Estimating Nonlinear Mixed-Effects Models by the Generalized Profiling Method and its Application to Pharmacokinetics

Liangliang Wang

Department of Mathematics and Statistics

McGill University, Montreal

August, 2007

This thesis is submitted to the Faculty of Graduate Studies and Research as a partial fulfillment of the requirements of the degree of Master of Science.

Abstract:

Several methods with software tools have been developed to estimate nonlinear mixed-effects models. However, fewer have addressed the issue when nonlinear mixed-effects models are implicitly expressed as a set of ordinary differential equations (ODE's) while these ODE's have no closed-form solutions. The main objective of this thesis is to solve this problem based on the framework of the *generalized profiling method* proposed by Ramsay, Hooker, Campbell, and Cao (2007).

Four types of parameters are identified and estimated in a cascaded way by a multiple-level nested optimization. In the outermost level, the smoothing parameter λ is selected by the criterion of generalized cross-validation (GCV). In the outer level, the structural parameters, including the fixed effects $\boldsymbol{\beta}$, the variance-covariance matrix for random effects Ψ, and the residual variance σ^2, are optimized by a criterion based on a first-order Taylor expansion of the nonlinear function. In the middle level, the random effects $\mathbf{b}$ are optimized by the penalized nonlinear least squares. In the inner level, the coefficients of basis function expansions $\mathbf{c}$ are optimized by penalized smoothing with the penalty defined by ODE's. Consequently, some types of parameters are expressed as explicit or implicit functions of other parameters. The dimensionality of the parameter space is reduced, and the optimization surface becomes smoother. The Newton-Raphson algorithm is applied to estimate parameters for each level of optimization with gradients and Hessian matrices worked out analytically with the Implicit Function Theorem.

Our method, along with MATLAB codes, is tested by estimating several compartment models in pharmacokinetics from both simulated and real data sets. Results are compared with the true values or estimates obtained by the package nlme in R, and it turns out that the generalized profiling method can achieve reasonable estimates without solving ODE's directly.

Résumé:

Il n'y a aucune solution de exacte pour beaucoup de modèles non-linéaires à effets mixtes exprimés comme un ensemble d'équations ordinaires (ODE's) en modèles de compartiment. Cette thèse passe en revue plusieurs méthodes et outils courants de logiciel pour modèles non-linéaires à effets mixtes, et explore une nouvelle manière d'estimer des effets mixtes non-linéaires en modèles de compartiment basée sur le cadre de la méthode de profilage généralisée proposée par Ramsay, Hooker, Campbell, et Cao (2007).

Quatre types de paramètres sont identifiés et estimés d'en cascade par une optimisation de multiple-niveau: le paramètre regularisateur est choisi par le critère de la contre-vérification généralisée (GCV); les paramètres structuraux, y compris les effets fixes, la matrice de variance-covariance pour les effets aléatoires, et la variance résiduelle sont optimisés par un critère basé sur une expansion de premier ordre de Taylor de fonction non-linéaire; les effets aléatoires sont optimisés par une methode des moindres carrés non-linéaires pénalisés; et les coefficients d'expansions de fonction de base sont optimisés par un lissage pénalisé avec la pénalité définie par l'equation differentielle. En conséquence, certains des paramètres sont exprimés en tant que fonctions explicites ou implicites d'autres paramètres. La dimensionnalité de l'espace des paramètres est réduite, et la surface d'optimisation devient plus lisse. L'algorithme de Newton-Raphson est appliqué aux paramètres d'évaluation pour chaque niveau d'optimisation, où le théorème des fonctions implicites est employé couramment pour établir les gradients et les matrices de Hessiennes de facon analytiques.

La méthode proposée et des codes de MATLAB sont examinés par des applications à plusieurs modèles de compartiment en pharmacocinétique sur des donnees simulées et vraies. Des résultats sont comparés aux valeurs ou aux évaluations vraies obtenues par paquet nlme du logiciel R, indiquant que la méthode de profilage généralisée peut réaliser des évaluations raisonnables sans résoudre l'equation differentielle.

Acknowledgements

I would like to take this opportunity to thank several individuals who have helped me to complete my master program and this thesis.

First and foremost, I would like to express my sincerest gratitude to my supervisor, Dr. James O. Ramsay, for his strong support in many aspects which enables me to complete my master program. I am thankful to him for his support in computer equipment, financial aid, and attending conferences. I also appreciate his many invitations to parties at his house. Without his knowledge, patience, encouragement, and pushes, I would never have finished this thesis.

I am sincerely thankful to Dr. Nandini Dendukuri for her enormous amount of help, encouragement, and support. She provided me good training in statistics, especially in Bayesian analysis. Her persistence and enthusiasm in research are always encouraging me.

I would like to thank Dr. Charles la Porte and Dr. Timothy O. Ramsay at Ottawa University. They provided me with the data to analyze and introduced pharmacokinetics to me.

I extend my thanks to the professors in the Department of Mathematics and Statistics and Department of Epidemiology, Biostatistics and Occupational Health. They have given me excellent lectures and all kinds of help and support. Particularly, I would like to thank Dr. James Hanley for being the external examiner of this thesis. I would like to thank the students of the two departments for creating a professional and friendly atmosphere. I express my sincere appreciation to Dr. Djivede Kelome for helping me translate the abstract of this thesis from English to French.

I deeply thank my parents and my sister who are always standing by me and offering endless love and support. Although they know little about my research topic, their listening ears and their faith in my ability really help me.

Finally, I would like to show my special appreciation to my husband, Dr. Jiguo Cao. I am thankful to him for proofreading my thesis and providing support throughout this process.

Contents

List of Tables

List of Figures

Chapter 1

Background and the organization of the thesis

1.1 Pharmacokinetics/pharmacodynamics

Pharmacology includes the study of interactions between an organism and substances residing within it to produce a change in functions. These substances can be various compounds, such as nutrients, metabolites, endogenous, and toxins, and the organism might be humans or animals. However, most applications focus on drug substances within humans. Pharmacokinetics and pharmacodynamics are two important branches in pharmacology; see Gabrielsson and Weiner (2000) and Sheiner and Wakefield (1999) for further references. For convenience, we often use the abbreviations "PK/PD" for "pharmacokinetics/pharmacodynamics" in this thesis.

The analysis of PK/PD data is concerned with the concentration-time curve, which describes the relationship between the dosing regimen and the body's exposure to a drug. Pharmacokinetics is dedicated to the study of the effect of the body on the drug. In other words, pharmacokinetic studies the time course of substances, and their mechanisms and kinetics within a biological system as a function of time. Concentrations over time are determined by the rate and extent of the processes of absorption, distribution, metabolism, and excretion, and these four processes are often abbreviated as ADME. Pharmacokinetics analysis is also concerned with determining

whether a variation in dose is required over subpopulations with different exposures based on individual-specific covariates, e.g. age, size, concomitant medications, and kidney function.

The information obtained by pharmacokinetic models has various uses, such as correlating drug doses with pharmacological and toxic responses and determining an optimum dose level for an individual. The ultimate goal of studying pharmacokinetics is to balance the efficacy and toxicity of a drug and optimize drug therapy, including dose, dosage regimen, and dosage form. In other words, knowledge of pharmacokinetic principles is essential to optimize the therapeutic effects of medicines and minimize the frequency of unwanted effects.

In contrast to pharmacokinetics, pharmacodynamics is concerned with the effects of a drug on the body. Pharmacodynamics associates dose-response relationships of drug effect with drug concentration, dose, time, patient characteristics, and the molecular mechanisms of drug activity. Applications in this thesis mainly focus on the pharmacokinetics data and their analyses.

ADME is an acronym in pharmacokinetics for absorption, distribution, metabolism, and excretion, which describes the disposition of a pharmaceutical compound within an animal or human body. Gabrielsson and Weiner (2000) is a general reference on PK/PD analysis.

Absorption is the process by which the intact drug proceeds from the site of administration to the blood circulation. Drugs are usually administrated through a subcutaneous injection or the oral route. For example, in the absorption process of an oral adminstration, the drug passes through intestinal membranes to enter the bloodstream where its concentration is usually measured.

The drug is carried to its site of action once it enters the bloodstream . *Distribution* describes this process during which the drug diffuses or is transferred from the bloodstream into various tissues and organs, which is important to ensure the drug efficacy. Drug uptake into the tissues might return to the general circulation system, i.e. systemic circulation. Distribution of the drug throughout the body is characterized by the volume of distribution which is defined as the amount of drug

in the body divided by the drug concentration.

The irreversible processes by which a drug is prevented from reaching its effector site are collectively referred to as elimination, including two sub-processes, *metabolism* and *excretion*. During the sub-process of metabolism, drugs are converted or transformed into compounds which are easier to eliminate. In most cases, metabolism inactivates the pharmacological response of a drug; however, in some cases, metabolites can be pharmacologically active as well. Although liver is the main metabolizing organ in the body, metabolism can also occur in intestine, blood, and other organs. The sub-process of excretion describes the removal of the drug and metabolites primarily by the kidneys and feces through forms of urine and bile, respectively.

Although ADME sub-processes generally follow the above sequence, they are not discrete events and may occur simultaneously; that is, one sub-process is still occurring while the next one begins. For example, a sustained release drug may still be distributing drug while previously absorbed drug is being eliminated.

The pharmacokinetic sub-processes ADME of a drug are specified in terms of measurable parameters such as plasma concentration, biological half-life, and rate constants. Since the precise site of action is often unknown, the concentration of the drug is usually monitored by following the drug concentration in the plasma and other suitable body fluids. These measurements are statistically correlated with the effects of drugs on humans.

1.2 Pharmacokinetic data

As an example of pharmacokinetic data, we consider a pharmacokinetic study focusing on protease inhibitor combinations of indinavir (IDV) and ritonavir (RTV). This study investigated the steady state pharmacokinetics of the 600/100 mg and 400/100 mg twice-daily IDV/RTV combinations in 16 healthy volunteers. Serial plasma sampling was performed at time points 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 4.0, 5.0, 6.0, 8.0, 10.0 and 12.0 hours after drug administration. A reference on this study is Wasmuth, la Porte, Schneider, Burger, and Rockstroh (2004).

Figure 1.1 shows the relationship between plasma concentrations of indinavir (IDV) and time since drug administration in a pharmacokinetic study of protease inhibitor combinations of 600 mg indinavir and 100 mg ritonavir (RTV) in 16 healthy volunteers. As can be seen in Figure 1.1, drug concentration-time profiles exhibit a similar profile for all subjects: the drug concentrations are rising rapidly before achieving their peaks, and then declining gradually. Nevertheless, peak concentration, rise, and decay vary significantly from subject to subject because of different age, sex, and other factors. In statistical analysis, unequally spaced samples, and missing data often cause some difficulties and might need special models. A noticeable feature about the data collection is that serial plasma sampling is not performed at equally spaced time points considering the high costs of experiments as well as the unique characteristics of the concentration curves. Drug concentrations are measured once every 0.5 hour in the first 3 hours, every 1 hour in the following 3 hours, and every 2 hours in the last 6 hours. In this way, not only the sharp curves shortly after drug administration can be captured through the intensive sampling, but also the costs can be reduced greatly because the sparse data will suffice for the declining curves. Another feature about this data is the presence of some missing data. One data point is missing for the 10th subject, and all the data are missing for the 12th subject.

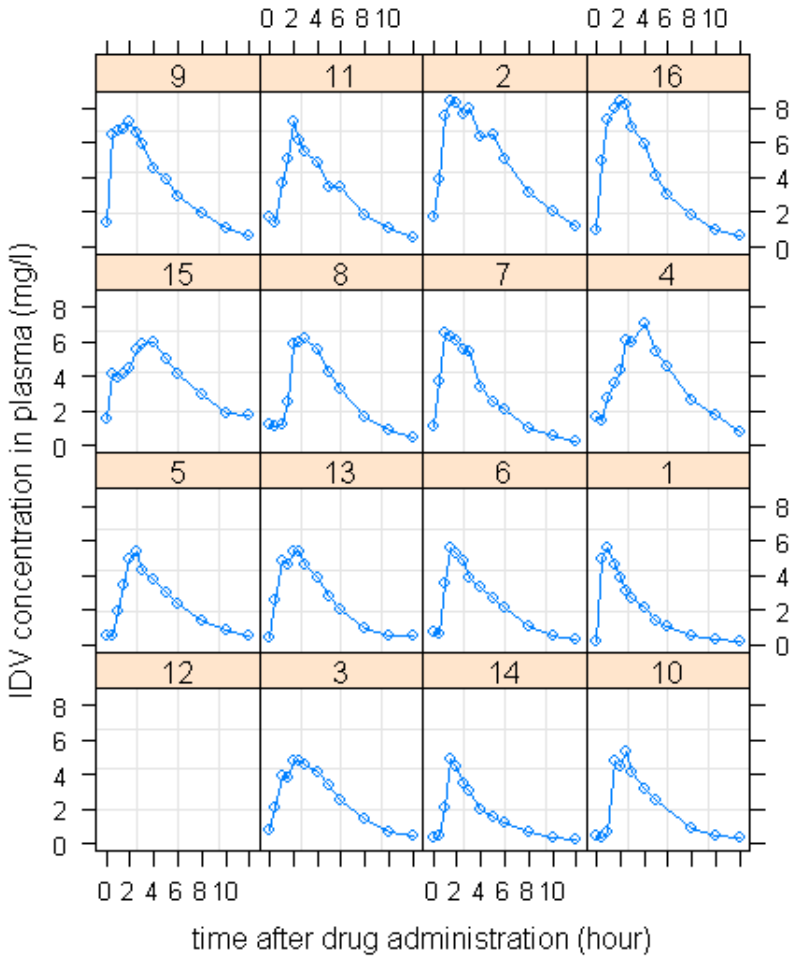

Figure 1.1: Plasma concentrations of indinavir versus time since administration of 600 mg indinavir plus 100 mg ritonavir

1.3 Compartment pharmacokinetic models

The compartment models represent a biological system describing pharmacokinetic behavior as a finite number of components referred to as compartments. Davidian and Giltinan (1995, 2003) and Seber and Wild (1988) are good references on this topic. A compartment is defined as a group of tissues that have a similar blood flow and drug affinity. Ideally, the compartments in the model would represent real,

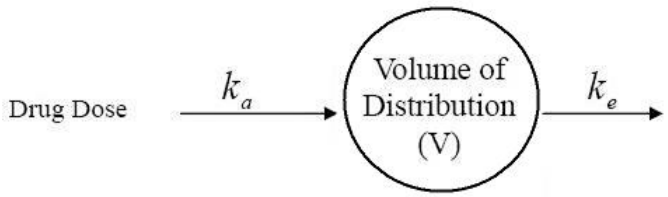

Figure 1.2: One-compartment model with first-order absorption and first-order elimination.

identifiable components of the body. For example, the systemic circulation in the body is identified as the blood compartment, i.e. the central compartment; organs or tissues with similar kinetic characteristics is amalgamated as the tissue compartment, i.e. the peripheral compartment. A compartment is described as open if it leaks to the environment, otherwise it is closed. In pharmacokinetics, most models have at least one open compartment with the first-order elimination.

All compartment models share the following assumptions, as described in Kinabo and McKellar (1989). First, the compartments communicate with each other by reversible processes. Second, rate constants are used to measure the rate of entry and exit of a drug from a compartment. Third, the drug is distributed rapidly and homogeneously within a compartment.

Mathematically, the compartment models are often expressed by a system of differential equations in terms of pharmacokinetic parameters describing the volume of specific compartments, the rate of change of the amounts or concentrations of substances in each compartment. Usually, the transfer rate of a drug from one compartment to another is assumed as the first-order kinetics, which means that the rate of change of a drug from a specific compartment to another is proportional to its concentration within the source compartment. The solutions of these differential equations provide a formal mathematical description of the amounts or concentrations in the compartments at any time as a function of the parameters. Since in compartment models the mathematical functions or differential equations are employed regardless

to any mechanistic aspects of the modeled biological system, great care must be exercised in extrapolating outside of the measured domain. Several methods have been proposed to estimate parameters in compartment models; see Bates and Watts (1988) for further references.

One-compartment models and two-compartment models are the most commonly used compartment models in pharmacokinetics. The simplest models are one-compartment models, in which the compartment represents the circulatory system and all the tissues are perfused rapidly by the drug. Elimination of a drug is assumed to occur only from this central compartment since the processes associated with elimination occur mainly in the plasma and the highly perfused tissues of the liver and kidney; that is, in one-compartment model, an agent is introduced into a single body compartment from which it is also eliminated. However, if drugs exhibit a slow equilibration with peripheral tissues, it is superior to use a two-compartment model to describe the distribution of drug from the central compartment to the tissue compartment.

For an orally-administered drug, the biological system can be described as a one-compartment model with first-order absorption and elimination rates, which can be represented graphically as a compartment or system diagram as in Figure 1.2. In this model, the body is represented as a single blood compartment. Let t be the time after administration of drug. At the initial time, $t = 0$, an oral dose D is instantaneously delivered into the blood compartment from a hypothetical absorption site, e.g. the stomach or the gut, resulting in a drug concentration $C_a(t)$ at time t. Meanwhile, we can measure the concentration of the drug in the blood compartment at time t. Drug transfers into the blood compartment at absorption rate constant k_a which is defined as the negative rate of change of the drug concentration at the absorption site. Drug is eliminated at elimination rate constant k_e, also called the fractional rate constant, which is associated with the drug concentration in the blood eliminated per unit time from the body. The bioavailability of the drug B describes the rate and extent of drug input, which is assumed to be 1 in this thesis. The volume of distribution V represents the apparent instantaneous dilution space of an instantaneously absorbed dose. Assuming the first-order kinetics, the drug concentrations at the absorption

site and in the blood, denoted by $[C_a(t), C(t)]^T$, can be described by the following linear system of differential equations:

$$\begin{aligned} \frac{\mathrm{d}C_a(t)}{\mathrm{d}t} &= -k_a C_a(t), \qquad C_a(0) = B\frac{D}{V} \\ \frac{\mathrm{d}C(t)}{\mathrm{d}t} &= k_a C_a(t) - k_e C(t), \quad C(0) = 0. \end{aligned} \tag{1.1}$$

Equation (1.2) gives the analytical solution of these two differential equations for $C(t)$. This function describes the nonlinear relationship between the drug concentration in the blood compartment and three pharmacokinetical parameters: the absorption rate k_a, the elimination rate k_e, and the volume of distribution V.

$$C(t) = \frac{BDk_a}{V(k_a - k_e)}[\exp(-k_e t) - \exp(-k_a t)]. \tag{1.2}$$

1.4 Population pharmacokinetics

Population pharmacokinetics modeling and analysis were first introduced in the context of the study of pharmaceutical agents (Sheiner, Rosenberg, and Melmon 1972), which focused primarily on mechanisms of pharmacokinetic behavior in the population rather than in any specific individual. The population pharmacokinetics developed based on the need to set dosing recommendations for a population. Actually, both the whole population and the variability among individuals are considered in the population pharmacokinetics. One of the purposes of population pharmacokinetics was to analyze routine clinical data for individualizing dosage regimens. Please refer to Sheiner, Rosenberg, and Marathe (1977), Beal and Sheiner (1982), and Sheiner and Ludden (1992).

A simple illustration can be seen in the data set of the drug indinavir in Figure 1.1, given orally in the same dose to all subjects. Although the drug concentration-time curves exhibit a similar profile for all subjects, peak concentration achieved,

rise, and decay vary significantly from subject to subject, which may be attributed to variability in the values of $[k_a,\ k_e,\ V]^T$ across individuals.

1.4.1 Three sub-models in population pharmacokinetics

Population pharmacokinetic models contain three sub-models: structural, statistical and covariate (the Food and Drug Administration 1999). The structural pharmacokinetic sub-model describes the overall trend in the data using parameters of fixed effects. The statistical sub-model accounts for variability by using two levels of random effects, inter-individual variability and residual variability. The covariate sub-model expresses relationships between covariates and model parameters.

The structural part of a mixed-effects model in pharmacokinetics data analysis describes the properties of a drug which are shared among all individuals in a population. For example, for a pharmacokinetic one-compartment model, the structural model includes fixed effects such as rates of absorption, elimination, and clearance that belong to an average individual in the population.

In population pharmacokinetics, the structural part (including residual error) of a mixed-effects model is written in Equation (1.3). Let y_{ij} be the j-th observation for the i-th individual; X_{ij} is the j-th value of the independent variables for the i-th individual, e.g. time and dose at certain time point j; $\boldsymbol{\beta}$ are the model parameters; $f(\cdot)$ is the function describing the structural PK model that relates the independent variables, X_{ij}, to the response given the parameters $\boldsymbol{\beta}$; ε_{ij} is the residual error for the j-th observation of the i-th individual; and σ^2 is the variance of the unexplained residual. The residual error is introduced to handle with unexplained variability.

$$y_{ij} = f(X_{ij}, \boldsymbol{\beta}, \varepsilon_{ij}), \quad \varepsilon_{ij} = N(0, \sigma^2). \tag{1.3}$$

The statistical sub-model considers the variability by using different levels of random effects, including inter-individual variability, intra-individual variability, and inter-occasion variability.

The inter-individual variability describes the non-explainable difference, or the

biological population variability of the model parameter between the individuals that cannot be explained solely in terms of the measurable independent variables. This means that the individual parameter value varies from the typical population parameter value to a random extent. In Equation (1.4), $\boldsymbol{\beta}_i$ denotes the parameters of the i-th individual, expressed as a function g of fixed effects parameters $\boldsymbol{\beta}$ and inter-individual random effects $\mathbf{b}_i$ describing the variability of the parameters between individuals. Random effects $\mathbf{b}_i$ are assumed to be normally distributed with mean zero and variance-covariance matrix Ψ.

$$\begin{aligned} y_{ij} &= f(X_{ij}, \boldsymbol{\beta}_i, \varepsilon_{ij}) \\ \boldsymbol{\beta}_i &= g(\boldsymbol{\beta}, \mathbf{b}_i) \\ \mathbf{b}_i &= N(0, \Psi). \end{aligned} \tag{1.4}$$

The inter-occasion variability defines another level of random effects accounting for variation between study occasions, which can be defined as each dosing interval in multiple dose studies or each treatment period of a cross-over study can be defined as an occasion. It is necessary to have more than one measurement per individual per occasion to assess the the inter-occasion variability of a specific parameter. Let η_{ik} denote the unexplained inter-occasion variability for the i-th individual and k-th occasion. In this case, the parameters of the i-th individual and k-th occasion, $\boldsymbol{\beta}_{ik}$, is expressed as a function, h, of $\boldsymbol{\beta}$, $\mathbf{b}_i$, and η_{ik}. The inter-occasion variability η_{ik} is assumed to follow a normal distribution with mean zero and variance-covariance matrix Ω.

$$\begin{aligned} \boldsymbol{\beta}_{ik} &= h(\boldsymbol{\beta}, \mathbf{b}_i, \eta_{ik}) \\ \eta_{ik} &= N(0, \Omega). \end{aligned} \tag{1.5}$$

Intra-individual variability, i.e. residual variability, ε_{ij} in Equation (1.3) and (1.4), describes the extent of deviation between the observed and the predicted value by

the model, including the inter-individual variability and the inter-occasion variability. The residual variability might be caused by errors in the documentation of the dosing and blood sampling times, analytical errors, structural model approximations in the models and other factors. The simplest model describing the residual variability is the additive error model in Equation (1.6):

$$y_{ij} = f(X_{ij}, \boldsymbol{\beta}_i) + \varepsilon_{ij}. \tag{1.6}$$

The covariate sub-model expresses the relationship between covariates and model parameters. Covariates are individual-specific variables that describe the patients' demographics, disease status or environmental factors. The parameter covariate relation can explain the variability in pharmacokinetic model parameters to a certain extent. Let z_i denote the individual-specific covariates for the i-th individual. The function g is expressed as follows:

$$\boldsymbol{\beta}_i \;\; = \;\; g(z_i, \boldsymbol{\beta}, \mathbf{b}_i). \tag{1.7}$$

1.4.2 Methods for population PK/PD modeling

The simplest method to perform population PK/PD modeling is to estimate population mean parameters by treating all data as if they arose from the same individual. That is, this method assumes that parameters for all individuals are all equal to their mean values in the population. Moreover, the entire structure of intra-individual and inter-individual variability is ignored. Therefore, Sheiner and Beal (1980) refer to this method as the *naive pooled data* method and present evidence demonstrating its poor performance in the context of compartment PK models.

Standard Two-Stage (Steimer et al. 1984) is one of the common methods to perform population PK/PD modeling. Given complete PK/PD profiles for a certain number of subjects, the model is identified separately for all subjects, obtaining individual estimates of the parameters, and the sample mean and variances are calculated using the individual estimated parameters. The Standard Two-Stage method is sim-

ple and intuitive, but it has several limits and drawbacks. This method is often used to analyze PK data from studies involving intensive sampling performed on a limited number of individuals; it ignores the precision of the individual estimates; it generally overestimates the population variance.

Bayesian estimation uses prior distribution of parameters in a population of subjects and data from an individual to estimate the individual's parameters. Since there are two sources of variability, inter- and intra- individuals, that are empirically observed in the data, it is a natural way to use the Bayesian hierarchical framework involving two stages. In the first stage, the relationship between concentration and time is modeled for each individual, addressing the intra-individual variability. In the second stage the inter-individual variability is accommodated through the specification of a model for the individuals' parameter vector. In the Bayesian hierarchical model, a third stage is incorporated to indicate the prior for the population parameters from the first and second stages. A comprehensive reference on Bayesian inference is Gelman et al. (2004).

Nonlinear mixed-effects models use a hierarchical model structure, which allows for simultaneous estimation of the inter- and intra-individual variability (random effects) as well as the influence of measured covariates on the fixed effects parameters. This method enables the analysis of PK/PD data from both sparse-sampled and unbalanced study designs. The nonlinear mixed-effects modeling approach is typically the preferred method in population PK/PD modeling because it provides reliable predictions of the variability and because it is the only practical method for analyzing data from multiple studies in a single data analysis (Davidian and Giltinan (1995, 2003)).

1.5 Objectives and organization of the thesis

The primary objectives of this thesis are to estimate nonlinear mixed-effects models with the generalized profiling method and to explore its application to pharmacokinetics. Specifically, this thesis investigates the nonlinear mixed effects models in forms

of ordinary differential equations (ODE's) generated from the compartment models in pharmacokinetics. Accordingly, general MATLAB programs are geared towards estimating parameters in ODE's from the compartment models. Methods and software tools are tested on both simulated and real data sets in pharmacokinetics. Results are verified and compared with true values or parameter estimates obtained by the package nlme in R/SPLUS. The structure of this thesis is organized as follows.

Chapter 2 provides a literature review of the nonlinear mixed effects models and some of existing software tools. We review several methods that approximate the integration of the likelihood, including the first-order method, the conditional first-order approximation, Laplacian approximation, and the Lindstrom and Bates algorithm. This chapter also introduces several widely used software tools in pharmacokinetics, including NONMEM, nlme/nlmeODE in R, PROC NLMIXED in SAS, and PKbugs.

Chapter 3 introduces how to estimate nonlinear mixed-effects models expressed by differential equations with the generalized profiling method. Preliminary techniques for functional data analysis are introduced, mainly including basis function expansions and the penalized smoothing. Based on the framework of the generalized profiling method, we propose a way to estimate nonlinear mixed effects in compartment models without solving ODE's. Four types of parameters are identified and are optimized by a multiple-level optimization in a cascaded way.

Chapter 4 implements simulations for three compartment pharmacokinetic models, including a one-compartment elimination model, a one-compartment model with first-order absorption and elimination, and a two-compartment open model. Data characteristics are described and results of estimates are compared with true values.

Applications to two real data sets in pharmacokinetics are provided in Chapter 5. Finally, Chapter 6 gives the conclusions and discusses some directions of the future work.

Chapter 2

Nonlinear mixed-effects models and software

2.1 Nonlinear mixed-effects models

The nonlinear mixed-effects (NLME) framework is widely used in modeling repeated measurements data, where measurements are obtained for a number of individuals at series of time points. Nonlinear mixed-effects models incorporate both the population and individual-specific characteristics, which are represented by fixed-effects parameters and random-effects parameters, respectively. Nonlinear mixed effects models were originally introduced in the population pharmacokinetic settings, and since then a lot of research has been done on this topic; see Pinheiro and Bates (2000) for further references.

Let y_{ij} denote the j-th observed response for the i-th individual measured at time point t_{ij} , $i = 1, 2, \cdots, N$, $j = 1, 2, \cdots, n_i$. In the pharmacokinetic settings, t_{ij} is the time after drug administration when the j-th drug concentration for the i-th subject is measured. This thesis considers time as the only dynamic explanatory variable in the model, but methodologies can be extended to a more general case with multiple dynamic covariates.

We consider the Gaussian-based two-stage nonlinear mixed-effects model expressed in Equation (2.1) and Equation (2.2). Let $\mathbf{y} = [y_{i1}, \cdots, y_{in_i}]^T$. The mean function

$f(\cdot)$ describes the within-individual behavior, which depends on a $r \times 1$ vector of individual-specific parameters for the i-th individual, $\boldsymbol{\beta}_i$, and a covariate vector $\mathbf{z}_i$. The parameter $\boldsymbol{\beta}_i$ is a r-dimensional function depending on an $p \times 1$ vector of fixed effects, $\boldsymbol{\beta}$, and a $q \times 1$ vector of random effects b_i associated with the i-th individual. A common special case is that there is a linear relationship between $\boldsymbol{\beta}_i$ and fixed and random effects. Equation (2.2) describes how element of $\boldsymbol{\beta}_i$ vary among individuals, due both to systematic association with individual attributes in A_i, a design matrix, and to unexplained variation in the population of individuals, e.g. biological variation, represented by $\mathbf{b}_i$. B_i is a design matrix typically involving only zeros and ones allowing some elements of $\boldsymbol{\beta}_i$ to have no associated random effects. The random effects $\mathbf{b}_i^{'}$s are assumed to arise from a common distribution with mean 0 and variance Ψ.

For the i-th individual, let ε_{ij} be the intra-individual error of the measurement uncertainty associated with the observed response at time point t_{ij}, and $\boldsymbol{\varepsilon}_i = [\varepsilon_{i1}, \varepsilon_{i2}, \cdots, \varepsilon_{in_i}]^T$. These random errors are assumed to be independently distributed with zero mean and constant variance across all measurements, shown in Equation (2.1).

$$\text{Individual} \quad \mathbf{y}_i = f(\mathbf{z}_i, \boldsymbol{\beta}_i) + \boldsymbol{\varepsilon}_i, \quad \boldsymbol{\varepsilon}_i \sim N(0, \sigma^2 I), \tag{2.1}$$

$$\text{Population} \quad \boldsymbol{\beta}_i = A_i\boldsymbol{\beta} + B_i\mathbf{b}_i; \ \mathbf{b}_i \sim N(0, \Psi), \quad i = 1, \cdots, N. \tag{2.2}$$

Assuming normality of the responses and random effects, the marginal likelihood of $\boldsymbol{\beta}$, Ψ, σ^2 for the i-th individual can be written as in Equation (2.3).

$$P(\mathbf{y}_i|\boldsymbol{\beta}, \sigma^2, \Psi) = \int P(\mathbf{y}_i|\mathbf{b}_i, \boldsymbol{\beta}, \sigma^2) P(\mathbf{b}_i|\Psi) d\mathbf{b}_i \tag{2.3}$$

$$= \int (2\pi\sigma^2)^{-N/2} \exp\left\{-\frac{1}{2}\left\{(\mathbf{y}_i - f(\mathbf{z}_i, \boldsymbol{\beta}_i))^T \sigma^{-2} (\mathbf{y}_i - f(\mathbf{z}_i, \boldsymbol{\beta}_i)) + \mathbf{b}_i^T \Psi^{-1} \mathbf{b}_i\right\}\right\} d\mathbf{b}_i.$$

We denote the log-likelihood of $\boldsymbol{\beta}$ and $\mathbf{b}_i$ as follows

$$l_i(\boldsymbol{\beta}, \Psi, \sigma^2, \mathbf{b}_i, \mathbf{y}_i) = \frac{1}{2}\left\{(\mathbf{y}_i - f(\mathbf{z}_i, \boldsymbol{\beta}_i))^T \sigma^{-2} (\mathbf{y}_i - f(\mathbf{z}_i, \boldsymbol{\beta}_i)) + \mathbf{b}_i^T \Psi^{-1} \mathbf{b}_i\right\}. \tag{2.4}$$

The empirical Bayes estimates $\hat{\mathbf{b}}_i$ satisfy the following equation:

$$\left.\frac{\partial l_i}{\partial \mathbf{b}_i}\right|_{\hat{\mathbf{b}}_i} = 0,\ i = 1, \cdots, N. \tag{2.5}$$

The parameters involved in the NLME model ideally can be estimated by maximizing this likelihood function, given by Equation (2.3). If $f(\cdot)$ is a linear function in terms of parameters $\boldsymbol{\beta}_i$, the integral in Equation (2.3) can be evaluated to obtain an analytic expression. However, more often in the case of NLME models, $f(\cdot)$ is a nonlinear function of parameters $\boldsymbol{\beta}_i$, making it impossible to obtain an analytic expression for the integral described in Equation (2.3) and therefore classical approach such as maximum likelihood method for parameter estimation becomes analytically intractable.

Many methods have been proposed to approximate the integration of the likelihood. A common approach to handle the integral with respect to $\mathbf{b}_i$ in Equation (2.3) involves linearization of the nonlinear model. Generally, these methods differ in their assumptions regarding the distribution of the random effects, including the inter- and intra-individual variability, and in approximations used to deal with inter-individual random effects. These methods includes the first-order method, the conditional first-order linearization, Laplacian approximation, Lindstrom and Bates algorithm, assuming various linear approximations to the nonlinear model. Roe (1997) provide a systematic comparisons of estimates of these methods.

The *First-order method* was first developed for analysing population pharmacokinetic data; see Beal and Sheiner (1982) for further references. It approximates the nonlinear model with a model that is linear in all iter-individual random effects, obtained by using a first-order Taylor expansion in all the inter-individual random effects $\mathbf{b}_i$ about 0; parameter estimates are obtained using extended least squares.

The *conditional first-order linearization* is roughly the first-order method, but the first-order Taylor series expansion is about conditional estimates, i.e. empirical Bayes estimates, of the inter-individual random effects, $\hat{\mathbf{b}}_i$, rather than about $\mathbf{b}_i = 0$. The contribution of the second-order partial derivatives is usually negligible compared to

that of the square of the first-order partial derivatives because the second-order Taylor expansion in Equation (2.6) is about the value of b_i which minimizes l_i. Lindstrom and Bates (1990) and Wolfinger (1993) are good references on this method.

The *Laplacian method* of evaluating the exact marginal likelihood consists of using a second-order Taylor expansion of l_i around the empirical Bayes estimate $\mathbf{b}_i$ which minimizes l_i, i.e. the mode of the posterior distribution for $\mathbf{b}_i$ given $\boldsymbol{\beta}, \Psi$, and σ^2.

We denote the first and second derivatives of l_i as follows:

$$\begin{aligned} l_i' &= \frac{\partial l_i}{\partial \mathbf{b}_i} \\ l_i'' &= \frac{\partial^2 l_i}{\partial \mathbf{b}_i \partial \mathbf{b}_i^T}. \end{aligned}$$

Thus

$$\begin{aligned} l_i &\approx l_i(\hat{\mathbf{b}_i}) + l_i'(\hat{\mathbf{b}_i})^T(\mathbf{b}_i - \hat{\mathbf{b}_i}) + \frac{1}{2}(\mathbf{b}_i - \hat{\mathbf{b}_i})^T l_i''(\hat{\mathbf{b}_i})(\mathbf{b}_i - \hat{\mathbf{b}_i}) \\ &= l_i(\hat{\mathbf{b}_i}) + \frac{1}{2}(\mathbf{b}_i - \hat{\mathbf{b}_i})^T l_i''(\hat{\mathbf{b}_i})(\mathbf{b}_i - \hat{\mathbf{b}_i}), \end{aligned} \tag{2.6}$$

Consequently, the integral in Equation (2.3) can be approximated by Equation (2.7); see Davidian and Giltinan (1995), Wolfinger (1993), and Vonesh (1996) for further readings.

$$\int \exp(-l_i) d\mathbf{b}_i \approx (2\pi)^{q/2} |l_i''(\hat{\mathbf{b}_i})|^{-1/2} \exp\{l_i(\hat{\mathbf{b}_i})\}. \tag{2.7}$$

The *Lindstrom and Bates algorithm* can be derived using Laplacian approximation. The method uses a first-order Taylor expansion about the conditional estimates of the inter-individual random effects. The estimation algorithm alternates between two steps: a penalized nonlinear least-squares (PNLS) step and a linear mixed-effects (LME) step. For the simplicity of calculation, we express the random effects variance-covariance matrix in terms of the relative precision factor Δ, which satisfies $\Psi^{-1} = \sigma^{-2}\Delta^T\Delta$.

In the PNLS step, the conditional random effects $\mathbf{b}_i$ and the conditional estimates

of the fixed effects $\boldsymbol{\beta}$ based on the current estimate of Ψ are obtained by minimizing the PNLS objective function as follows

$$O_{PNLS} = \sum_{i=1}^{N} (\mathbf{y}_i - f_i(\mathbf{z}_i, \boldsymbol{\beta}_i))^T (\mathbf{y}_i - f_i(\mathbf{z}_i, \boldsymbol{\beta}_i)) + \mathbf{b}_i^T \Delta^T \Delta \mathbf{b}_i, \tag{2.8}$$

In order to update the estimate of Ψ, the mean function $f(\cdot)$ is linearized in the LME step using a first-order Taylor expansion around the current estimates of $\boldsymbol{\beta}$ and the conditional estimates of the random effects $\mathbf{b}_i$ denoted by $\hat{\mathbf{b}_i}$. The approximative log-likelihood function for the estimation of Ψ in the LME step can thereby be written as

$$\begin{aligned} \log L_{LME}(\boldsymbol{\beta}, \sigma^2, \Delta) = -\frac{\sum_i^N n_i}{2} \log 2\pi\sigma^2 - \frac{1}{2} \sum_{i=1}^{N} \Bigg\{ \log \left(\left| I + \frac{\partial f_i}{\partial \mathbf{b}_i^T} \Delta^{-1} \Delta^{-T} \frac{\partial f_i^T}{\partial b_i^T} \right| \right) \\ + \left[\mathbf{y}_i - f_i + \frac{\partial f_i}{\partial \mathbf{b}_i^T} \hat{\mathbf{b}_i} \right]^T \left(I + \frac{\partial f_i}{\partial \mathbf{b}_i^T} \Delta^{-1} \Delta^{-T} \frac{\partial f_i^T}{\partial \mathbf{b}_i^T} \right)^{-1} \left[\mathbf{y}_i - f_i + \frac{\partial f_i}{\partial \mathbf{b}_i^T} \hat{\mathbf{b}_i} \right] \Bigg\} \end{aligned} \tag{2.9}$$

The *Bayesian approach* to population pharmacokinetic modeling computes the posterior distribution of the parameters given the observed data assuming hyperprior distributions of the parameters. This method is based on the calculation of the full conditional probability distribution of the parameters, including $\mathbf{b}$, Ψ, and σ^2. More references on Bayesian analysis and its applications to pharmacokinetics include Gelman et al. (2004), Davidian and Giltinan (2003), Lunn, Best, Thomas, Wakefield, and Spiegelhalter (2002), and Wakefield (1996).

In the Bayesian framework, prior knowledge based on previous studies can be introduced by specifying appropriate prior distributions for parameters. Bayesian inferences consider all the parameters, including both random effects and fixed effects, as random because they are taken to have probability distributions.

To implement the Bayesian inferences for nonlinear mixed-effects models, a prior distribution must be indicated for the structural parameters $\boldsymbol{\theta} = (\boldsymbol{\beta},\ \Psi,\ \sigma^2)$, which is denoted by $P(\boldsymbol{\theta}) = P(\boldsymbol{\beta},\ \Psi,\ \sigma^2)$. This prior is often called a hyperprior in the

context of population analysis. There is no prior specified for the random effects $\mathbf{b}_i,\ i = 1, \cdots, N$. The joint posterior density of all of $\boldsymbol{\beta},\ \Psi,\ \sigma^2,\ \mathbf{b}$ is given by:

$$P(\boldsymbol{\beta}, \Psi, \sigma^2, \mathbf{b}|\mathbf{y}) = \frac{\sum_{i=1}^{N} P(\mathbf{y}_i|\mathbf{b}_i; \sigma^2) P(\mathbf{b}_i|\boldsymbol{\beta}, \Psi) P(\boldsymbol{\beta}, \Psi, \sigma^2)}{\int\int\int\int \sum_{i=1}^{N} P(\mathbf{y}_i|\mathbf{b}_i; \sigma^2) P(\mathbf{b}_i|\boldsymbol{\beta}, \Psi) P(\boldsymbol{\beta}, \Psi, \sigma^2) d\mathbf{b} d\boldsymbol{\beta} d\Psi d\sigma^2} \quad (2.10)$$

2.2 Software

Many software tools have been developed and applied to pharmacokinetics. This section introduces NONMEM, nlme/nlmeODE in R, PROC NLMIXED in SAS, and PKbugs. While PKbugs uses the Bayesian method, NONMEM, nlme and PROC NLMIXED are parametric non-Bayesian likelihood approaches proposing different approximations of the population likelihood function. More references on PK/PD software tools can be found in Pillai, Mentre, and Steimer (2005).

2.2.1 NONMEM

Software NONMEM was developed by Beal and Sheiner (1980) and has been widely used by practitioners to implement PK/PD data analysis. The NONMEM program performs maximum likelihood estimation based on several approximation methods of the log-likelihood function, including first-order, first-order conditional estimation method, and the Laplacian method. A more recent reference is Beal and Sheiner (1994).

NONMEM has various attractive features, thus it is regarded as the gold-standard software for population PK/PD data analysis. The most prominent advantage is that it has many libraries for fitting standard PK/PD compartment models. Immediate analyses are easily implemented based on these models. Moreover, multiple and complicated dosing history can be specified in models. Another feature is its ability of fitting models expressed by ordinary differential equations. In addition, NONMEM runs in many types of operating systems, and computation is fast because it is written in the compiled language Fortran. Generally, NONMEM is quite accurate, stable,

flexible, and fast to fit PK/PD models.

However, models in NONMEM run in a batch mode, and this increases the difficulty of using it. In addition, graphics for diagnostics are line-printer style so that users usually have to employ some external packages.

2.2.2 nlme

The R/S-PLUS software package nlme, developed by J. Pinheiro, D. Bates and M. Lindstrom, is used to fit nonlinear mixed-effects models. The package nlme implements the Lindstrom and Bates algorithm described previously. A detailed description of nlme and some example functions are given in Pinheiro and Bates (2000).

The package nlme is not specifically designed for PK/PD models in which there are only two levels of random effects. Instead, the package nlme is capable of fitting multiple-level models in other fields. Another benefit of nlme is that many functions can be used to implement model checking and plotting. Moreover, models in nlme can be modified and updated due to the object-oriented programming.

One disadvantage of nlme is that it is difficult to fit multiple-dose models. nlme itself cannot handle compartment models expressed by ODE's without closed-form solutions. It was necessary to implement an ODE solver to be able to handle these nonlinear PK/PD models. For that purpose, the nlmeODE package (Tornϕ et al. 2004) was developed by combining nlme with the odesolve package (Setzer 2003) in R. The odesolve package provides an interface to the Fortan ODE solver Isoda (Petzold 1983), which can be used to solve initial value problems for systems of first-order ODE's. Computation times are usually significantly longer using nlme together with nlmeODE compared to NONMEM because R is an interpreted language.

2.2.3 SAS NLMIXED

The current version of statistical software SAS includes the procedure NLMIXED for fitting nonlinear mixed-effects models. The procedure NLMIXED can use the first-order approximation to approximate the log-likelihood function, but this procedure

uses the numerical integration approximation method, adaptive Gaussian quadrature, as the default method to optimize the objective function. This method can be viewed as giving the exact answer to the optimization problem, but it leads to a high computation load. All of the parameters, including the random effects and fixed effects, are optimized by an iterative process, finding the maximum likelihood estimates of the fixed effects for these values, re-estimating the random effects, then going back to the fixed effects. However, the NLMIXED procedure is quite sensitive to starting values and parameterization of the model (Pillai et al. 2005).

2.2.4 PKBugs

PKBugs was developed by Dave Lunn at the Department of Epidemiology and Public Health of Imperial College at St Mary's Hospital London; see PKBugs (2004) for further references. PKBugs is an efficient and user-friendly interface for specifying complex population PK/PD models within the widely-used WinBUGS (Bayesian inference Using Gibbs Sampling) software. The software and manuals for both PKBugs and WinBUGS are currently free, and are available at http://www.cran.r-project.org.

WinBUGS is a powerful tool of Bayesian modeling to analyze arbitrarily Bayesian full probability models using techniques like Markov chain Monte Carlo (MCMC). It is possible to specify population PK/PD models using WinBUGS directly, but high-level programming skills are required and models would be too complicated for many practitioners in the field of population PK/PD.

PKBugs alleviates these difficulties of model specification by providing a user-friendly interface between users and WinBUGS. Practitioners can specify a PK/PD model by a series of simple dialogue boxes and menu commands. PKBugs can parse the specified model and generate the pseudo-code of WinBUGS. It can also convert NONMEM data files into the format used in WinBUGS. In this way, WinBUGS will implement the rest analysis and return results.

Chapter 3

Estimating nonlinear mixed-effects models by the generalized profiling method

This chapter begins with a brief introduction to functional data analysis (FDA) because the generalized profiling method is developed to deal with functional data. Some techniques of FDA are briefly discussed, mainly including basis function expansions and the penalized smoothing.

The generalized profiling method was proposed by Ramsay, Hooker, Campbell, and Cao (2007). This method has been applied to several fields, such as adaptive penalized smoothing, generalized semiparametric additive models, estimating differential equations, and Bayesian analysis. Detailed applications can be found in Cao (2006). This thesis explores the application to nonlinear mixed-effects models expressed by ODE's based on the framework of the generalized profiling method. Different types of parameters are identified and optimized through a nested multiple-level optimization.

3.1 Introduction to functional data analysis

Functional data analysis converts a set of data into infinite-dimensional curves and surfaces instead of looking at data individually. These functional data might be complicated in some ways; for example, they might not be equally spaced. Even though there are only discrete observations, it is often desired to use functions to reflect a smooth variation in the measured variable. An important theme of FDA is its many uses of derivatives. Derivatives are often of direct interest, or some features of data can be better illustrated by derivatives of certain orders. Therefore, it is better to think of the records as functions instead of observations in discrete time. Ramsay and Silverman (2005) can be consulted for further details.

FDA represents noisy observations in discrete time points with a linear combination of basis functions. The commonly used basis systems include splines, Fourier bases, wavelet bases, polynomial bases, polygonal bases, step-function bases, and constant bases. Basis approximation provides a good estimation for functional data given that the basis functions can describe the essential characteristics of data. For instance, the Fourier basis system is often used for periodic data while the spline system is usually used for non-periodic data.

Smoothing is used to describe the process of converting discrete observations into functions, and the discrete values are assumed to be subject to observational errors. The *roughness penalized smoothing* method is widely used, which provides a good approximation to functional data as well as continuous control of the smoothness. It is especially useful to estimate derivatives. After smoothing, the dimension is reduced from the number of observations, n, per subject to the number of basis functions, J, used to represent functional data. However, the number of basis functions may be larger than the number of observations when the underlying functions are difficult to approximate because of sharp changes, discontinuity, or other features. In this case, a penalty term is used to control the roughness of estimated functions in order to avoid the problem of overfitting. The penalty term can be defined by a derivative of some order. Moreover, differential equations can also be applied to define the penalty

term in penalized smoothing, leading to better estimates for smooth functions and their derivatives Ramsay and Silverman (2002).

It is helpful to compare functional data analysis with time series analysis and longitudinal data analysis. Time series analysis usually requires time points between observations to be equally spaced. Differencing is widely used in time series analysis. In contrast, derivatives are the most popular elements in FDA. Compared with longitudinal data analysis, functional data analysis requires more frequent observations and the time variable itself usually does not appear as an explicit covariate in functional models while some covariates and parameters can often be functions of time.

3.2 Nonlinear mixed-effects models involving ODE's

In a general compartment model consisting of K components, we write concentrations of K compartments at time t for the i-th individual as $\mathbf{x}_i(t) = (x_{i1}(t), \cdots, x_{iK}(t))^T$. We denote fixed-effects parameters as $\boldsymbol{\beta} = [\beta_1, \cdots, \beta_p]^T$, and random-effects parameters $\mathbf{b}_i = [b_{i1}, \cdots, b_{iq}]^T \sim \texttt{Normal}(0, \Psi)$. Let M_i be the $K \times K$ system transfer matrix containing the rate constants. The concentrations satisfy the linear system of differential equations as follows

$$\frac{d\mathbf{x}_i(t)}{dt} = M_i \mathbf{x}_i(t). \tag{3.1}$$

For example, in the pharmacokinetic one-compartment model with first-order absorption and elimination, $\mathbf{x}_i(t)$ is a vector of the concentrations at the absorption site and in the blood at time t for the ith individual. The transfer matrix M_i contains both fixed effects and random effects. Let $\mathbf{b}_i = [b_{i1}, b_{i2}]^T$ be a vector of random-effects parameters and let $\boldsymbol{\beta} = [k_a, k_e]^T$ be a vector of fixed-effects parameters for the absorption rate constant and elimination rate constant. The transfer matrix M_i can be

expressed as follows

$$M_i = \begin{bmatrix} -(k_a + b_{i1}) & 0 \\ k_a + b_{i1} & -(k_e + b_{i2}) \end{bmatrix}. \tag{3.2}$$

Let y_{ijk}, $i = 1, \cdots, N$, $j = 1, \cdots, n_i$, $k = 1, \cdots, K$, be the observation of the k-th component at time t_j for the i-th subject. The nonlinear mixed-effects models involving ODE's can be expressed as follows:

$$\begin{aligned} y_{ijk} &= x_{ik}(t_{ij}) + \epsilon_{ij}, \\ \epsilon_{ij} &\sim \texttt{Normal}(0, \sigma^2) \\ i &= 1, \cdots, N;\ j = 1, \cdots, n_i;\ k = 1, \cdots, K. \end{aligned} \tag{3.3}$$

Let $\mathbf{y}_{ij} = [y_{ij1}, \cdots, y_{ijK}]^T$, Equation (3.3) can be rewritten as

$$\begin{aligned} \mathbf{y}_{ij} &= \mathbf{x}_i(t_{ij}) + \epsilon_i, \\ \epsilon_i &\sim \texttt{Normal}(0, \Sigma),\ \Sigma = \sigma^2 I \\ i &= 1, \cdots, N;\ j = 1, \cdots, n_i, \end{aligned} \tag{3.4}$$

It is easy to implement parameter estimation, model fitting and verification if ODE's can be solved analytically. However, very few real-world ODE's can be solved analytically, and numerical approximation is almost the only option in the large and realistic world of nonlinear ODE's and non-stationary processes.

Although many methods have been proposed to estimate parameters in ODE's without closed-form solutions, there are many drawbacks when fitting noisy data. A commonly used method is the nonlinear optimization procedure. Using this method, the computations are usually intensive because ODE's are repeatedly numerically solved when updating the parameter values and initial values of components. Moreover, initial values of components become additional parameters to estimate. Finally, this procedure relies heavily on the quality of the starting values for parameters and the initial values of components. As a result, the algorithms can be easily trapped in

local minima, and ODE's may not even be solvable in some cases. The other common method is the Bayesian approach, which can also handle mixed-effects models. However, this method shares the same shortcoming with many current methods in terms of the intensive computation. In addition, it is difficult to select appropriate initial values and priors.

In the generalized profiling method, data are smoothed by a linear combination of basis functions penalized by its fidelity to ODE's. A smoothing parameter reconciles the trade-off between fitting the data and fidelity to ODE's. Parameters are divided into several levels, and optimized by different criteria in different levels. Through a nested optimization, the dimensionality of the parameter space is reduced by treating one group of parameters as an explicit or implicit function of other parameters, and thereby the optimization surface becomes smoother.

The generalized profiling method has several advantages in comparison with other methods. First, ODE's do not have to be solved, and therefore the initial values of components are not needed. In addition, this method can also work satisfactorily when some components are not observable. Finally, the process of estimation is very stable because functional relationships among parameters are worked out analytically. Both the gradient and Hessian matrices for the optimization criteria can be calculated analytically using the Implicit Function Theorem, which definitely speeds up the computation.

3.3 Basis function expansions for ODE solutions

A basis function system is a set of known functions $\phi_l, \ \ l = 1, \cdots, J$, that are used to approximate any function by a weighted sum or linear combination on the condition that the number of functions J is large enough. Functions $\phi_l, \ \ l = 1, \cdots, J$, are called bases and are mathematically independent of each other, that is, none of them can be written as a linear combination of other bases in the set. The basis approximation for a function $\mathbf{z}(t)$ is a linear expansion in terms of J known basis functions ϕ_l and a

series of coefficients c_l for each basis, shown in Equation (3.5).

$$\mathbf{z}(t) = \sum_{l=1}^{J} c_l \phi_l(t). \tag{3.5}$$

Let $\mathbf{c}$ be a vector of length J with each element as one of the coefficients and $\boldsymbol{\phi}$ be a vector with its elements as one of the basis functions. In the vector and matrix version, Equation (3.5) can be expressed as

$$\mathbf{z}(t) = \mathbf{c}^T \boldsymbol{\phi}(t) = \boldsymbol{\phi}(t)^T \mathbf{c}. \tag{3.6}$$

For compartment models, we use the same basis function system to approximate all the components of the concentration function $\mathbf{x}_i$ for the i-th individual as follows,

$$x_{ik}(t) = \sum_{l=1}^{J_i} c_{ikl} \phi_{il}(t) = \mathbf{c}_{ik}^T \phi_i(t), \tag{3.7}$$

where J_i is the number of basis functions in vector $\phi_i(t)$, coefficients of basis functions $\mathbf{c}_{ik} = (c_{ik1}, \cdots, c_{ikJ})^T$, and basis functions $\phi_i(t) = (\phi_{i1}(t), \cdots, \phi_{iJ_i}(t))$. Define the $n_i \times J_i$ matrix Φ_i as containing the values of the basis functions at the times $t_{ij}, j = 1, \cdots, n_i$. We define $\boldsymbol{\Phi}_i$ as a matrix of $(Kn_i) \times (KJ_i)$ as follows

$$\boldsymbol{\Phi}_i = I_K \otimes \Phi_i,$$

where $\otimes$ is the Kronecker product. Let $\mathbf{c}_i = [\mathbf{c}_{i1}^T, \cdots, \mathbf{c}_{iK}^T]^T$. All the observations for the i-th individual can be approximated as $\mathbf{x}_i = \boldsymbol{\Phi}_i \mathbf{c}_i$. We define $\boldsymbol{\phi}_i$ as a matrix as follows

$$\boldsymbol{\phi}_i(t) = I_K \otimes \phi_i(t).$$

The values of the function $\mathbf{x}_i(t)$ at time t is obtained by

$$\mathbf{x}_i(t) = \boldsymbol{\phi}_i(t)^T \mathbf{c}_i.$$

For the future use, we define a differential operator L of $\mathbf{x}_i(t)$ as

$$L\mathbf{x}_i(t) = \frac{d\mathbf{x}_i(t)}{dt} - M_i\mathbf{x}_i(t). \tag{3.8}$$

In terms of basis functions, the above equation becomes

$$\begin{aligned} L_i\boldsymbol{\phi}_i(t)^T\mathbf{c}_i &= \frac{d\boldsymbol{\phi}_i(t)^T\mathbf{c}_i}{dt} - M_i\boldsymbol{\phi}_i(t)^T\mathbf{c}_i \\ \text{and } L_i\boldsymbol{\phi}_i(t) &= \frac{d\boldsymbol{\phi}_i(t)}{dt} - \boldsymbol{\phi}_i(t)M_i^T. \end{aligned} \tag{3.9}$$

For simplicity , let $\mathbf{y}_i = (\mathbf{y}_{i1}^T, \cdots, \mathbf{y}_{in_i}^T)^T$, $\mathbf{y} = (\mathbf{y}_1^T, \cdots, \mathbf{y}_N^T)^T$, $\mathbf{x} = (\mathbf{x}_1^T, \cdots, \mathbf{x}_N^T)^T$, and $L\mathbf{x}(t) = (L\mathbf{x}_1(t)^T, \cdots, L\mathbf{x}_N(t)^T)^T$. Furthermore, we denote $\boldsymbol{\Phi}$ as the block diagonal matrix of $\boldsymbol{\Phi}_1, \cdots, \boldsymbol{\Phi}_N$.

3.4 Estimation with the generalized profiling method

In the model expressed by Equation (3.4), there are four types of parameters to be estimated: coefficients $\mathbf{c}$ defining basis function expansions; random effects $\mathbf{b}$; fixed effects $\boldsymbol{\beta}$, a variance-covariance matrix for random effects, Ψ, and residual variance σ^2; a smoothing parameter λ. These parameters are treated hierarchically in four different classes of parameters that we refer to as a parameter cascade. The fixed effects $\boldsymbol{\beta}$, variance-covariance matrix Ψ, and residual variance σ^2 are *structural* in the sense of being of primary interest. The coefficients $\mathbf{c}$ and random effects $\mathbf{b}$ are considered *nuisance* parameters that are essential for fitting the data but usually not of direct concern. The sizes of nuisance parameters vary with the number of observations, and other aspects of the structure of the data; the number of nuisance parameters can be orders of magnitude larger than the number of structural parameters.

To simplify the computation, we define the relative precision factor Δ satisfying $\Delta^T\Delta = \Psi^{-1}\sigma^2$, and we estimate the parameters Δ and σ^2 instead of Ψ and σ^2. Structural parameters, including $\boldsymbol{\beta}$, Δ, and σ^2, are denoted as $\boldsymbol{\theta}$. Based on the above discussion, it is reasonable to define the four levels of parameters in a cascaded

way: coefficients of basis function expansions $\mathbf{c}$ are defined as functions $\hat{\mathbf{c}}(\mathbf{b}, \boldsymbol{\theta}, \lambda)$ depending on $\mathbf{b}$, $\boldsymbol{\theta}$, and λ; random effects $\mathbf{b}$ are defined as functions $\hat{\mathbf{b}}(\boldsymbol{\theta}, \lambda)$ conditional on $\boldsymbol{\theta}$ and λ; the structural parameters $\boldsymbol{\theta}$ are defined as functions $\hat{\boldsymbol{\theta}}(\lambda)$ conditional on λ. To realize these parameter estimates, four different criteria are optimized in our generalized profiling procedure, where the criteria are $J(\mathbf{c}|\mathbf{b}, \boldsymbol{\theta}, \lambda), H(\mathbf{b}|\boldsymbol{\theta}, \lambda)$, $G(\boldsymbol{\theta}|\lambda)$, and $F(\lambda)$ for the coefficients of basis function expansions, random effects, and structural parameters, and the smoothing parameter, respectively. We shall refer to $J(\mathbf{c}|\mathbf{b}, \boldsymbol{\theta}, \lambda)$ as the inner criterion at the inner optimization level; we refer to $H(\mathbf{b}|\boldsymbol{\theta}, \lambda)$ as the random-effects criterion at the middle optimization level; we refer to $G(\boldsymbol{\theta}|\lambda)$ the structural criterion at the outer optimization level; we refer to $F(\lambda)$ as the smoothing criterion at the outermost optimization level. The inner criterion $J(\mathbf{c}|\mathbf{b}, \boldsymbol{\theta}, \lambda)$ is based on an error sum of squares (SSE), or other suitable measures of the quality of the fit to data, plus a regularization or smoothing term that defines smoothness in terms of $\mathbf{c}$ through which the criterion will also depend on the random-effects parameters $\mathbf{b}$, structural parameters $\boldsymbol{\theta}$, and the complexity parameter λ. The nuisance parameter vector $\mathbf{c}$ is removed from the parameter space by defining the inner optimization conditional on $\mathbf{b}$, $\boldsymbol{\theta}$, and λ. Each time $\mathbf{b}$, $\boldsymbol{\theta}$, and λ are changed, the fitting criterion $J(\mathbf{c}|\mathbf{b}, \boldsymbol{\theta}, \lambda)$ is re-optimized with respect to $\mathbf{c}$ alone. The amount of regularization is controlled by the smoothing parameter λ. The random-effects criterion $H(\mathbf{b}|\boldsymbol{\theta}, \lambda)$ defines fit to the data conditional on $\boldsymbol{\theta}$, and λ. This criterion is also regularized, but the regularization term is defined as a function of $\mathbf{b}$ rather than as a function of $\mathbf{c}$, called a penalized nonlinear least squares. The structural criterion $G(\boldsymbol{\theta}|\lambda)$ is then optimized with respect to the structural parameters alone. The smoothing criterion $F(\lambda)$ is optimized with respect to λ.

In our generalized profiling method, the Newton-Raphson method is applied in each level of optimization, and the gradient and Hessian matrix are worked out analytically with the Implicit Function Theorem so that the optimization process converges quickly and stably.

As a consequence of these conditional optimizations, the nuisance parameter vector $\mathbf{c}$ is removed from the parameter space as an independent parameter by defining

it through the inner optimization as a function of $\mathbf{b}$, $\boldsymbol{\theta}$, and λ. In a similar way, the random effects vector $\mathbf{b}$ is removed from the parameter space through the random effects optimization as a function of $\boldsymbol{\theta}$ and λ, and the structural parameters $\boldsymbol{\theta}$ is removed from the parameter space through the structural optimization as a function of λ. The smoothing parameter λ is selected by the criterion of generalized cross-validation (GCV). Our final parameter estimates become the functional cascade $\hat{\mathbf{c}}(\hat{\mathbf{b}}(\hat{\boldsymbol{\theta}}(\hat{\lambda})))$, $\hat{\mathbf{b}}(\hat{\boldsymbol{\theta}}(\hat{\lambda}))$, $\hat{\boldsymbol{\theta}}(\hat{\lambda})$, and $\hat{\lambda}$ defined by the optimization with respect to criterion J, H, G, and F respectively.

In the following, we introduce our selection of the optimization criteria to fit nonlinear mixed-effects models and some computational techniques. After introducing basis function expansions for the solution functions of ODE's, we begin with the inner criterion defined by the penalized smoothing. Then we introduce the penalized nonlinear least squares for random effects. For the structural parameters, two situations are considered. In the first situation, Δ and σ^2 are assumed known and not of interest, and structural parameters only contain the fixed-effects parameters. A more complicated situation is to estimate not only random effects $\mathbf{b}$ and fixed effects $\boldsymbol{\beta}$, but also the relative precision factor Δ and σ^2. Finally, the criterion GCV for the smoothing parameter is described.

3.4.1 Penalized smoothing with the penalty defined by ODE's

The fitting function can be estimated by minimizing the sum of squared errors (SSE), which can be written as:

$$\text{SSE} = (\mathbf{y} - \mathbf{x})^T \mathbf{W} (\mathbf{y} - \mathbf{x}), \tag{3.10}$$

where $\mathbf{W}$ is a diagonal matrix with diagonal values of either 0 or 1; values are 0 for unobservable or missing observations, and 1 for measured observations.

The basis system must not only fit the data, but also have the capacity to approximate ODE solutions and derivatives involved in ODE's. To avoid the problem of over-fitting, smoothing often requires a penalty term to penalize the roughness of

the smooth function. For instance, in order to obtain a fitting function, the penalty term can be defined in terms of the differential operator $L\mathbf{x}(t)$.

$$\text{PEN}_1 = \int (L\mathbf{x}(t))^T (L\mathbf{x}(t)) dt. \tag{3.11}$$

The fitting criterion to estimate the fitting function is therefore composed of SSE and the penalty term. A smoothing parameter λ controls the trade off of fitting the data and infidelity to ODE's.

$$J(\mathbf{c}|\boldsymbol{\theta}, \mathbf{b}, \lambda) \;\; = \;\; (\mathbf{y} - \mathbf{x})^T \mathbf{W} (\mathbf{y} - \mathbf{x}) + \lambda \int (L\mathbf{x}(t))^T (L\mathbf{x}(t)) dt.$$

We define $\mathbf{R} = \int (L\boldsymbol{\phi}(t))^T (L\boldsymbol{\phi}(t)) dt$, the previous equation can be rewritten as

$$J(\mathbf{c}|\mathbf{b}, \boldsymbol{\theta}, \lambda) \;\; = \;\; (\mathbf{y} - \boldsymbol{\Phi}\mathbf{c})^T \mathbf{W} (\mathbf{y} - \boldsymbol{\Phi}\mathbf{c}) + \lambda \mathbf{c}^T \mathbf{R}(\mathbf{b}, \boldsymbol{\theta}, \lambda) \mathbf{c}.$$

In this case, we can minimize the fitting criterion $J(\mathbf{c}|\mathbf{b}, \boldsymbol{\theta}, \lambda)$ and derive the analytical form of the coefficient vector $\mathbf{c}$ as:

$$\hat{\mathbf{c}}(\mathbf{b}, \boldsymbol{\theta}, \lambda) = [\boldsymbol{\Phi}^T \mathbf{W} \boldsymbol{\Phi} + \lambda \mathbf{R}(\mathbf{b}, \boldsymbol{\theta}, \lambda)]^{-1} \boldsymbol{\Phi}^T \mathbf{W} \mathbf{y}. \tag{3.12}$$

If an explicit solution for $\mathbf{c}$ is not possible, we can still define this function implicitly by optimizing a penalized or regularized fitting function like $J(\mathbf{c}|\mathbf{b}, \boldsymbol{\theta}, \lambda)$ each time we change the value of random effects or fixed-effects parameters. When the functional relationship is implicit in this way, the Implicit Function Theorem permits the explicit gradient and Hessian calculations that are essential for fast optimization.

Numerical integration

When L is a nonlinear differential operator, we can approximate the penalty term as:

$$\text{PEN}_1 \approx \sum_{j=1}^{J} v_j (L\boldsymbol{\phi}(t_j))(L\boldsymbol{\phi}(t_j))^T, \tag{3.13}$$

where t_j is a quadrature point and v_j is the corresponding quadrature weight. Let ξ_l be a unique knot location, the evaluation points t_j can be chosen by dividing each interval $[\xi_l, \xi_{l+1}]$ into the odd number of equal-sized intervals, say r, and the quadrature weight $v_j = [1, 4, 2, 4, 1](\xi_{l+1} - \xi_l)/5$ by Simpson's rule. In our experience, the integrals can be satisfactorily approximated using $r = 5$. In practice, the total quadrature points and weights along with the corresponding basis function values can be saved at the beginning of the computation in order to save computation time. The speed of computation can be further improved by using the sparse matrix methods in MATLAB if a B-spline basis function is used.

3.4.2 Optimizing random effects b

The optimization criterion for the random effects $H(\mathbf{b}|\boldsymbol{\theta}, \lambda)$ is the penalized nonlinear least squares. This fitting criterion is composed of the sum of squared errors (SSE) conditional on $\boldsymbol{\theta}$ and λ, and a penalty term on the random effects $\mathbf{b}$.

$$\begin{aligned} & H(\mathbf{b}|\boldsymbol{\theta}, \lambda) = \mathrm{SSE}(\mathbf{b}|\boldsymbol{\theta}, \lambda) + \mathrm{PEN}_2(\mathbf{b}|\boldsymbol{\theta}, \lambda) \\ = \; & (\mathbf{y} - \mathbf{x})^T \mathbf{W} (\mathbf{y} - \mathbf{x}) + \mathbf{b}^T \Delta^T \Delta \mathbf{b} \\ = \; & \mathbf{y}^T [I - A(\mathbf{b}|\boldsymbol{\theta}, \lambda)]^T \mathbf{W} [I - A(\mathbf{b}|\boldsymbol{\theta}, \lambda)] \mathbf{y} + \mathbf{b}^T \Delta^T \Delta \mathbf{b}. \end{aligned}$$

where the smoothing matrix $A(\mathbf{b}|\boldsymbol{\theta}, \lambda) = \boldsymbol{\Phi}[\boldsymbol{\Phi}^T \mathbf{W} \boldsymbol{\Phi} + \lambda \mathbf{R}(\mathbf{b}|\boldsymbol{\theta}, \lambda)]^{-1} \boldsymbol{\Phi}^T \mathbf{W}$.

Since for each $\mathbf{b}$, $\boldsymbol{\theta}$, and λ, we can estimate $\mathbf{c}$ as an explicit function of $\mathbf{b}$, $\boldsymbol{\theta}$, and λ, the fitting criterion H can be regarded as a function of $\mathbf{b}$ alone conditional on $\boldsymbol{\theta}$ and λ. However, the estimate $\hat{\mathbf{b}}(\boldsymbol{\theta}, \lambda)$ can not be expressed as an explicit function of $\boldsymbol{\theta}$ and λ. By calculating the gradient and Hessian of $H(\mathbf{b}|\boldsymbol{\theta}, \lambda)$, the Newton-Raphson method is applied to find the estimate $\hat{\mathbf{b}}(\boldsymbol{\theta}, \lambda)$. Details of the calculations are provided in the Appendix.

3.4.3 Criterion when the structural parameters only have $\boldsymbol{\beta}$

If Δ and σ^2 are assumed known and not of interest, structural parameters only contain the fixed effects parameters; that is, $\boldsymbol{\theta} = \boldsymbol{\beta}$. In this case, the structural optimization criterion $G(\boldsymbol{\theta}|\lambda)$ is defined by the sum of squared errors (SSE) only.

$$\begin{aligned} G(\boldsymbol{\theta}|\lambda) &= \text{SSE}(\boldsymbol{\theta}|\lambda) \\ &= (\mathbf{y} - \mathbf{x})^T \mathbf{W} (\mathbf{y} - \mathbf{x}) \\ &= \mathbf{y}^T [I - A(\hat{\mathbf{b}}(\boldsymbol{\theta}, \lambda), \boldsymbol{\theta}|\lambda)]^T \mathbf{W} [I - A(\hat{\mathbf{b}}(\boldsymbol{\theta}, \lambda), \boldsymbol{\theta}|\lambda)] \mathbf{y}. \end{aligned} \tag{3.14}$$

where the smoothing matrix

$$A(\hat{\mathbf{b}}(\boldsymbol{\theta}, \lambda), \boldsymbol{\theta}|\lambda) = \boldsymbol{\Phi}[\boldsymbol{\Phi}^T \mathbf{W} \boldsymbol{\Phi} + \lambda \mathbf{R}(\hat{\mathbf{b}}(\boldsymbol{\theta}, \lambda), \boldsymbol{\theta}|\lambda)]^{-1} \boldsymbol{\Phi}^T \mathbf{W}. \tag{3.15}$$

Newton-Raphson method is used again to find the estimate $\hat{\boldsymbol{\beta}}$. Since the estimate $\hat{\mathbf{b}}(\boldsymbol{\theta}, \lambda)$ can not be expressed as an explicit function of $\boldsymbol{\theta}$ and λ, the Implicit Function Theorem is applied to calculate the gradient and Hessian matrix with respect to $\boldsymbol{\beta}$. The total derivative of $G(\boldsymbol{\beta}, \hat{\mathbf{b}}(\boldsymbol{\beta}|\lambda))$ with respect to $\boldsymbol{\beta}$ is as follows

$$\frac{dG(\boldsymbol{\beta}, \hat{\mathbf{b}}(\boldsymbol{\beta}, \lambda)|\lambda)}{d\boldsymbol{\beta}} = \frac{\partial G(\boldsymbol{\beta}, \hat{\mathbf{b}}(\boldsymbol{\beta}, \lambda)|\lambda)}{\partial \boldsymbol{\beta}} + \frac{\partial G(\boldsymbol{\beta}, \hat{\mathbf{b}}(\boldsymbol{\beta}, \lambda)|\lambda)}{\partial \hat{\mathbf{b}}} \cdot \frac{\partial \hat{\mathbf{b}}(\boldsymbol{\beta}, \lambda)}{\partial \boldsymbol{\beta}}.$$

The formula of $\frac{dG(\boldsymbol{\beta}, \hat{\mathbf{b}}(\boldsymbol{\beta}, \lambda)|\lambda)}{d\boldsymbol{\beta}}$ involves the term $\frac{\partial \hat{\mathbf{b}}(\boldsymbol{\beta}, \lambda)}{\partial \boldsymbol{\beta}}$. If the inner optimization leads to an explicit solution for $\hat{\mathbf{b}}(\boldsymbol{\beta}, \lambda)$, the gradient is readily available. But if not, the Implicit Function Theorem can be applied to find $\frac{\partial \hat{\mathbf{b}}(\boldsymbol{\beta}, \lambda)}{\partial \boldsymbol{\beta}}$.

Since the optimal parameter $\hat{\mathbf{b}}$ satisfying

$$\frac{\partial H(\mathbf{b}|\boldsymbol{\beta}, \lambda)}{\partial \mathbf{b}} = \mathbf{0}, \tag{3.16}$$

and $\hat{\mathbf{b}}$ is a function of $\boldsymbol{\beta}$, we can take the derivative with respect to $\boldsymbol{\beta}$ on Equation

3.16 about $\hat{\mathbf{b}}$ as follows:

$$\frac{d}{d\boldsymbol{\beta}}\left(\frac{\partial H(\mathbf{b}|\boldsymbol{\beta},\lambda)}{\partial \mathbf{b}}\bigg|_{\hat{\mathbf{b}}}\right) = \frac{\partial^2 H(\mathbf{b}|\boldsymbol{\beta},\lambda)}{\partial \mathbf{b}\partial\boldsymbol{\beta}}\bigg|_{\hat{\mathbf{b}}} + \frac{\partial^2 H(\mathbf{b}|\boldsymbol{\beta},\lambda)}{\partial \mathbf{b}\partial\mathbf{b}^T}\bigg|_{\hat{\mathbf{b}}} \cdot \frac{\partial \hat{\mathbf{b}}(\boldsymbol{\beta}|\lambda)}{\partial\boldsymbol{\beta}} = 0. \quad (3.17)$$

which holds since $\frac{\partial H(\mathbf{b}|\boldsymbol{\beta},\lambda)}{\partial \mathbf{b}}\Big|_{\hat{\mathbf{b}}}$ is a function of $\boldsymbol{\beta}$ that is identically $\mathbf{0}$. Assuming that $\left|\frac{\partial^2 H(\mathbf{b}|\boldsymbol{\beta},\lambda)}{\partial \mathbf{b}\partial\mathbf{b}^T}\Big|_{\hat{\mathbf{b}}}\right| \neq 0$, we obtain the following equation from the Implicit Function Theorem

$$\frac{\partial \hat{\mathbf{b}}(\boldsymbol{\beta}|\lambda)}{\partial\boldsymbol{\beta}} = -\left[\frac{\partial^2 H(\mathbf{b}|\boldsymbol{\beta},\lambda)}{\partial \mathbf{b}\partial\mathbf{b}^T}\bigg|_{\hat{\mathbf{b}}}\right]^{-1}\left[\frac{\partial^2 H(\mathbf{b}|\boldsymbol{\beta},\lambda)}{\partial \mathbf{b}\partial\boldsymbol{\beta}}\bigg|_{\hat{\mathbf{b}}}\right]. \quad (3.18)$$

Details of the gradient and Hessian matrices in each level are provided in the Appendix.

3.4.4 Criterion when structural parameters are $\boldsymbol{\beta}, \Delta$, and σ^2

This criterion is based on a first-order Taylor expansion of the nonlinear function $\mathbf{x}_i$ around the current value of $\boldsymbol{\beta}$ and the optimized $\mathbf{b}$ by the criterion $H(\mathbf{b}|\boldsymbol{\theta},\lambda)$.

We define

$$\Sigma(\Delta) = I + \frac{\partial \mathbf{x}_i}{\partial \mathbf{b}_i^T}\bigg|_{\hat{\mathbf{b}}} \Delta^{-1}\Delta^{-T} \frac{\partial \mathbf{x}_i}{\partial \mathbf{b}_i^T}\bigg|_{\hat{\mathbf{b}}}^T, \quad (3.19)$$

and the optimization criterion is a function of $\boldsymbol{\beta}$, Δ, and σ^2 conditional on the smoothing parameter λ, as shown in Equation 3.20.

$$G_{LME}(\boldsymbol{\theta}|\lambda) = -\frac{\sum_i^N n_i}{2}\log 2\pi\sigma^2 - \frac{1}{2}\sum_{i=1}^{N}\{\log(|\Sigma(\Delta)|) + \left[\mathbf{y}_i - \mathbf{x}_i + \frac{\partial \mathbf{x}_i}{\partial \mathbf{b}_i^T}\bigg|_{\hat{\mathbf{b}}}\hat{\mathbf{b}}_i\right]^T (\Sigma(\Delta))^{-1}\left[\mathbf{y}_i - \mathbf{x}_i + \frac{\partial \mathbf{x}_i}{\partial \mathbf{b}_i^T}\bigg|_{\hat{\mathbf{b}}}\hat{\mathbf{b}}_i\right]\Bigg\}. \quad (3.20)$$

We make the assumption that the variance-covariance matrix Ψ of random effects is positive-definite, which ensures that a Δ will exist. In the optimization process, we need to parameterize Δ to guarantee that it is positive-definite. Let S be a positive-

definite, symmetric matrix; it can be expressed as the matrix exponential of another symmetric matrix D as follows

$$S = e^D = I + D + \frac{D^2}{2!} + \frac{D^3}{3!} + \cdots \tag{3.21}$$

If S is called the matrix exponential of D, D is the matrix logarithm of S. One way of evaluating the matrix logarithm D is to calculate an eigenvalue-eigenvector decomposition. There is a nonsingular matrix of eigenvectors, U, and a diagonal matrix of eigenvalues, $\Lambda = \text{diag}(\lambda_1, \cdots, \lambda_q)$, such that

$$S = U\Lambda U^{-1}. \tag{3.22}$$

If S is positive-definite, then all the diagonal elements of Λ must be positive. The matrix logarithm of Λ is the diagonal matrix in which diagonal elements are the logarithms of the corresponding elements of Λ. We denote this by $\log \Lambda$ as follows

$$D = \log S = U \log \Lambda U^{-1}. \tag{3.23}$$

In this thesis, we only consider the simple situation where Ψ is a diagonal matrix, and therefore Δ is also a diagonal matrix. We parameterize Δ in terms of $\boldsymbol{\delta} = [\delta_1, \cdots, \delta_q]^T$, which are the logarithms of the diagonal values of the matrix Δ, shown in Equation 3.24. In this way, $\boldsymbol{\delta}$ become unconstrained parameters during the optimizations.

$$\Delta = \text{diag}(e^{\delta_1}, \cdots, e^{\delta_q}). \tag{3.24}$$

3.4.5 Criterion for the smoothing parameter λ

It is very important to choose an appropriate value for the smoothing parameter. When the smoothing parameter is too small, the fitted curves tend to be rough. On the other hand, the fitted curve is far from observations if the smoothing parameter is too large, since there is too much weight on the roughness penalty and the fitted curve

is forced to be very smooth. The generalized cross-validation (GCV) is a widely-used criterion to find the optimal value of the smoothing parameter, which minimizes mean square errors (MSE) between fitted curves and the data.

Let m be the total number of observations for all individuals, and define the degree of freedom as follows

$$\text{dfe}(\lambda) \quad = \quad m - \text{trace}\left(\mathbf{\Phi}(\mathbf{\Phi}^T\mathbf{\Phi} + \lambda\mathbf{R})^{-1}\mathbf{\Phi}^T\mathbf{W}\right).$$

The fourth-level optimization criterion GCV is

$$F(\lambda) = \text{GCV}(\lambda) = \left[\frac{m}{m - \text{dfe}(\lambda)}\right]\left[\frac{\text{SSE}}{m - \text{dfe}(\lambda)}\right],$$

where

$$\text{SSE}(\lambda) \quad = \quad \mathbf{y}^T[I - A(\lambda)]^T\mathbf{W}[I - A(\lambda)]\mathbf{y}.$$

3.5 Inference and predictions

With the optimization criterion $G_{LME}(\boldsymbol{\theta}|\lambda)$, the distribution of the maximum likelihood estimators $\hat{\boldsymbol{\beta}}$ of the fixed effects can be expressed in Equation (3.25) (Pinheiro and Bates 2000).

$$\boldsymbol{\beta} \sim \text{Normal}\left(\boldsymbol{\beta}, \sigma^2\left[\sum_{i=1}^{N}\left.\frac{\partial \mathbf{x}_i}{\partial \boldsymbol{\beta}^T}\right|_{\hat{\boldsymbol{\beta}}}^{T}\Sigma_i(\Delta)^{-1}\left.\frac{\partial \mathbf{x}_i}{\partial \boldsymbol{\beta}^T}\right|_{\hat{\boldsymbol{\beta}}}\right]^{-1}\right). \tag{3.25}$$

Through the optimization, the estimates of the coefficients for the basis functions are obtained as follows

$$\hat{\mathbf{c}}(\hat{\boldsymbol{\beta}}, \hat{\mathbf{b}}, \hat{\lambda}) = [\mathbf{\Phi}^T\mathbf{W}\mathbf{\Phi} + \lambda\mathbf{R}(\hat{\boldsymbol{\beta}}, \hat{\mathbf{b}}, \hat{\lambda})]^{-1}\mathbf{\Phi}^T\mathbf{W}\mathbf{y}. \tag{3.26}$$

Therefore, the predicted value at time t is

$$\hat{\mathbf{x}}(t) = \boldsymbol{\phi}(t)^T \hat{\mathbf{c}}.$$

Numerically solving ODE's relies on initial values, which are the values of ODE components at the first time point, usually time $t = 0$. A small change in initial values results in a large difference in the numerical ODE solutions. However, observations in real life, including the observed initial values, usually have some measurement error, and it is dangerous to use the first observations as the initial values directly. Moreover, some components in ODE's are not observable, in which case there is no way to observe the initial values for these components.

The byproduct of the generalized profiling method is that the fitted curves for all components can be obtained using the ODE parameter estimates. The estimates for initial values are obtained by evaluating the fitted curves for all components at the first time point. The initial values $\mathbf{x}(0)$ can be estimated as

$$\hat{\mathbf{x}}(0) = \boldsymbol{\phi}(0)^T \hat{\mathbf{c}}.$$

Chapter 4

Simulations

This chapter aims to validate the proposed nonlinear mixed-effects models by the generalized profiling method with simulated data sets for three compartment pharmacokinetic models. These models are introduced briefly together with their structural parameters. In the following, we estimate random-effects parameters $\mathbf{b}$, fixed-effects parameters $\boldsymbol{\beta}$ and the smoothing parameter λ with a fixed relative precision factor Δ and residual variance σ^2 by a four-level optimization for the one-compartment model with first-order absorption and elimination. Then we estimate $\mathbf{b}$, $\boldsymbol{\beta}$, Δ, and σ^2 by a three-level optimization for three compartment models with a fixed λ. Finally, this chapter gives some discussions on several practical issues, such as derived parameters, identifiability, and starting values.

4.1 Compartment pharmacokinetic models

Pharmacokinetics is an important application field of the nonlinear mixed-effects models expressed by ODE's. This section considers three pharmacokinetic models used in the following sections, including a one-compartment elimination model, a one-compartment model with first-order absorption and elimination, and a two-compartment open model.

The simplest compartment model is a *one-compartment elimination model* with the first-order elimination rate k_e. For injected drugs, the process of absorption tends

to be very short and neglectful, and all the tissues are perfused rapidly by the drug. Therefore, the whole body can be represented by a single compartment, the central compartment. Let $C(t)$ be the drug concentration in this central compartment at time t after drug administration. This model can be described by the differential equation as follows

$$\frac{dC(t)}{dt} = -k_e C(t). \tag{4.1}$$

With an initial value of the drug concentration $C(0)$, the solution to this differential equation is

$$C(t) = C(0) \cdot \exp(-k_e t). \tag{4.2}$$

A *one-compartment model with first-order absorption and elimination* can be used to represent more complicated pharmacokinetic processes. For orally-administered drugs, the process of absorption might take a long time to complete. It is reasonable to consider the absorption rate k_a besides the elimination rate k_e. Let $[C_a(t), C(t)]^T$ be a vector representing the drug concentration at the absorption site and in the blood at time t. The dose is known and denoted by D. This thesis assumes that the bioavailability of the drug, B, is always equal to 1. The volume of distribution V represents the apparent instantaneous dilution space of an instantaneously absorbed dose. The one-compartment model with the first-order absorption rate k_a and the first-order elimination rate k_e can be expressed by two differential equations as follows.

$$\begin{aligned} \frac{\mathrm{d}C_a(t)}{\mathrm{d}t} &= -k_a C_a(t), \qquad C_a(0) = B\frac{D}{V}, \\ \frac{\mathrm{d}C(t)}{\mathrm{d}t} &= k_a C_a(t) - k_e C(t). \end{aligned} \tag{4.3}$$

These two differential equations in Equation (4.3) can be solved analytically, as shown in Equation (4.4). In a general case, numerical methods are often needed to

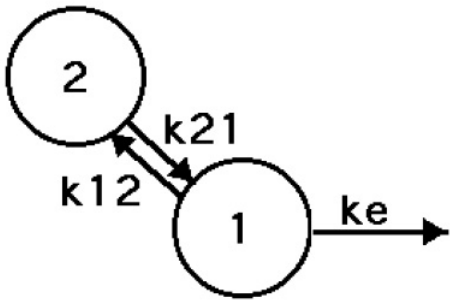

Figure 4.1: A diagram of a two-compartment model. Compartment 1 is the central compartment; compartment 2 is the tissue compartment.

obtain solutions of ODE's when differential equations cannot be solved analytically.

$$\begin{aligned} C_a(t) &= C_a(0)\exp(-k_a t) \\ C(t) &= C_a(0)\frac{k_a}{(k_a - k_e)}[\exp(-k_e t) - \exp(-k_a t)] + C(0)\exp(-k_e t). \end{aligned} \tag{4.4}$$

If drugs exhibit a slow equilibration with peripheral tissues, a *two-compartment model* can be used to describe the distribution of drugs from the central compartment to the tissue compartment. Figure 4.1 shows a diagram of a two-compartment model expressed by ODE's in Equation (4.5); Compartment 1 represents the central compartment with the drug concentration $C(t)$, and Compartment 2 symbolizes the peripheral compartment with the drug concentration $C_2(t)$. The process of distribution is represented by the rate constants k_{12} and k_{21}; the process of elimination is represented by the rate constant k_e. For this model, solutions are obtained using numerical methods in our simulations.

$$\begin{aligned} \frac{dC(t)}{dt} &= -(k_e + k_{12})C(t) + k_{21}C_2(t), \\ \frac{dC_2(t)}{dt} &= k_{12}C(t) - k_{21}C_2(t),\ C_2(0) = 0. \end{aligned} \tag{4.5}$$

A property of the compartment pharmacokinetic models is that the absorption

rate and the elimination rate must be positive, and concentrations must not be negative. Using logarithms of parameters in the model are a common and effective way to ensure positive values while keeping the optimization problem unconstrained. This thesis parameterizes models in terms of the logarithms of rate constants. As an example of this parameter transform, we consider the transfer matrix for the i-th individual in the one-compartment model with first-order absorption and elimination as follows

$$M_i = \begin{bmatrix} -\exp(\ln(k_a) + b_{i1}) & 0 \\ \exp(\ln(k_a) + b_{i1}) & -\exp(\ln(k_e) + b_{i2}) \end{bmatrix},$$

where random effects $\mathbf{b}_i = [b_{i1}, b_{i2}] \sim \text{Normal}(0, \Psi)$.

In addition, to ensure a positive-definite matrix, we parameterize the relative precision factor Δ, a diagonal matrix in this thesis, as the logarithms of its diagonal values, denoted by $\boldsymbol{\delta}$ or $[\delta_1, \cdots, \delta_q]^T$. For the same reason, we parameterize the residual variance σ^2 as its logarithm. Particularly, the parameters are $[\ln(k_e),\ \delta,\ \ln(\sigma^2)]^T$ in the one-compartment elimination model, $[\ln(k_a),\ \ln(k_e),\ \delta_1,\ \delta_2,\ \ln(\sigma^2)]^T$ in the one-compartment model with first-order absorption and elimination, and $[\ln(k_e),\ \ln(k_{12}),\ \ln(k_{21}),\ \delta_1,\ \delta_2,\ \delta_3,\ \ln(\sigma^2)]^T$ in the two-compartment open model.

4.2 Estimating mixed effects and the smoothing parameter with a fixed relative precision factor

This section investigates the estimates of coefficients of basis functions $\mathbf{c}$, random effects $\mathbf{b}$, fixed effects $\boldsymbol{\beta}$, and the smoothing parameter λ through a four-level optimization with a fixed relative precision factor Δ for the one-compartment model with first-order absorption and elimination. In the inner level, nuisance parameters $\mathbf{c}$ are optimized by penalized smoothing defined by ODE's; in the middle level, random effects $\mathbf{b}$ are optimized by a penalized nonlinear least squares; in the outer level, fixed effects $\boldsymbol{\beta}$ are optimized by the sum of squared errors; in the outermost level, the smoothing parameter λ is optimized by the criterion of generalized cross-validation

(GCV).

We assume that the data sets are generated in the following pharmacokinetical scenario. An oral drug is given to 16 healthy individuals with the same dose $D = 10$. After drug administration, the drug concentrations $C(t)$ in the blood are measured at 11 time points, 0, 0.25, 0.5, 0.75, 1, 2, 3, 4, 5, 6, 7, 9, and 12 hours. Although we can simulate the drug concentration $C_a(t)$ at the absorption site in simulations, it is impossible to measure them in real data sets.

For the population, we assume that the absorption rate k_a and the elimination rate k_e are 1.65 and 0.25, respectively. In other words, the fixed-effects parameters $\boldsymbol{\beta}$, logarithms of k_a and k_e, are $[0.5, -1.4]^T$. For all individuals, the initial values of the drug concentrations in the blood, $C(0)$, are assumed to be 0, and the initial values for the drug concentrations at the absorption site, $C_a(0)$, are supposed to be 10.

For each simulation, we assume that 16 individuals are drawn from the population randomly. The random-effects parameters $\mathbf{b}_i,\ i = 1, \cdots, 16$, are assumed to follow a normal distribution with mean 0 and variance-covariance Ψ. For simplicity, Ψ is assumed to be a diagonal matrix, with diagonal values $[0.04,\ 0.01]^T$ in this section. Data are supposed to be subject to measurement errors, which is simulated by adding noise from a normal distribution with mean 0 and variance 0.04.

Under these assumptions on the parameters and initial values, solutions to the differential equations in Equation (4.3) are given by red solid lines in Figure 4.2. As can be seen, the drug concentration at the absorption site, $C_a(t)$, decreases from 10 to 0 very quickly, and all drugs are absorbed after about 4 hours after drug administration. Meanwhile, the drug concentration in the blood, $C(t)$, is going up rapidly at the beginning of absorption, and reaches the maximum point within 2 hours. After that, it goes down slowly to a small value in 12 hours after drug administration. Figure 4.2 also illustrates one typical simulated data set based on the above setting of parameters. The blue circles are the simulated data points for each individual at particular time points. A connected blue curve represents all the observations for an individual. As can be seen, all the individuals share a similar shape represented by the population curve with certain variability.

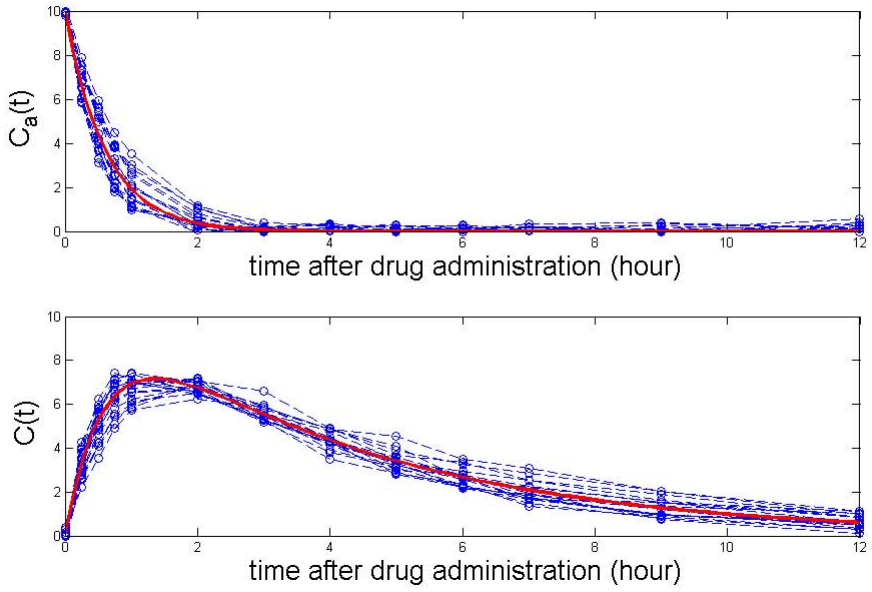

Figure 4.2: One typical simulated data for a one-compartment model with first-order absorption and elimination with $k_a = 1.65$, $k_e = 0.25$, $\sigma^2 = 0.04, \Psi = \text{diag}([0.04,\ 0.01])$.

It is important to select an appropriate smoothing parameter to obtain good estimates. We begin with investigating the behavior of GCV to validate the estimate of λ. For the typical simulated data set shown in Figure 4.2, Figure 4.3 shows values of GCV under different values of the smoothing parameter λ. These values of GCV are evaluated when λ is $2^r, r = 1, \cdots, 13$, with the true values of structural parameters. The left panel displays GCV versus $\log(\lambda)$ with both of the two components of data, the drug concentration at the absorption site $C_a(t)$ and the drug concentration in the blood $C(t)$. GCV is minimized when the smoothing parameter λ is around 512 with a minimum value 0.0427. In contrast, the right panel displays GCV versus $\log(\lambda)$ with only one component $C(t)$. The minimum value 0.0460 is obtained when λ is around 512.

For both of the two plots of GCV, there are relative large intervals for λ having similar values of GCV. This feature of flatting GCV might cause problems when we use it as the outermost criterion to estimate λ through a four-level optimization because it is highly likely that the process of optimization cannot converge to a unique

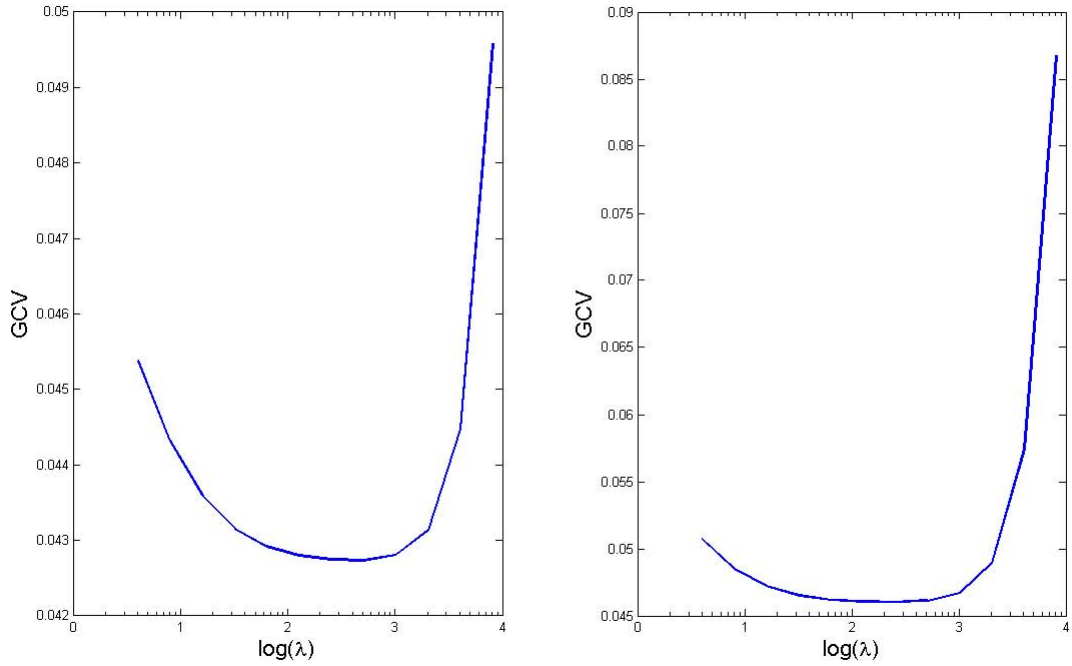

Figure 4.3: The left panel displays the generalized cross-validation (GCV) criterion versus $\log(\lambda)$ of the model with both of the two components; the right panel displays GCV versus $\log(\lambda)$ of the model with only the drug concentrations in the blood.

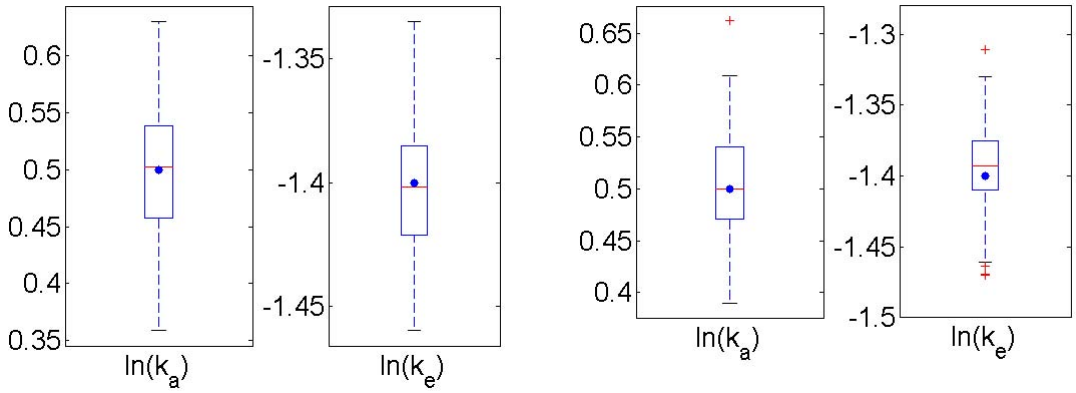

Figure 4.4: The Boxplots of the estimated $\ln(k_a)$ and $\ln(k_e)$ with a smoothing parameter chosen by GCV, where the dots represent the true values. The left two plots show the results with two components, and the right two plots show those with one component.

solution. To see this, we conduct the four-level optimization to estimate λ with the two components of data. The starting value for λ is 10; the random starting values for $\boldsymbol{\beta}$ are $[0.57, -1.36]^T$; random starting values are chosen from the standard normal distribution for the random-effects parameters $\mathbf{b}$. As a result, the estimate for the smoothing parameter, $\hat{\lambda}$, is 213. For comparison, we use the same data set to estimate λ again with a different set of starting values. The starting value for λ is set to be 5000; the random starting values of $\boldsymbol{\beta}$ are $[0.61, -1.33]^T$; starting values of $\mathbf{b}$ are selected randomly from a normal distribution. The new estimate $\hat{\lambda}$ is 1629, which is different from the previous estimate.

Based on the above results and discussion, it is not efficient to estimate the smoothing parameter λ by GCV in a four-level optimization. There is often a large interval instead of a unique point for the optimized λ based on GCV. Moreover, four levels of optimization are much more time-consuming than a three-level optimization because cascaded relationships among the parameters will become more complicated with the increase of levels of optimization. Each time of updating λ leads to a three-level optimization. Therefore, we choose a smoothing parameter either from the range from 100 to 10000 arbitrarily, or by minimizing values of GCV for a few numbers of λ. In this section, the smoothing parameter λ is chosen by minimizing GCV when λ is 2^r, $r = 1, \cdots, 13$.

The left two plots in Figure 4.4 show the boxplots for the estimated fixed-effects parameters $\boldsymbol{\beta}$ from 100 simulations with the two components of data. The medians of estimates for $\ln(k_a)$ and $\ln(k_e)$ are 0.50 and -1.40, respectively, which are almost identical to the true values. The estimates tend to be unbiased with only small biases, 0.0025 and 0.0018, respectively. The experimental 95% confidence intervals (CI's) are $(0.39, 0.60)$ for $\ln(k_a)$, and $(-1.45, -1.35)$ for $\ln(k_e)$. The standard deviations (SD's) are 0.054 and 0.027 for $\ln(k_a)$ and $\ln(k_e)$, respectively. Estimates also seem very stable based on the fact that the confidence intervals and SD's are relatively small.

In contrast, the right two plots in Figure 4.4 show the boxplots for the estimated fixed-effects parameters $\ln(k_a)$ and $\ln(k_e)$ with only one component of data, $C(t)$. The medians of estimates for $\ln(k_a)$ and $\ln(k_e)$ are 0.50 and -1.39, respectively, which

are also very close to the true values with small biases 0.0005 and 0.0071. The experimental 95% confidence intervals are $(0.41, 0.60)$ for $\ln(k_a)$, and $(-1.46, -1.34)$ for $\ln(k_e)$. The standard deviations over these 100 simulations are 0.053 and 0.030 for $\ln(k_a)$ and $\ln(k_e)$, respectively. Therefore, we can obtain similar results by using only one component of data.

4.3 Estimating random effects and all the structural parameters with a fixed smoothing parameter

This section tests our method of estimating random effects $\mathbf{b}$, fixed effects $\boldsymbol{\beta}$, the logarithms of the diagonal values of the relative precision factor, $\boldsymbol{\delta}$, and the logarithm of residual variance, $\ln(\sigma^2)$, with a fixed smoothing parameter λ through the three-level optimization. In the inner level, nuisance parameters $\mathbf{c}$ are optimized by penalized smoothing defined by ODE's; in the middle level, random effects $\mathbf{b}$ are optimized by a penalized nonlinear least squares; in the outer level, the structural parameters $\boldsymbol{\theta}$ are optimized by a criterion based on a first-order Taylor expansion of the nonlinear function. Based on the discussion in the previous section, we won't use a four-level optimization to choose λ. Instead, λ is chosen arbitrarily in the range from 100 to 10000. We consider simulations for the one-compartment elimination model, the one-compartment model with absorption and elimination, and the two-compartment open model.

4.3.1 The one-compartment elimination model

For the one-compartment elimination model, we estimate structural parameters $[\ln(k_e), \delta, \ln(\sigma^2)]^T$ with a fixed λ using simulations under three different settings in Table 4.1. This table shows the number of subjects, N; the initial values of drug concentration at absorption site, $C_a(0)$; structural parameters in the original scale, including k_e, Ψ, and σ^2; parameters in the process of optimization, including $\ln(k_e)$, δ, and $\ln(\sigma^2)$.

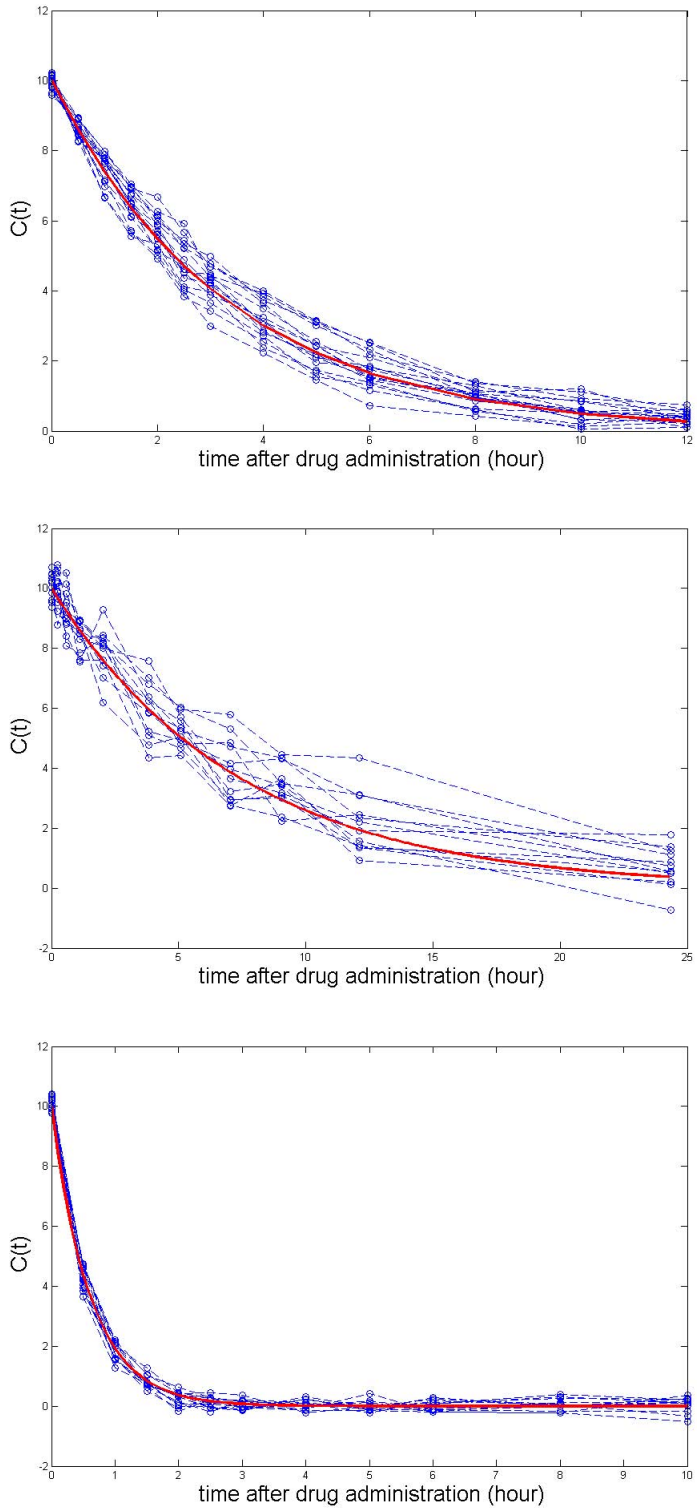

Figure 4.5: Typical simulated data sets of the one-compartment elimination model for Setting 1, Setting 2, and Setting 3 in Table 4.1 (from top to bottom).

For simplicity, we assume that drug concentrations are measured at the same series of time points for all subjects. However, the generalized profiling method can also consider measurements at different time points for different individuals, which is often the case for the real data sets.

Figure 4.5 illustrates graphically one typical simulated data set for each setting in Table 4.1. The red solid lines represent solutions of the differential equations with the assumed parameters, and the connected blue circles represent observations for an individual at a series of time points. As can be seen, all the individuals share a similar concentration-time curve with certain variability. Particularly, Setting 3 has the highest elimination rate and smallest variance for random effects. Setting 1 and Setting 2 have slower elimination rates, and they have the same variance for the random effects, while the residual variance of Setting 2 is much higher.

Table 4.1: Simulation settings for the one-compartment elimination model

Parameter	Setting 1	Setting 2	Setting 3
N	16	12	12
$C_a(0)$	10	10	10
k_e	0.3	0.14	1.65
Ψ	0.04	0.04	0.01
σ^2	0.04	0.36	0.04
$\ln(k_e)$	-1.2	-2	0.5
δ	0	1.10	0.69
$\ln(\sigma^2)$	-3.22	-1.02	-3.22
Time points			
Setting 1 (13)	0, 0.5, 1, 1.5, 2, 2.5, 3, 4, 5, 6 , 8, 10, 12		
Setting 2 (11)	0, 0.25, 0.57, 1.12, 2.02, 3.82, 5.1, 7.03, 9.05, 12.12, 24.37		
Setting 3 (12)	0, 0.5, 1, 1.5, 2, 2.5, 3, 4, 5, 6, 8, 10		

For each setting in Table 4.1, 100 simulated data sets are generated, and parameters are estimated with a fixed smoothing parameter $\lambda = 1000$. Table 4.2 gives true values, median estimates, experimental 95% confidence intervals (CI's), standard deviations (SD's), and estimated biases under the three settings of the one-compartment elimination model in Table 4.1; Figure 4.6 shows the boxplots of all the parameter estimates. Compared with the true values marked by blue dots, the median estimates

for the fixed-effects parameter $\ln(k_e)$ represented by the red horizontal lines are very accurate under the three settings. Moreover, with short confidence intervals and small standard deviations, estimates for $\ln(k_e)$ seem very stable. Estimates of δ are almost unbiased but with relatively larger intervals. The logarithms of the residual variance, $\ln(\sigma^2)$, are underestimated with short confidence intervals and small standard deviations.

Table 4.2: Parameter estimates for the one-compartment elimination model with $\lambda = 1000$

Setting 1					
Parameter	True	Estimates	95% CI	SD	Bias
$\ln(k_e)$	-1.2	-1.20	(-1.29, -1.10)	0.045	0.0047
δ	0	0	(-0.27, 0.49)	0.19	0.0087
$\ln(\sigma^2)$	-3.22	-3.30	(-3.55, -3.09)	0.12	0.080
Setting 2					
Parameter	True	Estimates	95% CI	SD	Bias
$\ln(k_e)$	-2	-2.00	(-2.11, -1.86)	0.060	0.0028
δ	1.10	1.12	(0.62, 1.88)	0.34	0.0086
$\ln(\sigma^2)$	-1.02	-1.13	(-1.45, -0.91)	0.14	0.11
Setting 3					
Parameter	True	Estimates	95% CI	SD	Bias
$\ln(k_e)$	0.5	0.50	(0.44, 0.56)	0.031	0.0011
δ	0.69	0.64	(0.29, 1.40)	0.30	0.046
$\ln(\sigma^2)$	-3.22	-3.33	(-3.58, -3.06)	0.14	0.11

4.3.2 The one-compartment model with first-order absorption and elimination

For the one-compartment model with first-order absorption and elimination, we test our method using simulations under three settings. Let $[\psi_1, \psi_2]^T$ be the diagonal values of the variance-covariance matrix Ψ for random effects, and let $[\delta_1, \delta_2]^T$ be the logarithms of the diagonal values of the relative precision factor Δ. Table 4.3 shows the number of subjects, N; the initial values of drug concentration at absorption site, $C_a(t)$; structural parameters in the original scale, including $k_a,\ k_e,\ \psi_1,\ \psi_2$, and σ^2, and parameters in the optimization, including $\ln(k_a),\ \ln(k_e),\ \delta_1,\ \delta_2$, and $\ln(\sigma^2)$.

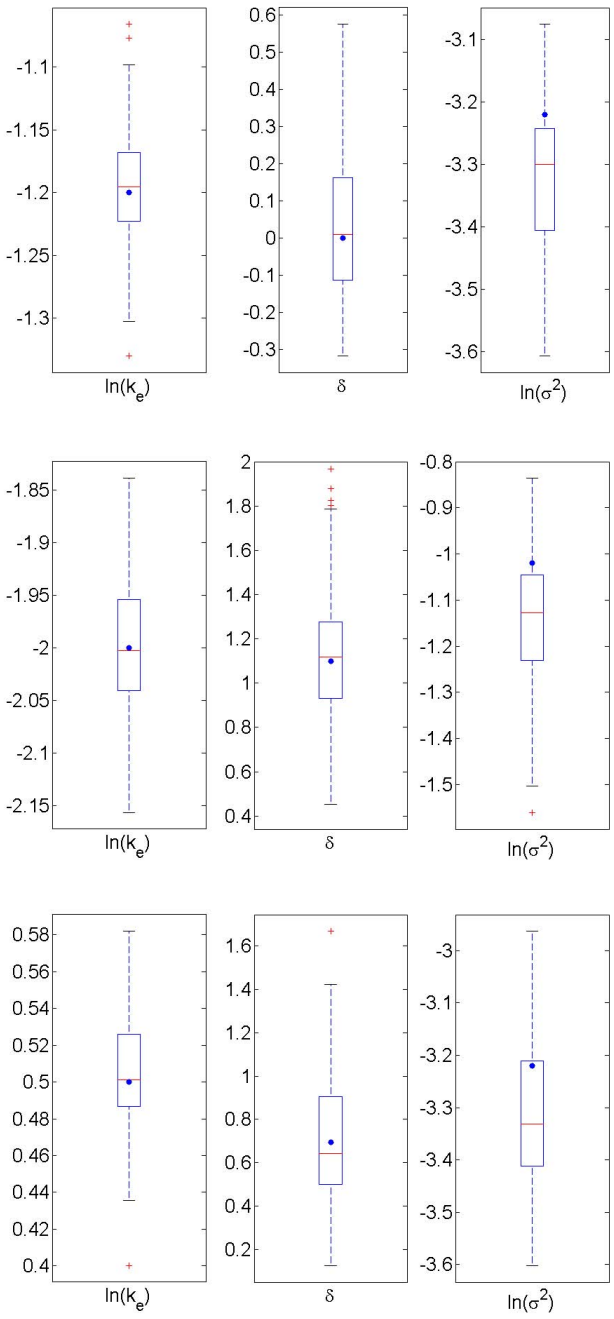

Figure 4.6: Boxplots of the estimated structural parameters for the one-compartment elimination model, where the dots represent the true values. From top to bottom: Setting 1, Setting 2, and Setting 3 in Table 4.1.

Parameters of Setting 1 are selected arbitrarily, while parameters of Setting 2 and Setting 3 are chosen to represent the characteristics of the real data sets in the next Chapter. For simplicity, we assume that drug concentrations are measured at the same series of time points for all individuals.

Table 4.3: Simulation settings for the one-compartment model with first-order absorption and the elimination

Parameter	Setting 1	Setting 2	Setting 3
N	12	12	16
$C_a(0)$	10	10	10
k_a	1.65	1.65	0.905
k_e	0.25	0.09	0.301
ψ_1	0.04	0.40	0.25
ψ_2	0.01	0.017	0.01
σ^2	0.04	0.46	0.04
$\ln(k_a)$	0.5	0.5	-0.1
$\ln(k_e)$	-1.4	-2.4	-1.2
δ_1	0	0.076	-0.92
δ_2	0.69	1.65	0.69
$\ln(\sigma^2)$	-3.22	-0.77	-3.22
Time points			
Setting 1 (12)	0, 0.5, 1, 1.5, 2, 2.5, 3, 4, 5, 6 , 8, 10.		
Setting 2 (11)	0, 0.25, 0.57, 1.12, 2.02, 3.82, 5.1, 7.03, 9.05, 12.12, 24.37.		
Setting 3 (13)	0, 0.5, 1.0, 1.5, 2.0, 2.5, 3, 4, 5, 6, 8, 10, 12.		

Figure 4.7 illustrates graphically one typical simulated data set for each setting in Table 4.3. The red solid lines represent solutions of the differential equations with the assumed parameters, and the connected blue circles represent observations for an individual at a series of time points. As can be seen, all the individuals share a similar concentration-time curve with certain variability. Setting 1 and Setting 2 have higher absorption rates than Setting 3. Setting 3 has the highest elimination rate while Setting 2 has the lowest value. In contrast to the absorption rates, all of the three settings have relative small variances for random effects of the elimination rates. Setting 2 owns the highest variances for random effects among the three settings while Setting 1 has the smallest values. Setting 3 has more individuals and more observations per subject than the other settings.

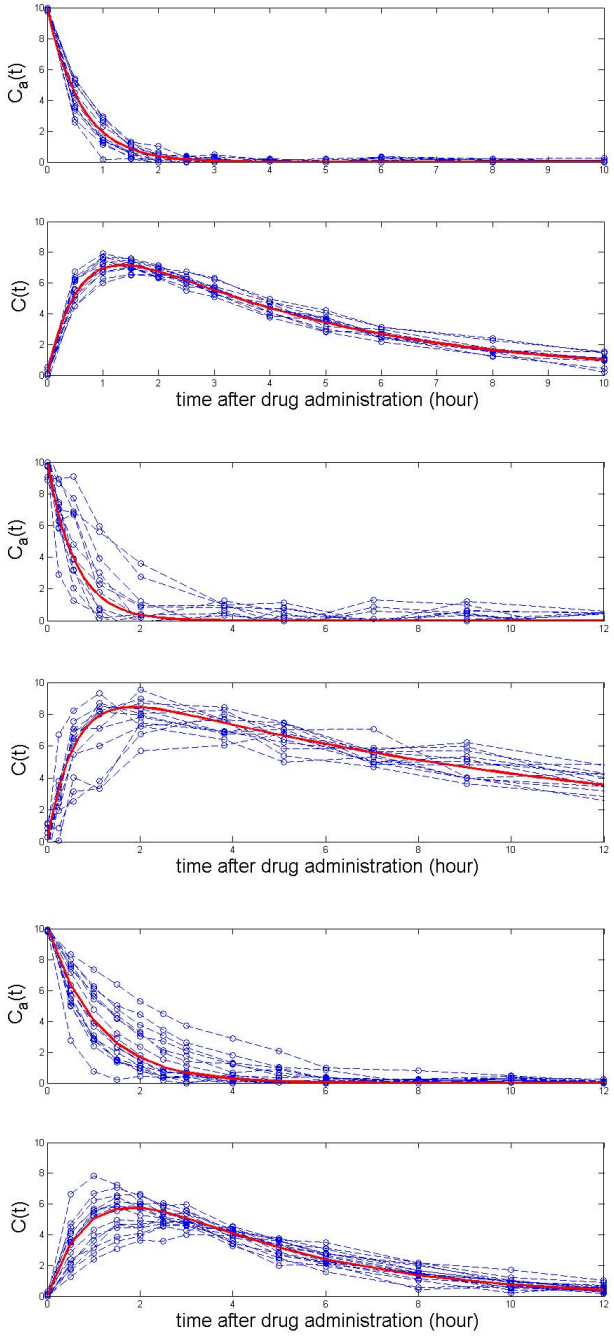

Figure 4.7: Typical simulated data sets of the one-compartment model with first-order absorption and elimination for Setting 1, Setting 2, and Setting 3 in Table 4.3 (from top to bottom).

Although we can use both of the two components of the simulated data to estimate parameters, it is more realistic to use only the drug concentration in the blood, $C(t)$, to test our method. For comparison, we estimate parameters under Setting 1 using two components and just one component, respectively. In this example, we estimate $\mathbf{b}$, $\boldsymbol{\delta}$, and $\ln(\sigma^2)$ using 100 simulated data sets with a fixed $\lambda = 100$. Table 4.4 shows true values, median estimates, experimental 95% confidence intervals, standard deviations, and estimated biases under Setting 1; Figure 4.8 illustrates the boxplots for the estimates. Except for the underestimated residual variance, both of the two sets of estimates are close to the true values. However, the estimates with two components of data are more accurate than those with only one component, and CI's are shorter and SD's are relatively smaller.

Table 4.4: Results of parameter estimates for setting 1 of the one-compartment model with first-order absorption and elimination with $\lambda = 100$

Two components					
Parameter	True	Estimate	95% CI	SD	Bias
$\ln(k_a)$	0.5	0.50	(0.40, 0.62)	0.058	0.0023
$\ln(k_e)$	-1.4	-1.40	(-1.47, -1.33)	0.033	0.0015
δ_1	0	-0.026	(-0.36, 0.48)	0.20	0.026
δ_2	0.69	0.70	(0.35, 1.33)	0.24	0.012
$\ln(\sigma^2)$	-3.22	-3.33	(-3.49, -3.20)	0.085	0.11
One component					
Parameter	True	Estimate	95% CI	SD	Bias
$\ln(k_a)$	0.5	0.51	(0.39, 0.62)	0.062	0.0097
$\ln(k_e)$	-1.4	-1.40	(-1.47, -1.35)	0.035	0.0036
δ_1	0	-0.041	(-0.46, 0.46)	0.24	0.041
δ_2	0.69	0.62	(0.14, 1.67)	0.42	0.066
$\ln(\sigma^2)$	-3.22	-3.46	(-3.71, -3.25)	0.12	0.24

To better validate estimates by the generalized profiling method with only one component of data, Table 4.5 summarizes results of 100 simulations for Setting 2 and Setting 3 in Table 4.3, and boxplots of these estimates are shown in Figure 4.9. Median estimates are close to the true values. The estimates of the elimination rate $\ln(k_e)$ are more accurate than those of the absorption rate $\ln(k_a)$. This is probably because the number of observations during the elimination process is larger than that

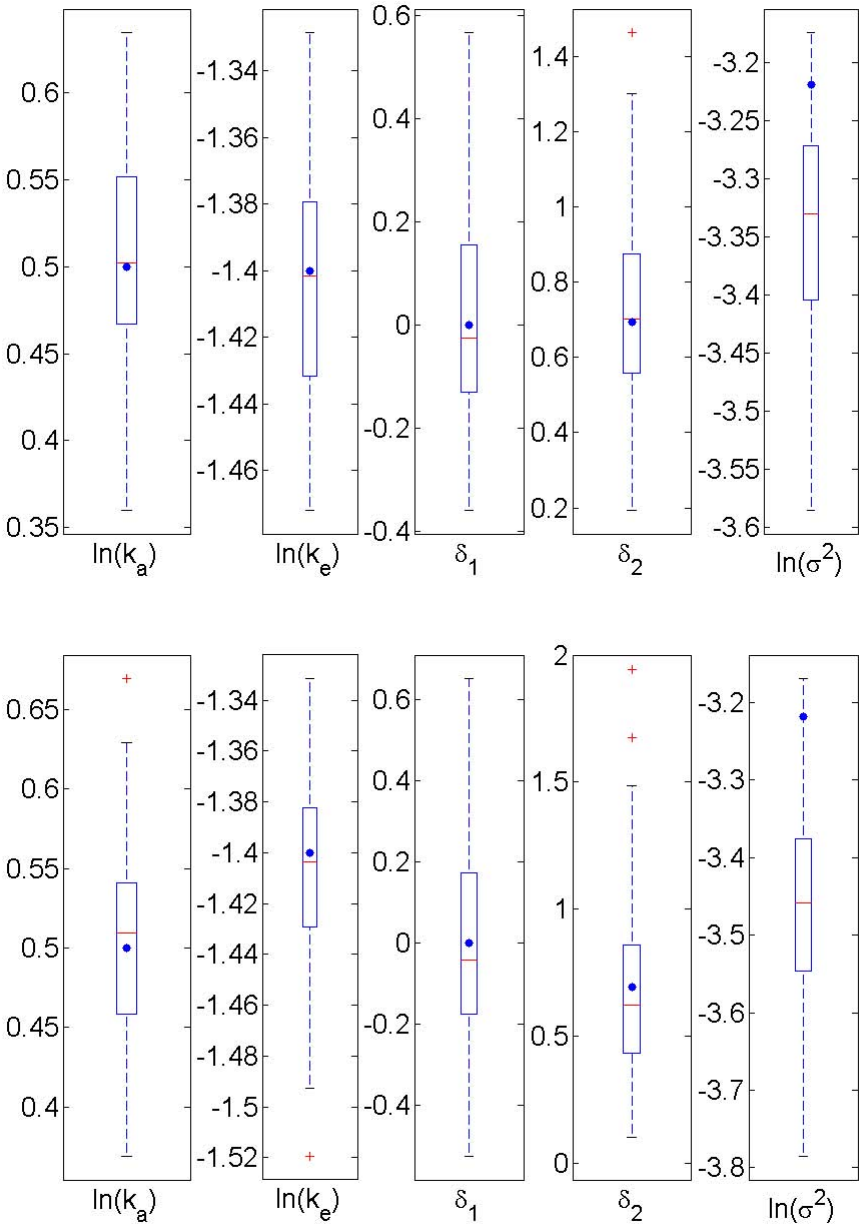

Figure 4.8: Boxplots of the estimated structural parameters for the one-compartment model with first-order absorption and elimination, where the dots represent the true values. The upper panel: with two components; the lower panel: with only one component.

during the absorption process. These results should be improved by choosing a more appropriate smoothing parameter λ or having more data points.

Table 4.5: Median estimates of parameters for Setting 2 and Setting 3 of the one-compartment model with first-order absorption and elimination with $\lambda = 1000$ using only one component of data

Setting 2					
Parameter	True	Estimate	95% CI	SD	Bias
$\ln(k_a)$	0.5	0.47	(0.14, 0.83)	0.19	0.028
$\ln(k_e)$	-2.4	-2.40	(-2.53, -1.25)	0.063	0.0038
δ_1	0.076	0.12	(-0.36, 0.63)	0.25	0.046
δ_2	1.65	1.52	(0.77, 7.69)	2.12	0.13
$\ln(\sigma^2)$	-0.77	-1.02	(-1.35, -0.70)	0.18	0.25
Setting 3					
Parameter	True	Estimate	95% CI	SD	Bias
$\ln(k_a)$	-0.1	-0.11	(-0.32, 0.13)	0.12	0.015
$\ln(k_e)$	-1.2	-1.19	(-1.24, -1.14)	0.024	0.012
δ_1	-0.92	-0.87	(-1.21, -0.54)	0.18	0.050
δ_2	0.69	0.66	(0.26, 1.23)	0.25	0.030
$\ln(\sigma^2)$	-3.22	-3.28	(-3.46, -3.17)	0.074	0.060

4.3.3 The two-compartment open model

We use one simulation setting to validate the performance of our method for the two-compartment open model. Let $[\delta_1, \delta_2, \delta_3]^T$ be logarithms of the diagonal values of the relative precision factor Δ. A total of 10 observations per subject are simulated for 12 individuals at time points, 0, 0.25, 0.57, 1.12, 2.02, 3.82, 5.1, 7.03, 9.05, and 12.12 hours. Initial values of drug concentration at absorption site $C_a(t)$ are assumed to be 8 for all individuals. Structural parameters in the optimization include $\ln(k_e)$, $\ln(k_{12})$, $\ln(k_{21})$, δ_1, δ_2, δ_3, and $\ln(\sigma^2)$.

With a fixed $\lambda = 1000$, Table 4.6 shows true values and median estimates, experimental 95% confidence intervals, standard deviations, and estimated biases of parameters from 100 simulations for the two-compartment open model. One set of typical simulated data is shown in Figure 4.10. Figure 4.11 illustrates the boxplots for these estimates. Only one component, $C(t)$, is used in estimating parameters. As

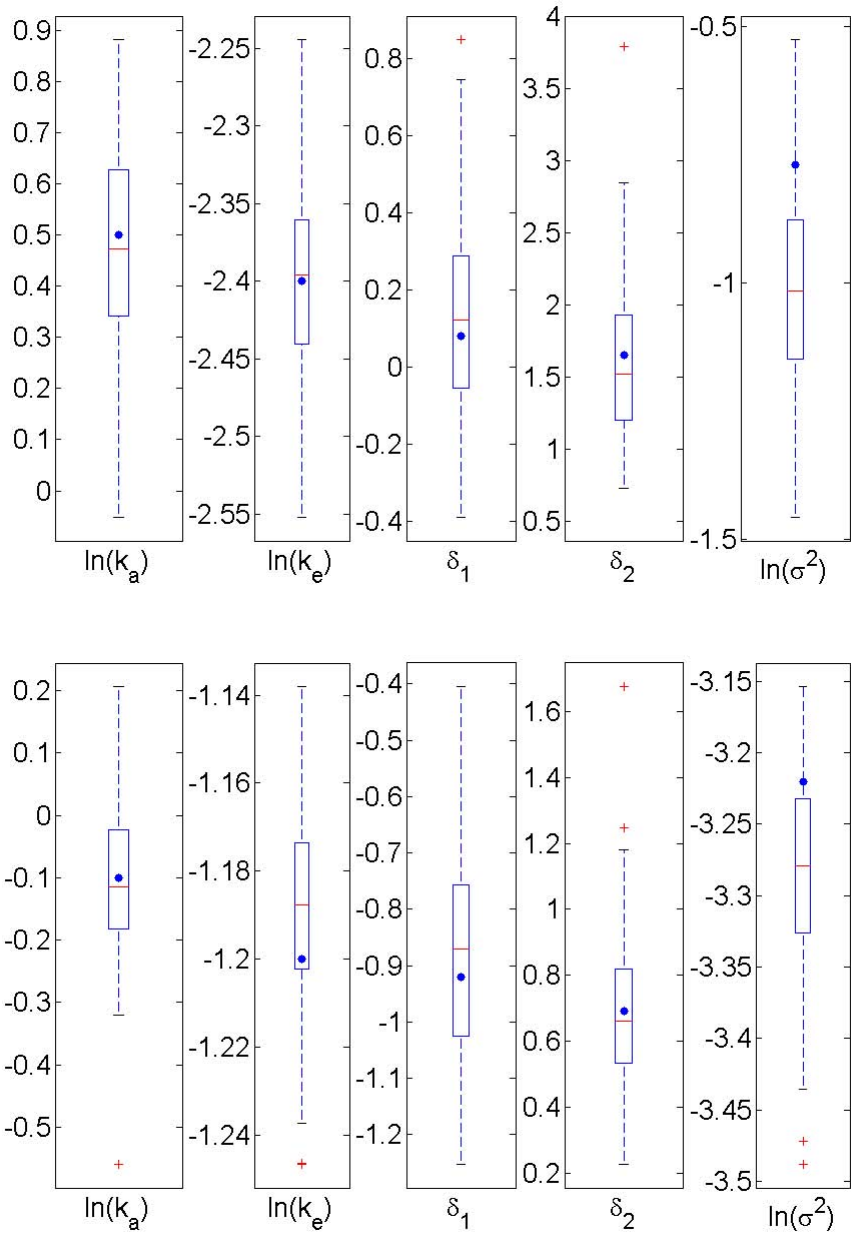

Figure 4.9: Boxplots of the estimated structural parameters for the one-compartment model with first-order absorption and elimination, where the dots represent the true values. The upper panel: Setting 2; the lower panel: Setting 3.

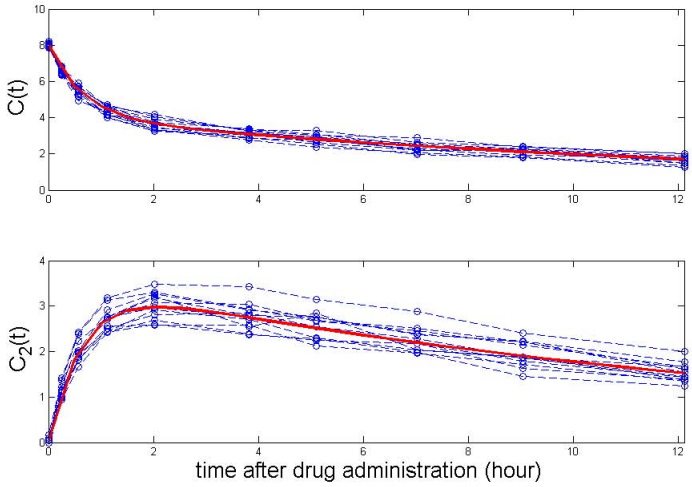

Figure 4.10: One typical simulated data sets of the two-compartment open model.

can be seen, median estimates for the fixed effects, $[\ln(k_e),\ \ln(k_{12}),\ \ln(k_{21})]^T$, are close to the true values and relatively stable. Median estimates for the logarithms of the diagonal values of the relative precision factor, $[\delta_1,\ \delta_2,\ \delta_3]^T$, are also close to the true values, but they are quite instable with large experimental confidence intervals and standard deviations. The residual variance is underestimated, but the estimate is stable.

Table 4.6: Parameter estimates for the two-compartment open model with $\lambda = 1000$

Parameter	True	Estimate	95% CI	SD	Bias
$\ln(k_e)$	-2	-1.99	(-2.17, -1.65)	0.11	0.011
$\ln(k_{12})$	-0.4	-0.40	(-0.59, -0.15)	0.11	0.0007
$\ln(k_{21})$	-0.2	-0.21	(-0.56, -0.042)	0.12	0.0076
δ_1	0	-0.011	(-0.66, 4.40)	1.31	0.011
δ_2	0	-0.073	(-0.68, 5.038)	1.59	0.073
δ_3	0	-0.24	(-1.01, 4.48)	1.66	0.24
$\ln(\sigma^2)$	-4.61	-4.86	(-5.18, -4.56)	0.15	0.25

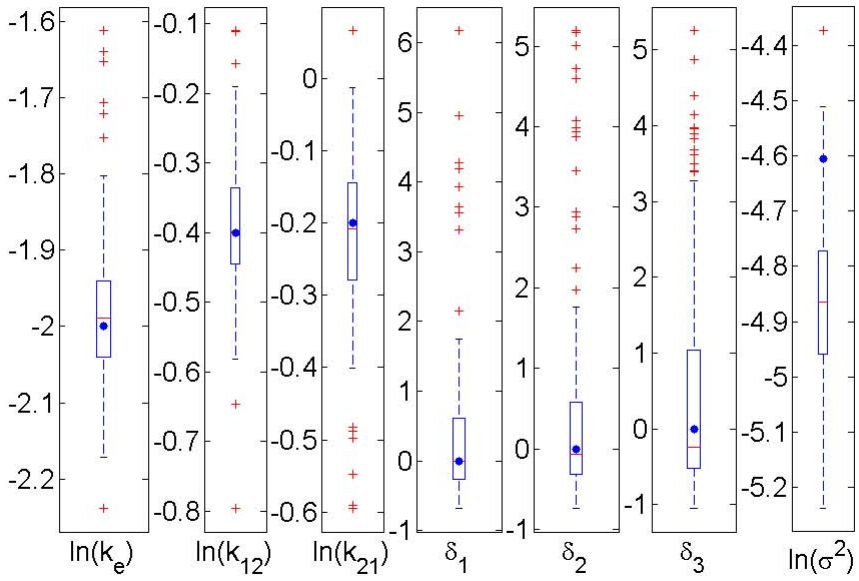

Figure 4.11: Boxplots of the estimated structural parameters for the two-compartment open model, where the dots represent the true values.

4.4 Discussion

The generalized profiling method of estimating nonlinear mixed-effects models expressed by ODE's has been validated by three compartment models in pharmacokinetics. Simulated data sets have been generated for each compartment model under different settings of parameters. We have compared the estimates with true values and illustrated boxplots for all the estimates. This section considers several practical issues related to fitting compartment models, including the problem of identifiability, derived parameters, and starting values for parameters. Further discussions on these topics can be found in Bates and Watts (1988).

4.4.1 Derived parameters

We consider several derived parameters of interest in pharmacokinetic practice for the one-compartment model with first-order absorption and elimination. For this model expressed by ODE's, there are only two fixed-effects parameters, the absorption rate

constant k_a and the elimination rate constant k_e, to be estimated directly. However, we can also estimate the initial values of the differential equations $C_a(0)$, which has been introduced in the previous chapter. Based on the relationship among parameters, some interesting parameters can be derived from these estimates.

The *half-life*, $t_{1/2}$, associated with the elimination rate constant k_e is

$$t_{1/2} = \frac{ln2}{k_e} \approx \frac{0.693}{k_e}. \tag{4.6}$$

The second derived quantity of interest in pharmacological studies is *the volume of distribution* V, which represents the apparent instantaneous dilution space of an instantaneously absorbed dose D. With a bolus injection, the dose D at the drug administration time is known, and the concentration $C_a(0)$ can be estimated. These are related by

$$C_a(0) = \frac{D}{V}. \tag{4.7}$$

Therefore, the volume of distribution V can be estimated by

$$V = \frac{D}{C_a(0)}. \tag{4.8}$$

The total body *clearance*, Cl, represents the volume of the fluid, blood or plasma, that is totally cleared of its content of drug per unit time. (e.g., mL/min or L/min).

$$Cl = \frac{Dk_e}{C_a(0)}. \tag{4.9}$$

Another derived quantity of interest in pharmacological studies is the *area under the curve* (AUC). For many simple compartment models this is equal to the initial concentration in the injection compartment, say $C_a(0)$, divided by the elimination rate, say k_e, so

$$\text{AUC} = \frac{C_a(0)}{k_e}. \tag{4.10}$$

Parameter C_{max} is the maximum or peak concentration of a drug observed after its administration. Parameter C_{min} is the minimum or trough concentration of a drug observed after its administration and just prior to the administration of a subsequent dose.

4.4.2 Identifiability

A noticeable problem when fitting the model with only drug concentrations in the blood is that different settings of the starting values might lead to exchangeable parameter estimates. Although these different sets of parameters give different predicted values for the unobserved component, they obtain the same predicted responses for the observed component(s). Since we cannot evaluate the unobserved components, problems occur when choosing correct estimates from discrete sets of parameters, which is called the problem of identifiability.

We study the problem of identifiability of the one-compartment model with first order absorption and elimination rates by investigating the ODE solution of the drug concentrations in the blood:

$$C(t) = \frac{Dk_ak_e}{Cl(k_a - k_e)}[\exp(-k_et) - \exp(-k_at)] + C(0)\exp(-k_et). \quad (4.11)$$

As can be seen, when $C(0)$ is 0, solutions $[k_a, k_e, Cl]^T$ and $[k_e, k_a, Cl]^T$ are exchangeable.

In order to solve this problem of identifiability, one common way is to put some constraints on the parameters. In practice, we usually have some prior knowledge about the parameters to be estimated. In this case, we have known that the elimination rate is smaller than the absorption rate. Therefore, we could force $0 \leq k_e \leq k_a$ with the following parameter transformations

$$k_e = \exp(\theta_e) \quad (4.12)$$

$$k_a = \exp(\theta_e)(1 + \exp(\theta_a)). \quad (4.13)$$

Alternatively, in this example, we can try different sets of starting values, and chose the estimates that match our prior knowledge.

4.4.3 Starting values

In the simulations of this section, starting values are usually chosen randomly from the standard normal distribution. For the one-compartment model with first-order absorption and elimination, a unique set of estimates is obtained using different starting values if both of the two components of data are used in the optimization. However, as discussed in the previous subsection, two discrete sets of estimates are achieved when different starting values are tried. In simulations, since the initial values of concentration in the blood are 0 for all individuals, the two sets of estimates are exchangeable. We start the optimization with random starting values, and exchange the estimates to ensure that the elimination rate is slower than the absorption rate. In the case of simulations, we can also use starting values that are close to the true values or results from the previous studies. Alternatively, we can always start with a simple model, such as a one-compartment model, and extend it, which is illustrated in Bates and Watts (1988).

Chapter 5

Applications to two real data sets

In this chapter, we analyze two real data sets using the one-compartment model with first-order absorption and elimination expressed by ODE's with the generalized profiling method. As an illustration, the first section analyzes the widely used data set in pharmacokinetics, the theophylline data set. The second section gives the results for antiretroviral drugs, combinations of indinavir (IDV) and ritonavir (RTV), to treat HIV-positive patients.

5.1 Theophylline data

For the theophylline data, twelve subjects are given oral doses of theophylline. Serum concentrations are measured at 11 time points over the next 25 hours after drug administration. This data set can be obtained from the R package nlme by the name Theoph, containing 5 columns as follows: theophylline concentration in the sample (mg/L), time since drug administration when the sample was drawn (hr), a factor identifying the subject, weight of the subject (Kg), and dose administered to the subject (mg/Kg).

Pinheiro and Bates (2000) have explored the theophylline data set by fitting it to the one-compartment open model with first-order absorption and elimination by the R package nlme. In contrast to the generalized profiling method, solutions of ODE's have to be worked out for the drug concentration $C(t)$ at time t after drug

administration before the analysis by nlme. In this case, $C(t)$ is a nonlinear function of the absorption rate k_a, the elimination rate k_e, and the total body clearance Cl, as shown in Equation (5.1).

$$C(t) = \frac{Dk_ak_e}{Cl(k_a - k_e)}[\exp(-k_et) - \exp(-k_at)]. \quad (5.1)$$

To ensure the positive rate constants, logarithms of parameters are estimated instead in the analysis; that is, the vector of fixed-effects parameters is $[\ln(k_a), \ln(k_e), \ln(Cl)]^T$. The package nlme gives the estimates for fixed effects, standard deviations for the fixed effects, random effects, and residuals. Based on these estimates and relationships among parameters, the estimates for $\boldsymbol{\delta}$ are derived for the purpose of comparison with those results by the generalized profiling method. Meanwhile, ψ_1 and ψ_2, variances of the absorption rate and the elimination rate, are the diagonal values of the matrix Ψ which can be obtained using the relationship $\Delta^T\Delta = \Psi^{-1}\sigma^2$.

Table 5.1: Estimates for the structural parameters with different λ by the generalized profiling method and estimates for two models in nlme

The generalized profiling method					nlme	
λ	100	400	1000	10000	Model 1	Model 2
GCV	0.57	0.55	0.56	0.86	-	-
$\ln(k_a)$	0.49	0.48	0.47	0.31	0.45	0.47
$\ln(k_e)$	-2.46	-2.45	-2.45	-2.42	-2.43	-2.45
δ_1	0.029	0.076	0.11	0.70	0.067	0.097
δ_2	1.88	1.76	1.76	5.72	1.65	10.48
$\ln(\sigma^2)$	-0.92	-0.89	-0.87	-0.46	-0.77	-0.69
k_a	1.63	1.62	1.61	1.36	1.57	1.59
k_e	0.086	0.086	0.086	0.089	0.089	0.086
ψ_1	0.38	0.35	0.34	0.16	0.41	0.41
ψ_2	0.0093	0.012	0.012	<1e-5	0.017	<1e-5
σ^2	0.40	0.41	0.42	0.63	0.46	0.50

We estimate $\ln(k_a)$, $\ln(k_e)$, δ_1, δ_2, and $\ln(\sigma^2)$ with different values of λ, as shown in Table 5.1. Initially, values of GCV are evaluated when λ is 100, 1000, and 10000, respectively, resulting that the best estimate of λ is 1000. A better λ is found by further searching around 1000, which results $\lambda = 400$. As can be seen, values of

GCV are quite similar in the range from 100 to 1000, and estimates for the structural parameters are close to each other. However, GCV becomes much larger when λ is 10000, which results in very different estimates of parameters. We also list estimates of two models by nlme in R in which all of the parameters, $\ln(k_a)$, $\ln(k_e)$, and $\ln(Cl)$, have mixed effects. For Model 1, the whole variance-covariance matrix for random effects, Ψ, is estimated. In contrast, Ψ is assumed to be a diagonal matrix in Model 2.

Based on the GCV criterion, we choose results when the smoothing parameter is 400 to compare with estimates by nlme. For the two models by nlme, estimates of the fixed effects k_a and k_e are closer to Model 2 because the generalized profiling method uses the same assumption on Ψ which is a diagonal matrix. However, the estimates of the relative precision factor are not very close to each other. In addition, the relative precision factor is estimated directly in the generalized profiling method, while nlme derives it using relationships among parameters. Therefore, further investigation is needed to compare results from the generalized profiling method and nlme. The diagonal values of the variance-covariance matrix for the random effects Ψ and the residual variance σ^2 are less than the estimates by nlme. These values are underestimated in the generalized profiling method because we use the maximum likelihood estimate while nlme employs the restricted maximum likelihood estimate (REML).

We can obtain the individual-specific parameters based on the estimates of both the fixed-effects parameters and random-effects parameters. For the fitted model with a smoothing parameter $\lambda = 400$, Table 5.2 shows the individual-specific estimates for some pharmacokinetic parameters of interest. Those parameters include the initial value for the drug concentration at the absorption site, $C_a(0)$; the absorption rate, k_a; the elimination rate, k_e; the clearance, Cl; the volume of distribution, V; the half-time, $T_{1/2}$; the area under the curve, AUC. As can be seen, there are certain variability for these parameters from individual to individual, which is very useful for individualizing dosage regimens.

The adequacy of the fitted model is better visualized by displaying the fitted and

Table 5.2: Estimates of parameters for individuals with $\lambda = 400$

Subject	$C_a(0)$	k_a	k_e	Cl	V	$T_{1/2}$	AUC
1	11.59	1.50	0.072	0.029	0.34	9.60	160.5
2	10.11	2.15	0.091	0.037	0.43	7.58	110.6
3	9.27	2.30	0.085	0.042	0.48	8.15	109.1
4	10.83	1.28	0.086	0.034	0.40	8.06	126.0
5	12.63	1.60	0.086	0.039	0.46	7.97	145.4
6	7.86	1.36	0.088	0.043	0.50	7.87	89.2
7	9.43	0.84	0.087	0.045	0.52	7.94	108.2
8	8.83	1.44	0.087	0.044	0.51	7.94	101.2
9	7.71	5.43	0.088	0.034	0.40	7.79	86.8
10	12.78	0.64	0.081	0.037	0.43	8.50	156.9
11	7.76	3.72	0.090	0.054	0.63	7.63	85.5
12	13.11	1.01	0.090	0.034	0.40	7.63	144.3

observed values in the same plot. Figure 5.1 displays both the population predictions obtained by setting the random effects to zero and the individual-specific predictions using both random effects and fixed effects. Observations marked by blue circles only exist for the plasma concentrations of theophylline, while the concentrations at the absorption site is unobservable. However, the unobserved components can be predicted in the generalized profiling method. The red solid lines represent individual-specific predictions while the black dashed lines represent the population predictions. As can be seen, the individual-specific predictions follow the observed closely, indicating that the nonlinear mixed-effects model explains the drug concentrations well.

After fitting the model to the data, it is important to examine if the two basic distributional assumptions underlying the nonlinear mixed-effects model remain valid. First, the within-group errors are independent and identically normally distributed with mean zero and variance σ^2, and they are independent of the random effects. Second, the random effects are normally distributed with mean zero and variance-covariance matrix Ψ and are independent for the different groups.

Because we usually have only a small number of individuals with a few observations, we cannot reliably test assumptions about the random-effects distribution and the independence of the within-group errors. However, a plot of the residuals

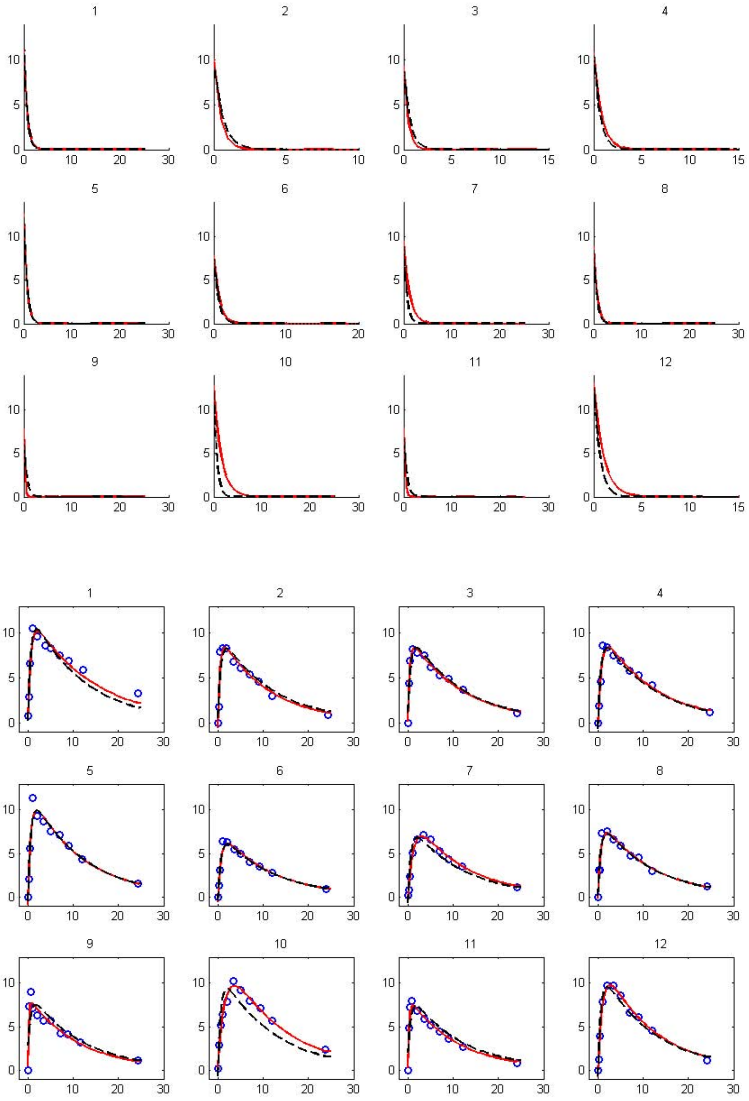

Figure 5.1: Observations (blue circles), population predictions (black dashed lines), and individual-specific predictions (red solid lines) of concentrations of theophylline versus time since administration. Top panel: concentrations at absorption site; bottom panel: plasma concentrations.

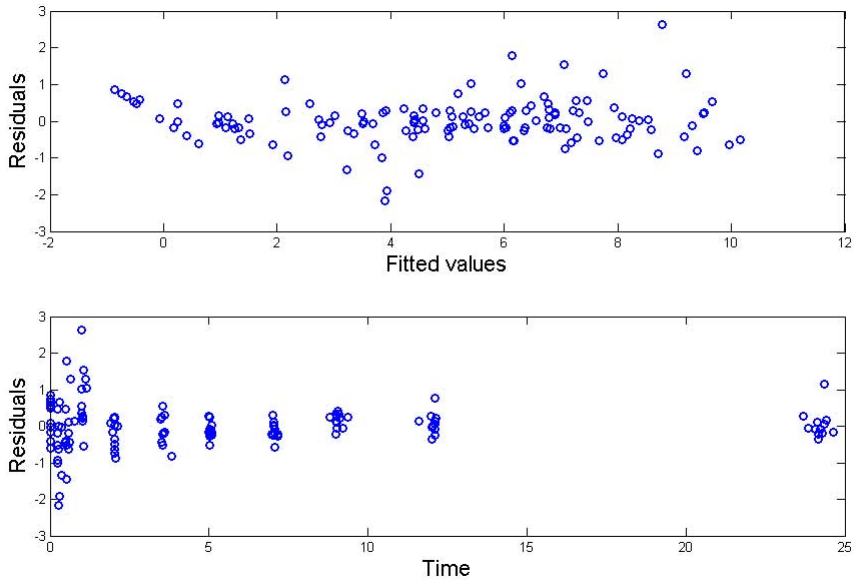

Figure 5.2: Residuals versus fitted values and residuals versus time.

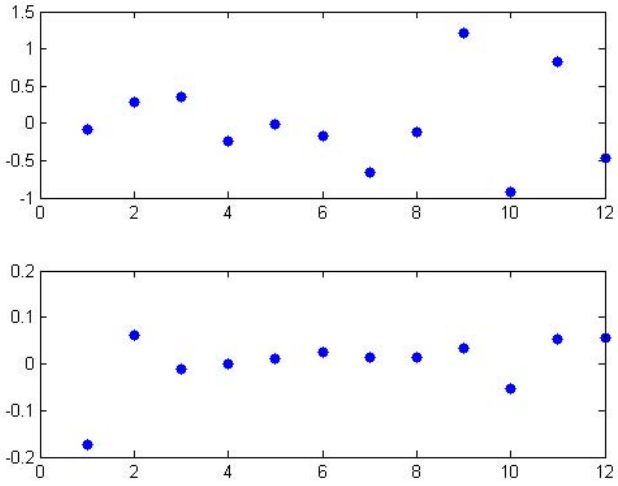

Figure 5.3: Estimates of random effects for $\ln(k_a)$ (top panel) and $\ln(k_e)$ (bottom panel) corresponding to 12 subjects.

versus the fitted values provides a useful tool for assessing the assumptions of normality and thus determine the adequacy of the model. Residuals represent the part of the observation that is not explained by the fitted model, which are defined as the difference between the observed values and predicted or fitted values. An easy way to check independence for data collected over time, i.e. a time series, is to plot the residuals against time indices and look for any suggestive patterns. Figure 5.2 shows the plots of residuals versus fitted values and residuals versus time. As can be seen, the residuals are generally distributed symmetrically around zero, with an approximately constant variance. The plot does not indicate any violations of the assumptions for the within-group error. There are no particular trends in the plot of residuals against time. A few large residuals within the first two hours after drug administration might be caused by intense observations in the process of absorption and sparse observations in the process of elimination. Moreover, no systematic pattern can be observed from Figure 5.3, showing the random effects $\ln(k_a)$ and $\ln(k_e)$ corresponding to 12 subjects.

5.2 Combinations of indinavir and ritonavir

This pharmacokinetical study of antiretroviral drugs intends to understand the widely used protease inhibitor combinations of indinavir (IDV) and ritonavir (RTV) to treat HIV-positive patients. The data sets, provided by Dr. Charles la Porte, were analyzed in Wasmuth, la Porte, Schneider, Burger, and Rockstroh (2004) using non-compartment analysis. For this example, we cannot give results by the package nlme to compare with the generalized profiling method because nlme indicates that there are errors of singular gradients.

Efficacy, toxicity, and costs of drugs are all important issues in clinical practice. The ultimate goal of pharmacokinetical studies is to balance the efficacy and toxicity of drugs, and therefore find the optimum dosage for particular patients. Although better efficacy can be usually obtained by increasing the drug dosage, it is not feasible for a successful long-period treatment due to drug toxicity, adherence issues, and high

costs. Currently HIV/AIDS needs a lifelong treatment, but antiretroviral drugs are not accessible to many HIV-positive patients because of the high costs. Consequently, it is important to study pharmacokinetics in order to reduce the toxicity and costs by lowering the amount of drug dosage, while the antiviral activity is preserved.

This pharmacokinetical study was designed to compare two combinations of indinavir and ritonavir, 400/100 mg IDV/RTV combination and 600/100 mg IDV/RTV combination. Pharmacokinetics and tolerability of IDV/RTV twice daily were assessed in a randomized crossover design in 16 healthy volunteers. Each dosage was taken twice daily for 2 weeks before 12 hours pharmacokinetics were obtained. Serial plasma sampling was performed at 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 4.0, 5.0, 6.0, 8.0, 10.0 and 12.0 hours after drug administration. Table B.1 gives the data set of 400/100 mg IDV/RTV combination, which is referred to as Treatment 1; Table B.2 provides the data set of 600/100 mg IDV/RTV combination, which is referred to as Treatment 2.

Table 5.3 and Table 5.4 give estimates of structural parameters and several derived parameters of indinavir and ritonavior for the two treatments, respectively, by the generalized profiling method. The smoothing parameter λ is selected from $100, 1000$, and 10000 by minimizing the values of GCV. Direct parameters include the logarithms of the population absorption rate and population elimination rate, $\ln(k_a)$ and $\ln(k_e)$; logarithms of the diagonal values of the relative precision factor, δ_1 and δ_2; the logarithm of the residual variance, $\ln(\sigma^2)$. For convenience of interpretation, pharmacokinetic parameters in the original scale are also listed in the table. Parameters ψ_1 and ψ_2 are variances for random effects of the absorption rate and elimination rate, respectively. The other pharmacokinetic parameters are calculated for individuals using the estimated fixed effects and random effects. Geometric means of individual estimates and ranges are given for the following parameters: initial values of the concentration at the absorption site, $C_a(0)$; the area under the curve, AUC; the trough concentration of the drug, C_{min}; the peak concentration of the drug, C_{max}; the half-life, $t_{1/2}$; the total body clearance, Cl; the volume of distribution, V.

For the drug indinavir, the absorption rate k_a for the population in Treatment 1 is higher than that in Treatment 2, while the elimination rate k_e remains the same

Table 5.3: Summary of steady-state pharmacokinetic parameters of indinavir

Parameter	Treatment 1 (400/100 mg)	Treatment 2 (600/100 mg)
λ	1000	10000
$\ln(k_a)$	-0.11	-0.54
$\ln(k_e)$	-1.08	-1.08
δ_1	-0.46	0.16
δ_2	1.08	2.35
$\ln(\sigma^2)$	-1.73	0.89
k_a	0.90	0.58
k_e	0.34	0.34
ψ_1	0.44	0.30
ψ_2	0.020	0.0037
σ^2	0.18	0.41
Geometric mean of individual estimates (range)		
$C_a(0)$ (mg/l)	6.50(3.62 - 10.94)	11.77 (6.69 - 18.00)
AUC (h.mg/l)	18.31 (9.80 - 31.90)	33.23(19.23 - 53.27)
C_{min} (mg/l)	0.24 (0.050 - 0.56)	0.59 (0.13 - 1.44)
C_{max} (mg/l)	3.65 (2.39 - 5.12)	5.77 (3.87 - 8.35)
$t_{1/2}$ (h)	2.03 (1.88 - 2.35)	2.03 (1.99 - 2.07)
Cl (l/h)	21.80 (12.48 - 37.67)	18.04 (11.36 - 30.57)
V (l)	63.88 (36.58 - 110.42)	52.96 (33.34 - 89.75)

in the two populations. Variances for the random effects of k_a and k_e are higher in Treatment 1 than in Treatment 2. For both Treatment 1 and Treatment 2, the variance of k_a is larger than that of k_e. The residual variance in Treatment 1 is less than that in Treatment 2. The geometric mean of the half-time, $t_{1/2}$, is identical for the two treatments, but Treatment 2 has a shorter range. Treatment 1 has slightly larger geometric means of the body clearance Cl and the volume of distribution V with wider ranges than Treatment 2.

The exposure to drug is measured by several pharmacokinetic parameters, including the highest plasma concentration C_{max}, the trough concentration C_{min}, and the area under the concentration-time curve AUC. As can be seen in Table 5.3, all of these three parameters show that the exposure to IDV is reduced after reduction of IDV dose from 600 mg to 400 mg twice daily. The geometric mean value of the individual-specific estimates of AUC has been reduced by almost a half from 33 h.mg/l.

The antiviral efficacy of IDV is assumed to be dependent on the trough concentration values, C_{min} , which should be kept above a certain threshold in order to obtain and maintain adequate suppression of viral replication. Trough concentrations below this threshold will lead to a high risk of possible viral replication and resistance development in case of HIV infection. Therefore, it is essential to control the resulting C_{min} value of IDV for the reduced dose. A threshold of 0.10 mg/l is usually assumed to be adequate for IDV in patients. From the results in Table 5.3, the mean values of C_{min} for both of the two treatments are above this threshold. All values of C_{min} for subjects with Treatment 2 are above 0.10 mg/l threshold, while 2 out of 16 subjects with Treatment 1 are below this threshold.

For the drug ritonavior, the absorption rate k_a for the population in Treatment 1 is much higher than that in Treatment 2, while the elimination rate k_e is slightly higher than that in Treatment 2. Variances for the random effects of k_a and k_e are higher in Treatment 1 than in Treatment 2. For both Treatment 1 and Treatment 2, the variance of k_a is larger than that of k_e. The residual variance in Treatment 1 is less than that in Treatment 2. Treatment 1 has a smaller geometric mean of the

half-time, $t_{1/2}$, than Treatment 2, but Treatment 2 has a shorter range. Treatment 1 has slightly smaller geometric means of the body clearance Cl and the volume of distribution V with shorter ranges than Treatment 2.

While antiviral efficacy depends on C_{min}, the development of IDV-related toxicity depends on the height of plasma levels, C_{max}. Higher peak plasma levels of IDV will lead to more side effects. IDV plasma concentrations above 8 mg/l are generally associated with severe side effects. From the results in Table 5.3, the mean values of C_{max} for both of the two treatments are below this threshold. None of values of C_{max} with Treatment 1 are above 8 mg/l threshold, while 2 out of 16 subjects with Treatment 2 are above this threshold.

Similar to IDV, the exposure to RTV is reduced after reduction of IDV dose because IDV has an inhibiting effect on the metabolism of RTV. As can be seen from Table 5.4, the trough concentration values C_{min} has been reduced by almost a half after reduction of IDV dose from 600 mg to 400 mg twice daily, and the area under the concentration-time curve AUC has been reduced by about 25%. However, values of C_{max} remain almost the same for the two treatments.

The fitted models are illustrated in Figure 5.4 - 5.7. If the individual-specific predictions follow the observed closely, the data can be fitted by the nonlinear mixed-effects model well. As can be seen, the one-compartment open model with first-order absorption and elimination can fit indinavir better than ritonavir. For individuals with a "delay" in the response time, the fitted curves are not as good as others, which might be improved by considering the delay time as parameters to estimate.

Table 5.4: Summary of steady-state pharmacokinetic parameters of ritonavior

Parameter	Treatment 1 (400/100 mg)	Treatment 2 (600/100 mg)
λ	1000	100
$\ln(k_a)$	-0.32	-0.94
$\ln(k_e)$	-1.59	-1.75
δ_1	-1.46	-0.92
δ_2	0.032	5.54
$\ln(\sigma^2)$	-3.35	-2.49
k_a	0.72	0.39
k_e	0.20	0.17
ψ_1	0.65	0.53
ψ_2	0.033	<1e-5
σ^2	0.035	0.083
Geometric mean of individual estimates (range)		
$C_a(0)$ (mg/l)	2.08 (0.87 - 3.36)	2.50 (1.18, 6.70)
AUC (h.mg/l)	9.47 (4.20 - 20.29)	12.68 (6.75 - 38.40)
C_{min} (mg/l)	0.32 (0.11 - 0.82)	0.61 (0.25 - 1.80)
C_{max} (mg/l)	1.51 (0.67 - 3.16)	1.60 (0.77 - 2.94)
$t_{1/2}$ (h)	3.36 (2.91 - 4.36)	3.98 (3.98 - 3.98)
Cl (l/h)	41.90 (24.37 - 94.24)	47.33 (15.63 - 88.84)
V (l)	204.52 (118.96 - 460.01)	271.43 (89.61, 509.48)

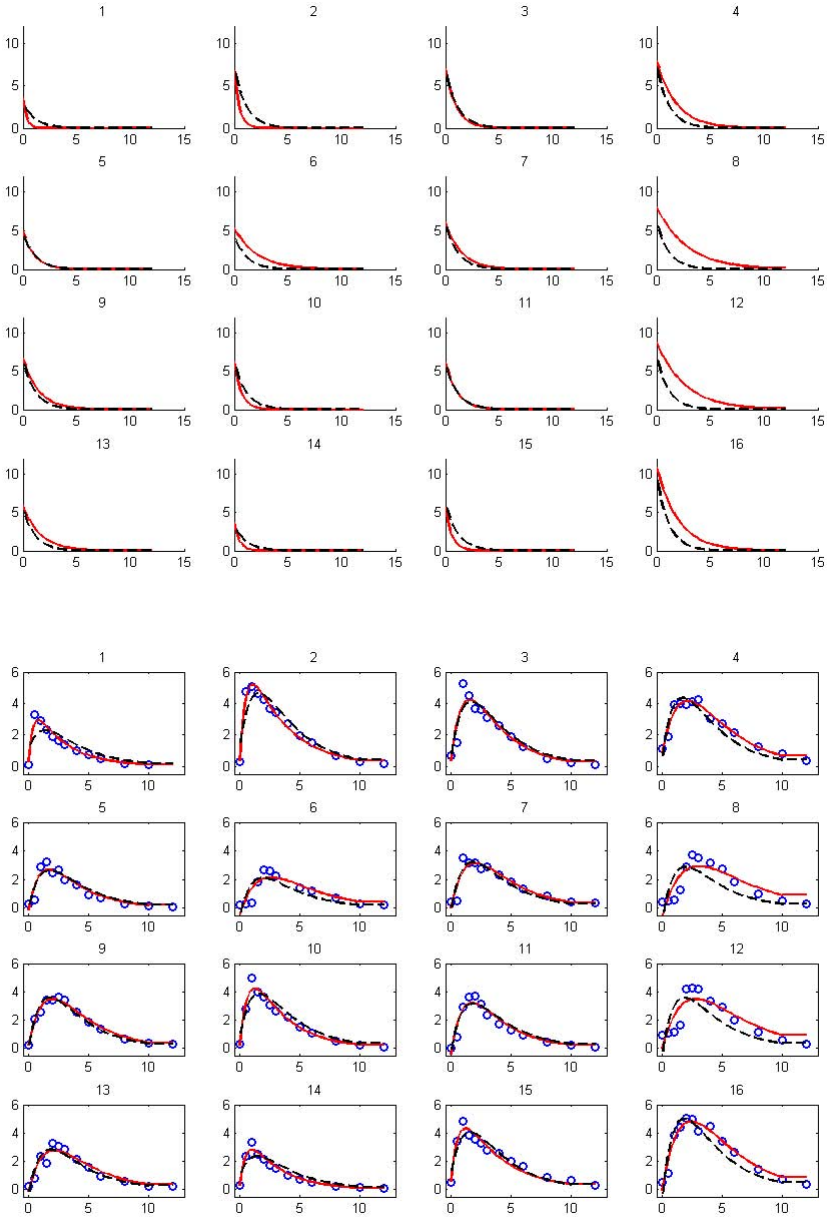

Figure 5.4: Observations (blue circles), population predictions (black dashed lines), and individual-specific predictions (red solid lines) of concentrations of indinavir versus time since administration of 400 mg indinavir plus 100 mg ritonavir. Top panel: concentrations at absorption site; bottom panel: plasma concentrations.

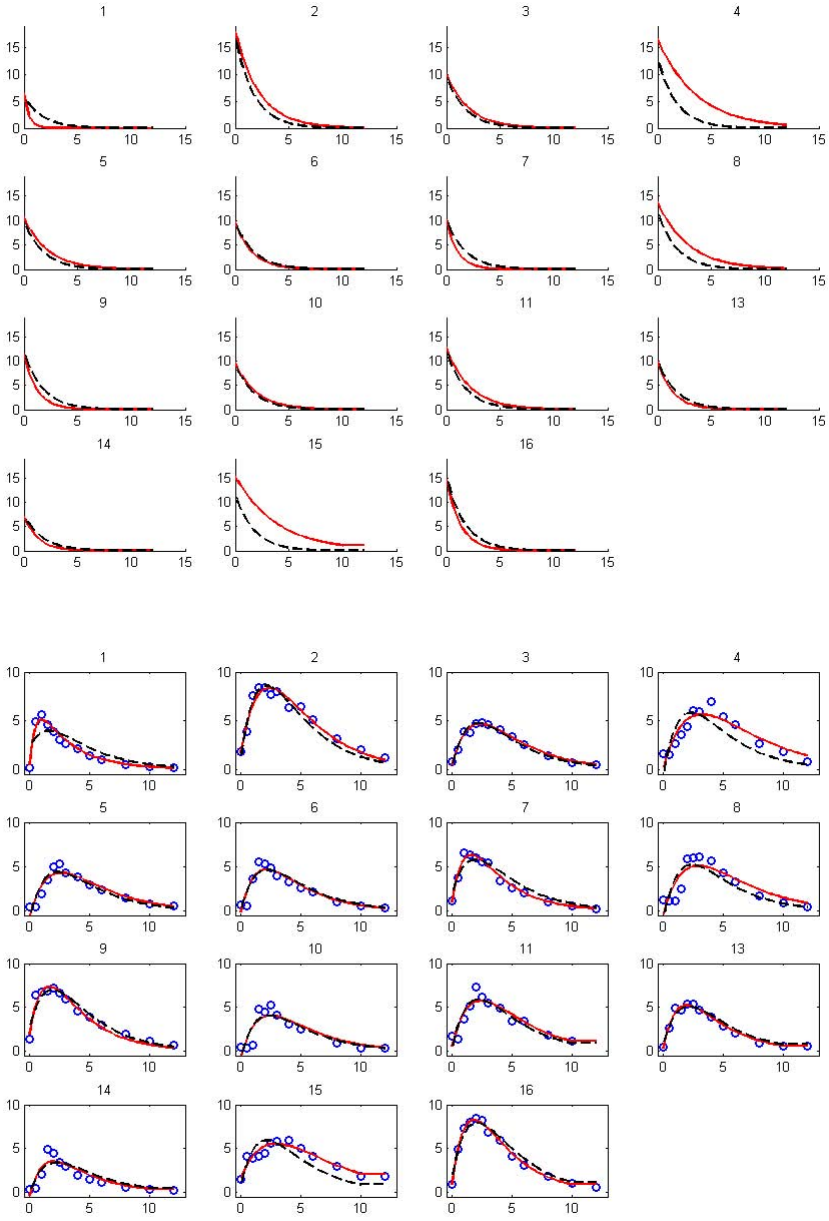

Figure 5.5: Observations (blue circles), population predictions (black dashed lines), and individual-specific predictions (red solid lines) of concentrations of indinavir versus time since administration of 600 mg indinavir plus 100 mg ritonavir. Top panel: concentrations at absorption site; bottom panel: plasma concentrations.

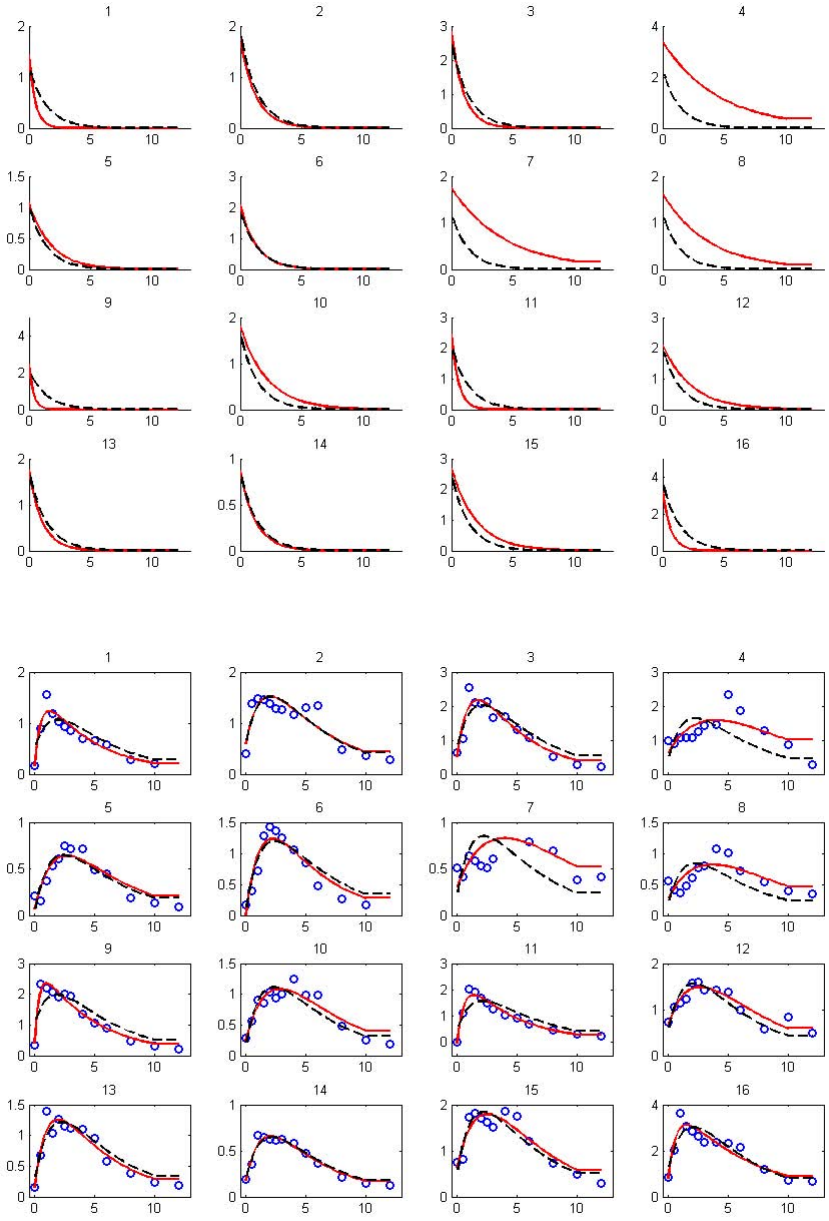

Figure 5.6: Observations (blue circles), population predictions (black dashed lines), and individual-specific predictions (red solid lines) of concentrations of ritonavir versus time since administration of 400 mg indinavir plus 100 mg ritonavir. Top panel: concentrations at absorption site; bottom panel: plasma concentrations.

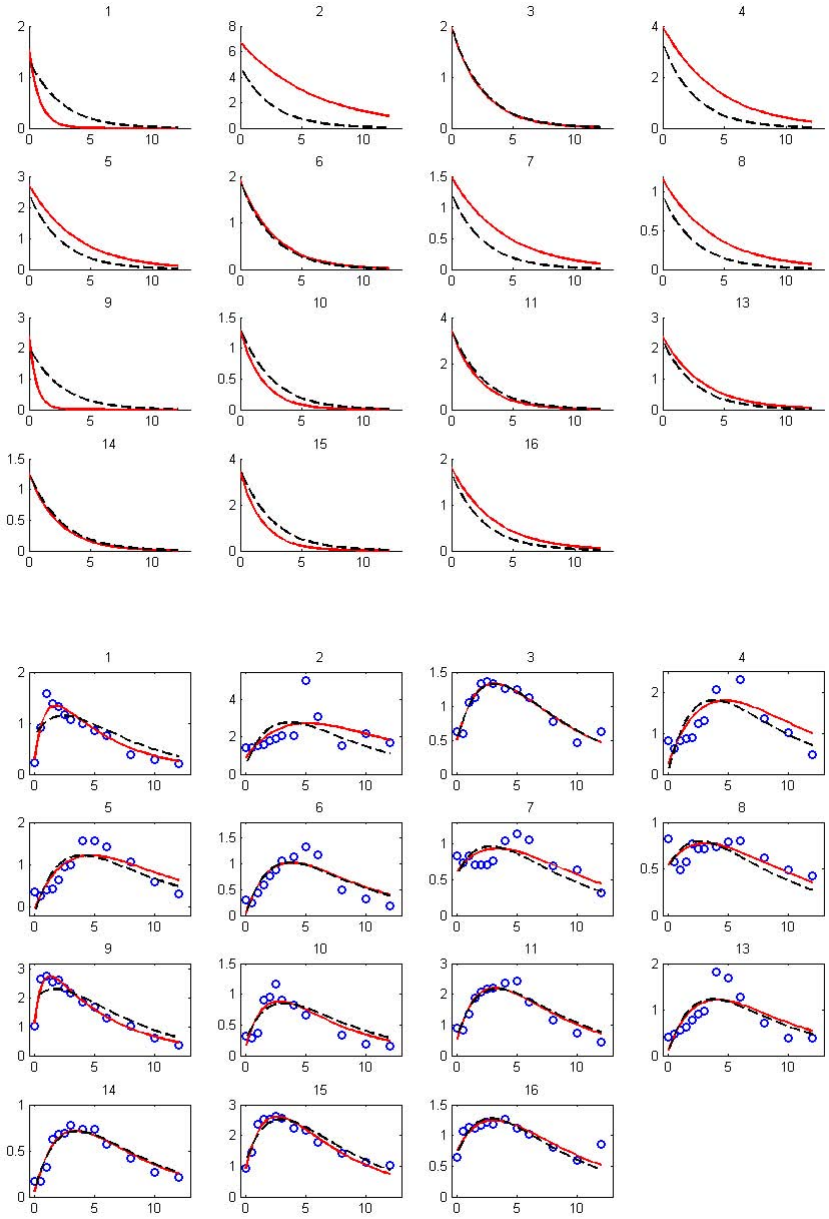

Figure 5.7: Observations (blue circles), population predictions (black dashed lines), and individual-specific predictions (red solid lines) of concentrations of ritonavir versus time since administration of 600 mg indinavir plus 100 mg ritonavir. Top panel: concentrations at absorption site; bottom panel: plasma concentrations.

Chapter 6

Conclusions and conjectures

This thesis has explored the nonlinear mixed-effects models (NLME) involving the first-order differential equations based on the framework of the generalized profiling method. The generalized profiling method has been validated by simulations for three compartment pharmacokinetic models, indicating that the estimates are generally unbiased for the fixed effects $\boldsymbol{\beta}$ and the relative precision factor Δ with an appropriate smoothing parameter λ, but the maximum likelihood estimate used in the generalized profiling method tends to underestimate the residual variance σ^2. The restricted maximum likelihood (REML) estimates should be preferred to obtain unbiased estimates in future work. For the real theophylline data, we have obtained similar estimates with those using the package nlme in R/S-PLUS except that we underestimate the residual variance σ^2; For the real IDV/RTV data set, we also acquire reasonable estimates while nlme fails with this data set.

Conventional methods for estimating nonlinear mixed-effects models require analytical solutions for ODE's. With the obtained ODE solutions, non-Bayesian approaches maximize the (restricted) likelihood of structural parameters based on a variety of approximation methods, such as the first-order method, the conditional first-order linearization, Laplacian approximation, and the Lindstrom and Bates algorithm. In contrast, the Bayesian method considers the full probability distribution of random effects $\mathbf{b}$ and structural parameters $[\boldsymbol{\beta},\ \Psi,\ \sigma^2]^T$, and obtains inferences by techniques like MCMC.

In contrast to the conventional methods, the generalized profiling method uses smoothing splines to approximate ODE solutions with linear combinations of basis functions. It then optimizes the coefficients to basis functions instead of solving ODE's analytically or numerically, which makes our method particularly appealing for ODE's without analytical solutions. In addition, the generalized profiling method can estimate both initial values and unobserved components of ODE's. This is particularly useful for the examples of compartment pharmacokinetic models, where we can only observe the drug concentrations in the central compartment. Furthermore, we can include some prior knowledge by choosing appropriate basis functions. For example, B-splines can have more knots in the interval where the curves are sharp and complicated. For simplicity, this thesis doesn't explore the effect of knot selection, which will leave for future work.

The estimation process in the generalized profiling method is different from conventional methods. We choose the similar optimization criteria when estimating random effects and structural parameters in the generalized profiling method as those in the two steps in the Lindstrom and Bates algorithm. In both methods, random effects $\mathbf{b}$ are optimized by a penalized nonlinear least squares, and structural parameters $\boldsymbol{\theta}$ are optimized by the likelihood based on a first-order Taylor expansion of the mean function around the current estimate of $\boldsymbol{\theta}$ and the optimized $\mathbf{b}$. However, the generalized profiling method implements a nested multiple-level optimization, while Lindstrom and Bates algorithm employs a process of alternating between two steps. Although we only analyze two real data sets, estimates by the generalized profiling method seem more robust and stable based on the fact that the package nlme has failed to analyze the data sets of IDV/RTV combinations. The robustness and stability of our estimates are probably caused by the analytic gradients and Hessian matrices worked out with the implicit function theorem.

The smoothing parameter λ is selected by the criterion GCV, which is a reasonable criterion as illustrated by examples. However, it is much more time-consuming to conduct a four-level optimization than a three-level one, since updating λ in each time leads to a three-level optimization. The Newton-Raphson algorithm is applied

to estimate parameters with the gradients and Hessian matrices written out analytically for each level of optimization. Some groups of parameters are expressed as an explicit or implicit function of other parameters by the implicit function theorem if necessary. However, cascaded relationships among the parameters will become much more complicated with the increase of levels of optimization. Moreover, the curve of GCV versus λ tends to have a flat region, which makes the process of optimization hard to converge to a unique estimate. In future work, better criteria for optimizing the smoothing parameter need to be explored.

Parameter transforms are widely used to estimate rate constants in compartment models to ensure positive values while keeping the optimization problem unconstrained. This thesis only uses the logarithms for rate constants, and it will be even better if we can also work on the logarithms of concentrations to avoid negative values.

In the examples of this thesis, we assume that all parameters have mixed effects, including random effects and fixed effects. It is desirable if we can indicate which group of parameters have random effects. The initial values of ODE components are estimated individually, but it is not clear how to estimate mixed effects for them. The generalized profiling method considers the structural sub-model and statistical sub-model while the covariate sub-model hasn't been included. In principle, covariates can be included into a compartment model by indicating them in its transfer matrix. More examples are needed to consider flexible mixed effects and covariates.

Although the current models and the corresponding MATLAB codes are developed for compartment models expressed by first-order ODE's, it is possible to improve them to fit more complicated ODE's, which only need more applications of Implicit Function Theorem to derive the total derivatives.

More parameters can be estimated to improve the accuracy of models in future work. As shown in the real data, time delay makes the analysis more challenging. Although it is still not clear how to estimate the time delay effectively, one possible way is to treat it as a local parameter for each individual and estimate it with random effects $\mathbf{b}$ at the same level of optimization. This thesis assumes that the variance-

covariance matrix for random effects, Ψ, is a diagonal matrix, but the whole matrix can be estimated through more complicated matrix transform to ensure a positive-definite matrix.

The most complicated compartment model in this thesis includes two differential equations, one of which is unobservable, with three fixed parameters and their corresponding random effects. We believe the generalized profiling method can be more desirable for more complicated ODE's. Moreover, more software tools besides R should be compared with the generalized profiling method for further evaluation.

Appendix A

Derivative Calculations

The Newton-Raphson algorithm is used to the optimization of random effects and structural effects. In the following, we write out the optimization criteria along with the major steps of the gradients and Hessian matrices.

A.1 The inner optimization level to estimate c

The coefficients of basis function expansions $\mathbf{c}$ are optimized by penalized smoothing with the penalty defined by ODE's, which is written as follows

$$J(\mathbf{c}|\mathbf{b},\boldsymbol{\theta},\lambda) = \sum_{i=1}^{N} \left\{ (\mathbf{y}_i - \mathbf{x}_i)^T \mathbf{W} (\mathbf{y}_i - \mathbf{x}_i) + \lambda \int (L\mathbf{x}_i(t))^T (L\mathbf{x}_i(t)) dt \right\}. \quad \text{(A.1)}$$

Then the coefficient vector for the i-th individual, $\mathbf{c}_i$, is obtained by minimizing the criterion $J(\mathbf{c}|\mathbf{b},\boldsymbol{\theta},\lambda)$ analytically as follows

$$\hat{\mathbf{c}}_i(\mathbf{b},\boldsymbol{\theta},\lambda) = [\boldsymbol{\Phi}_i^T \mathbf{W} \boldsymbol{\Phi}_i + \lambda \mathbf{R}_i(\mathbf{b},\boldsymbol{\theta},\lambda)]^{-1} \boldsymbol{\Phi}_i^T \mathbf{W} \mathbf{y}_i. \quad \text{(A.2)}$$

where

$$\mathbf{R}_i(\mathbf{b},\boldsymbol{\theta},\lambda) = \int L\boldsymbol{\phi}_i(t) L\boldsymbol{\phi}_i(t)^T dt. \quad \text{(A.3)}$$

and

$$L\boldsymbol{\phi}_i(t) \;=\; \frac{d\boldsymbol{\phi}_i(t)}{dt} - \boldsymbol{\phi}_i(t)M_i^T. \tag{A.4}$$

A.2 The middle optimization level to estimate b

The optimization criterion for the random effects, $H(\mathbf{b}|\boldsymbol{\theta},\lambda)$, is the penalized nonlinear least squares as follows

$$\begin{aligned} & H(\mathbf{b}|\boldsymbol{\theta},\lambda) && (A.5)\\ =\; & \sum_i^N \left\{ (\mathbf{y}_i - \mathbf{x}_i)^T \mathbf{W} (\mathbf{y}_i - \mathbf{x}_i) + \mathbf{b}_i^T \Delta^T \Delta \mathbf{b}_i \right\} \\ =\; & \sum_i^N \left\{ \mathbf{y}_i^T [I - A_i(\mathbf{b}|\boldsymbol{\theta},\lambda)]^T \mathbf{W} [I - A_i(\mathbf{b}|\boldsymbol{\theta},\lambda)] \mathbf{y}_i + \mathbf{b}_i^T \Delta^T \Delta \mathbf{b}_i \right\}. && (A.6) \end{aligned}$$

where the smoothing matrix $A_i(\mathbf{b}|\boldsymbol{\theta},\lambda) = \boldsymbol{\Phi}_i[\boldsymbol{\Phi}_i^T\mathbf{W}\boldsymbol{\Phi}_i + \lambda\mathbf{R}_i(\mathbf{b}|\boldsymbol{\theta},\lambda)]^{-1}\boldsymbol{\Phi}_i^T\mathbf{W}$.

The gradient of $H(\mathbf{b}|\boldsymbol{\theta},\lambda)$ with respect to $\mathbf{b}_i$ is

$$\frac{dH}{d\mathbf{b}_i} \;=\; \frac{d\mathrm{SSE}_i(\mathbf{b}|\boldsymbol{\theta},\lambda)}{d\mathbf{b}_i} + 2\Delta^T\Delta\mathbf{b}_i.$$

The Hessian matrix of $H(\mathbf{b}|\boldsymbol{\theta},\lambda)$ with respect to $\mathbf{b}_i$ is

$$\frac{d^2H}{d\mathbf{b}_i^2} \;=\; \frac{d\mathrm{SSE}_i^2(\mathbf{b}|\boldsymbol{\theta},\lambda)}{d\mathbf{b}_i^2} + 2\Delta^T\Delta.$$

A.3 The outer optimization level

A.3.1 Estimating $\boldsymbol{\beta}$

The optimization criterion for the fixed effects, $G(\boldsymbol{\beta}|\lambda)$, is the sum of squared errors as follows

$$
\begin{aligned}
G(\boldsymbol{\beta}|\lambda) &= \sum_{i=1}^{N} \mathrm{SSE}_i(\boldsymbol{\beta}|\lambda) \\
&= \sum_{i=1}^{N} (\mathbf{y}_i - \mathbf{x}_i)^T \mathbf{W} (\mathbf{y}_i - \mathbf{x}_i) \\
&= \sum_{i=1}^{N} \mathbf{y}_i^T [I - A_i(\hat{\mathbf{b}}(\boldsymbol{\beta}, \lambda), \boldsymbol{\beta}|\lambda)]^T \mathbf{W} [I - A_i(\hat{\mathbf{b}}(\boldsymbol{\beta}, \lambda), \boldsymbol{\beta}|\lambda)] \mathbf{y}_i. \quad \text{(A.7)}
\end{aligned}
$$

where the smoothing matrix

$$
A_i(\hat{\mathbf{b}}(\boldsymbol{\beta}, \lambda), \boldsymbol{\beta}|\lambda) = \boldsymbol{\Phi}_i [\boldsymbol{\Phi}_i^T \mathbf{W} \boldsymbol{\Phi}_i + \lambda \mathbf{R}_i(\hat{\mathbf{b}}(\boldsymbol{\beta}, \lambda), \boldsymbol{\beta}|\lambda)]^{-1} \boldsymbol{\Phi}_i^T \mathbf{W}. \quad \text{(A.8)}
$$

The gradient of $G(\boldsymbol{\beta}|\lambda)$ is calculated using

$$
\frac{d\mathrm{SSE}_i(\hat{\mathbf{b}}(\boldsymbol{\beta}, \lambda), \boldsymbol{\beta})|\lambda}{d\boldsymbol{\beta}} = \frac{\partial \mathrm{SSE}_i(\mathbf{b}, \boldsymbol{\beta}, \lambda)}{\partial \beta}\bigg|_{\hat{\mathbf{b}}} + \frac{\partial \mathrm{SSE}_i(\mathbf{b}, \boldsymbol{\beta}, \lambda)}{\partial \mathbf{b}_i}\bigg|_{\hat{\mathbf{b}}} \frac{d\hat{\mathbf{b}}_i(\boldsymbol{\beta}, \lambda)}{d\boldsymbol{\beta}}, \quad \text{(A.9)}
$$

where

$$
\begin{aligned}
&\frac{\partial \hat{\mathbf{b}}_i(\boldsymbol{\beta}, \lambda)}{\partial \boldsymbol{\beta}} \\
&= -\left[\frac{\partial^2 \mathrm{SSE}_i(\mathbf{b}, \boldsymbol{\beta}, \lambda)}{\partial \mathbf{b}_i^2}\bigg|_{\hat{\mathbf{b}}} + \frac{\partial \mathrm{PEN2}(\boldsymbol{\beta}, \mathbf{b}, \lambda)}{\partial \mathbf{b}_i^2}\bigg|_{\hat{\mathbf{b}}}\right]^{-1} \left[\frac{\partial^2 \mathrm{SSE}_i(\boldsymbol{\beta}, \mathbf{b}, \lambda)}{\partial \mathbf{b}_i \partial \boldsymbol{\beta}}\bigg|_{\hat{\mathbf{b}}}\right] \quad \text{(A.10)}
\end{aligned}
$$

The Hessian matrix of $G(\boldsymbol{\beta}|\lambda)$ is calculated by

$$
\begin{aligned}
&\frac{d^2 \mathrm{SSE}_i(\hat{\mathbf{b}}(\boldsymbol{\beta}, \lambda), \boldsymbol{\beta}|\lambda)}{d\beta_k d\beta_l} \\
&= \frac{\partial^2 \mathrm{SSE}_i(\mathbf{b}, \boldsymbol{\beta}, \lambda)}{\partial \beta_k \partial \beta_l}\bigg|_{\hat{\mathbf{b}}} + \frac{\partial \mathrm{SSE}_i^2(\hat{\mathbf{b}}(\boldsymbol{\beta}, \lambda), \boldsymbol{\beta}|\lambda)}{\partial \beta_k \partial \hat{\mathbf{b}}_i} \cdot \frac{\partial \hat{\mathbf{b}}_i(\boldsymbol{\beta}, \lambda)}{\partial \beta_l} \\
&\quad + \left[\frac{\partial \mathrm{SSE}_i^2(\hat{\mathbf{b}}(\boldsymbol{\beta}, \lambda), \boldsymbol{\beta}|\lambda)}{\partial \hat{\mathbf{b}}_i \partial \beta_l} + \frac{\partial \mathrm{SSE}_i^2(\hat{\mathbf{b}}(\boldsymbol{\beta}, \lambda), \boldsymbol{\beta}|\lambda)}{\partial \hat{\mathbf{b}}_i \partial \hat{\mathbf{b}}_i^T} \cdot \frac{\partial \hat{\mathbf{b}}_i(\boldsymbol{\beta}, \lambda)}{\partial \beta_l}\right] \frac{\partial \hat{\mathbf{b}}_i(\boldsymbol{\beta}, \lambda)}{\partial \beta_k} \\
&\quad + \frac{\partial \mathrm{SSE}_i(\hat{\mathbf{b}}(\boldsymbol{\beta}, \lambda), \boldsymbol{\beta}|\lambda)}{\partial \hat{\mathbf{b}}} \frac{\partial^2 \hat{\mathbf{b}}(\boldsymbol{\beta}, \lambda)}{\partial \beta_k \partial \beta_l}. \quad \text{(A.11)}
\end{aligned}
$$

where

$$\begin{aligned}
&\frac{\partial \hat{\mathbf{b}}_i^2}{\partial\beta_k\partial\beta_l} \\
= & -\left[\frac{\partial^2 \mathrm{SSE}_i}{\partial\hat{\mathbf{b}}_i\partial\hat{\mathbf{b}}_i}+\frac{\partial^2 \mathrm{PEN2}_i}{\partial\hat{\mathbf{b}}_i\partial\hat{\mathbf{b}}_i}\right]^{-1}\left[\frac{\partial^3 \mathrm{SSE}_i}{\partial\hat{\mathbf{b}}_i\partial\beta_k\partial\beta_l}+\frac{\partial^3 \mathrm{SSE}_i}{\partial\hat{\mathbf{b}}_i\partial\beta_k\partial\hat{\mathbf{b}}_i}\frac{\partial\hat{\mathbf{b}}_i}{\partial\beta_l}\right. \\
& \left.+\sum_{m=1}^{q}\left(\frac{\partial^3 \mathrm{SSE}_i}{\partial\hat{\mathbf{b}}_i\partial\hat{\mathbf{b}}_{im}\partial\hat{\mathbf{b}}_i}\cdot\frac{\partial\hat{\mathbf{b}}_i}{\partial\beta_l}+\frac{\partial^3 \mathrm{SSE}_i}{\partial\hat{\mathbf{b}}_i\partial\hat{\mathbf{b}}_{im}\partial\beta_l}\right)\cdot\frac{\partial\hat{\mathbf{b}}_{im}}{\partial\beta_k}\right]. \qquad \text{(A.12)}
\end{aligned}$$

A.3.2 Estimating $\boldsymbol{\theta} = [\boldsymbol{\beta}, \boldsymbol{\delta}, \sigma^2]^T$

We define

$$\boldsymbol{\omega}_i = \mathbf{y}_i - \mathbf{x}_i + \left.\frac{\partial \mathbf{x}_i}{\partial \mathbf{b}_i^T}\right|_{\hat{\mathbf{b}}} \hat{\mathbf{b}}_i, \qquad \text{(A.13)}$$

and

$$\Sigma(\Delta) = I + \left.\frac{\partial \mathbf{x}_i}{\partial \mathbf{b}_i^T}\right|_{\hat{\mathbf{b}}} \Delta^{-1}\Delta^{-T} \left.\frac{\partial \mathbf{x}_i}{\partial \mathbf{b}_i^T}\right|_{\hat{\mathbf{b}}}^T. \qquad \text{(A.14)}$$

The criterion $G_{LME}(\boldsymbol{\theta}|\lambda)$ is based on a first-order Taylor expansion of the nonlinear function $\mathbf{x}$ around the current value of $\boldsymbol{\beta}$ and the optimized $\mathbf{b}$.

$$\begin{aligned}
& G_{LME}(\boldsymbol{\theta}|\lambda) \\
= & \frac{\sum_i^N n_i}{2}\log 2\pi\sigma^2 + \frac{1}{2}\sum_{i=1}^{N}\left\{\log\left(|\Sigma(\Delta)|\right) + \sigma^{-2}\boldsymbol{\omega}_i^T\left(\Sigma(\Delta)\right)^{-1}\boldsymbol{\omega}_i\right\}. \qquad \text{(A.15)}
\end{aligned}$$

The gradient of $G_{LME}(\boldsymbol{\theta}|\lambda)$ with respect to σ^2 is

$$\frac{dG_{LME}(\boldsymbol{\theta}|\lambda)}{d\sigma^2} = \frac{\sum_i^N n_i}{2\sigma^2} - \frac{1}{2\sigma^4}\sum_{i=1}^{N}\boldsymbol{\omega}_i^T\left(\Sigma(\Delta)\right)^{-1}\boldsymbol{\omega}_i. \qquad \text{(A.16)}$$

The gradient of $G_{LME}(\boldsymbol{\theta}|\lambda)$ with respect to $\boldsymbol{\beta}$ is

$$\frac{dG_{LME}(\boldsymbol{\theta}|\lambda)}{d\boldsymbol{\beta}} = \frac{\partial G_{LME}(\boldsymbol{\theta}|\lambda)}{\partial \boldsymbol{\beta}} + \sum_{i=1}^{N} \frac{\partial G_{LME}(\boldsymbol{\theta}|\lambda)}{\partial \mathbf{b}_i} \frac{\partial \mathbf{b}_i}{\partial \boldsymbol{\beta}}. \tag{A.17}$$

The gradient of $G_{LME}(\boldsymbol{\theta}|\lambda)$ with respect to $\boldsymbol{\delta}$ is

$$\frac{dG_{LME}(\boldsymbol{\theta}|\lambda)}{d\boldsymbol{\delta}} = \frac{\partial G_{LME}(\boldsymbol{\theta}|\lambda)}{\partial \boldsymbol{\delta}} + \sum_{i=1}^{N} \frac{\partial G_{LME}(\boldsymbol{\theta}|\lambda)}{\partial \mathbf{b}_i} \frac{\partial \mathbf{b}_i}{\partial \boldsymbol{\delta}}. \tag{A.18}$$

The following formulas are needed to calculate the gradient of $G_{LME}(\boldsymbol{\theta}|\lambda)$.

$$\begin{aligned} & \frac{\partial G_{LME}(\boldsymbol{\theta}|\lambda)}{\partial \beta_j} \\ = \; & \frac{1}{2} \sum_{i=1}^{N} \left\{ \text{trace} \left(\Sigma(\Delta)^{-1} \frac{\partial \Sigma(\Delta)}{\partial \beta_j} \right) \right. \\ + \; & \left. \sigma^{-2} \left[\frac{\partial \boldsymbol{\omega}_i}{\partial \beta_j}^T \Sigma(\Delta)^{-1} \boldsymbol{\omega}_i + \boldsymbol{\omega}_i^T \frac{\partial \Sigma(\Delta)^{-1}}{\partial \beta_j} \boldsymbol{\omega}_i + \boldsymbol{\omega}_i^T \Sigma(\Delta)^{-1} \frac{\partial \boldsymbol{\omega}_i}{\partial \beta_j} \right] \right\}. \end{aligned} \tag{A.19}$$

$$\begin{aligned} & \frac{\partial G_{LME}(\boldsymbol{\theta}|\lambda)}{\partial \delta_j} \\ = \; & \frac{1}{2} \sum_{i=1}^{N} \left\{ \text{trace} \left(\Sigma(\Delta)^{-1} \frac{\partial \Sigma(\Delta)}{\partial \delta_j} \right) \right. \\ + \; & \left. \sigma^{-2} \left[\frac{\partial \boldsymbol{\omega}_i}{\partial \delta_j}^T \Sigma(\Delta)^{-1} \boldsymbol{\omega}_i + \boldsymbol{\omega}_i^T \frac{\partial \Sigma(\Delta)^{-1}}{\partial \delta_j} \boldsymbol{\omega}_i + \boldsymbol{\omega}_i^T \Sigma(\Delta)^{-1} \frac{\partial \boldsymbol{\omega}_i}{\partial \delta_j} \right] \right\}. \end{aligned} \tag{A.20}$$

$$\begin{aligned} & \frac{\partial G_{LME}(\boldsymbol{\theta}|\lambda)}{\partial b_{ij}} \\ = \; & \frac{1}{2} \left\{ \text{trace} \left(\Sigma(\Delta)^{-1} \frac{\partial \Sigma(\Delta)}{\partial b_{ij}} \right) \right. \\ + \; & \left. \sigma^{-2} \left[\frac{\partial \boldsymbol{\omega}_i}{\partial b_{ij}}^T \Sigma(\Delta)^{-1} \boldsymbol{\omega}_i + \boldsymbol{\omega}_i^T \frac{\partial \Sigma(\Delta)^{-1}}{\partial b_{ij}} \boldsymbol{\omega}_i + \boldsymbol{\omega}_i^T \Sigma(\Delta)^{-1} \frac{\partial \boldsymbol{\omega}_i}{\partial b_{ij}} \right] \right\}. \end{aligned} \tag{A.21}$$

$$\frac{\partial \hat{\mathbf{b}}_i}{\partial \boldsymbol{\beta}} = -\left[\frac{\partial^2 H_i(\mathbf{b}|\boldsymbol{\theta},\lambda)}{\partial \mathbf{b}_i^2}\bigg|_{\hat{\mathbf{b}}}\right]^{-1}\left[\frac{\partial^2 H_i(\mathbf{b}|\boldsymbol{\theta},\lambda)}{\partial \mathbf{b}_i \partial \boldsymbol{\beta}}\bigg|_{\hat{\mathbf{b}}}\right]. \tag{A.22}$$

$$\frac{\partial \hat{\mathbf{b}}_i}{\partial \boldsymbol{\delta}} = -\left[\frac{\partial^2 H_i(\mathbf{b}|\boldsymbol{\theta},\lambda)}{\partial \mathbf{b}_i^2}\bigg|_{\hat{\mathbf{b}}}\right]^{-1}\left[\frac{\partial^2 H_i(\mathbf{b}|\boldsymbol{\theta},\lambda)}{\partial \mathbf{b}_i \partial \boldsymbol{\delta}}\bigg|_{\hat{\mathbf{b}}}\right]. \tag{A.23}$$

A.4 The outermost optimization level to estimate λ

The smoothing parameter λ is optimized by generalized cross-validation (GCV) as follows:

$$F(\lambda) = \mathrm{GCV}(\lambda) = \left[\frac{m}{m - \mathrm{df}(\lambda)}\right]\left[\frac{\mathrm{SSE}}{m - \mathrm{df}(\lambda)}\right],$$

where m is the total number of observations for all individuals,

$$\mathrm{SSE}(\lambda) = \sum_{i=1}^{N} y_i^T [I - A_i(\lambda)]^T \mathbf{W} [I - A_i(\lambda)] y_i,$$

and the degree of freedom is

$$\mathrm{dfe}(\lambda) = m - \sum_{i=1}^{N} \mathrm{trace}\left(\Phi_i (\Phi_i^T \Phi_i + \lambda R_i)^{-1} \Phi_i^T \mathbf{W}\right).$$

The gradient of GCV is

$$\frac{d\mathrm{GCV}(\lambda)}{d\lambda} = m\left[\mathrm{dfe}\frac{d\mathrm{SSE}}{d\lambda} - 2\mathrm{SSE}\frac{d\mathrm{dfe}}{d\lambda}\right]\mathrm{dfe}^{-3}.$$

where

$$\begin{aligned}\frac{d\text{dfe}(\lambda)}{d\lambda} &= -\sum_{i=1}^{N}\text{Tr}\left(\frac{dA_i}{d\lambda}\right),\\ \frac{\partial\text{SSE}(\lambda)}{\partial\lambda} &= -\sum_{i=1}^{N}y_i^T\left(\left[\frac{dA_i}{d\lambda}\right]^T\mathbf{W}[I-A_i]+[I-A_i]^T\mathbf{W}\left[\frac{dA_i}{d\lambda}\right]\right)y_i.\end{aligned}$$

The Hessian matrix of GCV(λ) is:

$$\begin{aligned}\frac{d^2\text{GCV}(\lambda)}{d\lambda^2} &= \frac{m}{\text{dfe}^2}\cdot\frac{d^2\text{SSE}}{d\lambda^2}-\frac{2m\text{SSE}}{\text{dfe}^3}\cdot\frac{d^2\text{dfe}}{d\lambda^2}\\ &+ \frac{6m\text{SSE}}{\text{dfe}^4}\cdot\left[\frac{\text{d dfe}}{d\lambda}\right]^2-\frac{4m}{\text{dfe}^3}\cdot\frac{d\text{ dfe}}{d\lambda}\cdot\frac{d\text{ SSE}}{d\lambda},\end{aligned}\tag{A.24}$$

where

$$\frac{d^2\text{dfe}}{d\lambda^2}=-\sum_{i=1}^{N}\left[\text{Tr}\left(\frac{d^2A_i}{d\lambda^2}\right)\right],\tag{A.25}$$

and

$$\begin{aligned}&\frac{d^2\text{SSE}(\lambda)}{d\lambda^2}\\ &= \sum_{i=1}^{N}\mathbf{y}_i^T\left[2\left(\frac{dA_i}{d\lambda}\right)^T\left(\frac{dA_i}{d\lambda}\right)-\left(\frac{d^2A_i}{d\lambda^2}\right)^T(I-A_i)-(I-A_i)\left(\frac{d^2A_i}{d\lambda^2}\right)^T\right]\mathbf{y}_i.\end{aligned}\tag{A.26}$$

Appendix B

Data sets used in examples

Table B.1: Drug concentrations of the 400/100 mg indinavir/ritonavir combination

Concentrations of indinavir																
Subject	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
T = 0	0.08	0.28	0.70	1.14	0.28	0.23	0.41	0.45	0.27	0.29	0.00	0.95	0.21	0.29	0.55	0.53
0.5	3.28	4.75	1.48	1.90	0.55	0.28	0.48	0.50	2.08	2.80	0.84	1.07	0.78	2.37	3.41	1.14
1	2.89	5.07	5.28	3.93	2.91	0.37	3.49	0.60	2.60	4.98	2.92	1.16	2.35	3.38	4.85	3.85
1.5	2.35	4.64	4.51	3.99	3.22	1.85	3.20	1.28	3.41	4.01	3.62	1.68	1.89	2.50	3.88	4.41
2	1.92	4.25	3.69	3.93	2.50	2.68	3.16	2.89	3.44	3.66	3.71	4.21	3.29	2.14	3.55	5.08
2.5	1.63	3.65	3.59	4.13	2.67	2.61	2.75	3.74	3.60	3.07	3.16	4.27	3.09	1.72	3.25	4.98
3	1.36	3.45	3.09	4.24	2.00	2.26	2.87	3.53	3.40	2.62	2.34	4.19	2.89	1.53	2.76	4.16
4	0.97	2.75	2.58	3.03	1.62		2.34	3.14	2.57	2.25	1.70	3.37	2.13	1.03	2.55	4.51
5	0.72	1.96	1.86	2.69	0.96	1.39	1.83	2.78	1.89	1.50	1.30	2.92	1.58	0.71	2.01	3.45
6	0.50	1.49	1.28	2.14	0.69	1.19	1.38	1.84	1.39	1.11	0.93	2.02	0.94	0.51	1.68	2.65
8	0.20	0.68	0.48	1.23	0.28	0.69	0.87	1.02	0.70	0.49	0.46	1.15	0.60	0.25	0.86	1.43
10	0.12	0.30	0.21	0.80	0.18	0.32	0.45	0.54	0.39	0.22	0.25	0.62	0.26	0.16	0.65	0.74
12		0.16	0.11	0.33	0.09	0.23	0.35	0.32	0.30	0.13	0.13	0.34	0.24	0.08	0.33	0.36

Concentrations of ritonavir																
Subject	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
T=0	0.17	0.41	0.63	0.99	0.21	0.17	0.51	0.55	0.35	0.29	0.00	0.73	0.16	0.19	0.75	0.86
0.5	0.90	1.39	1.04	0.91	0.16	0.40	0.41	0.41	2.34	0.57	1.12	1.06	0.68	0.35	0.81	2.04
1	1.56	1.49	2.53	1.06	0.37	0.72	0.64	0.37	2.19	0.90	2.03	1.14	1.39	0.67	1.72	3.64
1.5	1.19	1.46	2.11	1.07	0.53	1.29	0.59	0.47	2.06	0.85	1.90	1.23	1.03	0.65	1.81	3.07
2	1.02	1.39	2.08	1.07	0.61	1.42	0.53	0.61	1.90	1.04	1.67	1.57	1.27	0.63	1.70	2.84
2.5	0.94	1.29	2.13	1.25	0.75	1.36	0.51	0.76	2.00	0.93	1.50	1.60	1.15	0.61	1.62	2.63
3	0.86	1.26	1.66	1.44	0.71	1.25	0.61	0.80	1.93	1.00	1.26	1.43	1.12	0.62	1.52	2.36
4	0.70	1.16	1.69	1.45	0.71	1.05	1.22	1.08	1.34	1.24	1.05	1.43	1.10	0.58	1.86	2.39
5	0.65	1.31	1.32	2.35	0.49	0.85	1.07	1.01	1.07	0.99	0.91	1.37	0.96	0.47	1.75	2.31
6	0.57	1.34	1.08	1.88	0.45	0.47	0.79	0.71	0.89	0.98	0.71	0.99	0.58	0.36	1.21	2.14
8	0.29	0.49	0.51	1.27	0.19	0.27	0.69	0.54	0.47	0.48	0.46	0.58	0.39	0.21	0.72	1.19
10	0.20	0.36	0.28	0.88	0.13	0.17	0.38	0.39	0.32	0.26	0.32	0.84	0.24	0.15	0.50	0.71
12		0.29	0.21	0.29	0.09		0.41	0.34	0.22	0.19	0.24	0.50	0.18	0.13	0.30	0.67

Table B.2: Drug concentrations of the 600/100 mg indinavir/ritonavir combination

Concentrations of indinavir																
Subject	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
T=0	0.20	1.78	0.78	1.62	0.50	0.71	1.11	1.27	1.44	0.43	1.71		0.44	0.32	1.55	0.97
0.5	4.95	3.87	2.03	1.49	0.51	0.59	3.76	1.17	6.44	0.38	1.38		2.61	0.43	4.12	4.93
1	5.62	7.53	3.90	2.69	1.95	3.58	6.53	1.18	6.73	0.71	3.67		4.89	2.02	3.91	7.37
1.5	4.60	8.38	3.80	3.61	3.47	5.59	6.35	2.48	6.85	4.78	5.10		4.65	4.93	4.17	7.95
2	3.89	8.36	4.73	4.37	4.99	5.29	6.04	5.87	7.22	4.45	7.29		5.40	4.41	4.52	8.42
2.5	3.09	7.65	4.82	6.07	5.37	4.81	5.53	5.98	6.58	5.26	6.16		5.38	3.45	5.57	8.20
3	2.66	7.98	4.61	5.98	4.28	3.91	5.39	6.16	5.94	4.16	5.51		4.65	3.01	5.83	6.91
4	2.14	6.33	4.10	7.01	3.81	3.33	3.41	5.60	4.56	3.12	4.89		3.87	2.01	5.95	5.96
5	1.43	6.43	3.35	5.39	2.97	2.65	2.58	4.29	3.91	2.47	3.47		2.83	1.52	5.05	4.08
6	1.04	5.13	2.54	4.59	2.34	2.18	2.09	3.23	2.96		3.47		2.06	1.17	4.10	3.07
8	0.47	3.17	1.39	2.59	1.43	1.04	1.00	1.71	1.95	0.92	1.81		0.95	0.62	2.97	1.81
10	0.25	2.06	0.70	1.80	0.82	0.54	0.55	0.91	1.12	0.41	1.13		0.53	0.30	1.87	1.01
12	0.16	1.15	0.50	0.80	0.54	0.30	0.28	0.50	0.65	0.32	0.55		0.54	0.22	1.82	0.64

Concentrations of ritonavir																
Subject	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
T=0	0.22	1.42	0.62	0.82	0.36	0.30	0.83	0.82	1.01	0.32	0.89		0.41	0.17	0.93	0.65
0.5	0.92	1.45	0.60	0.63	0.25	0.25	0.74	0.58	2.66	0.28	0.83		0.46	0.17	1.45	1.07
1	1.57	1.51	1.05	0.81	0.40	0.44	0.83	0.49	2.76	0.37	1.34		0.55	0.32	2.37	1.13
1.5	1.39	1.59	1.13	0.86	0.43	0.60	0.71	0.58	2.55	0.91	1.87		0.61	0.62	2.54	1.11
2	1.32	1.80	1.33	0.89	0.63	0.76	0.70	0.77	2.61	0.95	2.08		0.78	0.68	2.54	1.17
2.5	1.17	1.93	1.36	1.24	0.94	0.88	0.70	0.71	2.36	1.16	2.16		0.91	0.69	2.61	1.21
3	1.07	2.08	1.33	1.31	1.00	1.05	0.76	0.72	2.16	0.90	2.20		0.97	0.78	2.55	1.19
4	0.99	2.07	1.26	2.06	1.57	1.14	1.04	0.74	1.86	0.82	2.37		1.81	0.73	2.25	1.27
5	0.85	5.04	1.25	2.79	1.57	1.32	1.14	0.79	1.68	0.66	2.42		1.68	0.73	2.18	1.11
6	0.75	3.11	1.12	2.32	1.43	1.17	1.06	0.80	1.29		1.75		1.27	0.57	1.78	1.02
8	0.39	1.55	0.77	1.37	1.06	0.49	0.69	0.62	1.03	0.33	1.16		0.70	0.42	1.43	0.80
10	0.28	2.17	0.47	1.02	0.59	0.32	0.64	0.49	0.60	0.19	0.75		0.39	0.27	1.13	0.60
12	0.20	1.72	0.63	0.49	0.31	0.19	0.32	0.42	0.36	0.16	0.45		0.37	0.21	1.04	0.86

Bibliography

Bates, D. M. and D. B. Watts (1988). *Nonlinear Regression Analysis and Its Applications*. New York: Wiley.

Beal, S. and L. Sheiner (1982). Estimating population pharmacokinetics. *CRC Critical Reviews in Biomedical Engineering 8*, 195–222.

Beal, S. L. and L. B. Sheiner (1980). The nonmem system. *Amer. Statist. 34*, 118–119.

Beal, S. L. and L. B. Sheiner (1994). *NONMEM Project Group*. San Francisco: University of California.

Cao, J. (2006). Generalized profiling method and the applications to adaptive penalized smoothing, generalized semiparametric additive models and estimating differential equations. Technical report, Department of Mathematics and Statistics, McGill University.

Davidian, M. and D. Giltinan (1995). *Nonlinear Models for Repeated Measurement Data*. New York: Chapman and Hall.

Davidian, M. and D. M. Giltinan (2003). Nonlinear models for repeated measurement data: An overview and update. *Journal of Agricultural, Biological, and Environmental Statistics 8*, 387–419.

Gabrielsson, J. and D. Weiner (2000). *Pharmacokinetic/pharmacodynamic data analysis : concepts and applications*. Stockholm, Sweden: Apotekarsocieteten.

Gelman, A., J. Carlin, H. Stern, and D. Rubin (2004). *Bayesian Data Analysis, Second Edition*. Boca Raton: Chapman and Hall/CRC Press.

Kinabo, L. and Q. McKellar (1989). Current models in pharmacokinetics: Applications in veterinary pharmacology. *Veterinary Research Communications 13*, 141–157.

Lunn, D. J., N. Best, A. Thomas, J. Wakefield, and D. Spiegelhalter (2002). Bayesian analysis of population pk/pd models: General concepts and software. *Journal of Pharmacokinetics and Pharmacodynamics 29*, 271–307.

Petzold, L. R. (1983). Automatic selection of methods for solving stiff and and nonstiff systems of ordinary differential equations. *Siam J.Sci.Stat.Comput. 4*, 136C148.

Pillai, G. C., F. Mentre, and J.-L. Steimer (2005). Non-linear mixed effects modelling - from methodology and software development to driving implementation in drug development science. *Journal of Pharmacokinetics and Pharmacodynamics 32*, 161C183.

Pinheiro, J. and D. Bates (2000). *Mixed-Effects Models in S and Splus*. New York: Springer.

PKBugs (2004). *PKBugs: An Efficient Interface for Population PK/PD within WinBUGS*. London: Imperial College School of Medicine: http://www.winbugs-development.org.uk/pkbugs/home.html.

Ramsay, J., G. Hooker, D. Campbell, and J. Cao (2007). Parameter estimation for differential equations: A generalized smoothing approach. *Journal of the Royal Statistical Society (with discussion)*.

Ramsay, J. O. and B. W. Silverman (2002). *Functional Data Analysis* (First ed.). New York: Springer.

Ramsay, J. O. and B. W. Silverman (2005). *Functional Data Analysis* (Second ed.). New York: Springer.

Roe, D. J. (1997). Comparision of population pharmacokinetic modeling methods using simulated data: Results from the population modeling workgroup. *Statistics in Medicine 16*, 1241C1262.

Seber, G. and C. Wild (1988). *Nonlinear regression*. United States of America: John Wiley & Sons, Inc.

Setzer, R. W. (2003). *The odesolve Package. Solvers for Ordinary Differential Equations*. http://www.cran.r-project.org.

Sheiner, L. and S. Beal (1980). Evaluation of methods for estimating population pharmacokinetic paramters. i. michaelis-menten model: Routine linical pharmacokinetic data. *Journal of Pharmacokinetics and Biopharmaceutics 8*, 553–571.

Sheiner, L. and T. Ludden (1992). Population pharmacokinetics/dynamics. *Annual Review of Pharmacology and Toxicology 32*, 185–209.

Sheiner, L., B. Rosenberg, and V. Marathe (1977). Estimation of population characteristics of pharmacokinetic parameters from routine clinical data. *Journal of Pharmacokinetics and Biopharmaceutics 5*, 445–479.

Sheiner, L., B. Rosenberg, and K. Melmon (1972). Modelling of individual pharmacokinetics for computer-aided drug dosage. *Computers and Biomedical Research 5*, 441–459.

Sheiner, L. and J. Wakefield (1999). Population modelling in drug development. *Statistical Methods in Medical Research 8*, 183–193.

Steimer, J., A. Mallet, J. Golmard, and J. Boisvieux (1984). Alternative approaches to estimation of population pharmacokinetic parameters: Comparison with the nonlinear mixed effect model. *Drug Metabolism Reviews 15*, 265–292.

the Food and Drug Administration (1999). *Guidance for Industry: Population Pharmacokinetics*. http://www.fda.gov.

Tornϕ, C. W., H. Agersϕ, E. N. Jonsson, H. Madsen, and H. A. Nielsen (2004). Non-linear mixed-effects pharmacokinetic/pharmacodynamic modelling in nlme using differential equations. *Comput. Methods Programs Biomed. 76*.

Wakefield, J. (1996). The bayesian analysis of population pharmacokinetic models. *Journal of the American Statistical Association 91*, 62–75.

Wasmuth, J.-C., C. J. la Porte, K. Schneider, D. M. Burger, and J. K. Rockstroh (2004). Comparison of two reduced-dose regimens of indinavir (600 mg vs 400 mg twice daily) and ritonavir (100 mg twice daily) in healthy volunteers (coredir). *International Medical Press 2*, 1359–6535.

Wolfinger, R. (1993). Laplace's approximation for nonlinear mixed models. *Biometrika 80*, 791–795.