and the third column represents the simulated observations. There would be m entries against every subject number.

Step 4 - Estimation of $\mathbf{\Psi}$ and $\boldsymbol{b}_i$

The data file generated in Step 3 is then fed into the MATLAB® module *nlmefit* for the computation of $\widehat{\mathbf{\Psi}}$. Other function inputs are the model specified in Equation (7.2), the initial values and the selection of a method for estimation. As mentioned in Chapter 5, this module offers four methods from which we choose the LME method as it enables computation of not just $\widehat{\mathbf{\Psi}}$ but also $\widehat{\boldsymbol{b}}_i$, $i = 1, ..., c$.

Step 5 - Computation of $\widehat{\boldsymbol{\beta}}_i$ and $\boldsymbol{D}_i^*$

The vectors $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{b}}_i$ found in the previous step are then applied back into the Stage II model to get the subject specific PK parameters $\hat{\boldsymbol{\beta}}_i$. This is given in Equation (7.1). The c vectors of parameters $\widehat{\boldsymbol{\beta}}_i$ are then used as inputs to the ED algorithm. The algorithm is run c times, once for each subject, to get the individualised dose regimens for the subjects. In each run of the algorithm, the first dose level is restricted to the optimum level found in Step 2. Thus, corresponding to every subject, we have the individualised dose regimen.

The MATLAB® code for implementing this methodology is given in Appendix D.5. The ED algorithm given in Appendix D.1 is used as a submodule in this code. The submodule is called to determine the optimum initial dose d_1^* and then, after determination of individual PK parameters, c times to compute the individualised dose regimens.

7.1.3 Results

The above steps were carried out and the output is summarised in Table 7.1 and Figures 7.2 and 7.3. Subjects' PK parameters were obtained by applying formula (7.1)

with the estimated mean PK parameters for the group and the estimated individual random effects. The φ_A values for each subject are computed as described in Section 4.2. Subject-specific parameters $\widehat{\boldsymbol{\beta}}_i$ are used for this computation.

The second optimal dose d_2^* is considerably higher than d_1^* for subjects 6 and 10 while it is lower for all other subjects. The doses d_i^*, $i = 2, ..., 5$ which are administered to subjects 6 and 10 are almost of the maximum size which is 300 mg. The reason for this is that the estimated volume parameter, $\widehat{V}_i$, for these two subjects is very large: 41.6 L and 35.3 L. As can be seen in Equation (7.2), the concentration of the drug, $C(t)$, is inversely proportional this parameter. Also, it seems from Figure 7.2 that subjects 6 and 10 are highly under-dosed. Therefore, to maintain the concentration profile around C_{tgt}, the ED algorithm computes large doses for these two subjects.

Similarly, subjects 2, 4 and 9 seem to be significantly over-dosed in Figure 7.2. As a result, the subsequent doses for these two subjects have been substantially reduced by the algorithm.

The reason subject 6 appears under-dosed even after the optimised dose regimen is administered is because of the upper constraint of $d_{max} = 300$ mg on the dose sizes. The revised doses of subject 6 are all equal to 300 mg which are still inadequate for the subject, as can be seen from the figure.

The φ_A values for the subjects depend on their PK parameters and the initial dose administered. For example, a higher value of the elimination rate constant K_e generally results in greater under-exposure between the dosing time points which is reflected in an increased φ_A, as in the fitted concentration profiles of subjects 2 and 5. An inappropriate initial dose can also result in greater deviations from C_{tgt} which gets reflected in increased φ_A, as can be seen for subjects 6 and 10. The deviations from C_{tgt} are measured by the Δ-functions defined in Section 4.1.

It may appear counter-intuitive at first glance that the φ_A values for some subjects are even lower than the value $\varphi_A(\boldsymbol{\Delta}_o|\kappa = 2) = 17.81$ corresponding to the initial optimum dose regimen $\boldsymbol{D}_o$ derived for the general population. However, it must be remembered that the PK parameters used to compute the values for the subjects differ from the population parameters. The φ_A value computed by the ED algorithm is conditioned on the given value of the PK parameters $\boldsymbol{\beta}$. Thus, 17.81 is the lowest value of $\varphi_A(\boldsymbol{\Delta}_o|\kappa = 2)$ when the PK parameters are $\boldsymbol{\beta}_o$.

The larger φ_A values for the other subjects are due to the initial under- and over-dosing experienced by these subjects, which contributes to the deviations from C_{tgt}.

It can be observed from Figure 7.3 that there is a good compliance with C_{tgt} in every subject from the second dose onwards.

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Subject	Simulated concentrations			$\widehat{\boldsymbol{b}}_i$			$\widehat{\boldsymbol{\beta}}_i$			φ_A-efficient dose regimen					
No	$T_1^*=.1$	$T_2^*=3.66$	$T_3^*=10$	$\widehat{b}_{1i}$	$\widehat{b}_{2i}$	$\widehat{b}_{3i}$	$\widehat{K}_{a_i}$	$\widehat{K}_{e_i}$	$\widehat{V}_i$	d_{o1}	d_2^*	d_3^*	d_4^*	d_5^*	$\varphi_A(\boldsymbol{\Delta}^*\|\kappa=2)$
1	0.65	7.62	1.35	-0.06	.19	-.03	0.83	0.20	17.02	158	154	154	154	154	21.95
2	1.58	11.55	2.10	0.16	0.19	-0.51	1.04	0.20	10.57	158	86	94	92	92	24.60
3	1.45	4.99	2.84	.31	-0.04	-0.15	1.21	0.16	15.20	158	110	114	114	114	20.33
4	0.88	6.50	5.14	-0.11	-0.24	-0.26	0.79	0.13	13.56	158	80	90	90	90	16.09
5	1.84	6.03	1.42	0.41	0.27	-0.25	1.34	0.21	13.67	158	120	124	122	124	26.28
6	0.17	2.04	1.15	-0.42	-0.04	0.86	0.58	0.16	41.58	158	300	300	300	300	21.84
7	0.77	5.79	3.18	-0.06	-0.12	-0.08	0.84	0.14	16.21	158	114	118	118	118	17.17
8	0.61	5.19	3.14	-0.15	-0.15	0.01	0.76	0.14	17.82	158	128	128	128	128	16.23
9	0.91	7.50	4.44	-0.09	-0.18	-0.29	0.81	0.14	13.17	158	82	92	92	92	17.00
10	0.38	2.12	0.92	0.01	0.11	0.70	0.89	0.18	35.32	158	300	300	298	298	24.56
Initial parameters $\boldsymbol{\beta}_o$ and $\boldsymbol{D}_o$ for the general population							0.85	0.15	17.00	158	126	128	126	128	17.81
$\widehat{\boldsymbol{\beta}}$ and $\boldsymbol{D}^*$ for the group of $c=10$ subjects							0.89	0.16	17.59	166	136	138	136	138	19.04

Table 7.1: The data pertaining to the ten subjects whose concentrations were simulated at the times given in ξ_o. Columns 2 - 4 give the simulated concentrations. Columns 5 - 7 are the estimated random effects, obtained from the LME method. As explained before, columns 8 - 10 are obtained by multiplying the exponential of the estimated random effects with the estimated mean PK parameters of the group of c patients. Columns 11 - 15 give the individualised dose regimens and column 16 shows the efficiency measures computed by the ED algorithm for each subject using the parameter estimates given in columns 8 - 10.

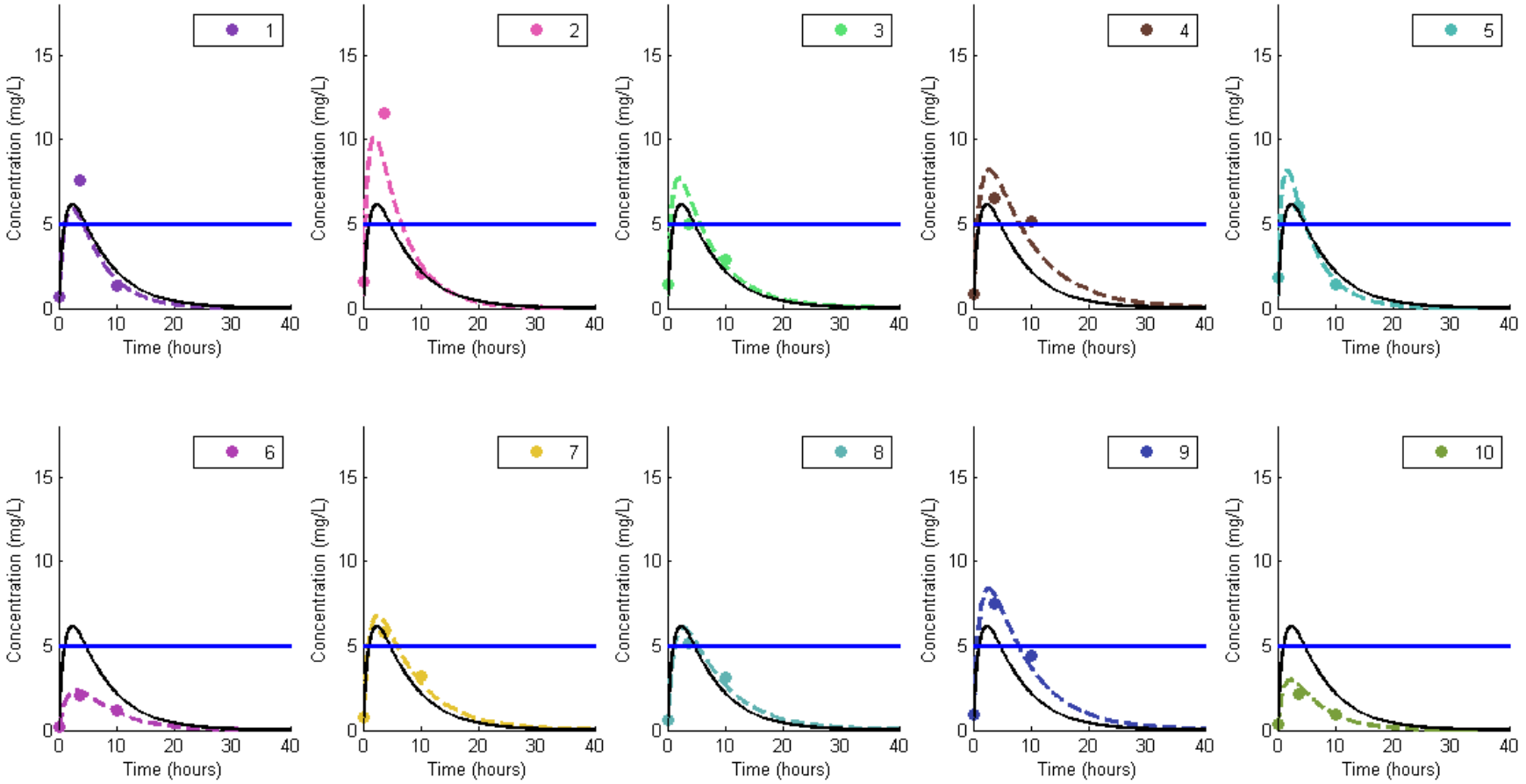


Figure 7.2: The solid curve represents the population concentration profile generated by the PK parameter values $\boldsymbol{\beta}_o$ and $d_{o1} = 158$ mg. The dotted curves represent the concentration profiles fitted for the $c = 10$ subjects on the basis of $m = 3$ blood samples collected at times ξ_o. The line parallel to the time axis represents the target concentration, C_{tgt}. Subjects 2 and 9 appear to be overdosed while subjects 6 and 10 appear to be underdosed. The purpose of our methodology is to correct this in the subsequent doses.

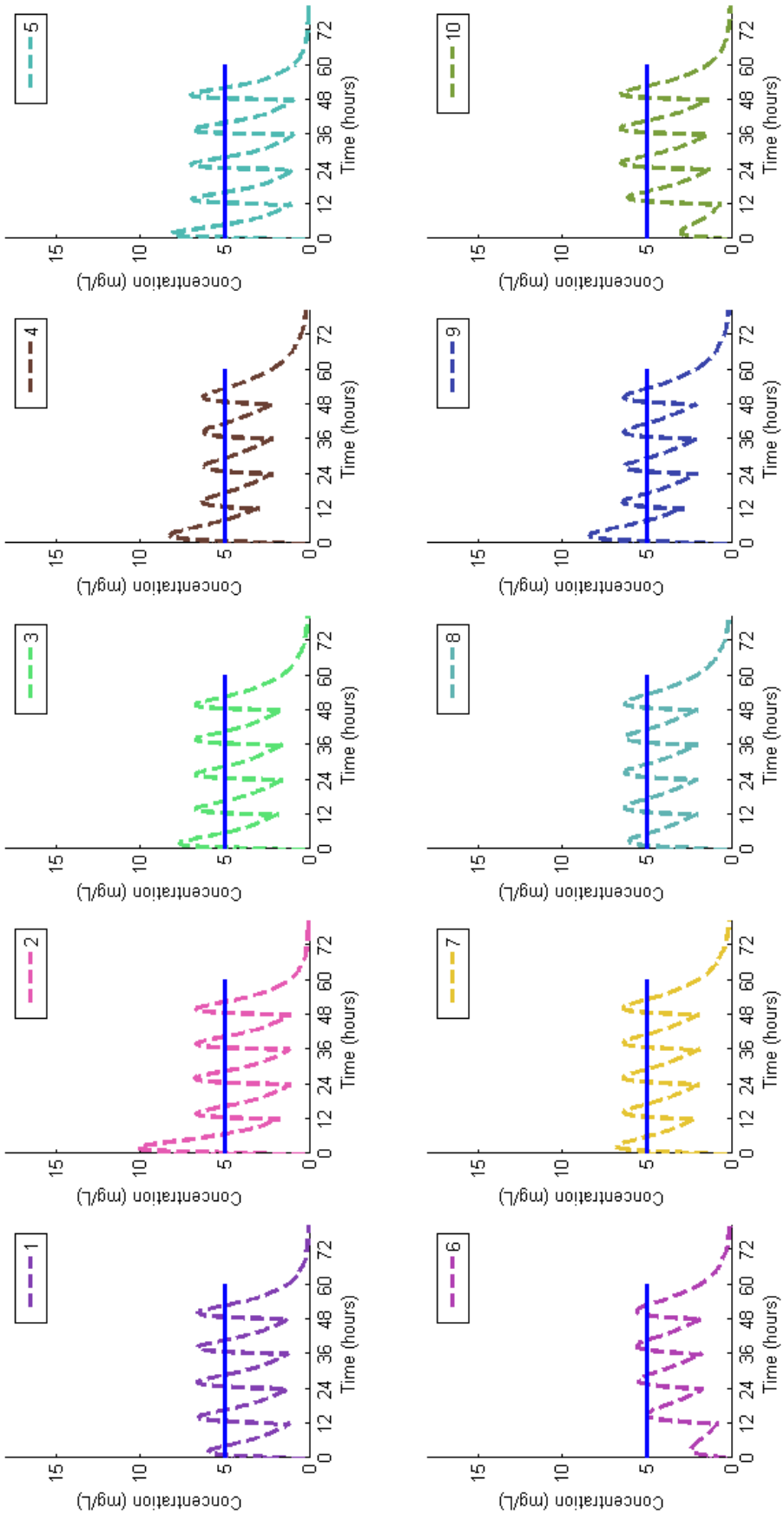


Figure 7.3: The curves represent the concentration profiles for the $c = 10$ subjects when administered the dose regimen $\boldsymbol{D}_i^*$, $i = 1, ..., 10$. There is a good compliance with C_{tgt} from the second dose onwards in every subject. The subjects under- or overdosed in the initial dose are subsequently given the appropriate doses.

7.1.4 Covariates and the ED Algorithm

The idea of covariates was introduced in Section 2.3. Covariates such as bodyweight, height and age can be used in the PK model to account for the inter-individual variability, thus leading to a reduction in the residual variability.

For example, the parameter V could be related to the bodyweight of the individual and thus inclusion of bodyweight as a covariate in the PK model can lead to a reduction in inter-individual variability in V.

Consider the second stage model defined in Equation (6.2). Suppose the weight of the i^{th} subject is denoted by W_i, $i = 1, ..., c$. This equation can then be re-written as:

$$\begin{pmatrix} K_{a_i} \\ K_{e_i} \\ V_i \end{pmatrix} = \begin{pmatrix} K_a \exp(b_{1i}) \\ K_e \exp(b_{2i}) \\ V \exp(b_{3i})(W_i/W_{std})^{\alpha} \end{pmatrix},$$

where W_{std} is the standard weight of an adult which is often taken to be 70 Kg. For weights above and below the standard weight, the V parameter is scaled accordingly depending on the parameter α. The overall model in the standard notation can then be found from Equation 6.3 as:

$$y_{ij} = \sum_{k=1}^{n} I_{\{T_j \geq t_k\}} \frac{d_k K_a e^{b_{1i}} \left(e^{-K_e e^{b_{2i}}(T_j - t_k)} - e^{-K_a e^{b_{1i}}(T_j - t_k)} \right) e^{\epsilon_{ij}}}{V(W_i/W_{std})^{\alpha} e^{b_{3i}} (K_a e^{b_{1i}} - K_e e^{b_{2i}})}. \tag{7.3}$$

The parameters of interest for this model are $\mathbf{\Psi} = (K_a,\ K_e,\ V,\ \alpha,\ \omega_1,\ \omega_2,\ \omega_3,\ \sigma^2)^T$.

Inclusion of covariates can have favourable effects on the performances of the adaptive procedure described in Chapter 6 and the dose individualisation method described in this section. In the adaptive procedure, we compute the best dose regimen for a cohort based on the current estimate of the vector of parameters, $\widehat{\mathbf{\Psi}}$, using the ED algorithm. Because of the inter-individual variability, the concentration profiles for the subjects in the cohort will deviate from the mean concentration profile. Therefore, including covariates in the PK model will lead to decreased inter-individual variability which will further enhance the efficiency of the dose regimen for the individual subjects in that cohort.

For example, for the model given in (7.3), the ED algorithm will be run (according to step 4 in Section (6.1.2) for each subject in the cohort using the subject's weight W_i and the vector $\widehat{\mathbf{\Psi}}$ to find the efficient dose regimen. This will result in more personalised dose regimens in the cohort as compared to every subject getting the

same dose regimen. By the same argument, for the problem of dose individualisation discussed in this section, inclusion of covariates in the model will increase the efficiency of the individualised dose regimens for each of the subjects.

Generally, in real life situations, there are several biologically plausible covariates available for inclusion in the model. However, having an excessive number of covariates increases the model's complexity and make it expensive especially when the clinical data are sparse. Statistically significant covariates are selected in the model using procedures such as stepwise model building, Wählby et al. (2002) and hierarchical Bayes modelling, George and McCulloch (1993).

7.1.5 Conclusions

In this section, we presented a new method which can be considered for individualisation of dose regimens. If the treatment duration T is large, the procedure can be split into daily or weekly schedules. For example, if $T = 6$ weeks, this procedure can be applied six times to update the individualised dose regimens at the beginning of each week. This can be done by collecting PK samples at the beginning of each week and using the observed concentrations, find the updated PK parameters for each subject and the individualised dose regimens.

The methodology is not necessarily an alternative to the well developed procedures of therapeutic drug monitoring and target concentration intervention discussed in Section 3.1.2 but it can also play a supplemental role to these methodologies. We discuss these two separate scenarios and our method's merits and drawbacks below:

As an individualisation method in its own right

In the example, we demonstrated the use of the method to individualise the dose regimens to patients' physical characteristics. The method is amenable to inclusion of covariates in the PK model, as discussed in Section 7.1.4. Inclusion of covariates will lead to decreased residual variability for every subject and will result in even more accurate subject-specific dose regimens. In therapeutic drug monitoring, samples are drawn at designated intervals of time to update the next dose. The same approach can be followed for our methodology, with the additional advantage of the use of an optimal design for blood sampling. Whenever a new sample is collected for a subject, the PK parameters can be updated and the ED algorithm re-run to get the next set of optimal doses.

The main disadvantage of our methodology is that it requires a group of subjects for estimation of the individualised PK parameters. As discussed in Section 3.1.2,

standard dose individualisation software can individualise the loading dose and the maintenance doses for a single patient using the patient's covariates and the sampled concentration values. This disadvantage can limit the use of our method in a hospital setting where generally the individualisation is done for a single patient at a time.

However, for individualisation in a clinical trial setting, the method could be used as the new treatment is generally tested on a group of volunteers. In early phase trials, where the PK parameters are not precisely known and the toxicity limit has not been fully established, this methodology can be combined with the adaptive scheme introduced in Chapter 6. The method would then be to individualise the dose regimens to the subjects in every cohort and update the information on the mean PK parameters from the samples collected from the current cohort. This would continue until the stopping rules are applied. This is in contrast to the available individualisation software which work only for commercially available drugs and not for a test drug under study. Our individualisation method can work for any medicinal product for which an analytical or numerical model is known. The features of the ED algorithm, such as computation of the optimal ratio for combination drugs, the choice of dose discretisation, measuring the degree of compliance with C_{tgt} or $(C^{-}_{tgt}, C^{+}_{tgt})$ and having a choice of objective functions $\vartheta(\mathcal{R})$ which can be optimised, are applicable to this method as well. To the best of our knowledge, these features are currently not available in the commonly used individualisation software.

Supplemental to the standard individualisation software

Even if the complete methodology described above is not applied, a part of it can still be useful, when used in conjunction with a standard individualisation software. The software can find the individualised PK parameters for the patient and then the ED algorithm can be used to find the individualised dose regimen. As pointed above, the additional features of the ED algorithm should widen the scope of the computation of the individualised dose regimen.

As mentioned previously, inclusion of covariates in our method can enhance the degree of personalisation of the dose regimens. In the future, we intend to conduct simulation studies to understand the effect of covariates on the performance of our methodology.

7.2 The ED Algorithm for a Pharmacodynamic Target

This application illustrates a possible solution to the Type 2 problem proposed in Section 3.3. Here we optimise the objective function $\vartheta(\mathcal{R}|\theta, \boldsymbol{t})$ with respect to the dose vector $\boldsymbol{D}$, the treatment duration T and the number of doses n. This is done in the context of anti-infective drugs whose pharmacodynamic action is elimination of parasitical load from the body.

To illustrate the idea, we firstly describe the PK-PD model given in Czock and Keller (2007). It models the number of parasites or microbes in the body as a function of the concentration of an anti-infective drug. The usual unit for measuring the parasite population is the colony-forming unit (CFU) per mL, which is an approximate number of bacterial cells which have the ability to multiply under controlled conditions in 1 mL of the sample.

The most common mechanistic model of antimicrobial effects is given by the differential equation:

$$\frac{\mathrm{d}\, N(t)}{\mathrm{d}\, t} = K_{r_o} E_r(C(t))N(t) - K_{d_o} E_d(C(t))N(t), \tag{7.4}$$

where

$N(t)$ is the number of microbes at time t (assumed to be a differentiable function of time) (CFU/ml),

K_{r_o} is the replication-rate constant of the microbes (without drug) (h^{-1}),

K_{d_o} is the death-rate constant of the microbes (without drug) (h^{-1}),

$C(t)$ is the serum concentration of the drug at time t,

$E_r(C(t))$ is the fractional decrease in the replication rate,

$E_d(C(t))$ is the fractional increase in the death rate.

For inhibition of rate of replication of the microbes, we use the sigmoid E_{max} model, introduced in Section 2.2,

$$E_r(C(t)) = 1 - \frac{I_{max} C^H(t)}{IC_{50}^H + C^H(t)},$$

where I_{max} is the maximum inhibition, IC_{50} is the concentration which results in half of the maximum inhibition and the Hill coefficient H describes the sigmoidicity.

Similarly, the increase in the rate of death of the microbes can be expressed as

$$E_d(C(t)) = 1 + \frac{E_{max} C^H(t)}{EC_{50}^H + C^H(t)}, \tag{7.5}$$

where E_{max} is the maximum stimulation of the death rate, EC_{50} is the concentration which results in half of the maximum increase and the Hill coefficient H describes the sigmoidicity.

It is difficult to separate the effects of replication inhibition and death stimulation. A more practicable form of the model can be obtained by assuming constant replication rate of the microbes (i.e. ignoring the effect of replication inhibition), as explained below.

Enhanced-Death Constant-Replication Model

Assuming a constant replication rate, Equation (7.4) becomes

$$\frac{\mathrm{d}N(t)}{\mathrm{d}t} = K_{r_o} N(t) - K_{d_o} E_d(C(t)) N(t).$$

Letting $K_g = K_{d_o} - K_{r_o}$ and the maximum kill rate $K_{kmax} = K_{d_o} E_{max}$,

$$\frac{\mathrm{d}N(t)}{\mathrm{d}t} = K_g N(t) - K_{kmax} \frac{C^H(t)}{EC_{50}^H + C^H(t)} N(t). \tag{7.6}$$

This model has an explicit solution for the one compartment model with zero order absorption as shown in, for example, Lipsitch and Levin (1997).

Kill Rate Curve

Using Equation 7.5, the death rate $K_d(C(t))$ can be expressed as

$$K_d(C(t)) = K_{d_o} E_d(C(t)).$$

To describe the effect of the anti-microbial, it is more intuitive to use the kill rate, $K_k(C(t)) = K_d(C(t)) - K_{d_o}$ which works out to be

$$K_k(C(t)) = K_{kmax} \frac{C^H(t)}{EC_{50}^H + C^H(t)}. \tag{7.7}$$

The area under the curve given in Equation 7.7 denoted by $\mathrm{AUC_K}(T)$, represents the total number of parasites killed in the interval $[0, T]$ by the treatment drug.

7.2.1 Efficient Dose Regimens for a PD target

We now extend the theory of efficient dose regimens for the models described above. So far, dose regimens were optimised for a given target concentration. Now, we find efficient dose regimens for a PD target. A standard approach in designing anti-infective drugs is to target the worst-case scenario which is to assume a total parasite burden that has to be eliminated. For example, for the malarial parasite Plasmodium falciparum, the worst case scenario is a parasite load of 10^{12} or 12 $\log_{10}$ units at time zero in an adult, Simpson et al. (2000). The treatment goal is the reduction of 12 $\log_{10}$ units i.e., from a load of 10^{12}, reduction to 10^0. We consider the case of this parasite to illustrate our methodology.

In case of anti-infectives, there are usually some constraints for the concentration of the drug. The concentrations should not exceed a toxic level, say, C^+_{tgt} but should be above a threshold, say, C^-_{tgt} for inhibition of the growth of parasites. As described in Section 3.2, the treatment duration is denoted by T. We will be optimising T simultaneously with the dose vector $\boldsymbol{D}$ and the number of doses n. For computational reasons, we impose a constraint on T as $T \leq T_{max}$, where T_{max} is the maximum allowed treatment duration.

Let $T_{cure}(\boldsymbol{D}, n)$ be the time point at which the total number of parasites killed is 12 $\log_{10}$ units i.e.,

$$\int_0^{T_{cure}(\boldsymbol{D},n)} K_k(C(t))\,\mathrm{d}\,t = 10^{12}. \tag{7.8}$$

For simplicity, we refer to $T_{cure}(\boldsymbol{D}, n)$ as T_{cure}.

The Problem: We assume that the estimates of the PK and PD parameters are available. The drug's concentration is to be restricted within the therapeutic range (C^-_{tgt}, C^+_{tgt}) as much as possible while eliminating the maximum possible parasitical load of 12 $\log_{10}$ units. The number of doses n should be minimised and the individual constraints on the doses are to be strictly imposed. As before, we impose the limit d_{max} on the maximum dose that can be administered to a subject. Finally, once the parasitical load has been eliminated, the exposure to the drug should be minimised as well.

In other words, $\boldsymbol{D}^*$, n^* and T^*_{cure} need to be found such that the following objectives are achieved simultaneously:

- any concentration $C(t)$ outside the range (C^-_{tgt}, C^+_{tgt}) is minimised,
- the time at which 12 $\log_{10}$ units of parasites are obliterated is achieved as soon as possible and the number of doses n is minimal, and

- once 12 $\log_{10}$ units of the parasites have been eliminated, the exposure to the drug is minimal.

Figure 7.4 illustrates the problem to be solved.

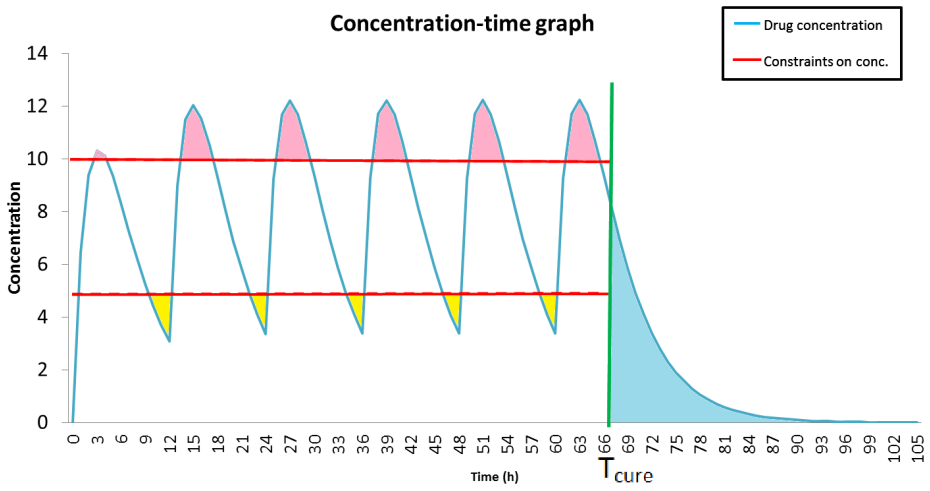


Figure 7.4: The problem is to find the dose regimen which minimises the area lying outside the constraints and the area beyond the cure point T_{cure}.

Here we define the objective function as

$$\vartheta(\mathcal{R}|\theta, \boldsymbol{t}) = \varphi(\boldsymbol{\Delta}^{\pm}) + \mathrm{AUC}_{\Delta} + HI_K, \tag{7.9}$$

where the three components of this function are explained below:

- $\varphi(\boldsymbol{\Delta}^{\pm})$ measures the area lying outside the two constraints C^{+}_{tgt} and C^{-}_{tgt}. It is the same as what was defined in Section 4.4.3 with the only difference that in this case we do not compute the average area by dividing by the number of doses administered. We just compute the total area lying outside the constraints in the time interval $[0, T_{cure}]$, i.e.,

$$\varphi(\boldsymbol{\Delta}^{\pm}) = \nu \int_0^{T_{cure}} \max(0, C^{-}_{tgt} - C(t, \boldsymbol{D}))\,\mathrm{d}\,t + (1-\nu) \int_0^{T_{cure}} \max(0, C(t, \boldsymbol{D}) - C^{+}_{tgt})\,\mathrm{d}\,t,$$

  where $C(t, \boldsymbol{D})$ is the concentration profile associated with the administered dose vector $\boldsymbol{D}$. As before, ν can be used to adjust the relative importance of the upper and the lower constraint.

- AUC_{Δ} is the area under the concentration-time curve $C(t, \boldsymbol{D})$ over the interval $[T_{cure}, T_{max}]$.

  Inclusion of this term is necessary to minimise the exposure to the drug once the goal of 12 $\log_{10}$ units has been achieved. It will also be computationally useful for selection of the number of doses required for treatment. This will be explained in the computational aspects of the minimisation algorithm.

- HI_K is a penalising function. H is a large positive number and I_K is an indicator variable such that,

$$I_K = \begin{cases} 1 & \text{if } \mathrm{AUC_K}(T) < 10^{12} \\ 0 & \mathrm{AUC_K}(T) \geq 10^{12}, \end{cases} \tag{7.10}$$

  where $\mathrm{AUC_K}(T)$ is the area under the curve given in Equation 7.7.

Its role in the objective function is to reject those dose vectors $\boldsymbol{D}$ for which $T_{cure} \notin [0, T_{max}]$. For such dose vectors, I_K will be 1 and the associated penalty will make the dose regimen ineligible to be considered any further.

The objective function $\vartheta(\mathcal{R}|\theta, \boldsymbol{t})$ for this problem is denoted by $\Upsilon(\boldsymbol{D})$ and the corresponding efficiency criterion is defined below:

Definition 6 (Υ-efficiency)**.** *A dose regimen $\mathcal{R}$ is called Υ-efficient if the function*

$$\Upsilon(\boldsymbol{D}) = \varphi(\boldsymbol{\Delta}^{\pm}) + AUC_{\Delta} + HI_K$$

is simultaneously minimised by n^* for all $n \in \mathcal{N}$, $\boldsymbol{D}^*$ for all $\boldsymbol{D} \in [0, d_{max}]^{n^*}$ and T^*_{cure} for all $T_{cure} \in [0, T_{max}]$.

7.2.2 The Optimisation Algorithm

The ED algorithm can be readily adapted for solving this problem. As before, t_j is the dosing time point for the j^{th} dose, where $j = 1, ..., n$ and $t_1 = 0$. These time points are contained in the vector $\boldsymbol{t}$. Also, $\tau_j = t_{j+1} - t_j$, $j = 1, ..., n-1$, is the time interval between two successive doses. As it is a multiple dose problem, we apply Equation 5.1 to determine the concentration at any time t.

To start the algorithm, an initial value of n, say n_o, needs to be specified. We discuss the appropriate choice of n_o later. The optimisation algorithm for this problem is run in two stages. In the first stage, an approximation of $\boldsymbol{D}^*$ is found using the discretised version of the ED algorithm. The optimal number of doses n^* is also computed at this stage. In the second stage, the approximate solution found in the first stage is refined to find the optimal $\boldsymbol{D}^*$. This is explained below.

Stage I

In Stage I, the ED algorithm is run in the discretised doses version having resolution κ. The objective function for the ED algorithm is defined in Equation (7.9). The first term in this equation has already been explained and handled in Chapter 4. The second and third terms depend on the time of parasite clearance, T_{cure} which requires a method for its numeric computation which is explained later. In each iteration

of the ED algorithm, the dose vector $\boldsymbol{D}$ which minimises the objective function in Equation (7.9) is selected to form the dose sets for the next iteration using κ, as explained in Section 4.3. The objective function has been formulated in such a way that any excess doses will cause it to increase (because of the over-exposure, as measured by $\varphi(\boldsymbol{\Delta}^{\pm})$ and AUC_{Δ}) while insufficient doses will lead to the rejection of that dose vector because of the indicator function I_K.

The discretised version of the ED algorithm will determine that $\boldsymbol{D}$ which minimises $\Upsilon(\boldsymbol{D})$ at that iteration. During the course of the iterations, the unnecessary doses are driven to 0 by subtraction of κ from them in each iteration. This is the reason of taking the resolution κ at this stage and not δ as in the latter case the doses are multiplied and divided by δ and driving them to 0 is mathematically not possible.

Once the algorithm converges at the discrete resolution κ, $n_o(\geq 0)$ doses would have been driven to 0 by the algorithm. Let the number of non-zero doses be denoted by n^*. This is the optimal number of doses required for the problem. Thus, at the end of stage I, n^* is already available. The optimal dose vector at this stage will be given by $\boldsymbol{D} = (d_1, ..., d_{n^*})^T$, with the $n - n^*$ doses removed from the vector. The algorithm now switches into stage II.

Stage II

In stage II, the dose vector $\boldsymbol{D} = (d_1, ..., d_{n^*})^T$ is further refined using the ED algorithm when doses are allowed to be real numbers. The resolution δ is used to form dose sets, as explained in Chapter 4. Stage II is essentially repeating stage I with this change. Stage II is terminated when the ED algorithm has converged according to the resolution δ. The dose vector associated with the minimum value of the objective function is denoted as $\boldsymbol{D}^*$ which consists of n^* doses. The optimal time of parasite clearance associated with this dose vector is denoted by T^*_{cure}. The optimised value of the objective function is denoted as $\Upsilon(\boldsymbol{D}^*|\kappa, \delta)$.

Computation of T_{cure}

In both stages of the above optimisation algorithm, the ED algorithm determines the best dose vector in each iteration - the one associated with the least value of $\Upsilon(\boldsymbol{D})$. This requires the computation of the T_{cure} values associated with all the l^n dose vectors in stage I and all the l^{n^*} dose vectors in stage II, where l is the size of the dose set.

For the purpose of determining T_{cure}, we divide the interval $[0, T_{max}]$ into a grid of step size 0.1. For example, for $T_{max} = 120$ h, 1200 points are considered in the interval [0,120]. We compute AUC_{K} for each of these points and we deem a point as

T_{cure} as shown below:

$$\int_0^u K_k(C(t))\,\mathrm{d}\,t = \begin{cases} < 10^{12} & \text{if } u \in [0, T_{cure}) \\ \geq 10^{12} & \text{if } u \in [T_{cure}, \infty). \end{cases} \tag{7.11}$$

Thus, the smallest point in the interval $[0, T_{max}]$ at which the reduction in the parasitical load is at least 12 $\log_{10}$ units, is designated as T_{cure}. By constructing a grid of step size 0.1, we manage to remain computationally efficient while restricting the maximum error in the true treatment time to less than 6 minutes (1 hour/10 points = 6 minutes per point). The step size can be further decreased if more precision is required.

Given the current value of T_{cure}, AUC_Δ can be computed by integrating $C(t, \boldsymbol{D})$ over the interval $[T_{cure}, T_{max}]$.

Choice of n_o and T_{max}

As mentioned before, n_o, the initial number of doses needs to be provided. The ED algorithm is able to drive the unnecessary doses to 0 and eventually find the optimal number of doses $n^*(\leq n)$. However, if n_o happens to be smaller than n^*, n_o cannot be increased in the course of the algorithm to reach n^*. For this reason, it is advisable to take a reasonably large value of n_o so that any excess doses are driven to zero. If it is found at the end that $n^* = n_o$ and any of the n^* optimal doses in $\boldsymbol{D}^*$ are computed as d_{max}, it indicates that a better solution can be obtained by increasing n_o. In that case, the algorithm should be run again with a higher value of n_o.

By the same argument, the maximum treatment time allowed, T_{max}, should be chosen to be a reasonably large number. The algorithm will not work if $T_{cure} \notin [0, T_{max}]$.

The MATLAB® code for implementing this methodology is given in Appendix D.6.

7.2.3 Example

We consider the concentration profile given by the one-compartmental model:

$$C(t, d) = \frac{FdK_a}{V(K_a - K_e)}(e^{-K_e t} - e^{-K_a t})\,, \tag{7.12}$$

where K_a denotes the absorption rate constant, K_e denotes the elimination rate constant, V is the volume of distribution and F is the bioavailability.

For the calculations we take the following values of the PK parameters as their estimates: $\widehat{K}_a = .46$ h^{-1}, $\widehat{K}_e = 0.17$ h^{-1}, $\widehat{V} = 14$ L and $\widehat{F} = .95$.

We assume that the maximum dose which can be administered at an occasion is $d_{max} = 250$ mg. n_o is taken to be 6 and the dosing time points are $\boldsymbol{t} = (0, 12, 24, 36, 48, 60)^T$.

The concentration profile of this drug is desired to be maintained between the range $(C^-_{tgt}, C^+_{tgt}) = (3, 8)$ mg/L. Equal importance is attached to both the constraints in the range, i.e., $\nu = 0.5$. The maximum treatment time allowed is $T_{max} = 120$ h.

The estimates of PD parameters are $\widehat{EC_{50}} = 3.6$ mg/L and $\widehat{H} = 1$; we consider three different values of the maximum rate of killing parasites, $\widehat{K}_{kmax}$.

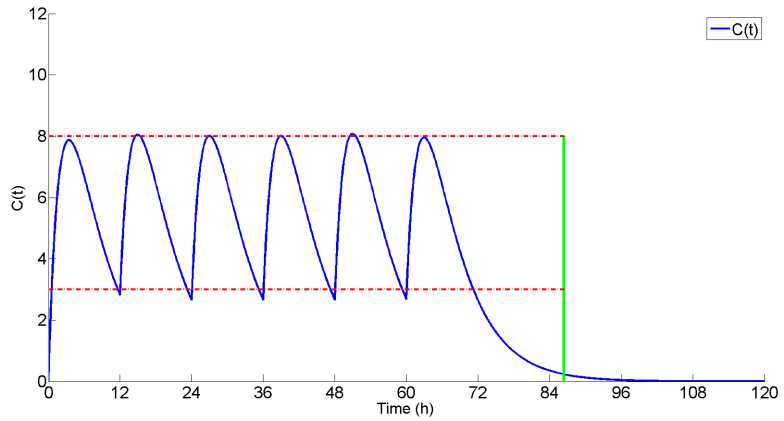
We consider dose sets of size $l = 3$ and take the three starting values for the dose sets as $\{40, 120, 250\}$. At $\kappa = 5$ and $\delta = 0.99$, we obtained $\Upsilon(\boldsymbol{D}^*|5, 0.99)$ and T_{cure} for different values for $\widehat{K}_{kmax}$. The optimisation algorithm was run as described above and T_{cure} was found by dividing the interval [0,120] h into a grid of step size 0.1 h. The results are summarised in Table 7.2.

	$\Upsilon-$efficient dose regimen			
$\widehat{K}_{kmax}$	n^*	T^*_{cure}	$\boldsymbol{D}^*$	$\Upsilon(\boldsymbol{D}^*\|5, 0.99)$
$10^{10.34}$	6	86.40	$(207.90, 168.35, 170.00, 170.00, 171.72, 168.30)^T$	3.90
$10^{10.50}$	4	62.80	$(212.12, 168.30, 171.72, 173.25, 0.00, 0.00)^T$	3.90
$10^{10.70}$	3	44.60	$(204.06, 168.30, 49.50, 0.00, 0.00, 0.00)^T$	3.94

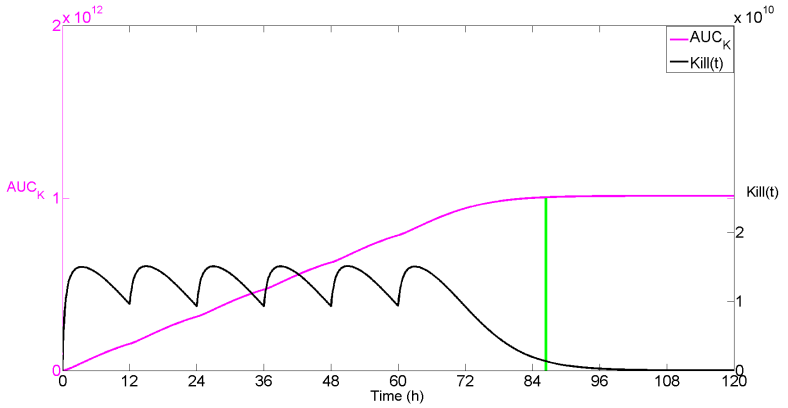
Table 7.2: Comparison of optimal dose regimens computed for different values of $\widehat{K}_{kmax}$.

The output from the optimisation algorithm for these cases is shown in Figures 7.5, 7.6 and 7.7. In these graphs, the green coloured line perpendicular to the time axis is where T^*_{cure} occurs. The shapes of the concentration-time curve and the kill rate curve are similar because of the nature of their relationship defined in Equation (7.7).

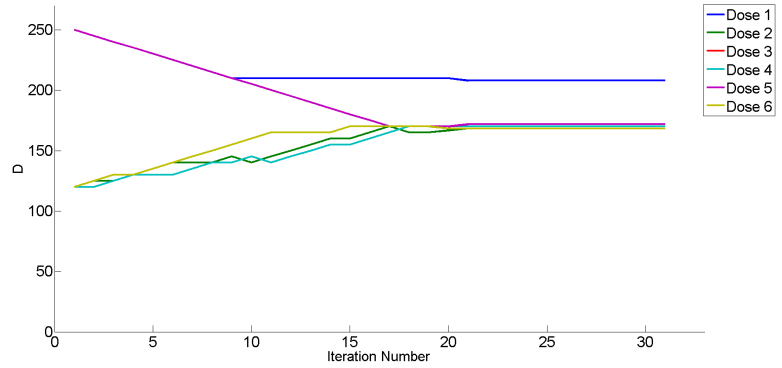
As expected, a higher value of $\widehat{K}_{kmax}$, the maximum rate of killing parasites per hour, results in lower values of n^* and T_{cure}. The simultaneous optimisation of the objective function defined in (7.9) enables obliteration of the target microbial load while using minimum number of doses and maintaining the blood concentration of the drug within the desired therapeutic range.



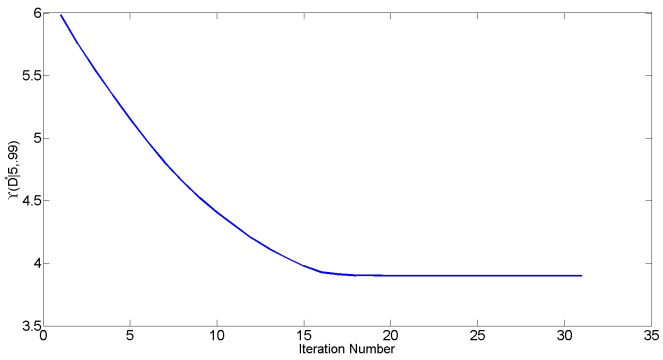
(a) The Concentration-time graph



(b) The kill curve and AUC_K resulting from $\boldsymbol{D}^*$

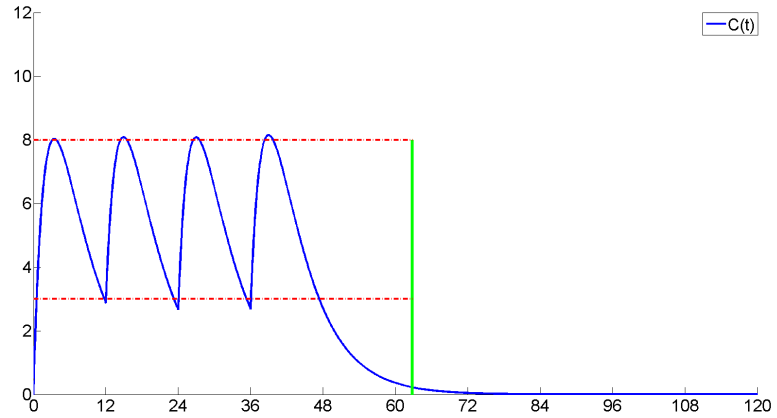


(c) Convergence of $\boldsymbol{D}^* = (d_1, ..., d_6)^T$

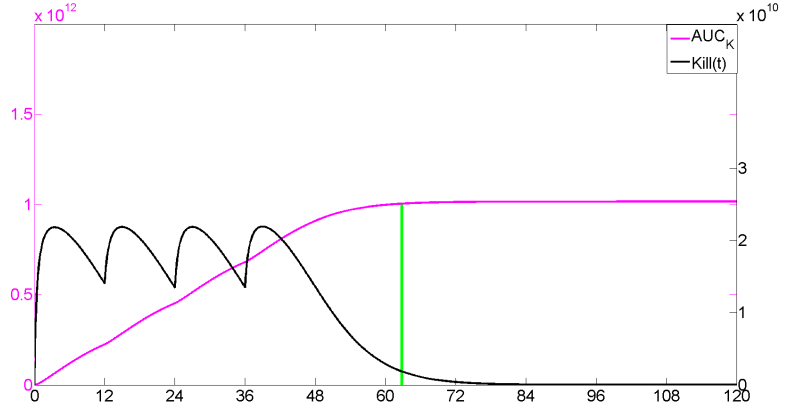


(d) Convergence of $\Upsilon(\boldsymbol{D}^*|5, 0.99)$

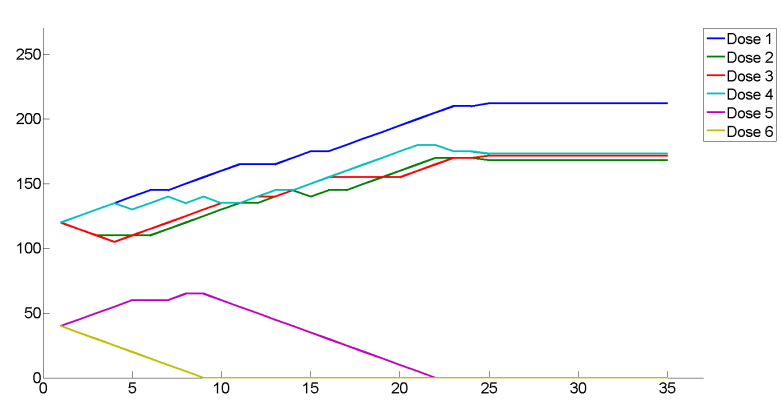
Figure 7.5: Output from the algorithm for $\widehat{K}_{kmax} = 10^{10.34}$. The algorithm converged at the 22^{nd} iteration but was allowed to run for another 10 iterations to demonstrate convergence. The line perpendicular to the time axis represnts T^*_{cure}.



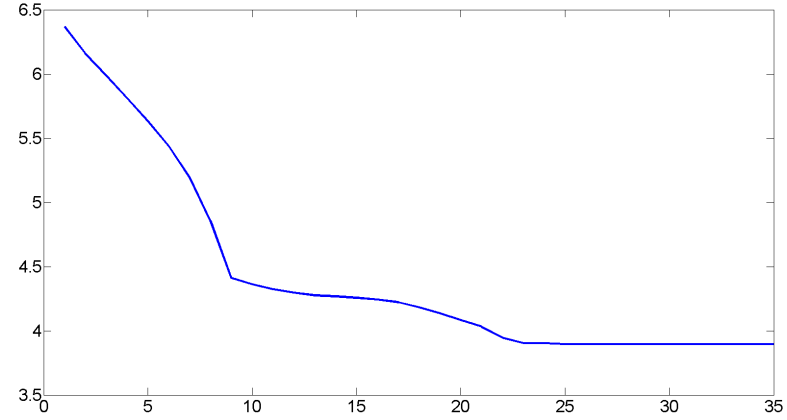
(a) The Concentration-time graph



(b) The kill curve and AUC_K resulting from $\boldsymbol{D}^*$

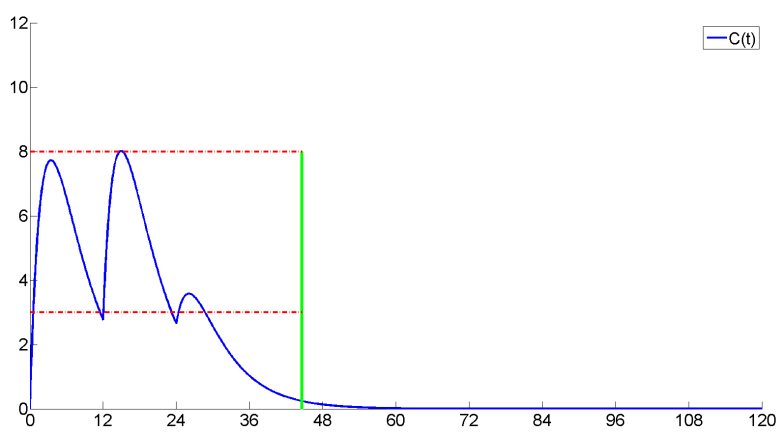


(c) Convergence of $\boldsymbol{D}^* = (d_1, ..., d_6)^T$

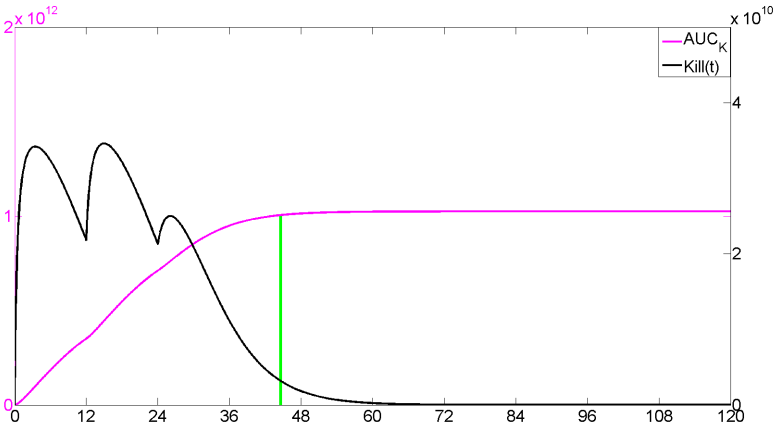


(d) Convergence of $\Upsilon(\boldsymbol{D}^*|5, 0.99)$

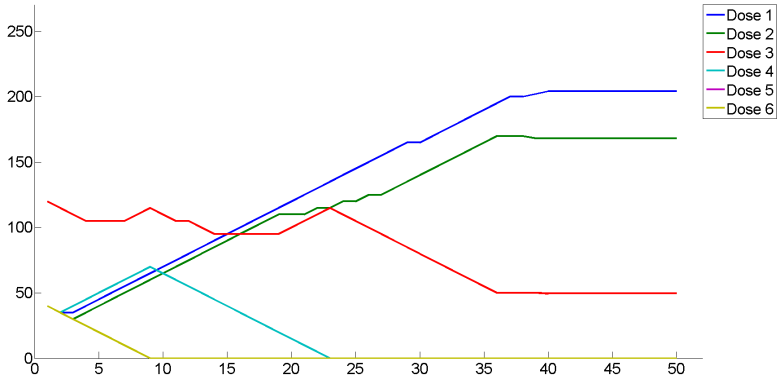
Figure 7.6: Output from the algorithm for $\widehat{K}_{kmax} = 10^{10.50}$. The algorithm converged at the 26^{th} iteration but was allowed to run for another 10 iterations to demonstrate convergence. The last 2 doses got driven to 0, thus, only 4 doses would suffice for a successful treatment, i.e. $n^* = 4$.



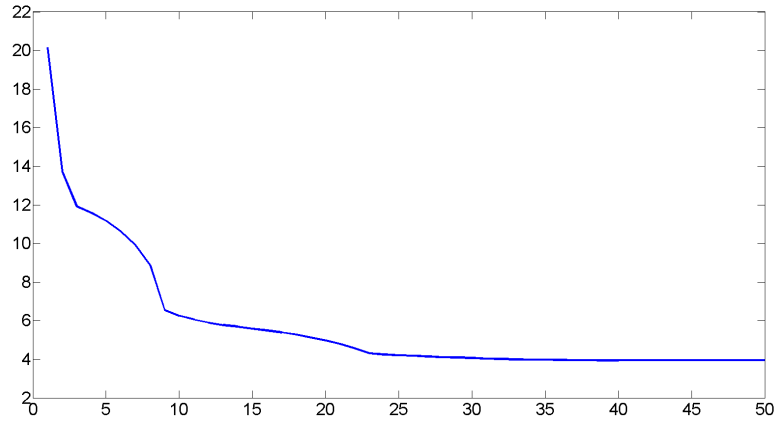
(a) The Concentration-time graph



(b) The kill curve and AUC_K resulting from $\boldsymbol{D}^*$



(c) Convergence of $\boldsymbol{D}^* = (d_1, ..., d_6)^T$



(d) Convergence of $\Upsilon(\boldsymbol{D}^*|5, 0.99)$

Figure 7.7: Output from the algorithm for $\widehat{K}_{kmax} = 10^{10.70}$. The algorithm converged at the 40^{th} iteration but was allowed to run for another 10 iterations to demonstrate convergence. The last 3 doses got driven to 0, thus, only 3 doses would suffice for a successful treatment, i.e., $n^* = 3$.

7.2.4 A Note on Optimisation of Antibiotics' Dose Regimens

Here we consider the three types of antibiotics discussed in Section 3.1.1. The problem is to optimise the dose regimens for drugs of each of these three types. Let n and T be pre-fixed and consider a single drug therapy so that $\theta = 0$. We briefly discuss these types one by one:

Type 1

As the objective is to maximise the peak concentration, C_{max}, it is obvious that that the optimal dose vector $\boldsymbol{D}^*$ should consist of the maximum possible doses which can be administered to the subject. The optimal dosing time points $\boldsymbol{t}^*$ should be as close as practically possible to each other. This can also been seen, for example, from Equations (2.6) and (2.9). As the steady-state concentration is a decreasing function of the dosing interval (τ), a smaller value of τ ensures that higher concentrations are achieved.

Type 2

For antibiotics of Type 2, the goal is to maximise the time for which the concentration remains above the MIC. By the same argument as given above, the optimal dosing time points $\boldsymbol{t}^*$ should be as close as practically possible to each other. However, if $\boldsymbol{t}$ is fixed, the optimal dose vector $\boldsymbol{D}^*$ and the number of doses n^* can be computed using the methodology described in this section. For this, C^-_{tgt} can be set to the MIC and C^+_{tgt} can be taken as the toxicity constraint (or as ∞, if there is none). The weight ν should be taken to be close to 1 to ensure maximisation of the time the concentration remains above the MIC, unless it is equally important to restrict the concentration below C^+_{tgt}. If the concentration profile resulting from the optimised dose vectors is not found to be satisfactory, the dosing time points contained in $\boldsymbol{t}$ can be adjusted using a trial and error approach. The proper way to do it would be simultaneous optimisation of $\boldsymbol{t}$ and $\boldsymbol{D}$ which is beyond the scope of the current work.

Type 3

For this class of antibiotics, the treatment goal is to maximise the AUC. Like the Type 1 antibiotics, the optimal dose vector $\boldsymbol{D}^*$ is given by the maximum doses which can be administered to the patient. The total area under the curve corresponding to $\boldsymbol{D}$, however, is invariant with respect to the dosing time points $\boldsymbol{t}$. This is proved below.

Property 2. *Let AUC^{∞} be the total area under the curve corresponding to the dose vector $\boldsymbol{D} = (d_1, ..., d_n)^T$ which is administered at times $\boldsymbol{t} = (0, t_2, ..., t_n)^T$ and AUC_i^{∞} be the total area under the curve corresponding to the single dose d_i, $i = 1, ..., n$. Then AUC^{∞} does not depend on $\boldsymbol{t}$ and it can be expressed as $AUC^{\infty} = \sum_{i=1}^{n} AUC_i^{\infty}$.*

Proof. The total AUC^{∞} corresponding to the dose vector $\boldsymbol{D}$ is computed as:

$$AUC^{\infty} = \int_{T=0}^{\infty} C(T; \boldsymbol{D}) \, \mathrm{d}\, T.$$

Consider the multiple dose model given in (5.1) for the i^{th} subject. Dropping the argument $\boldsymbol{\beta}$ and the subscripts i and j for the sake of convenience, we get the concentration at time T as:

$$C(T; \boldsymbol{D}) = \sum_{k=1}^{n} I_{\{T \geq t_k\}} C(T - t_k; d_k). \tag{7.13}$$

Therefore,

$$AUC^{\infty} = \int_{T=0}^{\infty} \left(\sum_{k=1}^{n} I_{\{T \geq t_k\}} C(T - t_k; d_k) \right) \mathrm{d}\, T = \sum_{k=1}^{n} \left(\int_{T=0}^{\infty} I_{\{T \geq t_k\}} C(T - t_k; d_k) \right) \mathrm{d}\, T.$$

The integration and the summation operators in the above equation are interchangeable since the n integrands are all positive. This follows from the Fubini-Tonelli theorem presented in, for example, Howard (2015) which says that if $f_n(x) \geq 0$ for all n and x, then $\sum \int f_n(x) \, \mathrm{d}\, x = \int \sum f_n(x) \, \mathrm{d}\, x$.

The integral on the RHS can be split using the definition of the indicator function as:

$$AUC^{\infty} = \int_{T=0}^{\infty} C(T; d_1) \, \mathrm{d}\, T + \int_{T=t_2}^{\infty} C(T - t_2; d_2) \, \mathrm{d}\, T + ... + \int_{T=t_n}^{\infty} C(T - t_n; d_n) \, \mathrm{d}\, T$$

or,

$$AUC^{\infty} = \sum_{i=1}^{n} AUC_i^{\infty}.$$

□

Note that this property holds only when the total AUC is considered, that is, the limits of the integral are $[0, \infty)$.

Thus, the area under the concentration-time curve for multiple doses is a function of the dose vector $\boldsymbol{D}$ only and not of the dosing time points $\boldsymbol{t}$. However, it may still be necessary to keep the dosing time points close to each other to prevent development of resistance to the drug and recrudescence of the infection.

7.2.5 Conclusions

In this section, we provided a solution to the Type 2 problem proposed in Section 3.3. We showed that the ED algorithm can be adapted for optimising the dose regimen for a target reduction in the microbial load, given the population mean PK and PD parameters. Knowledge of the therapeutic range, (C^-_{tgt}, C^+_{tgt}), and the target parasite load is also assumed.

Other features of the ED algorithm, as described in Chapter 4, such as discretisation of doses, optimisation of the combination ratio (the EED algorithm) can be considered for this case also.

If the PK or PD parameters are unknown, an adaptive scheme can be used, as described in Chapter 6. In that case, blood samples need to be collected not just for measuring the drug's concentration but also to measure the parasite population for estimation of PK as well as PD parameters. Nielsen et al. (2007) and Tam et al. (2005), for example, discuss fitting of PK-PD models for antibacterial therapies.

For some drugs, a PK parameter is found to be a function of a component of the dose regimen. For example, Plaisance et al. (1987) found that bioavailability of the antibiotic ciprofloxacin varies with the size of the administered dose. For accurate optimisation of dose regimens of such drugs, it is important that such relationships are incorporated in the optimisation procedure.

In the model described in this example, we have ignored the aspect of drug resistance. Resistance is generally modelled by assuming that subpopulations of microorganisms have different grades of susceptibility and resistance to the anti-infective and these are then expresses using separate functions, Czock and Keller (2007). The optimisation algorithm proposed by us is expected to work when resistance is taken into account in the PD model although some modifications may be necessary. We intend to take up this work in future.

Chapter 8

Discussion

The components of a dose regimen determine the drug's concentration in the body. The pharmacological action of the drug is a function of the concentration and the time for which the therapeutic levels are sustained. The choice of the dose regimen is therefore critical to derive the maximum benefit from the therapy. In statistical literature, dose regimens are usually optimised for a specific drug or a class of drugs and for a narrow, problem-specific criterion. Our approach to the problem is mathematically formal, it is for a general criterion and can be applied to any drug for which the mechanistic model is known. The main objective of our research has been two-fold. Firstly, to develop a theoretical base for formulating the problem of optimisation mathematically so that the performance of competing dose regimens can be quantified according to the desired criterion. Secondly, to develop a general algorithm for solving the so-formulated optimisation problem. Our methodology can be instrumental in designing dose regimens which could achieve the best balance between efficacy and safety. The algorithm can also be used to conduct simulation studies to see how different components of the dose regimen influence the objective function. It can also be used to evaluate what-if scenarios to study how patient non-compliance with the recommended dose regimen can affect the treatment's outcome.

8.1 Conclusions

In Chapter 3, we expounded the need for optimisation of dose regimens. Optimisation may be required at the individual level of a patient, or it could be required for a typical person in the population or a subpopulation. To attempt the problem of optimisation at the various levels, we first expressed the notion of a dose regimen mathematically in Section 3.2 as $\mathcal{R} = \{n, \boldsymbol{D}, \boldsymbol{t}, T, \theta\}$. This is helpful in formulating the optimisation problem for a general criterion. The dose regimens which optimise the given objective

function are called efficient and the value of the optimised objective function is called the efficiency.

Two specific types of problems are derived from the general optimisation problem described in Section 3.3. Type 1 problem entails minimisation of $\vartheta(\mathcal{R}|n, \boldsymbol{t}, T)$, i.e., minimisation with respect to the dose vector $\boldsymbol{D}$ and the combination ratio θ (in case of combination therapies) keeping n, $\boldsymbol{t}$ and T fixed. Type 2 problem is about minimisation of $\vartheta(\mathcal{R}|\boldsymbol{t}, \theta)$, i.e., with respect to $\boldsymbol{D}$, n and T keeping $\boldsymbol{t}$ and θ fixed.

The ED algorithm in Chapter 4 was developed to solve optimisation problems of Types 1 and 2. It is an iterative algorithm converging to the efficient dose regimen in successive iterations and, as shown in Theorem 1, to the optimal regimen when the resolution tends to one. The algorithm requires the knowledge of the mechanistic model followed by the drugs' concentrations. The doses can be either permitted to be real numbers or they can be enforced to take integer values. The criteria we discussed in this chapter were about minimising overexposure and underexposure to the drug by maintenance of the concentration around a target level or within a therapeutic range. Furthermore, we extend the ED algorithm to find the optimal combination ratio and the efficient dose vector of a fixed dose combination unit so that a linear combination of the objective functions for the two drugs can be optimised. In general, the algorithm is reasonably robust for mild misspecification in the model. For moderate to severe misspecification, the administered dose regimen could be much less efficient. We also found that in some cases, misspecified parameters can balance out each others' effect resulting in minimal decrease in the efficiency of the dose regimen or they can reinforce their effects and seriously decrease the efficiency. Thus, the effects of misspecification are also dependent on which parameters are misspecified and in which direction. This chapter presents our contribution to the subject of dose regimen optimisation. The algorithm is programmed in MATLAB® and the codes are given in Appendices D.1, D.2 and D.3.

There is some scope to improve the computational performance of the ED algorithm. To speed up the convergence of the algorithm, the resolution can be chosen to be adaptive in nature. We demonstrated in Section 4.4 that, in case of real doses, setting the resolution δ closer to 1 leads to an increased number of iterations before convergence is obtained and an improved value of φ. Therefore, the ED algorithm can be programmed in such a way that in the initial iterations, δ is small so that the dose sets move towards the optimal doses quickly and in the latter iterations, δ can be set closer to 1 to obtain precise values of $\boldsymbol{D}^*$. For the same reason, for the case of discretised doses, the resolution κ could be large in the initial iterations and as small as permitted in the final iterations.

The non-linear mixed effects models described in Chapter 5 are used to extend

the scope of the ED algorithm in finding the efficient dose regimen when the model parameters are not known and are treated as random variables to account for the population variability. This is done in an adaptive setting wherein blood samples are collected from the cohorts at optimal sampling time points and the mean model parameters and their variances are estimated. The computation of population D-optimal sampling times is done using PopED software. The adaptive setting is presented in Chapter 6.

The simulation studies confirmed the effectiveness of our methodology in achieving the twin objective of PK estimation and dose regimen optimisation simultaneously. We also concluded from our study that given a fixed number of volunteers in a trial, it is better to have larger cohorts rather than a greater number of smaller cohorts. However, if the safety profile of the drug is not fully established, a better approach could be to start with smaller cohorts and increase the cohort size progressively.

We proposed three stopping rules to terminate the adaptive procedure. SR1 utilises all the available cohorts in the trial, while SR2 and SR3 terminates the trial according to two differently defined stability properties in the optimised dose vectors for the successive cohorts. SR2 terminates the trial when the cumulative dose for the current cohort differs from the cumulative dose of the previous cohort by less than 5% of the latter, whereas for SR3, this condition is applied on every dose contained in the dose regimen individually. According to the problem at hand, other stopping rules could be more relevant. For example, when optimising the combination ratio and dose regimen of a fixed dose combination unit, the stopping rule should also consider value of the ratio.

The adaptive procedure described in this thesis may suffer from some procedural imperfections when implemented in practice. These imperfections are primarily missing response data and non-compliance to the recommended dose regimen by the subjects in the trial. Examples of non-compliance that we considered are not following the prescribed dosing schedule and skipping of the prescribed doses by the patients. The procedure was found to be moderately robust against missing response data and non-compliance to dosing schedules. However, the methodology was found to be quite sensitive to the instances of dose skipping. Therefore, avoidance of dose omission is critical to the success of the methodology. Furthermore, we found that large values of the error variance can have an adverse effect on the quality of some of the parameters' estimates.

Another important aspect for making the methodology successful is the choice of initial values. While we found that mild deviations from the initial values do not significantly affect the performance, very strong deviations may introduce serious bias in some of the parameters' estimates. However, the accumulating data and the interim

analyses provide an opportunity of course-correction and the adaptive procedure is able to approximately determine the optimal dose regimen towards the end of the trial. If reliable information about the initial values is not available, evaluation of all the available cohorts is recommended instead of early termination of the trial.

The adaptive procedure was determined to be reasonably robust against misspecification of the underlying PK model by approximately determining the correct dose regimen over the course of the trial. We observed in the simulation study that the estimated parameters of the assumed model 'adapt' to the response in such a way that the concentration profile generated by them resembles the true concentration profile. The adaptive procedure was also observed to be reasonably robust against deviation from the assumed error structure, although there could be large bias in the estimate of the error variance.

The difference between error in dosing at the population level and for an individual was highlighted in Section 6.2.4. At the population level, the error is measured around the optimal dose regimen based on the mean PK parameters. At the individual level, the error is measured around the optimal dose regimen based on the subject's own PK parameters.

The adaptive methodology was also compared with a non-adaptive approach. We found that if the initial values of the model parameters are very close to the true values, the non-adaptive approach is superior. In case there is a wide difference between them, the adaptive approach seems to be more ethical to use. We also evaluated a two-stage design which was found to be a good balance between the adaptive and non-adaptive approaches. Choice of a small first cohort and a much larger second cohort was found to be ethical as well as effective in achieving the trial's objectives.

As mentioned before, combination therapies are highly efficacious for certain diseases and our methodology can help in determination of the optimal ratio and the optimum dose regimen during the course of the trial. Pre-requisites of the adaptive procedure are: knowledge of the PK and PD models, availability of initial values of the parameters and correct choice of the targets in the objective function. The methodology may not be suitable in cases where the target concentration levels have not been established.

Chapter 7 presented two of many other possible applications of the ED algorithm. In the first, we showed how the algorithm can be used for dose individualisation by administering a single dose to every subject in the cohort and revising the subsequent doses with the help of the ED algorithm. In contrast to the standard dose individualisation software which only individualise a limited number of drugs, our method can work for any drug for which the mechanistic model is known. A possible disad-

vantage of our method is that it requires a cohort of subjects to be treated at the same time. This may be feasible in cases of diseases such as malaria, where the numbers of affected people are large. Our method can also be useful for the randomised concentration-controlled trials (RCCTs) which entail random assignment of the volunteers to predetermined levels of a target concentration. This is achieved in practice by an individualised PK controlled dosing scheme, Sanathanan and Peck (1991). Our methodology can help in this assignment by individualising the dose regimen for every volunteer in the group.

In the second application in the chapter, we presented one possible solution to the Type 2 problem. For treatment of infectious diseases, maintenance of the drug's concentration above a certain threshold level is critical to prevent recrudescence and resistance. We demonstrated the use of the algorithm for optimising the dose regimen for obliteration of a predefined level of bacterial load while keeping the concentration profile within a therapeutic range. We hope that this method can be used to design dose regimens for anti-microbial drugs. However, if the PK or PD parameters are not known, an adaptive scheme as explained in Chapter 6 may have to be considered.

In summary, we present a method for computer assisted dose finding by explicit optimisation of a target criterion. This approach is likely to supersede 'brute force' techniques based exclusively on simulation. To the best of our knowledge, there does not exist a flexible and universal method for explicit optimisation of dose regimens, which computes the best values of dose vectors and the combination ratios and also quantifies the degree of deviation from the desired PK or PD targets. By flexible we mean that the method suits different medical criteria and should be able to reconcile with the practical constraints that confront dose administration, e.g., skipped doses. By universal, we imply the ability of the method to compute optimised dose regimens for any drug, irrespective of the kind of PK/PD model followed by it. Our work is a step towards filling these gaps. Furthermore, we extend the problem of optimisation to combination therapies.

We think that our algorithm and the associated methodology could make the following contribution to the science of finding the best dose regimen:

- In early phase clinical trials, when PK information tends to be sparse, the algorithm can aid pharmaceutical companies in estimation of the mean PK parameters while administering the optimum dose regimen to the cohorts of patients. The adaptive procedure can be extended for clinical trials which involve combination therapies so that the optimal ratio can also be determined during the course of the trial.

- For drugs for which reliable PK estimates are already available, the ED algo-

rithm can be used to for conducting 'what-if' analyses which help in the quantitative assessment of the adherence of different dose regimens and combination ratios to the desired criteria. Furthermore, the algorithm can be used to quantify the effect of patient non-compliance with the prescribed dose regimen on the objective function. Patient's non-compliance is a significant medical challenge with rates of non-compliance being about 45% for long-term and about 25% for short-term therapies, Jin et al. (2008). Scenarios such as skipping doses, taking incomplete doses and overdosing can be quantitatively analysed.

- For dose individualisation: Apart from individualising a dose regimen of a single drug, our method comes with an additional feature of individualising the optimal combination ratio for combination therapies. This ensures that the partner drugs' concentrations are maintained in the desired ranges. Also, our method can be applied for any drug for which the mechanistic model is available in numerical or analytic form and is therefore not restricted to any library of a few selected drugs. As mentioned before, the individualisation method can also be applied in RCCTs to allocate volunteers to levels of target concentration.

8.2 Future Work

In the future, we would like to further develop our research from both perspectives - academic interest and clinical applications. This is discussed below.

An assumption of the ED algorithm that the mechanistic model followed by the drug is known to the user can be prohibitive in certain cases. Since the ED algorithm optimises an objective function of the model, the accuracy of the algorithm is dependent on the goodness of fit of the model to the actual data. As mentioned in Section 2.1, non-compartmental analysis does not require development and validation of a model. In the future, we would like to develop the algorithm such that optimisation of dose regimens can be done without assuming an underlying mechanistic model for the concentration values. Furthermore, the examples in this thesis used compartmental models which have a closed-form solution. It would be useful to make the algorithm work when the concentrations are expressed through a system of differential equations, as most industry-standard models are mathematically more sophisticated and do not have a closed form solution, for example, the PBPK models described in Section 2.1.

The ED algorithm can be developed further to make it amenable to some of the practical challenges. For instance, for optimisation of the combination ratio, drug-drug interactions can be incorporated into the Extended ED algorithm. Interactions

between drugs can be synergistic - the net PD effect is increased, or antagonistic - the net PD effect is subdued. In some cases, a new effect may occur which is produced by neither of the drugs in isolation. Models are available to explain such interactions, for example, in Fahmi and Ripp (2010). They can be used to jointly model the concentrations of drug A with that of drug B and the EED algorithm can be applied to optimise the combination ratio of such drugs. An important goal to accomplish would be to optimise the dose regimens and the combination ratio when the PK parameters of the models are not known. This can also be done in an adaptive setting as it was done for a single drug in Chapter 6.

Furthermore, as already discussed in Section 7.1.4, inclusion of covariates is important to improve the performance of the methodology. Inclusion of covariates in the assumed model is expected to reduce the inter-subject variability. The potential benefit of this is not only enhancement of the efficiency of the computed dose regimen for the population but as discussed in Section 6.3 could also mitigate to some extent the adverse effect of departure from the model assumptions. We plan to do more research in this direction in the future.

The ED algorithm computes the optimum dose regimen conditional on the mean PK parameters as inputs. By using the variance parameters of the model, inter-individual variability can be simulated to generate distributions of the PK parameters, as we did in the simulation study of Chapter 6. Distributions of the covariates can be simulated as well using data from past surveys. For example, Kuczmarski et al. (2002) present bodyweight distributions at different ages of the American children. For each simulated vector of parameters and the simulated vector of covariates, the ED algorithm can be run to obtain distribution of the optimum dose regimen for the population. This distribution can be useful in dividing the population into subpopulations based on covariates such as bodyweight or body surface area.

In this thesis, we solved Type 1 problem which consists of optimisation of ϑ as a function of $\mathcal{R}$ with respect to $\boldsymbol{D}$ and θ and Type 2 problem in which ϑ was optimised with respect to n, $\boldsymbol{D}$ and T. It would be interesting to study other cases of the general optimisation problem. For example, optimisation of ϑ not only with respect to $\boldsymbol{D}$ and θ but also some other design variable, such as the dosing time points $\boldsymbol{t}$. This would be challenging, but a simpler problem to consider might be to optimise $\vartheta(\mathcal{R}|n, \boldsymbol{D}, T, \theta)$. Just like dose sets were created in each iteration to drive the ED algorithm for solving for the optimal dose vector $\boldsymbol{D}^*$, *time sets* can possibly be used to find the optimal dosing time points $\boldsymbol{t}^*$. However, a potential problem that may arise is that the compartmental models are generally nonlinear functions of time in contrast with dose, which occurs linearly in these models. Because of the nonlinearity, a change in value of time will not be proportionally reflected in the value of concentration. It will be interesting to study the behaviour of the algorithm in this case.

The criteria considered by us in this thesis were deterministic in nature. It would be interesting to explore whether the ED algorithm can optimise an objective function which consists of random variables. For example, instead of maintenance of the concentration within a therapeutic range, a composite criterion can be used which enforces the concentration to remain above a lower limit and minimises the likelihood of toxicity.

This PhD project was co-sponsored by Novartis Pharma AG and in our interactions with their staff, we discussed the applicability of the algorithm on real-world models and parameters. It is envisioned that the MATLAB® programs, presented in Appendices D.1 - D.6 for implementing the ED algorithm and its extensions, will be developed into a software tool which can be leveraged by clinicians and pharmaceutical companies to design dose regimens for single and combination therapies on stronger quantitative grounds.

Our joint publication, Soeny et al. (2016), is primarily based on Chapter 4. At the moment three more manuscripts are being developed from this thesis for publication:

- *Simultaneous PK estimation and dose regimen optimisation in clinical trials.* This manuscript will be mainly based on Chapter 6.
- *A novel method of individualisation of dose regimens.* This will be based on Section 7.1.
- *Optimisation of dose regimens of anti-microbials for a target load.* This will be developed from Section 7.2.

The form of the optimisation problem we proposed in Chapter 3, the ED algorithm introduced in Chapter 4, expression of the multiple dose model in Equation (5.1) and the theory of non-linear mixed effects models in Chapter 5 will be useful in supplementing the above manuscripts with the necessary theoretical base.

Appendices

Appendix A

Multiple Dose Formula for One Compartment Model

Let an individual be given a dose d at times $t = 0, \tau, 2\tau, ..., (n-1)\tau$. Thus, n doses are administered at an interval of τ units. We are interested in finding the concentration of the drug at time t after the n^{th} dose is administered. For the model defined in (2.4), the concentration at time t after the n^{th} dose is given as (by the principle of superposition):

$$C(t;\tau,n) = C_1(t+(n-1)\tau) + C_2(t+(n-2)\tau) + ... + C_{n-1}(t+\tau) + C_n(t),$$

where $C_i(.)$ is the concentration of the drug in blood plasma contributed by the i^{th} dose.

$$\Rightarrow C(t;\tau,n) = \frac{FdK_a}{V_1(K_a-K_e)}\left(e^{-K_e(t+(n-1)\tau)} - e^{-K_a(t+(n-1)\tau}\right)$$

$$+ \ldots + \frac{FdK_a}{V_1(K_a-K_e)}\left(e^{-K_e(t+\tau)} - e^{-K_a(t+\tau)}\right) + \frac{FdK_a}{V_1(K_a-K_e)}\left(e^{-K_et} - e^{-K_at}.\right)$$

$$\Rightarrow C(t;\tau,n) = \frac{FdK_a}{V_1(K_a-K_e)}\Bigg(e^{-K_et}(e^{-K_e(n-1)\tau} + e^{-K_e(n-2)\tau} + ... + e^{-K_e\tau} + 1)$$

$$- \ e^{-K_at}(e^{-K_a(n-1)\tau} + e^{-K_a(n-2)\tau} + ... + e^{-K_a\tau} + 1)\Bigg).$$

Now, the sum of n terms of a geometric progression with first term a and common ratio r is $a + ar + ar^2 + ... + ar^{n-1} = a(1-r^n)/(1-r)$.

Using the above formula, the final expression of the concentration at time t after the n^{th} dose is administered is,

$$C(t;\tau,n) = \frac{FdK_a}{V_1(K_a-K_e)}\left(\frac{1-e^{-nK_e\tau}}{1-e^{-K_e\tau}}e^{-K_et} - \frac{1-e^{-nK_a\tau}}{1-e^{-K_a\tau}}e^{-K_at}\right). \tag{A.1}$$

Appendix B

Two Compartment Model With First Order Absorption

B.1 Derivation of the Model for Single Dose

Here, we provide the proof to the model stated in Equation (2.7). In this model, the dose d is absorbed into the central compartment at the rate of K_a. The drug gets absorbed into the peripheral compartment from the central compartment at the rate of K_{12} and it gets re-absorbed from the peripheral compartment to the central compartment at the rate of K_{21}. The drug can be eliminated from the body only through the central compartment which takes place at the rate of K_e. Let the volume of the central and the peripheral compartments be V_1 and V_2 respectively. $X_1(t), X_3(t)$ and $X_2(t)$ are respectively the amounts of drug in the central compartment, peripheral compartment and not yet absorbed in the central compartment at time t.

The differential equations describing the process can be written as:

$$\frac{\mathrm{d}}{\mathrm{d}\,t}X_1(t) = K_a X_2(t) + K_{21}X_3(t) - (K_{12} + K_e)X_1(t), \tag{B.1}$$

$$\frac{\mathrm{d}}{\mathrm{d}\,t}X_2(t) = -K_a X_2(t), \tag{B.2}$$

$$\frac{\mathrm{d}}{\mathrm{d}\,t}X_3(t) = K_{12}X_1(t) - K_{21}X_3(t) \tag{B.3}$$

with the initial conditions $X_1(0) = 0$, $X_3(0) = 0$ and $X_2(0) = d$.

Solving Equation (B.2) and plugging it in Equation (B.1) we get a system of simultaneous differential equations.

Letting $\mathbb{D} \equiv \frac{\mathrm{d}}{\mathrm{d}t}$, the system can be written as,

$$(\mathbb{D} + K_{12} + K_e)X_1(t) - K_{21}X_3(t) = dK_a e^{-K_a t}, \tag{B.4}$$

$$(\mathbb{D} + K_{21})X_3(t) - K_{12}X_1(t) = 0. \tag{B.5}$$

Eliminating $X_3(t)$ from Equations B.4 and B.5 and solving simultaneously we get,

$$f(\mathbb{D})X_1(t) = dK_a(K_{21} - K_a)e^{-K_a t}. \tag{B.6}$$

The auxiliary equation is, $f(m) = 0$ where,

$$m^2 + m(K_{12} + K_{21} + K_e) + K_e K_{21} = 0. \tag{B.7}$$

The roots of this quadratic equation are given as,

$$m = \frac{-(K_{12} + K_{21} + K_e) \pm \sqrt{(K_{12} + K_{21} + K_e)^2 - 4K_e K_{21}}}{2}. \tag{B.8}$$

Let $S = K_{12} + K_{21} + K_e$ and $R = \sqrt{S^2 - 4K_{21}K_e}$. Then the roots of Equation (B.7) are given by $-\lambda$ and $-\mu$ where, $\lambda = \frac{1}{2}(S + R)$ and $\mu = \frac{1}{2}(S - R)$.

The complimentary function (CF) is therefore, $CF = c_1 e^{-\lambda t} + c_2 e^{-\mu t}$, where c_1 and c_2 are constants to be determined.

The particular integral (PI) is given as: $PI = \frac{1}{f(\mathbb{D})} dK_a(K_{21} - K_a)e^{-K_a t}$.

This is of the form $\frac{1}{f(\mathbb{D})}e^{at}$. The solution is given as $\frac{1}{f(a)}e^{at}$ provided $f(a) \neq 0$.

Therefore,

$$PI = \frac{1}{f(-K_a)} dK_a(K_{21} - K_a)e^{-K_a t} = \frac{dK_a(K_{21} - K_a)e^{-K_a t}}{(K_a - \lambda)(K_a - \mu)}.$$

The general solution can now be written as

$$X_1(t) = c_1 e^{-\lambda t} + c_2 e^{-\mu t} + \frac{dK_a(K_{21} - K_a)e^{-K_a t}}{(K_a - \lambda)(K_a - \mu)}. \tag{B.9}$$

Using the initial condition $X_1(0) = 0$,

$$c_1 + c_2 = -\frac{dK_a(K_{21} - K_a)}{(K_a - \lambda)(K_a - \mu)}. \tag{B.10}$$

From Equation (B.1) we have,

$$X_1'(0) = K_a X_2(0) + K_{21} X_3(0) - (K_{12} + K_e) X_1(0) = dK_a,$$

where $X_1'(0)$ is the first derivative of $X_1(t)$, evaluated at $t = 0$.

Also, from Equation (B.9),

$$X_1'(0) = -\lambda c_1 - \mu c_2 - \frac{dK_a^2(K_{21} - K_a)}{(K_a - \lambda)(K_a - \mu)}.$$

This gives,

$$\lambda c_1 + \mu c_2 = -dK_a - \frac{dK_a^2(K_{21} - K_a)}{(K_a - \lambda)(K_a - \mu)}. \tag{B.11}$$

Solving Equations (B.10) and (B.11) we get,

$$c_1 = \frac{FdK_a(K_{21} - \lambda)}{(K_a - \lambda)(\mu - \lambda)} \quad \text{and} \quad c_2 = \frac{FdK_a(K_{21} - \mu)}{(K_a - \mu)(\lambda - \mu)}.$$

To find the concentration $C(t)$ at time t, we divide by the volume of the compartment V_1 to get

$$C(t) = Ae^{-\lambda t} + Be^{-\mu t} - (A + B)e^{-K_a t}, \tag{B.12}$$

where

$$A = \frac{FdK_a(K_{21} - \lambda)}{V_1(K_a - \lambda)(\mu - \lambda)} \quad \text{and} \quad B = \frac{FdK_a(K_{21} - \mu)}{V_1(K_a - \mu)(\lambda - \mu)}.$$

B.2 Derivation of Multiple Dose Formula

Let a subject be given n doses of size d at times $t = 0, \tau, ..., (n-1)\tau$. We are interested in finding the concentration of the drug at time t after the n^{th} dose is administered. For the model defined in Equation (B.12), the concentration at time t after the n^{th} dose is given by the principle of superposition as:

$$C(t; \tau, n) = C_1(t + (n-1)\tau) + C_2(t + (n-2)\tau) + ... + C_{n-1}(t + \tau) + C_n(t).$$

where $C_i(.)$ is the concentration of the drug in blood plasma contributed by the i^{th} dose.

Analogously to the derivation shown in Appendix (A), we get the expression as

$$C(t; \tau, n) = A\frac{1 - e^{-n\lambda\tau}}{1 - e^{-\lambda\tau}}e^{-\lambda t} + B\frac{1 - e^{-n\mu\tau}}{1 - e^{-\mu\tau}}e^{-\mu t} - (A + B)\frac{1 - e^{-nK_a\tau}}{1 - e^{-K_a\tau}}e^{-K_a t}. \tag{B.13}$$

Appendix C

Expressing Variability in PK Parameters as Coefficients of Variation

It is a standard practice to express the variability in the PK parameters of a population as coefficients of variation (CV) instead of the variance or the standard deviation. The reason for this stems from the relationship between the normal and the lognormal distributions. Owing to the positivity and the skewed distributions of PK parameters, they are generally assumed to have a lognormal distribution. We now prove that the standard deviation of the logarithm of a PK parameter is approximately equal to its coefficient of variation in the original scale.

Let $X \sim \text{logN}(\mu, \sigma^2)$ where logN denotes lognormal distribution.

Then $y = \log(X) \sim \text{N}(\mu, \sigma^2)$.

The r^{th} central moment of X is given by $m_r = \text{E}[X^r]$.

Since $X = e^Y$ we have,

$$m_r = \text{E}[e^{ry}] = \exp\left\{\mu r + \frac{\sigma^2 r^2}{2}\right\},$$

where the function on the R.H.S. is the moment generating function of a normally distributed variable.

Therefore,

$$\text{E}[X] = m_1 = e^{\mu + \frac{\sigma^2}{2}},$$

$$\text{E}[X^2] = m_2 = e^{2\mu + 2\sigma^2},$$

$$\mathrm{Var}[X] = \mathrm{E}[X^2] - \mathrm{E}[X]^2 = e^{2\mu+\sigma^2}\left(e^{\sigma^2} - 1\right).$$

The coefficient of variation (CV) of X is given as:

$$CV = \frac{\sqrt{\mathrm{Var}(X)}}{\mathrm{E}[X]} = \sqrt{e^{\sigma^2} - 1} \approx \sigma,$$

with the approximation holding well only for a certain range of values of σ.

Figure C.1 plots the relationship between the standard deviation σ and CV $= \sqrt{e^{\sigma^2} - 1}$.

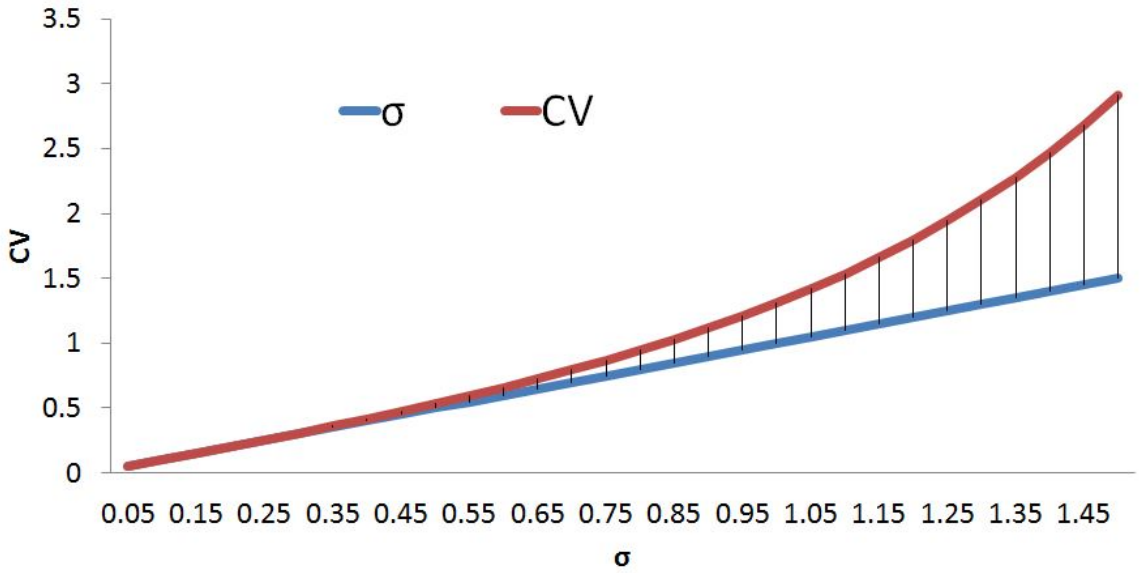


Figure C.1: Plot showing the range of values over which the coefficient of variation is a good approximation of σ.

It can be observed from the figure that a standard deviation of 0.70 or less of the logarithm of a PK parameter can be reasonably approximated by the CV of the parameter in original scale. Beyond this range, σ significantly underestimates the CV.

Therefore, in a certain range, the standard deviation of a normal variate is approximately equal to the CV of the corresponding log-normal variate.

Appendix D

MATLAB® Programs

D.1 The ED Algorithm for a C_{tgt}

```
function main
% PK parameters
Ka = .37; Ke = 0.2; V = 24; F = .95;
C_tgt = 3;   %Specify Ctgt here
D_MAX = 250; %the maximum dose that can be given
epsilon  = .99; %the resolution
kappa = 5; %for discretised doses
% Dose options
D = ones(7,3);
D(:,1) = 40;
D(:,2) = 120;
D(:,3) = 240;
dose_set = [0*ones(7,1),8*ones(7,1),D(:,1),D(:,2),D(:,3)];
compare = 10000;
% Dosing Interval
time = 6; %equal dose interval
tcure = 42; %T
n = 7; %number of doses
last_time = tcure-(n-1)*time;
history = ones(10000,7);
varphi = ones(10000,1);
x=1;
stop = 0;
while (x <10000 && stop < 7)
stop = 0;
```

```
lim = size(D(1,:),2);
for i = 1:lim
   conc1 = @(t) abs (fterm(D(1,i)) *(exp(-Ke*t) - exp(-Ka*t))
        - C_tgt );
   obj1(i) = integral(conc1,0,time,'AbsTol',10^-2,'RelTol'
      ,10^-2)  ;
   data1(i,:) = [D(1,i)  obj1(i) ];
end
grid2 = cartprod(D(1,:) , D(2,:) );
lim2 = size(grid2,1);
for i =1:lim2
     conc2 = @(t) abs ( fterm(grid2(i,1))* (exp(-Ke*(time + t
        ))-exp(-Ka*(time + t)))+fterm(grid2(i,2))*(exp(-Ke*t)
        -exp(-Ka*t))-C_tgt);
     obj2(i) = integral(conc2,0,time,'AbsTol',10^-2,'RelTol'
        ,10^-2);
end
grid3 = cartprod(D(1,:) , D(2,:) , D(3,:) );
lim3 = size(grid3,1);
for i =1:lim3
    conc3 = @(t) abs ( fterm(grid3(i,1)) * (exp(-Ke*(2*time +
        t)) - exp(-Ka*(2*time + t))) + fterm(grid3(i,2))* (
       exp(-Ke*(time + t))- exp(-Ka*(time + t))) +  fterm(
       grid3(i,3))*(exp(-Ke*t) - exp(-Ka*t)) - C_tgt);
    obj3(i) = integral(conc3,0,time,'AbsTol',10^-2,'RelTol'
       ,10^-2);
end
grid4 = cartprod(D(1,:) , D(2,:) , D(3,:), D(4,:) );
lim4 = size(grid4,1);
for i =1:lim4
   conc4 = @(t) abs ( fterm(grid4(i,1))* (exp(-Ke*(3*time + t
      )) - exp(-Ka*(3*time + t)))+ fterm(grid4(i,2))* (exp(-
      Ke*(2*time + t))- exp(-Ka*(2*time + t)))+fterm(grid4(i
      ,3))*(exp(-Ke*(time + t)) - exp(-Ka*(time + t)))+fterm(
      grid4(i,4))*(exp(-Ke*t)-exp(-Ka*t))-C_tgt);
   obj4(i) = integral(conc4,0,time,'AbsTol',10^-2,'RelTol'
      ,10^-2);
end
grid5 = cartprod(D(1,:) , D(2,:) , D(3,:), D(4,:) , D(5,:));
lim5 = size(grid5,1);
for i =1:lim5
```

```
      conc5 = @(t) abs (fterm(grid5(i,1))* (exp(-Ke*(4*time +
         t)) - exp(-Ka*(4*time + t)))+ fterm(grid5(i,2))* (exp
         (-Ke*(3*time + t)) - exp(-Ka*(3*time + t))) + fterm(
         grid5(i,3))* (exp(-Ke*(2*time + t)) - exp(-Ka*(2*time
          + t)))+ fterm(grid5(i,4))* (exp(-Ke*(time + t)) -
         exp(-Ka*(time + t))) + fterm(grid5(i,5))* (exp(-Ke*t)
          - exp(-Ka*t)) - C_tgt );
      obj5(i) = integral(conc5,0,time,'AbsTol',10^-2,'RelTol'
         ,10^-2)  ;
end
grid6 = cartprod(D(1,:) , D(2,:) , D(3,:), D(4,:) , D(5,:), D
   (6,:) );
lim6 = size(grid6,1);
for i =1:lim6

       conc6 = @(t) abs ( fterm(grid6(i,1))* (exp(-Ke*(5*time
          + t)) - exp(-Ka*(5*time + t)))+ fterm(grid6(i,2))*(
          exp(-Ke*(4*time + t)) - exp(-Ka*(4*time + t))) +
          fterm(grid6(i,3))* (exp(-Ke*(3*time + t)) - exp(-Ka
          *(3*time + t)))+ fterm(grid6(i,4))* (exp(-Ke*(2*time
           + t)) - exp(-Ka*(2*time + t))) + fterm(grid6(i,5))*
           (exp(-Ke*(time + t)) - exp(-Ka*(time + t))) + fterm
          (grid6(i,6))* (exp(-Ke*t) - exp(-Ka*t)) - C_tgt );
       obj6(i) = integral(conc6,0,time,'AbsTol',10^-2,'RelTol'
          ,10^-2)  ;
end
grid7 = cartprod(D(1,:) , D(2,:) , D(3,:), D(4,:) , D(5,:), D
   (6,:),D(7,:));
lim7 = size(grid7,1);
for i =1:lim7
 conc7 = @(t) abs ( fterm(grid7(i,1))* (exp(-Ke*(6*time + t))
     - exp(-Ka*(6*time + t)))+ fterm(grid7(i,2))* (exp(-Ke
    *(5*time + t)) - exp(-Ka*(5*time + t))) + fterm(grid7(i
    ,3))* (exp(-Ke*(4*time + t)) - exp(-Ka*(4*time + t)))+
    fterm(grid7(i,4))* (exp(-Ke*(3*time + t)) - exp(-Ka*(3*
    time + t))) + fterm(grid7(i,5))* (exp(-Ke*(2*time + t)) -
     exp(-Ka*(2*time + t))) + fterm(grid7(i,6))* (exp(-Ke*(
    time + t)) - exp(-Ka*(time + t))) + fterm(grid7(i,7))* (
    exp(-Ke*t) - exp(-Ka*t)) - C_tgt );
      obj7(i) = integral(conc7,0,last_time,'AbsTol',10^-2,'
         RelTol',10^-2) ;
```

```
        j = mod(i,lim);if (j ==0)
                          j = lim;
                       end
       k = mod(i,lim2);if (k ==0)
                          k = lim2;
                       end
       l = mod(i,lim3);if (l ==0)
                          l = lim3;
                       end
       m = mod(i,lim4);if (m ==0)
                         m = lim4;
                       end
       n = mod(i,lim5); if (n ==0)
                          n = lim5;
                        end
        p = mod(i,lim6);if (p ==0)
                          p = lim6;
                        end
       tp = [obj1(j) , obj2(k) , obj3(l) , obj4(m) , obj5(n) ,
          obj6(p), obj7(i)];
         cum = mean(tp); %\varphi_A criterion
         data7(i,:) = [ grid7(i,1) obj1(j)  grid7(i,2)  obj2(k)
             grid7(i,3)  obj3(l) grid7(i,4)  obj4(m) grid7(i,5)
             obj5(n) grid7(i,6)  obj6(p) grid7(i,7) obj7(i) cum
            ];
end
sort_data = sortrows(data7, 15);
sort2 = sort_data(1,:);
reg = sort2(:,[1,3,5,7,9,11,13])
history(x,:) = reg;
if x>1
    for i=1:7
         if abs(reg(i) - history(x-1,i))== 0
            stop = stop + 1;    %Checking for convergence
         end;
    end;
end;
%if doses are not discretised, use the block below
for i = 1:size(reg,2)
        if reg(i) == D(i,1)
         D(i,:) = min(D_MAX,[D(i,1)*epsilon,D(i,1), D(i,1)*(1/
```

```
            epsilon)]);

        elseif reg(i) == D(i,2)
         D(i,:) = min(D_MAX,[D(i,2)*epsilon,D(i,2), D(i,2)
            *(1/epsilon)]);

            elseif reg(i) == D(i,3)
          D(i,:) = min(D_MAX,[D(i,3)*epsilon,D(i,3), D(i,3)
             *(1/epsilon)]);
       end
end
%if doses are discretised, instead of above, use the one
   below.
% for i = 1:size(reg,2)
%       if reg(i) == D(i,1)
%           D(i,:) = max(0,min(D_MAX,[D(i,1)- kappa, D(i,1), D
   (i,1) + kappa]));
%       elseif reg(i) == D(i,2)
%           D(i,:) = max(0,min(D_MAX,[D(i,2) - kappa,D(i,2), D
   (i,2) + kappa]));   %%%% Discrete
%       elseif reg(i) == D(i,3)
%           D(i,:) = max(0,min(D_MAX,[D(i,3) - kappa, D(i,3),
   D(i,3)+ kappa]));
%       end
% end
compare = sort_data(1,15)
varphi(x) = compare;
x=x+1;
end
end
function ft = fterm(D)
Ka = .37; Ke = 0.2; V = 24; F = .95;
ft = (F*D*Ka)/(V*(Ka - Ke));
end
```

D.2 The EED Algorithm for Combination Therapies

```
function FINAL_Coartem_uneq_tau
%%% PK parameters Drug A
Ka = .37; Ke = 0.829; V = 217; F = .7;
%%% PK parameters Drug B
CL = 15;
V1 = 215;
ka = .13;
Q = 13.4;
V2 = 1043;
%Computed Parameters
k12 = Q/V1;
k21 = Q/V2;
ke = CL/V1;
summ = k12 + k21 + ke;
root = sqrt(summ^2 - 4*ke*k21);
alpha = (summ+root)*0.5;
beta = (summ-root)*0.5;
epsilon  = .99;
C_tgt_A = .037;
C_tgt_B = 1.14;
w = 0.5;
dmax = 1000;
dmax_B = 1000;
% Dosing Intervals
time1 = 8;time2 = 16;time3 = 12;time4 = 12;time5 = 12;tcure =
    168;
time6 = tcure - (time1+time2+time3+time4+time5);
compare = 10000;current_ld = 1000;ld_final = 1000;
D = ones(6,3);D(:,1) = 40;D(:,2) = 60;D(:,3) = 80;LD =
   [7,8,9];
history = ones(10000,6);
varphi = ones(10000,1);
com = 10000*ones(1,size(LD,2));
mu = 10000;
y=1
reg = ones(size(LD,2),6);
while (mu>0)
```

```
for I = 1:size(LD,2)
    x=1;
     stop = 0;
     while (x <10000 && stop < 6)
    stop = 0;
     lim = size(D(1,:),2);
     for i = 1:lim
         conc1 = @(t) abs ( fterm(D(1,i)) *(exp(-Ke*t) -
            exp(-Ka*t)) - C_tgt_A);
          conc1b = @(t) abs (Alum(LD(I)*D(1,i))*exp(-alpha
             *t) + Blum(LD(I)*D(1,i))*exp(-beta*t) - (Alum
             (LD(I)*D(1,i)) + Blum(LD(I)*D(1,i)))*exp(-ka*
             t) - C_tgt_B) ;
          obj1a(i) = quadl(conc1,0,time1,.1) ;
          obj1b(i) = quadl(conc1b,0,time1,.1) ;
     end
     grid2 = cartprod(D(1,:) , D(2,:) );
     lim2 = size(grid2,1);
    for i =1:lim2
        conc2 = @(t) abs ( fterm(grid2(i,1))* ...
            (exp(-Ke*(time1 + t)) - exp(-Ka*(time1 + t))
               ) + fterm(grid2(i,2))* (exp(-Ke*t) - exp(-
               Ka*t)) - C_tgt_A);
        conc2b = @(t) abs(Alum(LD(I)*grid2(i,1))*exp(-
           alpha*(time1+t)) + Blum(LD(I)*grid2(i,1))*exp(-
           beta*(time1+t)) - (Alum(LD(I)*grid2(i,1))+Blum
           (LD(I)*grid2(i,1))) *exp(-ka*(time1 + t)) +
           Alum(LD(I)*grid2(i,2))*exp(-alpha*t) + Blum(LD(
           I)*grid2(i,2))*exp(-beta*t) - (Alum(LD(I)*grid2
           (i,2))+Blum(LD(I)*grid2(i,2)))*exp(-ka*t) -
           C_tgt_B);
        obj2a(i) = quadl(conc2,0,time2,.1);
        obj2b(i) = quadl(conc2b,0,time2,.1);
    end
    grid3 = cartprod(D(1,:) , D(2,:) , D(3,:) );
    lim3 = size(grid3,1);
     for i =1:lim3
            conc3 = @(t) abs ( fterm(grid3(i,1)) * (exp(-
               Ke*(time1+time2 + t)) - exp(-Ka*(time1+
               time2 + t))) + fterm(grid3(i,2))* (exp(-Ke
               *(time2 + t)) - exp(-Ka*(time2 + t))) +
```

```
            fterm(grid3(i,3))* (exp(-Ke*t) - exp(-Ka*t
            )) - C_tgt_A);
        conc3b = @(t) abs ( Alum(LD(I)*grid3(i,1))*
            exp(-alpha*(time1+time2 + t)) + Blum(LD(I)
            *grid3(i,1))*exp(-beta*(time1 + time2 + t)
            )- (Alum(LD(I)*grid3(i,1))+Blum(LD(I)*
            grid3(i,1)))*exp(-ka*(time1 + time2 + t))
            + Alum(LD(I)*grid3(i,2))*exp(-alpha*(time2
             + t)) + Blum(LD(I)*grid3(i,2))*exp(-beta
            *(time2 + t))- (Alum(LD(I)*grid3(i,2)) +
            Blum(LD(I)*grid3(i,2)))*exp(-ka*(time2 + t
            ))    + Alum(LD(I)*grid3(i,3))*exp(-alpha*
            t) + Blum(LD(I)*grid3(i,3))*exp(-beta*t) -
             (Alum(LD(I)*grid3(i,3))+Blum(LD(I)*grid3(
            i,3)))*exp(-ka*t)  - C_tgt_B) ;
        obj3a(i) = quadl(conc3,0,time3,.1)  ;
        obj3b(i) = quadl(conc3b,0,time3,.1)  ;
end
grid4 = cartprod(D(1,:) , D(2,:) , D(3,:), D(4,:) );
lim4 = size(grid4,1);
for i =1:lim4
        conc4 = @(t) abs ( fterm(grid4(i,1))* (exp(-
            Ke*(time1+time2+time3 + t)) - exp(-Ka*(
            time1+time2+time3 + t)))+ fterm(grid4(i,2)
            )* (exp(-Ke*(time2+time3 + t)) - exp(-Ka*(
            time2+time3 + t))) +  fterm(grid4(i,3))* (
            exp(-Ke*(time3 + t)) - exp(-Ka*(time3 + t)
            )) + fterm(grid4(i,4))* (exp(-Ke*t) - exp
            (-Ka*t)) - C_tgt_A );
        conc4b = @(t) abs ( Alum(LD(I)*grid4(i,1))*
            exp(-alpha*(time1+time2+time3+t)) + Blum(
            LD(I)*grid4(i,1))*exp(-beta*(time1+time2+
            time3+t))- (Alum(LD(I)*grid4(i,1))+Blum(LD
            (I)*grid4(i,1)))*exp(-ka*(time1+time2+
            time3 + t))   +     Alum(LD(I)*grid4(i,2))
            *exp(-alpha*(time2+time3+t)) + Blum(LD(I)*
            grid4(i,2))*exp(-beta*(time2+time3+t))- (
            Alum(LD(I)*grid4(i,2))+Blum(LD(I)*grid4(i
            ,2)))*exp(-ka*(time2+time3 + t))  +
            Alum(LD(I)*grid4(i,3))*exp(-alpha*(time3+t
            )) + Blum(LD(I)*grid4(i,3))*exp(-beta*(
```

```
                time3+t))- (Alum(LD(I)*grid4(i,3))+Blum(LD
                (I)*grid4(i,3)))*exp(-ka*(time3 + t)) +
                Alum(LD(I)*grid4(i,4))*exp(-alpha*t) +
                Blum(LD(I)*grid4(i,4))*exp(-beta*t) - (
                Alum(LD(I)*grid4(i,4))+Blum(LD(I)*grid4(i
                ,4)))*exp(-ka*t) - C_tgt_B) ;
            obj4a(i) = quadl(conc4,0,time4,.1) ;
            obj4b(i) = quadl(conc4b,0,time4,.1) ;
    end
grid5 = cartprod(D(1,:) , D(2,:) , D(3,:), D(4,:) , D
    (5,:));
lim5 = size(grid5,1);
for i =1:lim5
    conc5 = @(t) abs ( fterm(grid5(i,1))* (exp(-Ke*(
        time1+time2+time3+time4 + t)) - exp(-Ka*(time1+
        time2+time3+time4 + t)))+ fterm(grid5(i,2))* (
        exp(-Ke*(time2+time3+time4 + t)) - exp(-Ka*(
        time2+time3+time4 + t))) + fterm(grid5(i,3))*
        (exp(-Ke*(time3+time4 + t)) - exp(-Ka*(time3+
        time4 + t)))+ fterm(grid5(i,4))* (exp(-Ke*(
        time4 + t)) - exp(-Ka*(time4 + t))) + fterm(
        grid5(i,5))* (exp(-Ke*t) - exp(-Ka*t)) -
        C_tgt_A );
     conc5b = @(t) abs ( Alum(LD(I)*grid5(i,1))*exp(-
        alpha*(time1+time2+time3+time4 + t)) + Blum(LD
        (I)*grid5(i,1))*exp(-beta*(time1+time2+time3+
        time4 + t))- (Alum(LD(I)*grid5(i,1))+Blum(LD(I
        )*grid5(i,1)))*exp(-ka*(time1+time2+time3+
        time4 + t)) + Alum(LD(I)*grid5(i,2))*exp
        (-alpha*(time2+time3+time4+t)) + Blum(LD(I)*
        grid5(i,2))*exp(-beta*(time2+time3+time4 + t))
        - (Alum(LD(I)*grid5(i,2))+Blum(LD(I)*grid5(i
        ,2)))*exp(-ka*(time2+time3+time4 + t)) +
        Alum(LD(I)*grid5(i,3))*exp(-alpha*(time3+time4
        +t)) + Blum(LD(I)*grid5(i,3))*exp(-beta*(time3
        +time4+t))- (Alum(LD(I)*grid5(i,3))+Blum(LD(I)
        *grid5(i,3)))*exp(-ka*(time3+time4 + t)) +
        Alum(LD(I)*grid5(i,4))*exp(-alpha*(time4+t)) +
         Blum(LD(I)*grid5(i,4))*exp(-beta*(time4+t))-
        (Alum(LD(I)*grid5(i,4))+Blum(LD(I)*grid5(i,4))
        )*exp(-ka*(time4 + t)) + Alum(LD(I)*grid5(i,5)
```

```
        )*exp(-alpha*t) + Blum(LD(I)*grid5(i,5))*exp(-
        beta*t) - (Alum(LD(I)*grid5(i,5))+Blum(LD(I)*
        grid5(i,5)))*exp(-ka*t) - C_tgt_B) ;
    obj5a(i) = quadl(conc5,0,time5,.1) ;
    obj5b(i) = quadl(conc5b,0,time5,.1) ;
end
grid6 = cartprod(D(1,:) , D(2,:) , D(3,:), D(4,:) , D
   (5,:), D(6,:) );
 lim6 = size(grid6,1);
 for i =1:lim6
    conc6 = @(t) abs ( fterm(grid6(i,1))* (exp(-Ke*(
        time1+time2+time3+time4+time5 + t)) - exp(-Ka
        *(time1+time2+time3+time4+time5 + t)))+ fterm
        (grid6(i,2))* (exp(-Ke*(time2+time3+time4+
        time5 + t)) - exp(-Ka*(time2+time3+time4+
        time5 + t))) + fterm(grid6(i,3))* (exp(-Ke*(
        time3+time4+time5 + t)) - exp(-Ka*(time3+time4
        +time5 + t)))+ fterm(grid6(i,4))* (exp(-Ke*(
        time4+time5 + t)) - exp(-Ka*(time4+time5 + t))
        )+ fterm(grid6(i,5))* (exp(-Ke*(time5 + t)) -
        exp(-Ka*(time5 + t))) + fterm(grid6(i,6))* (
        exp(-Ke*t) - exp(-Ka*t)) - C_tgt_A );
    conc6b = @(t) abs ( Alum(LD(I)*grid6(i,1))*exp(-
        alpha*(time1+time2+time3+time4+time5+t)) +
        Blum(LD(I)*grid6(i,1))*exp(-beta*(time1+time2+
        time3+time4+time5+t))- (Alum(LD(I)*grid6(i,1))
        +Blum(LD(I)*grid6(i,1)))*exp(-ka*(time1+time2+
        time3+time4+time5 + t)) + Alum(LD(I)*
        grid6(i,2))*exp(-alpha*(time2+time3+time4+
        time5+t)) + Blum(LD(I)*grid6(i,2))*exp(-beta*(
        time2+time3+time4+time5+t))- (Alum(LD(I)*grid6
        (i,2))+Blum(LD(I)*grid6(i,2)))*exp(-ka*(time2+
        time3+time4+time5 + t)) + Alum(LD(I)*
        grid6(i,3))*exp(-alpha*(time3+time4+time5+t))
        + Blum(LD(I)*grid6(i,3))*exp(-beta*(time3+
        time4+time5+t))- (Alum(LD(I)*grid6(i,3))+Blum(
        LD(I)*grid6(i,3)))*exp(-ka*(time3+time4+time5
        + t)) + Alum(LD(I)*grid6(i,4))*exp(-alpha*(
        time4+time5+t)) + Blum(LD(I)*grid6(i,4))*exp(-
        beta*(time4+time5+t))- (Alum(LD(I)*grid6(i,4))
        +Blum(LD(I)*grid6(i,4)))*exp(-ka*(time4+time5
```

```
+ t)) + Alum(LD(I)*grid6(i,5))*exp(-alpha*(
time5+t)) + Blum(LD(I)*grid6(i,5))*exp(-beta*(
time5+t))- (Alum(LD(I)*grid6(i,5))+Blum(LD(I)*
grid6(i,5)))*exp(-ka*(time5 + t)) + Alum(LD(I
)*grid6(i,6))*exp(-alpha*t) + Blum(LD(I)*grid6
(i,6))*exp(-beta*t) - (Alum(LD(I)*grid6(i,6))+
Blum(LD(I)*grid6(i,6)))*exp(-ka*t) - C_tgt_B)
;
obj6a(i) = quadl(conc6,0,12,.1) ;
obj6b(i) = quadl(conc6b,0,time6,.1) ;

j = mod(i,lim) ;
if (j ==0)
    j = lim;
end
k = mod(i,lim2);
if (k ==0)
    k = lim2;
end
l = mod(i,lim3);
if (l ==0)
    l = lim3;
end
m = mod(i,lim4);
if (m ==0)
    m = lim4;
end
n = mod(i,lim5);
if (n ==0)
    n = lim5;
end
  tpa = [obj1a(j) , obj2a(k) , obj3a(l) ,
    obj4a(m) ,obj5a(n) , obj6a(i)];
  tpb = [obj1b(j) ,obj2b(k) ,obj3b(l) ,obj4b
    (m), obj5b(n) ,obj6b(i)];
  cuma = 1000*mean(tpa);
  cumb = mean(tpb);
  cumab = w*cuma + (1-w)*cumb;
  data7(i,:) = [ grid6(i,1) obj1a(j) grid6(i
    ,2) obj2a(k) grid6(i,3) obj3a(l) grid6
    (i,4) obj4a(m) grid6(i,5) obj5a(n)
```

```
                    grid6(i,6)   obj6a(i)   cumab];
        end
        sort_data = sortrows(data7, 13);
        sort2 = sort_data(1,:);
        reg(I,:) = sort2(:,[1,3,5,7,9,11]);
        reg(I,:) = min(dmax,reg(I,:));
        reg_B(I,:) = min(dmax_B,LD(I)*reg(I,:));
        varphi = sort2(1,13)
        history(x,:) = reg(I,:);
        if x>1
            for i=1:6
                if abs(reg(I,i) - history(x-1,i)) == 0
                    stop = stop + 1;
                end;
            end;
        end;
        for i = 1:size(reg,2)
            if reg(I,i) == D(i,1)
              D(i,:) = max(0,min(dmax,[D(i,1)*epsilon, D(i
                  ,1), D(i,1)/epsilon]));
           elseif reg(I,i) == D(i,2)
                D(i,:) = max(0,min(dmax,[D(i,2)*epsilon,D(i
                    ,2), D(i,2)/epsilon]));
            elseif reg(I,i) == D(i,3)
                D(i,:) = max(0,min(dmax,[D(i,3)*epsilon, D(
                    i,3), D(i,3)/epsilon]));
             end
    end

        compare = sort_data(1,13);
        varphi(x) = compare;
        x=x+1
        end
    com(I)    = compare;
end
[a,b] = min(com);
if y>1
    hist_phi = [hist_phi a]
else
    hist_phi = a
end
```

```
if y>1
    hist_ld = [hist_ld LD(b)];
else
    hist_ld = LD(b);
end
hist_ld
mu = abs(current_ld - LD(b));
current_ld = LD(b)
LD = [LD(b)*epsilon,LD(b),LD(b)/epsilon];
y = y+1;
end
doseregA = reg(b,:);doseregB = reg_B(b,:);
end
function ft = fterm(D)
Ka = .37; Ke = 0.829; V = 217; F = .7;ft = (F*D*Ka)/(V*(Ka -
   Ke));
end
function ft = Alum(D)
CL = 15;V1 = 215;ka = .13;Q = 13.4;V2 = 1043;f = 1;k12 = Q/V1
   ;k21 = Q/V2;
ke = CL/V1;sum = k12 + k21 + ke;root = sqrt(sum^2 - 4*ke*k21)
   ;
alpha = (sum+root)*0.5;beta = (sum-root)*0.5;
ft = (f*D*ka*(k21 - alpha))/(V1*(ka - alpha)*(beta-alpha));
end
function ft = Blum(D)
CL = 15;V1 = 215;ka = .13;Q = 13.4;V2 = 1043;f = 1;k12 = Q/V1
   ;k21 = Q/V2;
ke = CL/V1;sum = k12 + k21 + ke;root = sqrt(sum^2 - 4*ke*k21)
   ;
alpha = (sum+root)*0.5;beta = (sum-root)*0.5;
ft = (f*D*ka*(k21 - beta))/(V1*(ka - beta)*(alpha - beta));
end
```

D.3 The ED Algorithm for Therapeutic Windows

```
% PK parameters
Ka = .37; Ke = 0.2; V = 24; F = .95;
C_tgt_l = 2.5;% Lower Ctgt
C_tgt_h = 3.5;% Upper Ctgt
nu = .95;      % nu
D_MAX = 250;
D = ones(7,3);
D(:,1) = .1*D_MAX*D(:,1);D(:,2) = .50*D_MAX*D(:,2);D(:,3) =
   D_MAX*D(:,3);
compare = 10000;
time = 6;
history = ones(10000,7);varphi = ones(10000,1);
x=1;stop = 0;
while (x <10000 && stop < 7)
stop = 0;
lim = size(D(1,:),2);
for i = 1:lim
          conc1 = @(t)     nu*max(0,C_tgt_l -  fterm(D(1,i)) *(
             exp(-Ke*t) - exp(-Ka*t))) +  (1-nu)*max(0,fterm(D
             (1,i)) *(exp(-Ke*t) - exp(-Ka*t))-C_tgt_h ) ;
          obj1(i) = quad(conc1,0,time)  ;
         data1(i,:) = [D(1,i)  obj1(i) ];
end
grid2 = cartprod(D(1,:) , D(2,:) );
lim2 = size(grid2,1);
for i =1:lim2
                  conc2 = @(t) nu*max(0,  C_tgt_l - (fterm(
                     grid2(i,1))* (exp(-Ke*(time + t)) - exp(-
                     Ka*(time + t)) ) +  fterm(grid2(i,2))* (
                     exp(-Ke*t) - exp(-Ka*t)) )) + (1-nu)*max
                     (0, fterm(grid2(i,1))* (exp(-Ke*(time + t)
                     ) - exp(-Ka*(time + t)) ) +  fterm(grid2(i
                     ,2))* (exp(-Ke*t) - exp(-Ka*t)) - C_tgt_h)
                     ;
                  obj2(i) = quad(conc2,0,time);
end
grid3 = cartprod(D(1,:) , D(2,:) , D(3,:) );
lim3 = size(grid3,1);
for i =1:lim3
```

```
                conc3 = @(t) nu*max (0, C_tgt_l - (fterm(
                    grid3(i,1)) * (exp(-Ke*(2*time + t)) - exp
                    (-Ka*(2*time + t))) + fterm(grid3(i,2))* (
                    exp(-Ke*(time + t)) - exp(-Ka*(time + t)))
                     + fterm(grid3(i,3))* (exp(-Ke*t) - exp(-
                    Ka*t)))) + (1-nu)*max (0, fterm(grid3(i
                    ,1)) * (exp(-Ke*(2*time + t)) - exp(-Ka
                    *(2*time + t))) + fterm(grid3(i,2))* (exp
                    (-Ke*(time + t)) - exp(-Ka*(time + t))) +
                     fterm(grid3(i,3))* (exp(-Ke*t) - exp(-Ka*
                    t)) - C_tgt_h );
                obj3(i) = quad(conc3,0,time)  ;
end
grid4 = cartprod(D(1,:) , D(2,:) , D(3,:), D(4,:) );
lim4 = size(grid4,1);
for i =1:lim4
                conc4 = @(t) nu*max(0, C_tgt_l - (fterm(grid4
                    (i,1))* (exp(-Ke*(3*time + t)) - exp(-Ka
                    *(3*time + t)))+ fterm(grid4(i,2))* (exp(-
                    Ke*(2*time + t)) - exp(-Ka*(2*time + t)))
                    + fterm(grid4(i,3))* (exp(-Ke*(time + t))
                     - exp(-Ka*(time + t))) + fterm(grid4(i,4)
                    )* (exp(-Ke*t) - exp(-Ka*t))) ) + (1-nu)*
                    max(0, fterm(grid4(i,1))* (exp(-Ke*(3*time
                     + t)) - exp(-Ka*(3*time + t)))+ fterm(
                    grid4(i,2))* (exp(-Ke*(2*time + t)) - exp
                    (-Ka*(2*time + t))) + fterm(grid4(i,3))*
                    (exp(-Ke*(time + t)) - exp(-Ka*(time + t))
                    ) + fterm(grid4(i,4))* (exp(-Ke*t) - exp(-
                    Ka*t))-C_tgt_h )      ;
                obj4(i) = quad(conc4,0,time)  ;
end
grid5 = cartprod(D(1,:) , D(2,:) , D(3,:), D(4,:) , D(5,:));
lim5 = size(grid5,1);
for i =1:lim5
          conc5 = @(t) nu*max(0, C_tgt_l - (fterm(grid5(i,1))
              * (exp(-Ke*(4*time + t)) - exp(-Ka*(4*time + t))
              )+ fterm(grid5(i,2))* (exp(-Ke*(3*time + t)) -
              exp(-Ka*(3*time + t))) + fterm(grid5(i,3))* (
              exp(-Ke*(2*time + t)) - exp(-Ka*(2*time + t)))+
              fterm(grid5(i,4))* (exp(-Ke*(time + t)) - exp(-
```

```
            Ka*(time + t))) + fterm(grid5(i,5))* (exp(-Ke*t)
             - exp(-Ka*t))) ) +  (1-nu)*max(0,  fterm(grid5(
            i,1))* (exp(-Ke*(4*time + t)) - exp(-Ka*(4*time
            + t)))+ fterm(grid5(i,2))* (exp(-Ke*(3*time + t)
            ) - exp(-Ka*(3*time + t))) +  fterm(grid5(i,3))*
             (exp(-Ke*(2*time + t)) - exp(-Ka*(2*time + t)))
            + fterm(grid5(i,4))* (exp(-Ke*(time + t)) - exp
            (-Ka*(time + t))) + fterm(grid5(i,5))* (exp(-Ke*
            t) - exp(-Ka*t)) -C_tgt_h);
        obj5(i) = quad(conc5,0,time)  ;
end
grid6 = cartprod(D(1,:) , D(2,:) , D(3,:), D(4,:) , D(5,:), D
   (6,:) );
lim6 = size(grid6,1);
for i =1:lim6
      conc6 = @(t) nu*max(0, C_tgt_l - (fterm(grid6(i,1))* (
         exp(-Ke*(5*time + t)) - exp(-Ka*(5*time + t)))+
         fterm(grid6(i,2))* (exp(-Ke*(4*time + t)) - exp(-Ka
         *(4*time + t))) +  fterm(grid6(i,3))* (exp(-Ke*(3*
         time + t)) - exp(-Ka*(3*time + t)))+ fterm(grid6(i
         ,4))* (exp(-Ke*(2*time + t)) - exp(-Ka*(2*time + t))
         )+ fterm(grid6(i,5))* (exp(-Ke*(time + t)) - exp(-Ka
         *(time + t))) + fterm(grid6(i,6))* (exp(-Ke*t) - exp
         (-Ka*t)))) +  (1-nu)*max(0, (fterm(grid6(i,1))* (exp
         (-Ke*(5*time + t)) - exp(-Ka*(5*time + t)))+ fterm(
         grid6(i,2))* (exp(-Ke*(4*time + t)) - exp(-Ka*(4*
         time + t))) +  fterm(grid6(i,3))* (exp(-Ke*(3*time +
          t)) - exp(-Ka*(3*time + t)))+ fterm(grid6(i,4))* (
         exp(-Ke*(2*time + t)) - exp(-Ka*(2*time + t)))+
         fterm(grid6(i,5))* (exp(-Ke*(time + t)) - exp(-Ka*(
         time + t))) + fterm(grid6(i,6))* (exp(-Ke*t) - exp(-
         Ka*t)))-C_tgt_h)          ;
      obj6(i) = quad(conc6,0,time)  ;
end
grid7 = cartprod(D(1,:) , D(2,:) , D(3,:), D(4,:) , D(5,:), D
   (6,:), D(7,:) );
lim7 = size(grid7,1);
for i =1:lim7
         conc7 = @(t) nu*max(0,C_tgt_l - ( fterm(grid7(i,1)
            )* (exp(-Ke*(6*time + t)) - exp(-Ka*(6*time + t
            )))+ fterm(grid7(i,2))* (exp(-Ke*(5*time + t))
```

```
    - exp(-Ka*(5*time + t))) +  fterm(grid7(i,3))*
    (exp(-Ke*(4*time + t)) - exp(-Ka*(4*time + t)))
    + fterm(grid7(i,4))* (exp(-Ke*(3*time + t)) -
    exp(-Ka*(3*time + t)))+ fterm(grid7(i,5))* (exp
    (-Ke*(2*time + t)) - exp(-Ka*(2*time + t))) +
    fterm(grid7(i,6))* (exp(-Ke*(time + t)) - exp(-
    Ka*(time + t))) + fterm(grid7(i,7))* (exp(-Ke*t
    ) - exp(-Ka*t)) ))  +      (1-nu)*max(0, ( fterm(
    grid7(i,1))* (exp(-Ke*(6*time + t)) - exp(-Ka
    *(6*time + t)))+ fterm(grid7(i,2))* (exp(-Ke
    *(5*time + t)) - exp(-Ka*(5*time + t))) +
    fterm(grid7(i,3))* (exp(-Ke*(4*time + t)) - exp
    (-Ka*(4*time + t)))+ fterm(grid7(i,4))* (exp(-
    Ke*(3*time + t)) - exp(-Ka*(3*time + t)))+
    fterm(grid7(i,5))* (exp(-Ke*(2*time + t)) - exp
    (-Ka*(2*time + t))) + fterm(grid7(i,6))* (exp(-
    Ke*(time + t)) - exp(-Ka*(time + t))) + fterm(
    grid7(i,7))* (exp(-Ke*t) - exp(-Ka*t)) )-
    C_tgt_h)          ;
  obj7(i) = quad(conc7,0,time)   ;
  j = mod(i,lim);if (j ==0)
                    j = lim;
                  end
  k = mod(i,lim2);if (k ==0)
                    k = lim2;
                  end
 l = mod(i,lim3);if (l ==0)
                    l = lim3;
                  end
m = mod(i,lim4); if (m ==0)
                    m = lim4;
                  end
n = mod(i,lim5);if (n ==0)
                    n = lim5;
                  end
p = mod(i,lim6);if (p ==0)
                    p = lim6;
                    end
tp = [obj1(j) , obj2(k) , obj3(l) , obj4(m) , obj5(
   n) , obj6(p), obj7(i)];
cum = mean(tp);
```

```
            data7(i,:) = [ grid7(i,1) obj1(j)  grid7(i,2)  obj2
               (k) grid7(i,3)  obj3(l) grid7(i,4)  obj4(m)
               grid7(i,5)  obj5(n) grid7(i,6)  obj6(p) grid7(i
               ,7) obj7(i) cum];
end
sort_data = sortrows(data7, 15);
sort2 = sort_data(1,:);
reg = sort2(:,[1,3,5,7,9,11,13])
history(x,:) = reg;
if x>1
    for i=1:7
         if abs(reg(i) - history(x-1,i))<.01
            stop = stop + 1;
         end;
    end;
end;
epsilon  = .99;
for i = 1:size(reg,2)
     if reg(i) == D(i,1)
         D(i,:) = min(D_MAX,[D(i,1)*epsilon,D(i,1), D(i,1)
            *(1/epsilon)]);
     elseif reg(i) == D(i,2)
            D(i,:) = min(D_MAX,[D(i,2)*epsilon,D(i,2), D(i,2)
               *(1/epsilon)]);


        elseif reg(i) == D(i,3)
            D(i,:) = min(D_MAX,[D(i,3)*epsilon,D(i,3), D(i,3)
               *(1/epsilon)]);
     end
end
compare = sort_data(1,15);varphi(x) = compare;
x=x+1;
end
```

D.4 The ED Algorithm in an Adaptive Trial

```
function master_big_picture_design_nocov
N_SIMUL = 1000;
ncohorts = 10;
rows = N_SIMUL*ncohorts;
sr_record = ones(N_SIMUL,4);
 beta_record = ones(rows,8);
   dose_reg_record  =  ones(rows,6);
   sam_time_record  = ones(rows,4);
   phi_record = ones(rows,2);
for i = 1:N_SIMUL
    start = tic;
    [ beta_est, sam_time, dose_reg, phi,sr] = simulator(
       ncohorts) ;
   beta_record([1 + (i-1)*ncohorts: i*ncohorts],:) = [i*ones(
      ncohorts,1) beta_est];
   dose_reg_record([1 + (i-1)*ncohorts: i*ncohorts],:)  =
          [i*ones(ncohorts,1) dose_reg];
   sam_time_record([1 + (i-1)*ncohorts: i*ncohorts],:)  = [i*
      ones(ncohorts,1) sam_time];
   phi_record([1 + (i-1)*ncohorts: i*ncohorts],:)  = [i*ones(
      ncohorts,1) phi];
   sr_record(i,:)  = [i sr];
   time_to_comp = toc(start);
end
 save('bigpicnocovrnd');
end
function  [ beta_est, sam_time, dose_reg, phi,sr] = simulator
   (ncohorts)
ndose = 5;   %%% No. of doses
mi = 3; %%% Number of observations per subject
 N = 10;  %number of subjects per cohort
sr = ncohorts*ones(1,3);
 fid = fopen('PK_samples.m', 'w');
 srflag = zeros(1,3);
%%% PK parameters
%%%   Prior values of the parameters
Ka = 1; Ke = .2; V = 20;omega1 = .05;omega2 = .15;omega3 =
   .05; sigma = .15;
% True values of the parameters
```

```
tr_Ka = .85;tr_Ke = .15;tr_V = 17;;tr_omega10 = .1;tr_omega20
    = .1;tr_omega30 = .1; tr_sigma0 = .1;
true_params = [tr_Ka,tr_Ke,tr_V ,tr_omega10,tr_omega20,
   tr_omega30, tr_sigma0];
dose_reg = zeros(ncohorts,ndose);
phi = zeros(ncohorts,1);
beta_est = zeros(ncohorts,7);
sam_time = zeros(ncohorts,mi);
ctr = 0;
for i = 1:ncohorts
            [no_need dose_reg(i,:)] = ED_algorithm(Ka,Ke,V,
               dose_reg(i,:),0);
             [phi(i) no_need2] = ED_algorithm(tr_Ka,tr_Ke,tr_V
                ,dose_reg(i,:),1);
             phi_and_dosereg = [phi dose_reg];
             fid_tp= fopen('C:\Users\Kabir.Soeny\Desktop\PHD\
                Big_picture\current_dose.m', 'w');
             fprintf(fid_tp,'%f %f %f %f %f',dose_reg(i,:));
             fclose(fid_tp);
          %%% Stopping rule 1 %%
             if i>1 && srflag(1) == 0
                 if abs(sum(dose_reg(i,:)) - sum(dose_reg(i -
                    1,:)))/sum(dose_reg(i - 1,:)) <= .05
                     sr(1) = i;
                     srflag(1) = 1;
                 end
             end
       %%% Stopping rule 2 %%%
        if i>1 && srflag(2) == 0
                 if abs((dose_reg(i,:)) - (dose_reg(i - 1,:)))
                     <= .05*dose_reg(i - 1,:)
                     sr(2) = i;
                     srflag(2) = 1;
                 end
         end
         %%% Find D - optimal sampling time points %%%%%
            [popedOutput,globalStructure,strRunDirectoryName]=
               poped(input_ed_algo(dose_reg(i,:),Ka,Ke,V,omega1
               ,omega2,omega3,sigma,N));
            sam_time(i,:) = sort(popedOutput.xt)
            draw_samples(dose_reg(i,:),i, sam_time(i,:));
```

```
            [Ka,Ke,V,omega1,omega2,omega3,sigma] = PK_estimates(
                dose_reg(i,:),Ka,Ke,V,omega1,omega2,omega3,sigma,
                i);
            beta_est(i,:) = [Ka,Ke,V,omega1,omega2,omega3,sigma
                ];
        end
        fclose all;
end
function [compare dose_reg] = ED_algorithm(Ka,Ke,V,fixed_dose
    ,measure)
C_tgt = 5;
D_MAX = 200;
ndose = 5;
epsilon  = .99;
kappa = 10;
D = ones(ndose,3);
D(:,1) = .1*D_MAX*D(:,1);
D(:,2) = .50*D_MAX*D(:,2);
D(:,3) = D_MAX*D(:,3);
if measure ==1
    kappa = 0;
    for m = 1:5
        D(m,:) = fixed_dose(:,m);
    end
end
compare = 10000;
time = 8;
history = ones(10000,5);
varphi = ones(10000,1);
x=1;
stop = 0;
while (x <10000 && stop < ndose)
stop = 0;
lim = size(D(1,:),2);
for i = 1:lim
          conc1 = @(t) abs (     fterm_big(D(1,i),Ka,Ke,V) *(
              exp(-Ke*t) - exp(-Ka*t))  - C_tgt );
          obj1(i) = quadl(conc1,0,time)  ;
end
grid2 = cartprod(D(1,:) , D(2,:) );
lim2 = size(grid2,1);
```

```
for i =1:lim2
        conc2 = @(t) abs ( fterm_big(grid2(i,1),Ka,Ke,V)*(exp
           (-Ke*(time + t)) - exp(-Ka*(time + t)) ) +
           fterm_big(grid2(i,2),Ka,Ke,V)* (exp(-Ke*t) - exp(-
           Ka*t))- C_tgt);
        obj2(i) = quadl(conc2,0,time);
end

grid3 = cartprod(D(1,:) , D(2,:) , D(3,:) );
lim3 = size(grid3,1);
for i =1:lim3
       conc3 = @(t) abs ( fterm_big(grid3(i,1),Ka,Ke,V)*(exp
          (-Ke*(2*time + t)) - exp(-Ka*(2*time + t))) +
          fterm_big(grid3(i,2),Ka,Ke,V)* (exp(-Ke*(time + t))
          - exp(-Ka*(time + t))) + fterm_big(grid3(i,3),Ka,
          Ke,V)* (exp(-Ke*t)- exp(-Ka*t)) - C_tgt);
                obj3(i) = quadl(conc3,0,time) ;
end
grid4 = cartprod(D(1,:) , D(2,:) , D(3,:), D(4,:) );
lim4 = size(grid4,1);
for i =1:lim4
                conc4 = @(t) abs ( fterm_big(grid4(i,1),Ka,Ke
                   ,V)* (exp(-Ke*(3*time + t)) - exp(-Ka*(3*
                   time + t)))+ fterm_big(grid4(i,2),Ka,Ke,V)
                   *(exp(-Ke*(2*time + t)) - exp(-Ka*(2*time
                   + t))) + fterm_big(grid4(i,3),Ka,Ke,V)*(
                   exp(-Ke*(time + t))- exp(-Ka*(time + t)))
                   + fterm_big(grid4(i,4),Ka,Ke,V)* (exp(-Ke*
                   t) - exp(-Ka*t)) - C_tgt );
                obj4(i) = quadl(conc4,0,time) ;
end
grid5 = cartprod(D(1,:) , D(2,:) , D(3,:), D(4,:) , D(5,:));
lim5 = size(grid5,1);
for i =1:lim5
                conc5 = @(t) abs ( fterm_big(grid5(i,1),Ka,
                   Ke,V)* (exp(-Ke*(4*time + t)) - exp(-Ka
                   *(4*time + t)))+ fterm_big(grid5(i,2),Ka,
                   Ke,V) * (exp(-Ke*(3*time + t)) - exp(-Ka
                   *(3*time + t))) + fterm_big(grid5(i,3),Ka,
                   Ke,V) * (exp(-Ke*(2*time + t)) - exp(-Ka
                   *(2*time + t))) + fterm_big(grid5(i,4),Ka,
```

```
                    Ke,V)*(exp(-Ke*(time + t)) - exp(-Ka*(time
                     + t))) + fterm_big(grid5(i,5),Ka,Ke,V) *
                    (exp(-Ke*t) - exp(-Ka*t)) - C_tgt );
                obj5(i) = quadl(conc5,0,time)  ;
                j = mod(i,lim); if (j ==0)
                                    j = lim;
                                end
                k = mod(i,lim2); if (k ==0)
                                     k = lim2;
                                end
                l = mod(i,lim3); if (l ==0)
                                     l = lim3;
                                end
                m = mod(i,lim4);if (m ==0)
                                      m = lim4;
                                end
                tp = [obj1(j) , obj2(k) , obj3(l) , obj4(m) ,
                     obj5(i)];
                cum = mean(tp);
                data7(i,:) = [ grid5(i,1) obj1(j)  grid5(i,2)
                      obj2(k) grid5(i,3) obj3(l)  grid5(i,4)
                    obj4(m) grid5(i,5) obj5(i)  cum];
end
sort_data = sortrows(data7, 11);
sort2 = sort_data(1,:);
reg = sort2(:,[1,3,5,7,9]);
history(x,:) = reg;
if x>1
    for i=1:5
        if abs(reg(i) - history(x-1,i))== 0
            stop = stop + 1;
        end;
    end;
end;
for i = 1:size(reg,2)
     if reg(i) == D(i,1)
         D(i,:) = max(0,min(D_MAX,[D(i,1)- kappa, D(i,1), D(i
             ,1) + kappa]));
     elseif reg(i) == D(i,2)
         D(i,:) = max(0,min(D_MAX,[D(i,2) - kappa,D(i,2), D(i
             ,2) + kappa]));   %%% Discrete
```

```
      elseif reg(i) == D(i,3)
          D(i,:) = max(0,min(D_MAX,[D(i,3) - kappa, D(i,3), D(
              i,3)+ kappa]));
      end
end
compare = sort_data(1,11);
varphi(x) = compare; x=x+1;
end
dose_reg = reg;
 end
function ft = fterm_big(D,Ka,Ke,V)
ft = ((D*Ka)/(V*(Ka - Ke)));
end
function draw_samples(dose,cohort_id,sampling_time)
%%% True PK parameters %%%
 ka = .85; ke = .15; v = 17;phi = [ka, ke, v];
        N = 10;
 tau = 8;
varcov = [  .1     0     0     ;
             0     .1   0     ;
             0      0    .01];
sigma = .05;
sampling_time = sampling_time';
n = size(sampling_time,1);
tot_pat = N*n;
phi_mat = repmat(phi,tot_pat,1);
sigma_sq = sqrt(sigma);
 no_dose = size(dose,2);
  for i = 1:N
   rnd = mvnrnd([0,0,0],varcov);
    for j=1:n
        b_rnd((i-1)*n + j,:) = rnd;
    end
 end
sam_time = repmat(sampling_time,N,1);
phi_rand =  phi_mat.*exp(b_rnd);
sim_conc = zeros(tot_pat,1);
 parfor k = 1:tot_pat
     concn = 0;
       for j=1:no_dose
            tp1 =  (phi_rand(k,3)*(phi_rand(k,1) - phi_rand(k
```

```
                ,2))) ;
            tp2 = (dose(j)*phi_rand(k,1));
            tp = tp2/tp1;
            concn = concn + ((sam_time(k)>=((j-1)*tau))) *
                ...
                 (tp* (exp(-phi_rand(k,2)*abs(sam_time(k)-(j
                    -1)*tau)) - ...
                 exp(-phi_rand(k,1)*abs(sam_time(k)-(j-1)*tau)
                    )));
end
sim_conc(k) = concn * exp(normrnd(0,sigma_sq)); % For
    proportional error model
%sim_conc(k) =  concn + normrnd(0,sigma);   % For additive
    model
end
subject = (cohort_id-1)*N + 1;
parfor i = 1:N
      for j = 1:n
          if (i ==1 && j==1)
              continue;
          else
          subject = [subject (cohort_id-1)*N + i];
          end
      end
  end
  subject = subject';
  data = [subject sam_time sim_conc];
  fid = fopen('C:\Users\Kabir.Soeny\Desktop\PHD\Big_picture\
     PK_samples.m', 'a');
  fprintf(fid,'%d  %8.4f %8.4f\n',data(:,1:3)');
end
function [Ka, Ke, V,omega1,omega2,omega3,sigma] =
    PK_estimates(dose,Ka,Ke,V,omega1,omega2,omega3,sigma,
    cohort_id)
 phi0 = [Ka, Ke, V];
phi0 = log(phi0);
 tau = 8;
 N = 10;
 n = 3;
mod_conc1 = @(phi,t)  ((dose(1)*phi(1))/(phi(3)*(phi(1) - phi
    (2)))) * (exp(-phi(2)*t) - exp(-phi(1)*t)) + (t >= tau)
```

```
    .*(((dose(2)*phi(1))/(phi(3) * (phi(1) - phi(2)))) * (exp
    (-phi(2)*(t-tau)) - exp(-phi(1)*(t-tau)))) + (t>= (2*tau))
    .*(((dose(3)*phi(1))/(phi(3)*(phi(1) - phi(2)))) *(exp(-
    phi(2)*(t-2*tau)) - exp(-phi(1)*(t-2*tau)))) + (t>=3*tau)
    .*(((dose(4)*phi(1))/(phi(3)*(phi(1) -phi(2)))) *(exp(-phi
    (2)*(t-3*tau)) - exp(-phi(1)*(t-3*tau)))) + (t>=4*tau).*
    (((dose(5)*phi(1))/(phi(3)*(phi(1) - phi(2))))*(exp(-phi
    (2)*(t-4*tau)) - exp(-phi(1)*(t-4*tau))));
tplot = 0:0.01:80;
load C:\Users\Kabir.Soeny\Desktop\PHD\Big_picture\PK_samples.
   m;
subject = PK_samples(:,1);
time = PK_samples(:,2);
conc = PK_samples(:,3);
dp = [subject conc];
P = [1 0 0;0 1 0;0 0 1];
xform = [1 1 1];
options = statset('nlmefit');
options = statset(options,'TolX',1e-8,'FunValCheck','Off');
 set(gcf,'visible','off')
[phi,PSI,stats,b] = nlmefit(time,conc,subject, [],mod_conc1,
   phi0, 'REParamsSelect',[1 2 3],'CovPattern',P,'ErrorModel'
   ,'exponential','ParamTransform',xform,'Options',options,'
   ApproximationType','LME')
     set(0,'DefaultFigureVisible','off');
    Ka = exp(phi(1));
    Ke = exp(phi(2));
    V = exp(phi(3));
    sigma =  stats.mse;
end
%%% Adapted from PopED software %%%%
 function [popedInput] = input_ed_algo(dose,Ka,Ke,V,omega1,
    omega2,omega3,sigma,Nc)
popedInput.strPopEDVersion='2.13';popedInput.ng=3;popedInput.
   nbpop=3;popedInput.nb=3;
popedInput.ndocc=0;popedInput.nx=0;popedInput.na=0;
popedInput.NumOcc=0;popedInput.m=1;popedInput.maxni=3;
popedInput.minni=3;popedInput.design.groupsize=Nc;
popedInput.design.maxgroupsize=Nc;popedInput.design.
   mingroupsize=Nc;
popedInput.design.maxtotgroupsize=Nc;popedInput.design.
```

```
    mintotgroupsize=Nc;
popedInput.d_switch=1;popedInput.iApproximationMethod=0;
popedInput.iFOCENumInd=1000;popedInput.iEDCalculationType=0;
popedInput.bUseRandomSearch=1;popedInput.
    bUseStochasticGradient=1;
popedInput.bUseLineSearch=1;popedInput.bUseExchangeAlgorithm
    =0;
popedInput.bUseBFGSMinimizer=0;popedInput.ofv_calc_type=1;
popedInput.prior_fim=zeros(1,0)';popedInput.optsw=[ 0 1 0 0
    0];
popedInput.line_opta=zeros(1,0)';popedInput.line_optx=zeros
    (1,0)';
popedInput.dSeed=-1;popedInput.design.groupsize=1;popedInput.
    design.maxgroupsize=1;
popedInput.design.mingroupsize=1;popedInput.design.
    maxtotgroupsize=0;
popedInput.design.mintotgroupsize=0;popedInput.design.sigma=[
    sigma 0;0 .00001];
popedInput.design.bpop=[ 0 Ka 0; 0 Ke 0; 0 V 0];
popedInput.design.d=[ 0 omega1 0;0 omega2 0; 0 omega3 0];
popedInput.design.covd=[ 0 0 0];popedInput.design.docc=zeros
    (3,0)';
popedInput.design.covdocc=zeros(0,1)';
popedInput.design.ni=3;popedInput.design.xt=[ 1 20 40];
popedInput.design.maxxt=[ 42 42 42];popedInput.design.minxt=[
     .1 .1 .1];
popedInput.design.x=zeros(0,1)';popedInput.design.discrete_x=
    cell(0,1)';
popedInput.design.a=zeros(0,1)';popedInput.design.maxa=zeros
    (0,1)';
popedInput.design.mina=zeros(0,1)';popedInput.design.
    model_switch=[ 1 1 1];
popedInput.notfixed_bpop=[ 1 1 1];popedInput.notfixed_d=[ 1 1
     1];
popedInput.notfixed_covd=[ 0 0 0];
popedInput.notfixed_docc=zeros(0,1)';popedInput.
    notfixed_covdocc=zeros(0,1)';
popedInput.notfixed_sigma=[ 1 0];popedInput.notfixed_covsigma
    =0;
popedInput.bUseGrouped_xt=0;popedInput.design.G=[ 1 2 3];
popedInput.bUseGrouped_a=0;popedInput.design.Ga=zeros(0,1)';
```

```
popedInput.bUseGrouped_x=0;popedInput.design.Gx=zeros(0,1)';
popedInput.ff_file='C:\Users\Kabir.Soeny\Desktop\PHD\
    Big_picture\model_input_big_picture.m';
popedInput.fg_file='C:\Users\Kabir.Soeny\Desktop\PHD\
    Big_picture\sfg.m';
popedInput.fError_file='C:\Users\Kabir.Soeny\Desktop\PHD\
    Big_picture\expo_error_model.m';
popedInput.strUserDistributionFile='';
popedInput.strEDPenaltyFile='';
popedInput.strAutoCorrelationFile='';
popedInput.modtit='One_comp_mul_dose_ed_algo';
popedInput.bShowGraphs=0;popedInput.use_logfile=0;
popedInput.output_file='C:\Users\Kabir.Soeny\Desktop\PHD\
    Big_picture\Poped\output_LS_2.txt';
popedInput.output_function_file='function_output';
popedInput.strIterationFileName='';
popedInput.strRunFile='';popedInput.m1_switch=0;popedInput.
    m2_switch=0;
popedInput.hle_switch=0;popedInput.gradff_switch=0;
popedInput.gradfg_switch=0;popedInput.bLHS=0;
popedInput.ourzero=1e-001;popedInput.rsit_output=100;
popedInput.sgit_output=100;popedInput.hm1=0.001;
popedInput.hlf=0.001;popedInput.hlg=0.001;
popedInput.hm2=0.001;popedInput.hgd=0.001;
popedInput.hle=0.001;popedInput.AbsTol=1e-01;
popedInput.RelTol=1e-01;popedInput.iDiffSolverMethod=0;
popedInput.bUseMemorySolver=0;popedInput.iFIMCalculationType
    =0;
popedInput.rsit=125;popedInput.sgit=75;popedInput.intrsit
    =100;popedInput.intsgit=50;
popedInput.maxrsnullit=100;popedInput.convergence_eps=1e-001;
popedInput.rslxt=4;popedInput.rsla=4;popedInput.cfaxt=0.01;
popedInput.cfaa=0.01;popedInput.bGreedyGroupOpt=0;popedInput.
    EACriteria=1;
popedInput.EAStepSize=0.01;popedInput.EANumPoints=0;
popedInput.EAConvergenceCriteria=1e-010;
popedInput.bEANoReplicates=0;popedInput.
    BFGSConvergenceCriteriaMinStep=1e-001;
popedInput.BFGSProjectedGradientTol=0.0001;
popedInput.BFGSTolerancef=0.001;popedInput.BFGSToleranceg
    =0.9;
```

```
popedInput.BFGSTolerancex=0.1;popedInput.ED_samp_size=45;
popedInput.ED_diff_it=30;popedInput.ED_diff_percent=10;
popedInput.line_search_it=10;popedInput.
    iNumSearchIterationsIfNotLineSearch=10;
popedInput.CriterionOptions.ds_index=[ 0 0 0 0 0 0 0];
popedInput.parallelSettings.iCompileOption=-1;
popedInput.parallelSettings.iUseParallelMethod=1;
popedInput.parallelSettings.
    strAdditionalMCCCompilerDependencies='';
popedInput.parallelSettings.strExecuteName='calc_fim.exe';
popedInput.parallelSettings.iNumProcesses=2;
popedInput.parallelSettings.iNumChunkDesignEvals=-2;
popedInput.parallelSettings.strMatFileInputPrefix='
    parallel_input';
popedInput.parallelSettings.strMatFileOutputPrefix='
    parallel_output';
popedInput.parallelSettings.strExtraRunOptions='';
popedInput.parallelSettings.dPollResultTime=1.000000e-001;
popedInput.parallelSettings.strFunctionInputName='
    function_input';
popedInput.parallelSettings.bParallelRS=0;popedInput.
    parallelSettings.bParallelSG=0;
popedInput.parallelSettings.bParallelLS=0;
popedInput.parallelSettings.bParallelMFEA=0;popedInput.
    user_data={};
 end
```

D.5 Dose Individualization Using the ED Algorithm

```
function master_big_picture
clear all;
ndose = 5;   % No. of doses
mi = 3; % Number of observations per subject
N = 10; % Group Size
% PK parameters
Ka = .89; Ke = .16; V = 17.59;omega1 = .1;omega2 = .1;omega3
   = .1;sigma = .1;
dose_reg = zeros(1,ndose);
beta_est = zeros(2,7);
sam_time = zeros(1,mi);
ctr = 0;
beta_est(1,:) = [Ka,Ke,V,omega1,omega2,omega3,sigma];
[dose_reg(:,:) tp] = ED_algorithm(Ka,Ke,V,0,0)
fid_tp = ...
fopen('C:\Users\Kabir.Soeny\Desktop\PHD\Big_picture\
   current_dose.m', 'w');
fprintf(fid_tp,'%f %f %f %f %f',dose_reg(:,:));
fclose(fid_tp);
sam_time(:,:) = [.1 3.66 10];
draw_samples(dose_reg(:,:),1, sam_time(:,:),Ka,Ke,V);
[Ka,Ke,V,omega1,omega2,omega3,sigma,PHI,col] = ...
PK_estimates(dose_reg(1,:),Ka,Ke,V,omega1,omega2,omega3,sigma
   ,1);
beta_est(2,:) = [Ka,Ke,V,omega1,omega2,omega3,sigma]
 for k = 1:N
    [ind_dose(k,:) varphi(k)] = ED_algorithm(PHI(1,k),PHI(2,k
       ),PHI(3,k),1,dose_reg(1,1));
 end
out =  [ind_dose varphi']
RES = zeros(mi,N);
figure(2)
tplot = 0:0.1:80;
tau = 12;
ctgt = 5;
for I = 1:N
   fitted_model = @(t)  ((ind_dose(I,1)*PHI(1,I))/(PHI(3,I)*(
```

```
        PHI(1,I) - PHI(2,I)))) *(exp(-PHI(2,I)*t) - exp(-PHI(1,
        I)*t)) + (t > tau).*(((ind_dose(I,2)*PHI(1,I))/(PHI(3,I
        )*(PHI(1,I) - PHI(2,I))))*(exp(-PHI(2,I)*(t-tau))- exp
        (-PHI(1,I)*(t-tau)))) +(t>2*tau).*(((ind_dose(I,3)*PHI
        (1,I))/(PHI(3,I)*(PHI(1,I) - PHI(2,I)))) *(exp(-PHI(2,I
        )*(t-2*tau)) - exp(-PHI(1,I)*(t-2*tau)))) +(t>3*tau)
        .*(((ind_dose(I,4)*PHI(1,I))/(PHI(3,I)*(PHI(1,I) - PHI
        (2,I)))) *(exp(-PHI(2,I)*(t-3*tau)) - exp(-PHI(1,I)*(t
        -3*tau)))) +(t>4*tau).*(((ind_dose(I,5)*PHI(1,I))/(PHI
        (3,I)*(PHI(1,I) - PHI(2,I)))) *(exp(-PHI(2,I)*(t-4*tau)
        ) - exp(-PHI(1,I)*(t-4*tau))));
subplot(2,5,I)
set(gca,'XTick',[0:12:72],'fontsize',10)
hold all;
plot(tplot,fitted_model(tplot),'--','Color',col(I,:),'
   linewidth',2.5)
plot([0,60],[ctgt,ctgt],'b','linewidth',2.5)
hold off
axis([0 80 0 18])
xlabel('Time (hours)')
ylabel('Concentration (mg/L)')
legend(num2str(I))
end
save('wrksp');
fclose all;
end
function [dose_reg compare] = ED_algorithm(Ka,Ke,V,fit,d1)
% Target Conc %
C_tgt = 5;
% MAX DOSE and no. of doses %
D_MAX = 300;
ndose = 5;
% Resolution %
epsilon = .99;
kappa = 2;
D = ones(ndose,3);
D(:,1) = .1*D_MAX*D(:,1);
D(:,2) = .50*D_MAX*D(:,2);
D(:,3) = D_MAX*D(:,3);
compare = 10000;
time = 12;
```

```
history = ones(10000,5);
varphi = ones(10000,1);
x=1;
stop = 0;
while (x <10000 && stop < ndose)
    stop = 0;
    lim = size(D(1,:),2);
    for i = 1:lim
        if fit == 1
            D(1,i) = d1;
        end
        conc1 = @(t) abs (     fterm_big(D(1,i),Ka,Ke,V) *(exp
           (-Ke*t)- exp(-Ka*t))  - C_tgt );
        obj1(i) = quad(conc1,0,time)  ;
    end
   grid2 = cartprod(D(1,:) , D(2,:) );
    lim2 = size(grid2,1);
    for i =1:lim2
        conc2 = @(t) abs ( fterm_big(grid2(i,1),Ka,Ke,V)* ...
        (exp(-Ke*(time + t)) - exp(-Ka*(time + t)) ) + ...
         fterm_big(grid2(i,2),Ka,Ke,V)* (exp(-Ke*t) - exp(-Ka
            *t) )...
           - C_tgt);
         obj2(i) = quad(conc2,0,time);
end
    grid3 = cartprod(D(1,:) , D(2,:) , D(3,:) );
    lim3 = size(grid3,1);
for i =1:lim3
          conc3 = @(t) abs ( fterm_big(grid3(i,1),Ka,Ke,V) *
             (exp(-Ke*(2*time + t)) - exp(-Ka*(2*time + t)))
             + fterm_big(grid3(i,2),Ka,Ke,V)* (exp(-Ke*(time
             + t))- exp(-Ka*(time + t))) +  fterm_big(grid3(i
             ,3),Ka,Ke,V)* (exp(-Ke*t)- exp(-Ka*t)) - C_tgt);
          obj3(i) = quad(conc3,0,time)  ;
end
grid4 = cartprod(D(1,:) , D(2,:) , D(3,:), D(4,:) );
lim4 = size(grid4,1);
for i =1:lim4
          conc4 = @(t) abs ( fterm_big(grid4(i,1),Ka,Ke,V)* (
             exp(-Ke*(3*time + t)) - exp(-Ka*(3*time + t)))+
             fterm_big(grid4(i,2),Ka,Ke,V)* (exp(-Ke*(2*time
```

```
            + t)) - exp(-Ka*(2*time + t))) + fterm_big(grid4
            (i,3),Ka,Ke,V)*(exp(-Ke*(time + t))- exp(-Ka*(
            time + t))) + fterm_big(grid4(i,4),Ka,Ke,V)* (
            exp(-Ke*t) - exp(-Ka*t)) - C_tgt );
        obj4(i) = quad(conc4,0,time)  ;
end
grid5 = cartprod(D(1,:) , D(2,:) , D(3,:), D(4,:) , D(5,:));
lim5 = size(grid5,1);
for i =1:lim5
        conc5 = @(t)  abs ( fterm_big(grid5(i,1),Ka,Ke,V)*
            (exp(-Ke*(4*time + t)) - exp(-Ka*(4*time + t)))+
             fterm_big(grid5(i,2),Ka,Ke,V) * (exp(-Ke*(3*
            time + t)) - exp(-Ka*(3*time + t))) + fterm_big(
            grid5(i,3),Ka,Ke,V) * (exp(-Ke*(2*time + t)) -
            exp(-Ka*(2*time + t))) + fterm_big(grid5(i,4),Ka
            ,Ke,V)*(exp(-Ke*(time + t)) - exp(-Ka*(time + t)
            )) + fterm_big(grid5(i,5),Ka,Ke,V) * (exp(-Ke*t)
             - exp(-Ka*t)) - C_tgt );
               obj5(i) = quad(conc5,0,time)  ;
        j = mod(i,lim);if (j ==0)
                          j = lim;
                       end
        k = mod(i,lim2); if (k ==0)
                          k = lim2;
                        end
        l = mod(i,lim3); if (l ==0)
                          l = lim3;
                        end
        m = mod(i,lim4);if (m ==0)
                          m = lim4;
                        end
       tp = [obj1(j) , obj2(k) , obj3(l) , obj4(m) , obj5(i)
          ];
              cum = mean(tp);
       data7(i,:) = [ grid5(i,1) obj1(j)  grid5(i,2)  obj2(k)
           grid5(i,3) obj3(l)  grid5(i,4)  obj4(m) grid5(i,5)
           obj5(i)  cum];
end
sort_data = sortrows(data7, 11);
sort2 = sort_data(1,:);
reg = sort2(:,[1,3,5,7,9]);
```

```
history(x,:) = reg;
if x>1
    for i=1:5
        if abs(reg(i) - history(x-1,i))== 0
            stop = stop + 1;
        end;
    end;
end;

for i = 1:size(reg,2)
    if reg(i) == D(i,1)
        D(i,:) = max(0,min(D_MAX,[D(i,1)- kappa, D(i,1), D(i
            ,1) + kappa]));
    elseif reg(i) == D(i,2)
        D(i,:) = max(0,min(D_MAX,[D(i,2) - kappa,D(i,2), D(i
            ,2) + kappa]));
    elseif reg(i) == D(i,3)
        D(i,:) = max(0,min(D_MAX,[D(i,3) - kappa, D(i,3), D(i
            ,3)+ kappa]));
     end
end
compare = sort_data(1,11);
varphi(x) = compare;
x=x+1;
end
dose_reg = reg;
 end
function ft = fterm_big(D,Ka,Ke,V)
ft = ((D*Ka)/(V*(Ka - Ke)));
end
function draw_samples(dose,cohort_id,sampling_time,ka,ke,v)
% True PK parameters %
phi = [ka, ke, v];
N = 10;
tau = 12;
varcov = [ .1    0    0    ;
           0    .1   0    ;
           0     0   .1];
sigma = .1;
sampling_time = sampling_time';
n = size(sampling_time,1);
```

```
tot_pat = N*n;
phi_mat = repmat(phi,tot_pat,1);
sigma_sq = sqrt(sigma);
no_dose = size(dose,2);
for i = 1:N
   rnd = mvnrnd([0,0,0],varcov);
   for j=1:n
       b((i-1)*n + j,:) = rnd;
   end
 end
 sam_time = repmat(sampling_time,N,1);
 phi_rand =  phi_mat.*exp(b);
 sim_conc = zeros(tot_pat,1);

 for k = 1:tot_pat
     concn = 0;
       for j=1:no_dose
            tp1 =  (phi_rand(k,3)*(phi_rand(k,1) - phi_rand(k
               ,2))) ;
            tp2 = (dose(j)*phi_rand(k,1));
            tp = tp2/tp1;
            concn = concn + ((sam_time(k)>=((j-1)*tau))) * (
               tp* (exp(-phi_rand(k,2)*abs(sam_time(k)-(j-1)*
               tau)) -exp(-phi_rand(k,1)*abs(sam_time(k)-(j
               -1)*tau))));
       end
 sim_conc(k) = concn * exp(normrnd(0,sigma_sq));
 % For proportional error model; For additive model use
    sim_conc(k) =  concn + normrnd(0,sigma);
 end
subject = (cohort_id-1)*N + 1;
  for i = 1:N
      for j = 1:n
          if (i ==1 && j==1)
              continue;
          else
          subject = [subject (cohort_id-1)*N + i];
          end
      end
  end
  subject = subject';
```

```
  data = [subject sam_time sim_conc];
 fid =...
   fopen('C:\Users\Kabir.Soeny\Desktop\PHD\Big_picture\
      PK_samples.m', 'w');
  fprintf(fid,'%d  %8.4f %8.4f\n',data(:,1:3)');
  fclose(fid);
end
function [Ka, Ke, V,omega1,omega2,omega3,sigma,PHI,col] =
   PK_estimates(dose,Ka,Ke,V,omega1,omega2,omega3,sigma,
   cohort_id)
 phi0 = [Ka, Ke, V];
 phi0 = log(phi0);
 tau = 12;
 N = 10;
n = 5;
mi = 3;
ctgt = 5;
mod_conc1 = @(phi,t)  ((dose(1)*phi(1))/(phi(3)*(phi(1) - phi
   (2)))) * (exp(-phi(2)*t) - exp(-phi(1)*t));
tplot = 0:0.1:80;
load C:\Users\Kabir.Soeny\Desktop\PHD\Big_picture\PK_samples.
   m;
subject = PK_samples(:,1);
time = PK_samples(:,2);
conc = PK_samples(:,3);
dp = [subject conc];
P = [1 0 0;0 1 0;0 0 1];
xform = [1 1 1];
options = statset('nlmefit');
options = statset(options,'TolX',1e-9,'FunValCheck','off','
   MaxIter',1000);
[phi,PSI,stats,b] = nlmefit(time,conc,subject,[],mod_conc1,
   phi0, 'REParamsSelect',[1 2 3],'CovPattern',P,'ErrorModel'
   ,'exponential', 'ParamTransform',xform,'Options',options,'
   ApproximationType','LME')
   phi = exp(phi)     ;
PHI = repmat(phi,1,N*cohort_id) .* ...                    %
   Fixed effects
       [exp(b(1,:));exp(b(2,:));exp(b(3,:))]    % Random
          effects
 colors = rand(1,N);
```

```
for I = 1:N
    fitted_model = @(t)  ((dose(1)*PHI(1,I))/(PHI(3,I)*(PHI
       (1,I) -PHI(2,I)))) *(exp(-PHI(2,I)*t) - exp(-PHI(1,I)*
       t));
    tI = time(subject == I);
    cI = conc(subject == I);
    subplot(2,5,I)
  col(I,:) = [min(max(rand,.1),.9)...
      min(max(rand,.25),.9) min(max(rand,.2),.7)];
   scatter(tI,cI,50,col(I,:),'filled')
  hold all
  plot(tplot,fitted_model(tplot),'--','Color',col(I,:),'
     linewidth',2.5)
    plot(tplot,mod_conc1(phi,tplot),'k','linewidth',1.5)
    plot([0,60],[ctgt,ctgt],'b','linewidth',2.5)
    hold off
    axis([0 40 0 18])
    xlabel('Time (hours)')
    ylabel('Concentration (mg/L)')
    legend(num2str(I))
end
phi = log(phi);
    Ka = exp(phi(1));
    Ke = exp(phi(2));
    V = exp(phi(3));
    sigma = stats.mse;
    omega1 = PSI(1,1);
    omega2 = PSI(2,2);
    omega3 = PSI(3,3);
end
```

D.6 The ED Algorithm for a PD Target

```
function kill_curve_max_diffeq_method_window_n6
T = 120;
timerange = 0:.1:T;
initial = [0,0,0];
tau = 12;
D_MAX = 250;
ctgtp = 8;
ctgtm = 3;
epsilon  = .99;
kappa = 5;
D = ones(6,3);
D(:,1) = 250;
D(:,2) = 120;
D(:,3) = 40;
extend = 0; %indicates when to switch from the discretized to
    the non-discretized doses during the run of the algorithm
ec50 = 3.6;
H = 1;
kkillmax = 10^10.34;
history = ones(10000,6);
varphi = ones(10000,1);
x=1;
stop = 0;
while (x < 1000 &&  extend < 12 )
    stop = 0;
    grid2 = cartprod(D(1,:) , D(2,:) );
    grid3 = cartprod(D(1,:) , D(2,:) , D(3,:) );
    grid4 = cartprod(D(1,:) , D(2,:) , D(3,:), D(4,:) );
    grid5 = cartprod(D(1,:) , D(2,:) , D(3,:), D(4,:) , D
       (5,:));
    grid6 = cartprod(D(1,:) , D(2,:) , D(3,:), D(4,:) , D
       (5,:), D(6,:) );
    lim6 = size(grid6,1);
for s =1:lim6
            for l=1:6
                DOSE(l) = grid6(s,l);
            end
        options = odeset('RelTol',1e-5,'Refine',1);
        [t, N] = ode45(@pd_model_kill, timerange,initial,
```

```
            options,DOSE,tau);
        size_N = size(N,1);
        for i = 1:size_N
            if N(i,1) >=   10^12 %Determination of the time
               of parasite clearance
               break;
            end
        end
         j=i;
         parasite_clear_time = t(j);
         flag = 0;
         AUC_kill = N(size_N,1);
         deltapm = N(timerange == t(j),2)/6;
         AUC_conc = N(size_N,3);
         AUC_conc_tgt = N(timerange == t(j),3);
         deltaauc = AUC_conc - AUC_conc_tgt;
         cono = zeros(size(N,1),2);
         for l=1:size(N,1)
              cono(l,1) = C(t(l),DOSE,tau);
              cono(l,2) = kkillmax*((C(t(l),DOSE,tau)^H)/(
                 ec50^H + C(t(l),DOSE,tau)^H));
         end
         criteria = deltapm + deltaauc + 100000000*(AUC_kill
            <10^12) ;
          data6(s,:)=[ grid6(s,1) grid6(s,2) grid6(s,3) grid6
             (s,4)...
          grid6(s,5) grid6(s,6)    AUC_kill
             parasite_clear_time criteria];
end
    sort_data = sortrows(data6, 9);
    sort2 = sort_data(1,:);
    reg = sort2(:,[1,2,3,4,5,6])
    parasite_clear_time = sort2(1,8)
    history(x,:) = reg;
    if x>1
        for l=1:6
            if abs(reg(l) - history(x-1,l)) == 0
                stop = stop + 1;
            end;
        end;
    end;
```

```
 if stop == 6
    extend = extend + 1;
 end
 if extend == 0
   for l = 1:size(reg,2)
    if reg(l) == D(l,1)
      D(l,:) = max(0,min(D_MAX,[D(l,1)- kappa, D(l,1), D(l,1)
          + kappa]));
    elseif reg(l) == D(l,2)
      D(l,:) = max(0,min(D_MAX,[D(l,2)- kappa,D(l,2), D(l,2)
         + kappa]));
    elseif reg(l) == D(l,3)
      D(l,:) = max(0,min(D_MAX,[D(l,3)- kappa, D(l,3), D(l,3)
          + kappa]));
    end
 end
end
if extend == 1
    for l = 1:size(reg,2)
         if reg(l) == D(l,1)
          D(l,:) = min(D_MAX,[D(l,1)*epsilon,D(l,1), D(l,1)
              *(1/epsilon)]);
        elseif reg(l) == D(l,2)
          D(l,:) = min(D_MAX,[D(l,2)*epsilon,D(l,2), D(l,2)
              *(1/epsilon)]);
        elseif reg(l) == D(l,3)
          D(l,:) = min(D_MAX,[D(l,3)*epsilon,D(l,3), D(l,3)
              *(1/epsilon)]);
        end
 end
end
compare = sort_data(1,9)
varphi(x) = compare; x = x+1;
end
DOSE = reg;
for l=1:size(N,1)
                  cono(l,1) = C(t(l),DOSE,tau);
                 cono(l,2) = kkillmax*((C(t(l),DOSE,tau)^H)/(
                    ec50^H + ...
                   C(t(l),DOSE,tau)^H));
```

```
end
end
function c = C(t,DOSE,tau)
            F = .95; ka = .46; ke = .17; v = 14; phi = [ka,
                ke, v];
            no_dose = size(DOSE,2);
            c = 0;

        for j=1:no_dose
            tp1 =  (phi(3)*(phi(1) - phi(2))) ;
            tp2 = (F*DOSE(j)*phi(1));
            tp = tp2/tp1;
            c = c + ((t>=((j-1)*tau)))* (tp* (exp(-phi(2)*abs
                (t-(j-1)*...
                tau)) - exp(-phi(1)*abs(t-(j-1)*tau))));
          end
end
function pcount = pd_model_kill(t,N,DOSE,tau)
ec50 = 3.6;
H = 1;
kkillmax = 10^10.34;
nu = .5;
ctgtp = 8;
ctgtm = 3;
pcount(1) = kkillmax*((C(t,DOSE,tau)^H)/(ec50^H + C(t,DOSE,
   tau)^H));
pcount(2) = (1 - nu)*max(0,C(t,DOSE,tau) - ctgtp) +  nu*max
   (0,ctgtm - C(t,DOSE,tau));
pcount(3) = C(t,DOSE,tau);
pcount = [pcount(1); pcount(2);pcount(3)];
end
```

Bibliography

Aarnoutse, R., J. Schapiro, C. Boucher, Y. Hekster, and D. Burger (2003). Therapeutic drug monitoring. *Drugs 63*(8), 741–753.

Almohaimeed, B. and A. Donev (2014). Experimental designs for drug combination studies. *Computational Statistics & Data Analysis 71*, 1077 – 1087.

Anderson, B. and N. Holford (2008). Mechanism-based concepts of size and maturity in pharmacokinetics. *Annual Review of Pharmacology and Toxicology 48*, 303–332.

Anderson, B. J. and G. H. Meakin (2002). Scaling for size: some implications for paediatric anaesthesia dosing. *Pediatric Anesthesia 12*(3), 205–219.

Andrews, J. M. (2001). Determination of minimum inhibitory concentrations. *Journal of Antimicrobial Chemotherapy 48*, 5–16.

Aronson, J. K. (2005). Adjusting therapeutic dosage regimens to optimise the balance of benefit to harm. *Clinical Medicine 5*(1), 16–19.

Atkinson, A., A. Donev, and R. Tobias (2007). *Optimum Experimental Designs, with SAS*. Oxford Statistical Science Series. Oxford University Press. Oxford, United Kingdom.

Atkinson, A. C. and A. N. Donev (1992). *Optimum Experimental Designs*. Oxford Statistical Science Series. Oxford University Press. Oxford, United Kingdom.

Battiti, R. and F. Masulli (1990). BFGS optimization for faster and automated supervised learning. In *International Neural Network Conference*, pp. 757–760. Springer Netherlands.

Bauer, R. and S. Guzy (2004). Monte carlo parametric expectation maximization (MC-PEM) method for analyzing population pharmacokinetic/pharmacodynamic data. In D. DArgenio (Ed.), *Advanced Methods of Pharmacokinetic and Pharmacodynamic Systems Analysis Volume 3*, Volume 765 of *The International Series in Engineering and Computer Science*, pp. 135–163. Springer US.

Bazzoli, C., S. Retout, and F. Mentré (2009). Design evaluation and optimisation in multiple response nonlinear mixed effect models: PFIM 3.0. *Computer Methods and Programs in Biomedicine 98*(1), 55–65.

Beal, S. L. and L. B. Sheiner (1989). NONMEM Users Guide Part I: Users Basic Guide. *NONMEM Project Group, University of California, San Francisco*.

Beal, S. L. and L. B. Sheiner (1998). NONMEM Users Guide Part VII: Conditional Estimation Methods. *NONMEM Project Group, University of California, San Francisco*.

Benet, L. Z. (1984). Pharmacokinetics: Basic principles and its use as a tool in drug metabolism. *Drug Metabolism and Drug Toxicity, JR Mitchell and MG Horning (eds.), Raven Press, New York*.

Bertrand, J. and F. Mentré (2008). Mathematical expressions of the pharmacokinetic and pharmacodynamic models implemented in the monolix software. *INSERM U738, Paris Diderot University*.

Bois, F. Y., M. Jamei, and H. J. Clewell (2010). PBPK modelling of inter-individual variability in the pharmacokinetics of environmental chemicals. *Toxicology 278*(3), 256 – 267.

Brabec, M., O. Konár, E. Pelikán, and M. Malý (2008). A nonlinear mixed effects model for the prediction of natural gas consumption by individual customers. *International Journal of Forecasting 24*(4), 659 – 678. Energy Forecasting.

Burton, M. (2006). *Applied Pharmacokinetics & Pharmacodynamics: Principles of Therapeutic Drug Monitoring*. Lippincott Williams & Wilkins. Philadelphia, United States.

Cascone, S., G. Lamberti, G. Titomanlio, and O. Piazza (2013). Pharmacokinetics of remifentanil: a three-compartmental modeling approach. *Translational Medicine @ UniSa: Official Journal of the Medical School of the University of Salerno 7*, 18–22.

Chao, Y.-S., L. Brunel, P. Faris, and P. J. Veugelers (2013, Oct). The Importance of Dose, Frequency and Duration of Vitamin D Supplementation for Plasma 25-Hydroxyvitamin D. *Nutrients 5*(10), 4067–4078.

Chow, S.-C. and M. Chang (2008). Adaptive design methods in clinical trials – a review. *Orphanet Journal of Rare Diseases 3*(1), 1–13.

Citron, M. L., D. A. Berry, C. Cirrincione, C. Hudis, E. P. Winer, W. J. Gradishar, N. E. Davidson, S. Martino, R. Livingston, J. N. Ingle, E. A. Perez, J. Carpenter, D. Hurd, J. F. Holland, B. L. Smith, C. I. Sartor, E. H. Leung, J. Abrams, R. L. Schilsky, H. B. Muss, and L. Norton (2003). Randomized trial of dose-dense versus conventionally scheduled and sequential versus concurrent combination chemotherapy as postoperative adjuvant treatment of node-positive primary breast cancer: First report of intergroup trial C9741/Cancer and Leukemia Group B trial 9741. *Journal of Clinical Oncology 21*(8), 1431–1439.

Czock, D. and F. Keller (2007). Mechanism-based pharmacokineticpharmacodynamic modeling of antimicrobial drug effects. *Journal of Pharmacokinetics and Pharmacodynamics 34*(6), 727–751.

D. Du Bois and E. F. Du Bois (1916). A formula to estimate the approximate surface area if height and weight be known. *Archives of Internal Medicine, Chicago 863871.*

Davidian, M. and D. Giltinan (1995). *Nonlinear Models for Repeated Measurement Data*. Monographs on Statistics and Applied Probability. Chapman & Hall, United Kingdom.

Delyon, B., M. Lavielle, and E. Moulines (1999). Convergence of a stochastic approximation version of the EM algorithm. *The Annals of Statistics 27*(1), 94–128.

Dempster, A., N. Laird, and D. Rubin (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B 39*(1), 1–38.

Dickinson, J. E. and S. F. Evans (2002, Mar). The optimization of intravaginal misoprostol dosing schedules in second-trimester pregnancy termination. *American Journal of Obstetrics and Gynecology 186*(3), 470–474.

DiNicolantonio, J. J. and V. L. Serebruany (2013). Challenging the FDA black box warning for high aspirin dose with ticagrelor in patients with diabetes. *Diabetes 62*(3), 669–671.

Dziura, J. D., L. A. Post, Q. Zhao, Z. Fu, and P. Peduzzi (2013). Strategies for dealing with missing data in clinical trials: From design to analysis. *Yale J Biol Med 86*(3), 343–358.

Elfving, G. (1952). Optimum allocation in linear regression theory. *The Annals of Mathematical Statistics 23*(2), 255–262.

EMEA (2004, December). Discussion paper on the impact of renal immaturity when investigating medicinal products intended for paediatric use. *European Medicines Agency, London CPMP/PEG/35132/03.*

EMEA-CHMP (2005, July). Concept paper on the impact of liver immaturity when investigating medicinal products intended for neonatal use. *Committee for Medicinal Products for Human Use, European Medicines Agency, London EMEA/CHMP/PEG/194605/2005.*

EMEA-CHMP (2006a, March). Concept paper on the impact of lung and heart immaturity when investigating medicinal products intended for neonatal use. *Committee for Medicinal Products for Human Use, European Medicines Agency, London EMEA/CHMP/114218/2006.*

EMEA-CHMP (2006b, June). Guideline on the role of pharmacokinetics in the development of medicinal products in the paediatric population. *Committee for Medicinal Products for Human Use (CHMP), European Medicines Agency, London*.

Ezzet, F., R. Mull, and J. Karbwang (1998). Population pharmacokinetics and therapeutic response of CGP 56697 (artemether+benflumetol) in malaria patients. *British Journal of Clinical Pharmacology 46*(6), 553–561.

Fahmi, O. A. and S. L. Ripp (2010). Evaluation of models for predicting drug-drug interactions due to induction. *Expert Opinion on Drug Metabolism & Toxicology 6*(11), 1399–1416.

FDA (1999). Guidance for industry - population pharmacokinetics. Technical report, U.S. Department of Health and Human Services, Food and Drug Administration, Rockville.

FDA (2010). Adaptive design clinical trials for drugs and biologics. Technical report, U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER).

FDA (2015). Pharmacometrics at FDA. http://www.fda.gov/AboutFDA/CentersOffices/OfficeofMedicalProductsandTobacco/CDER/ucm167032.htm.

Fedorov, V. V. (1972). *Theory of Optimal Experiments*. Academic Press, New York, United States of America.

Fedorov, V. V. and P. Hackl (1997). *Model-Oriented Design of Experiments*, Volume 125 of *Lecture Notes in Statistics*. Springer. New York, United States of America.

Fedorov, V. V. and S. L. Leonov (2014). *Optimal Design for Nonlinear Response Models*. CRC Press. Florida, United States of America.

Felton, T. W., J. A. Roberts, T. P. Lodise, M. Van Guilder, E. Boselli, M. N. Neely, and W. W. Hope (2014). Individualization of piperacillin dosing for critically ill patients: Dosing software to optimize antimicrobial therapy. *Antimicrobial Agents and Chemotherapy 58*(7), 4094–4102.

Foracchia, M., A. Hooker, P. Vicini, and A. Ruggeri (2004). PopED, a software for optimal experiment design in population kinetics. *Computer Methods and Programs in Biomedicine 74*(1), 29 – 46.

Fuchs, A., C. Csajka, Y. Thoma, T. Buclin, and N. Widmer (2013). Benchmarking therapeutic drug monitoring software: A review of available computer tools. *Clinical Pharmacokinetics 52*(1), 9–22.

Gabrielsson, J. and D. Weiner (2012). Non-compartmental analysis. In B. Reisfeld and A. N. Mayeno (Eds.), *Computational Toxicology*, Volume 929 of *Methods in Molecular Biology*, pp. 377–389. Humana Press.

George, E. I. and R. E. McCulloch (1993). Variable selection via Gibbs sampling. *Journal of the American Statistical Association 88*(423), 881–889.

Gervasini, G., J. Bentez, and J. Carrillo (2010). Pharmacogenetic testing and therapeutic drug monitoring are complementary tools for optimal individualization of drug therapy. *European Journal of Clinical Pharmacology 66*(8), 755–774.

Gibaldi, M. and G. Levy (1976). Pharmacokinetics in clinical practice: I. concepts. *The Journal of the American Medical Association 235*(17), 1864–1867.

Ginsberg, G., D. Hattis, A. Russ, and B. Sonawane (2004). Physiologically based pharmacokinetic (PBPK) modeling of caffeine and theophylline in neonates and adults: Implications for assessing children's risks from environmental agents. *Journal of Toxicology and Environmental Health, Part A 67*(4), 297–329.

Groot, E. H. d., N. Post, R. Boellaard, N. R. Wagenaar, A. T. Willemsen, and J. A. van Dalen (2013). Optimized dose regimen for whole-body FDG-PET imaging. *EJNMMI Research 3*(1), 1–11.

Gurney, H. (2002). How to calculate the dose of chemotherapy. *British Journal of Cancer 86*, 1297 – 1302.

Halpern, S. A. (1988). *American Pediatrics: The Social Dynamics of Professionalism, 1880-1980*. University of California Press. California, United States of America.

Hill, A. V. (1910). The possible effects of the aggregation of the molecules of haemoglobin on its dissociation curves. *Journal of Physiology 40*, iv–vii.

Holford, N. H. G. (1999). Target concentration intervention: beyond Y2K. *British Journal of Clinical Pharmacology 48*(1), 9–13.

Holmquist, B. (1988). Moments and cumulants of the multivariate normal distribution. *Stochastic Analysis and Applications 6*(3), 273–278.

Howard, R. M. (2015). *A Signal Theoretic Introduction to Random Processes* (First ed.). Wiley.

ICON plc (2015). NONMEM® . http://www.iconplc.com/innovation/solutions/nonmem/.

Jaki, T. and M. J. Wolfsegger (2011). Estimation of pharmacokinetic parameters with the R package PK. *Pharmaceutical Statistics 10*(3), 284–288.

Jaki, T. and M. J. Wolfsegger (2012). Non-compartmental estimation of pharmacokinetic parameters for flexible sampling designs. *Statistics in Medicine 31*(11-12), 1059–1073.

Jamei, M., G. L. Dickinson, and A. Rostami-Hodjegan (2009). A framework for assessing inter-individual variability in pharmacokinetics using virtual human populations and integrating general knowledge of physical chemistry, biology, anatomy, physiology and genetics: A tale of 'bottom-up' vs 'top-down' recognition of covariates. *Drug Metabolism and Pharmacokinetics 24*(1), 53–75.

Jin, J., G. E. Sklar, V. Min Sen Oh, and S. Chuen Li (2008, Feb). Factors affecting therapeutic compliance: A review from the patient's perspective. *Therapeutics and Clinical Risk Management 4*(1), 269–286.

Johnson, T. N. (2008). The problems in scaling adult drug doses to children. *Archives of Disease in Childhood 93*(3), 207–211.

Kang, J.-S. and M.-H. Lee (2009). Overview of therapeutic drug monitoring. *The Korean Journal of Internal Medicine 24(1).*

Kearns, G. L., S. M. Abdel-Rahman, S. W. Alander, D. L. Blowey, J. S. Leeder, and R. E. Kauffman (2003). Developmental pharmacology drug disposition, action, and therapy in infants and children. *New England Journal of Medicine 349*(12), 1157–1167.

Kenna, L. A., L. Labbé, J. S. Barrett, and M. Pfister (2005, Jun). Modeling and simulation of adherence: Approaches and applications in therapeutics. *AAPS (Official Journal of the American Association of Pharmaceutical Scientists) 7*(2), E390–E407.

Kiefer, J. and J. Wolfowitz (1959, 06). Optimum designs in regression problems. *The Annals of Mathematical Statistics 30*(2), 271–294.

Kiefer, J. and J. Wolfowitz (1960, February). The equivalence of two extremum problems. *Canadian Mathematical Society 12*, 363–366.

Kuczmarski, R., C. Ogden, S. Guo, L. G.-S. KM, Flegal, Z. Mei, R. Wei, L. Curtin, A. Roche, and C. Johnson (2002). 2000 CDC growth charts for the United States: methods and development. *Vital and Health Statistics*.

Kuti, J. L., P. K. Dandekar, C. H. Nightingale, and D. P. Nicolau (2003). Use of monte carlo simulation to design an optimized pharmacodynamic dosing strategy for meropenem. *The Journal of Clinical Pharmacology 43*(10), 1116–1123.

Leamer, E. E. (1983, March). Let's take the con out of econometrics. *American Economic Review 73*(1), 31–43.

Lindstrom, M. J. and D. M. Bates (1990). Nonlinear mixed effects models for repeated measures data. *Biometrics 46*, 673–687.

Lipsitch, M. and B. R. Levin (1997). The population dynamics of antimicrobial chemotherapy. *Antimicrobial Agents and Chemotherapy 41*(2), 363–73.

Little, R. J., R. D'Agostino, M. L. Cohen, K. Dickersin, S. S. Emerson, J. T. Farrar, C. Frangakis, J. W. Hogan, G. Molenberghs, S. A. Murphy, J. D. Neaton, A. Rotnitzky, D. Scharfstein, W. J. Shih, J. P. Siegel, and H. Stern (2012). The prevention and treatment of missing data in clinical trials. *New England Journal of Medicine 367*(14), 1355–1360.

Little, R. J. A. and D. B. Rubin (2002). *Statistical Analysis with Missing Data, Second Edition*. John Wiley & Sons, Inc.

Little, T. D. (Ed.) (2013). *The Oxford Handbook of Quantitative Methods in Psychology: Vol. 2: Statistical Analysis* . Oxford Library of Psychology.

Ma, Q. and A. Y. H. Lu (2011). Pharmacogenetics, pharmacogenomics, and individualized medicine. *Pharmacological Reviews 63*(2), 437–459.

Magnus, J. R. (1978). The moments of products of quadratic forms in normal variables. *Statistica Neerlandica 32*, 201–210.

Martinez, M. N., M. G. Papich, and G. L. Drusano (2012). Dosing regimen matters: the importance of early intervention and rapid attainment of the pharmacokinetic/pharmacodynamic target. *Antimicrobial Agents and Chemotherapy 56*(6), 2795–2805.

MathWorks (2015). Statistics and machine learning toolbox. http://uk.mathworks.com/help/stats/nlmefit.html.

McCullagh, P. (2002). What is a statistical model? *The Annals of Statistics 30*(5), 1225–1310.

Mehvar, R. (1998). Pharmacokinetic-based design and modification of dosage regimens. *American Journal of Pharmaceutical Education 62*, 189–195.

Mentré, F. and M. Lavielle (2008, March). Stochastic EM algorithms in population PKPD analyses. In *American Conference on Pharmacometrics*.

Nielsen, E. I., A. Viberg, E. Löwdin, O. Cars, M. O. Karlsson, and M. Sandström (2007, Jan). Semimechanistic pharmacokinetic/pharmacodynamic model for assessment of activity of antibacterial agents from time–kill curve experiments. *Antimicrobial Agents and Chemotherapy 51*(1), 128–136.

Nyberg, J., C. Bazzoli, K. Ogungbenro, A. Aliev, S. Leonov, S. Duffull, A. C. Hooker, and F. Mentré (2015). Methods and software tools for design evaluation in population pharmacokinetics-pharmacodynamics studies. *British Journal of Clinical Pharmacology 79*(1), 6–17.

Ogungbenro, K., G. Graham, I. Gueorguieva, and L. Aarons (2005). The use of a modified Fedorov exchange algorithm to optimise sampling times for population pharmacokinetic experiments. *Computer Methods and Programs in Biomedicine 80*(2), 115–125.

Pandey, U. B. and C. D. Nichols (2011). Human disease models in drosophila melanogaster and the role of the fly in therapeutic drug discovery. *Pharmacological Reviews 63*(2), 411–436.

Pázman, A. (1986). *Foundations of Optimum Experimental Design*. Springer. Netherlands.

Pinheiro, J. and D. Bates (2000). *Mixed Effects Models in S and S-Plus*. Statistics and Computing Series. Springer-Verlag. Berlin, Germany.

Pinheiro, J. C. and D. M. Bates (1995). Approximations to the log-likelihood function in the nonlinear mixed-effects model. *Journal of Computational and Graphical Statistics 4*(1), 12–35.

Plaisance, K. I., G. L. Drusano, A. Forrest, C. I. Bustamante, and H. C. Standiford (1987). Effect of dose size on bioavailability of ciprofloxacin. *Antimicrobial Agents and Chemotherapy 31*(6), 956–958.

PopED Manual (2012). Release version 2.13. http://poped.sourceforge.net/manual.pdf.

Proost, J. H. and D. K. Meijer (1992). Mw/pharm, an integrated software package for drug dosage regimen calculation and therapeutic drug monitoring. *Computers in Biology and Medicine 22*(3), 155–163.

Pukelsheim, F. (1993). *Optimal Design of Experiments*. Classics in Applied Mathematics. Society for Industrial and Applied Mathematics. Philadelphia, United States of America.

Ravina, B., J. Cummings, and M. McDermott (Eds.) (2012). *Clinical Trials in Neurology: Design, Conduct, Analysis*. Cambridge University Press.

Reeve, R. (1996). The randomized concentration-controlled trial:mathematical definitions, a dose-adjusting algorithm, and sample size efficiency. *Communications in Statistics - Theory and Methods 25*(9), 2169–2188.

Roden, D. M. and A. L. George Jr (2002, Jan). The genetic basis of variability in drug responses. *Nature Reviews Drug Discovery 1*(1), 37–44.

Rowland, M. and T. Tozer (2011). *Clinical Pharmacokinetics and Pharmacodynamics: Concepts and Applications*. Wolters Kluwer Health/Lippincott William & Wilkins. Philadelphia,, United States of America.

Rowland, M. and T. N. Tozer (1994). *Clinical Pharmacokientics: Concepts and Applications* (3rd ed.). Williams and Wilkins. Philadelphia, United States of America.

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. John Wiley & Sons.

Sacks, L., H. Shamsuddin, Y. Yasinskaya, K. Bouri, M. Lanthier, and R. Sherman (2014). Scientific and regulatory reasons for delay and denial of FDA approval of initial applications for new drugs, 2000-2012. *The Journal of the American Medical Association 311*(4), 378–384.

Saltelli, A., M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola (2008). *Global Sensitivity Analysis. The Primer*. John Wiley and Sons. Hoboken, New Jersey, United States of America.

Sanathanan, L. P. and C. C. Peck (1991). The randomized concentration-controlled trial: An evaluation of its sample size efficiency. *Controlled Clinical Trials 12*(6), 780 – 794.

Santos, R. G., M. A. Giulianotti, C. T. Dooley, C. Pinilla, J. R. Appel, and R. A. Houghten (2011). Use and implications of the harmonic mean model on mixtures for basic research and drug discovery. *ACS Combinatorial Science 13*(3), 337–344.

Shah, K. R. and B. K. Sinha (1989). *Theory of Optimal Designs*. Springer. New York, United States of America.

Sheiner, L. and S. Beal (1980). Evaluation of methods for estimating population pharmacokinetic parameters. I. Michaelis-Menten model: Routine clinical pharmacokinetic data. *Journal of Pharmacokinetics and Biopharmaceutics 8*, 553–571.

Sheiner, L. and S. Beal (1981). Evaluation of methods for estimating population pharmacokinetic parameters II. biexponential model and experimental pharmacokinetic data. *Journal of Pharmacokinetics and Biopharmaceutics 9*(5), 635–651.

Sheiner, L. B. and S. L. Beal (1985). Pharmacokinetic parameter estimates from several least squares procedures: Superiority of extended least squares. *Journal of Pharmacokinetics and Biopharmaceutics 13*(2), 185–201.

Simpson, J. A., E. R. Watkins, R. N. Price, L. Aarons, D. E. Kyle, and N. J. White (2000). Mefloquine pharmacokinetic-pharmacodynamic models: Implications for dosing and resistance. *Antimicrobial Agents and Chemotherapy 44*(12), 3414–3424.

Smith, K. (1918). On the standard deviations of adjusted and interpolated values of an observed polynomial function and its constants and the guidance they give towards a proper choice of the distribution of observations. *Biometrika 12*(1/2), 1–85.

Soeny, K., B. Bogacka, B. Jones, and T. Bouillon (2016). Optimizing dose regimens and fixed dose combination ratios in clinical trials. *Journal of Biopharmaceutical Statistics 26*(3), 432–451.

Stirnemann, J., A. Samson, and J. Thalabard (2012). Individual predictions based on nonlinear mixed modeling: application to prenatal twin growth. *Statistics in Medicine 31*(18), 1986–1999.

Tam, V. H., A. N. Schilling, and M. Nikolaou (2005). Modelling time-kill studies to discern the pharmacodynamics of meropenem. *Journal of Antimicrobial Chemotherapy 55*(5), 699–706.

Teorell, T. (1937). Kinetics of distribution of substances administered to the body. I. The extravascular modes of administration. *Archives Internationales de Pharmacodynamie et de Thérapie 57*, 205–225.

Thompson, G. A. (1992). Dosage regimen design: A pharmacokinetic approach. *The Journal of Clinical Pharmacology 32*(3), 210–214.

Thomson, A. H. and B. Whiting (1992). Bayesian parameter estimation and population pharmacokinetics. *Clinical Pharmacokinetics 22*(6), 447–467.

Tokdar, S. T. and R. E. Kass (2010). Importance sampling: a review. *Wiley Interdisciplinary Reviews: Computational Statistics 2*(1), 54–60.

Urquhart, J. (1996). Patient non-compliance with drug regimens: measurement, clinical correlates, economic impact. *European Heart Journal 17 (Supplement A)*, 8–15.

Vogel, C. L., M. A. Cobleigh, D. Tripathy, J. C. Gutheil, L. N. Harris, L. Fehrenbacher, D. J. Slamon, M. Murphy, W. F. Novotny, M. Burchmore, S. Shak, S. J. Stewart, and M. Press (2002). Efficacy and safety of trastuzumab as a single agent in first-line treatment of HER2-overexpressing metastatic breast cancer. *Journal of Clinical Oncology 20*(3), 719–726.

Wählby, U., E. N. Jonsson, and M. O. Karlsson (2002). Comparison of stepwise covariate model building strategies in population pharmacokinetic-pharmacodynamic analysis. *AAPS (Official Journal of the American Association of Pharmaceutical Scientists) 4*(4), 68–79.

Wald, A. (1943). On the efficient design of statistical investigations. *The Annals of Mathematical Statistics 14*(2), 134–140.

Weisstein, E. W. (2016). Skewness. From MathWorld – A Wolfram Web Resource. http://mathworld.wolfram.com/Skewness.html.

WHO (2005). WHO expert committee on specifications for pharmaceutical preparations. Technical Report 929, WHO Technical Report Series.

Wikipedia (2011). Physiologically based pharmacokinetic modelling. https://en.wikipedia.org/wiki/Physiologically_based_pharmacokinetic_modelling.

Wise, R. (2003). Maximizing efficacy and reducing the emergence of resistance. *Journal of Antimicrobial Chemotherapy 51*, 37–42.

Wolfinger, R. (1993). Laplace's approximation for nonlinear mixed models. *Biometrika 80*(4), 791–795.