SEMIPARAMETRIC MIXED-EFFECTS ANALYSIS ON PK/PD MODELS USING DIFFERENTIAL EQUATIONS

by

Yi Wang

A DISSERTATION

Presented to the Faculty of

The Graduate College at the University of Nebraska

In Partial Fulfillment of Requirements

For the Degree of Doctor of Philosophy

Major: Statistics

Under the Supervision of Professors Kent M. Eskridge and Shunpu Zhang

Lincoln, Nebraska

December, 2007

UMI Number: 3288807

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

SEMIPARAMETRIC MIXED-EFFECTS ANALYSIS ON PK/PD MODELS USING DIFFERENTIAL EQUATIONS

Yi Wang, Ph.D.

University of Nebraska, 2007

Advisers: Kent M. Eskridge and Shunpu Zhang

This dissertation consists of three papers written on semiparametric mixed-effects analysis in the presence of structural model misspecification. In the first paper, we develop a new semiparametric modeling approach to address potential structural model misspecification. Specifically, we use a set of ordinary differential equations (ODEs) with form $\dot{x} = A(t)x + B(t)$ where $B(t)$ is a nonparametric function vector estimated using penalized splines. The inclusion of nonparametric functions in ODEs makes identification of structural model misspecification feasible by quantifying the model uncertainty and provides flexibility for accommodating possible structural model deficiencies. The resulting model will be implemented in a nonlinear mixed-effects modeling setup for population analysis. We illustrate the method with an application to cefamandole data and evaluate its performance through simulations.

The second paper proposes an alternative method to the stochastic differential equation (SDE)-based method in developing a proper parametric ODE model. In SDE a Gaussian diffusion term is introduced into an ODE to quantify the system noise. However, in our proposed method $B(t)$ in a system of ODE with form $\dot{x} = A(t)x + B(t)$ is assumed to be a nonparametric function, which constructs a quantitative measure of model uncertainty so that information on the proper model structure can be derived

directly from data. By means of the two examples with simulated data, we find that our method, which is a data-driven, distribution-free and intensive computation-free method, can perform model diagnostic and provide a basis for systematic model development similar to the SDE-based method.

In the third paper, Bayesian analysis is provided for fitting a spline-enhanced ODE mixed model. Since the Lindstrom and Bates' estimation algorithm used in the first paper could perform poorly when the degree of nonlinearity is high, Bayesian analysis without linearization approximation needed is a good alternative for fitting a spline-enhanced ODE mixed model. We illustrate and evaluate Bayesian analysis on our proposed model with an application to cefamandole data and a Ditropan study using simulated data.

ACKNOWLEGEMENTS

I would like to thank my advisors Dr. Kent M. Eskridge and Dr. Shunpu Zhang for their guidance, patience and many helpful suggestions. Their support has genuinely enhanced my education. I would also like to thank Dr. David Skoug, Dr. Stephen Kachman and Dr. Dong Wang for serving on my PhD. supervisory committee. I dedicate this work to my wife, my parents, and our family who provided support and encouragement.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

1.1. Background Information

1.1.1. Parametric Nonlinear Mixed-Effects Models

Compartmental analysis is a method of biomathematical modeling which assumes that a biological system can be divided into a series of homogeneous compartments where the compartments interact by exchanging material. It has been widely used to analyze population pharmacokinetic data. Pharmacokinetics is the study of the time course of a drug and its metabolites following its introduction into the body. Population PK studies consider the pharmacokinetics of a number of individuals and are becoming increasingly important as an aid to drug development. Population data are generated by observing a number of individuals repeatedly under differing experimental conditions where the individuals are assumed to constitute a random sample from a population of interest. For the reader familiar with split-plot designs, each individual can be thought of as a whole plot and each observation as a subplot. However, unlike split-plot models, models used for population analysis usually include an underlying "functional" relationship between at least one of the predictor variables and the observations within individuals. The data from such studies typically consist of dose histories, drug concentrations with associated sampling times, and often covariates measurements such as the age and weight of each subject. These studies aim to provide an understanding of the pharmacokinetics of the drug in question, by identifying sources of and quantifying the remaining intra- and inter-individual variability, and so lead to an informed choice of dosage regime.

Models including random effects which account for the inter-individual variability enjoy an increasing popularity. In these models it is assumed that all responses follow a similar functional form, but with parameters that vary among individuals. Mixed-effects modeling is the most used method for analysis of population PK/PD data. Since in the mixed-effects models, intra-individual variability characterizes the error between the true and the observed drug concentrations, while inter-individual variability deals with the differences between the PK model parameters of each individual, this will often lead to more precise estimation of population parameters. However, in compartmental analysis continuous biological processes are often described by systems of ordinary differential equations (ODE), which leads to pharmacokinetic models that are nonlinear in the parameters and not straightforward to estimate. A number of inferential methods for the nonlinear mixed effects model have been developed since the pioneering work of Beal and Sheiner (1982).

As an example to illustrate the use of the nonlinear mixed-effects modeling, we consider the cefamandole data, which was first used in Aziz et al. (1977). Figure 1.1 shows data from a pilot study to investigate the pharmacokinetics of cefamandole, a potentially useful cephalosporin antibiotic with an extended spectrum of activity. In this experiment, a dose of 15mg/kg body weight of cefamandole was administered by ten-minute intravenous infusions to six healthy male volunteers. Blood samples were collected from each subject at each of 14 time points post-dose. Drug concentrations in plasma were determined for each sample by high performance liquid chromatography (HPLC). Aziz et al. (1977) suggested that cefamandole followed a two-compartment body kinetics rather than one-compartment, and using the nonlinear least-squares method

the data were fitted to the biexponential equation $C_t = \beta_1 e^{-\beta_2 t} + \beta_3 e^{-\beta_4 t}$ with $\beta_2 > \beta_4 > 0$ for identifiability of the model. This equation resulted from the assumption that cefamandole kinetics was best described by the following first-order differential equations system,

$$\begin{cases} x' = -(k_{12} + k_{10})x(t) + k_{21}y(t), & x(0) = D \\ y' = k_{12}x(t) - k_{21}y(t) \quad , & y(0) = 0 \end{cases}$$

where x and y are drug concentration of blood system and tissue system at time t after the initial dose D administered, respectively. The fitted line is shown in Figure 1.2 with the goodness of fit R^2=0.917. Nevertheless, standard nonlinear regression analysis for cefamandole data is likely not appropriate, since the analysis ignores the correlation between repeated measures pertaining to the same subject. Another drawback of using a standard nonlinear regression model with multi-subject data is that it prevents us from understanding the true structure of the data and from considering different sources of variability among and within individuals, which is precisely what mixed-effects models are designed to do.

The advantage of using nonlinear mixed-effects models (NLMMs) is that population mean parameters are obtained simultaneously with estimates of the inter- and intra-individual variability. Nonlinear mixed-effects models can be thought of as a hierarchical model structure where the variability in response is split into inter- and intra-individual variability. The hierarchical model consists of the following two stage hierarchy. At the first-stage of modeling, the NLMM assumes a common parametric function that relates the conditional mean of the response variable to the independent variables and a set of individual parameters. That is,

$$C_{ij} = x(\phi_i; t_{ij}) + \varepsilon_{ij}, \ \varepsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2),$$

where x is the solution of the differential equations for cefamandole data while ϕ is an 4-dimensional vector of unknown model parameters. At the second-stage of modeling, the parameters of the different individuals are modeled with fixed and random effects, and various variance-covariance structures are incorporated into the second stage model, namely,

$$\phi_i = \begin{bmatrix} lk_{10i} \\ lk_{12i} \\ lk_{21i} \\ lD_i \end{bmatrix} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix} + \begin{bmatrix} b_{1i} \\ b_{2i} \\ b_{3i} \\ b_{4i} \end{bmatrix} = \beta + b_i ,$$

where lk_{10}=log(k_{10}), lk_{12}=log(k_{12}), lk_{21}=log(k_{21}), and lD=log(D). For the model to be meaningful, all four parameters must be positive. To ensure positive estimates while keeping the optimization problem unconstrained, we reparameterize the model listed above in terms of the logarithm. The pharmacokinetic parameters ϕ_i are allowed to vary with subject. The fixed effects β represent the mean values of the parameters ϕ_i in the population and the random effects b_i represent the deviations of the ϕ_i from their mean values. Fitting the nonlinear mixed-effects model to the cefamandole data is done by running the "nlmeODE" function in R, which combines "odesolve" with the "nlme" function in R to handle the first-order ODE's. The computational algorithm used in nlmeODE will be elaborated in the literature review section. A graphical assessment of the adequacy of the cefamandole model is given in Figure 1.3, which shows both the population predictions (corresponding to random effects equal to zero) and the within-group predictions (obtained using the estimated random effects). Note that the within-group predictions are in close agreement with the observed concentrations, illustrating that the NLME model can accommodate individual effects.

1.1.2. Nonparametric Models

Linear regression analysis is one of the most commonly used techniques in statistics. The aim of the analysis is to explore the association between dependent and independent variables, to assess the contribution of the independent variables and to identify their impact on the dependent variable. However, there are many data sets such as cefamandole data where it is clearly inappropriate to fit a straight line model and where a model of the form $y_i = f(x_i) + \varepsilon_i$ is called for, where f is a curve of some sort. The classical approach is to use a low order polynomial for f, the coefficients of which are estimated by least squares. While this approach has been widely used, polynomial regression suffers from various drawbacks. One of these is that individual observations can exert an influence, in unexpected ways, on remote parts of the curve. Another difficulty is that improved model fit by increasing the polynomial degree is achieved in discrete steps and cannot be controlled continuously. A third point is that polynomial functions do not allow the data to determine the fitted model in a very flexible way. We now consider an example in which polynomial regression is not very predictive. LIDAR (abbreviated from light detection and ranging) is a technique that uses the reflection of laser-emitted light to detect chemical compounds in the atmosphere. The LIDAR data have been used for illustrations of nonparametric regression methods by Ruppert et al. (2003). In Figure 1.4 the relationship between range and logratio is nonlinear, where range is the distance traveled before the light is reflected back to its source while logratio is the logarithm of the ratio of received light from two laser sources. Figure 1.5 depicts a degree 10 polynomial model fit to the LIDAR data, which generally fit the data well. But wriggles from fitting high-degree polynomials are a big problem with the LIDAR data,

since the first derivative of the fitted function and itself are the interest of study. In fact, it seems that the relationship between logratio and range is decreasing and therefore the derivative should be nonpositive. However, wriggles could make the derivatives of the polynomial fit in Figure 1.5 quite large and positive, for example those at the beginning and end of range.

As with the LIDAR data, many data sets should not be modeled using traditional parametric techniques. For example, many of the problems described in Ruppert et al. (2003) involve nonlinear effects that are difficult to model parametrically. In these cases, nonparametric modeling techniques are required. Otherwise, when the parametric model does not fit the data well, the resulting estimates may be biased. To overcome the difficulty caused by the restrictive assumption of a parametric form of the regression function, one may relax the assumptions on parametric forms. This approach leads to nonparametric regression. In nonparametric regression, the function f is some unspecified smooth function that needs to be estimated from the (x_i, y_i). The method that we focus on, penalized splines, has the attractiveness of being a relatively straightforward extension of linear regression modeling. The unknown smooth function could be modeled using natural cubic splines, B-splines, truncated polynomials, radial splines etc. Various degrees of the splines, bases, and penalties can be used in penalized splines' fitting. As discussed in Ruppert et al. (2003), penalized splines can be viewed as best linear unbiased predictions (BLUPs) in a mixed model framework, which allows us to make systematic inference on all model components within a unified parametric mixed model framework and can be easily implemented by fitting a working linear mixed model using existing statistical software. This provides a foundation for the joint estimation procedure

for the variance and smoothing parameters and the amount of smoothing is chosen automatically in the mixed model representation of smoothers. For example, using penalized linear spline smoothing the underlying model for $f(x)$ is the mixed model

$$f(x) = \alpha_0 + \alpha_1 x + \sum_{k=1}^{K} u_k (x - k_k)_+ = T\alpha + Zu ,$$

where k_k are knots; $(x - k_k)_+$ is the spline basis function called the positive part of the function $x - k_k$; α and u are the coefficients of the polynomial functions and truncated line functions, respectively. Corresponding to these two vectors, define

$$T = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \text{ and } Z = \begin{bmatrix} (x_1 - k_1)_+ & \cdots & (x_1 - k_k)_+ \\ \vdots & \ddots & \vdots \\ (x_n - k_1)_+ & \cdots & (x_n - k_k)_+ \end{bmatrix}$$

By treating the u as a set of random coefficients distributed normally with mean 0 and $\text{cov}(u) = (\lambda\sigma_\varepsilon^2)^{-1} I = \sigma_u^2 I$, the mixed model representation of penalized linear spline smoothers allows for automatic fitting using the R linear mixed model function lme(). Smoothing parameter selection can be done via restricted maximum likelihood (REML) and $\hat{f}(x)$ can be obtained via estimated best linear unbiased prediction (EBLUP) (e.g. Robinson, 1991). Figure 1.6 shows the linear spline regression fit to LIDAR data using 35 knots equally selected among the range under the likelihood approach, which provides a good fit without overfitting.

The basic idea of nonparametric regression techniques is to remove the restriction that the regression function belongs to a parametric family and let the data determine the most suitable form of the functions. However, nonparametric and parametric techniques should not be regarded as competitors, instead they complement each other. In some situations, nonparametric techniques can be used to validate or suggest a parametric

model. A combination of both nonparametric and parametric methods is more powerful than any single method in many practical applications.

1.2 Problem Statement

The nonlinear mixed effects model has gained broad acceptance as a suitable framework for data in the form of continuous, repeated measurements on each of a number of individuals, and analyses based on this model can be used to characterize pharmacological processes within the body that dictate the time-concentration relationship for individual subjects and the population of subjects. It is no wonder that population PK/PD data analysis using NLMMs has become an increasingly important tool in drug development. Nevertheless, extensions and modifications of the model to handle new substantive challenges continue to emerge. Note that the model functional form in a NLMM is assumed to be known, the parametric NLMM may be too restrictive and the resulting model may or may not be appropriate for a dataset, since the variability is most often not predictable owing to the governing processes not being fully understood or too complex to model deterministically. For example, the variability from the true model may be somewhere in between that suggested by a two-compartment model and a three-compartment model. In such circumstances, the problem of model misspecification arises. As the popularity of NLMMs increase, more sophisticated methods and diagnostic techniques are needed to handle models with structural model misspecification.

Continuous biological processes are often described by systems of ordinary differential equations (ODEs), which leads to pharmacokinetic models that are nonlinear in the parameters and difficult to model parametrically. There is a clear imperative to be able to handle such nonlinear relationships effectively through more flexible techniques.

It is hence of substantial interest to develop a semiparametric model for pharmaceutical data by incorporating nonparametric functions into NLMMs. Due to flexibility of the nonparametric functions, this will allow more flexible functional dependence of the outcome variable on the covariates so that the structural model deficiencies incurred by model misspecification can be overcome.

This research is motivated by the works of Ke and Wang (2001) and Li et al. (2002). Ke and Wang (2001) propose a generalization of the nonlinear mixed-effects model where the model function is allowed to depend on a completely unspecified function of time and elements of model parameters. The authors suggest that their model provides flexibility for accommodating possible model misspecification and may be used as a diagnostic tool for assessing the form of time dependence in a fully parametric nonlinear mixed model. However, as commented upon by Bates et al. (2001), one of the widest areas of applications of NLMMs is in population pharmacokinetics, and most pharmacokinetic or pharmacodynamic models, especially those that handle multiple dosing, cannot be expressed in the form of semiparametric nonlinear mixed-effects models (SNMMs) of Ke and Wang (2001). Nevertheless, SNMMs provide a unified framework for a large class of models and bring together the advantages of nonlinear mixed effects, smoothing splines, and self-modeling to cover a wide class of problems. Li et al. (2002) study related methods in the context of pharmacokinetic analysis and extend such models to allow for some time-dependent PK parameters to be nonparametric functions of time using splines among the three dosing periods. Although the spline enhanced pharmacokinetic model (SEPK) used by Li et al. is able to quantify the drug concentration over time and describe time-dependent behavior for the drug absorption

and disposition processes, the method they used is rather limited. In their proposed model the PK parameters are assumed to be time dependent, which can be used to address the cases with parameter misspecification but cannot be used to fix the problems with structural model misspecification. In fact, for most PK/PD data analyses structural model misspecification occurs more often than parameter misspecification. Additionally, the model with time-dependent PK/PD rates is hard to identify and interpret.

We propose a new method designed for PK/PD models with potential structural model misspecification, which hasn't been addressed by Ke and Wang (2001) and Li et al. (2002). In contrast to the deterministic ODEs with form $\dot{x} = A(t)x + B(t)$ commonly used in PK/PD modeling, in our proposed method $B(t)$ is treated as a nonparametric function vector that will be decided by data and estimated using penalized splines. The inclusion of splines, which constitutes the primary difference between a spline-enhanced ODE model and the corresponding ODE model, allows the spline-enhanced ODE model to explain a larger portion of the variation in a given data set than a parametric model. This setup makes identification of structural model misspecification feasible by quantifying the model uncertainty and provides flexibility for accommodating possible structural model deficiencies. The resulting model will be implemented in a nonlinear mixed-effects modeling setup for population analysis. Overall, it is expected that the proposed spline-enhanced ODE model will have the following advantages: (1) the method extends the SNMMs by allowing the mean response to depend nonlinearly on both the parameters and the nonparametric functions; (2) the approach can be used to identify possible structural model misspecification by quantifying the model uncertainty; (3) it overcomes possible structural model deficiencies incurred by unidentified variation

in a given data set; (4) this setup can be used for PK/PD model development; (5) the estimation of the spline-enhanced ODE model is unified within a parametric mixed model framework and can be easily implemented in nlmeODE in R, which could potentially boost the use of the spline-enhanced ODE model within PK/PD.

Nevertheless, estimation and inference are challenging even if the existing software can be used for implementation of the proposed model. As the complexity of ODE systems increases and more nonparametric functions are introduced into the model, the estimates of variance components could become more and more biased and inaccurate, since linearization required by the Lindstrom and Bates' algorithm in nlmeODE in R is not a good approximation for functions with a high degree of nonlinearity (Bennett et al. 1996). We have noticed that inaccurate estimates of variance components particularly those associated with the splines may cause serious problems with smoothing, since the smoothing parameters are treated as variance components in our approach and undoubtedly the quality of smoothing heavily relies on the precision of estimation of the corresponding variance components. Therefore, more efficient algorithms developed for fitting the spline-enhanced ODE model will be preferred. Bayesian analysis without resorting to any approximation is a good alternative, and the implementation of Bayesian analysis using the spline-enhanced ODE model can be conducted within a unified Bayesian parametric mixed model framework.

1.3 Research Objectives

The following research objectives will be addressed in this study:

1. To develop a semiparametric nonlinear mixed-effects model called a spline-enhanced ODE mixed model that provides flexibility for accommodating possible model misspecification in population PK/PD analysis;
2. To illustrate how a spline-enhanced ODEs model is formulated and implemented in a nonlinear mixed-effects modeling setup;
3. To demonstrate its capability to account for the model uncertainty due to structural model misspecification by its application to some PK data examples and evaluate its performance through simulation;
4. To study how the proposed method can provide the basis for systematic model development by performing model diagnostics;
5. To provide Bayesian analysis on the proposed model and conduct a comparison on the estimation of the population parameters between classic and Bayesian analyses.