Flexible Modeling Tools for Continuous Longitudinal Data

Proefschrift voorgelegd tot het behalen van de graad van
Doctor in de Wetenschappen, richting Wiskunde
te verdedigen door

JAN SERROYEN

Promotor: Prof. dr. Geert Molenberghs

Contents

1

Introduction

In a longitudinal study individuals are measured repeatedly over time. Since observations gathered on the same subject tend to be more alike than observations from different subjects, they are *correlated*. This correlation must be taken into account when analyzing the collected data in order to draw valid conclusions. Although this fact has been long recognized, it was not until quite recently that the wide availability and advances in computer power have permitted the development of appropriate statistical techniques for the analysis of correlated measurements.

The prime advantage of a longitudinal study is its effectiveness for studying change (Diggle *et al* 1994). Namely, it can distinguish changes over time within individuals (longitudinal effects) from differences among people at baseline (cross-sectional effects).

In addition to longitudinal data, other types of correlated data arise naturally in a broad range of scientific disciplines, such as biological, health or social sciences. Correlated data is actually a generic term, covering a multitude of data structures, such as multivariate observations, clustered data, and spatially correlated data.

It is essential to consider the type of outcome when modeling correlated data. Methods for continuous correlated data are well developed and implemented in software, mainly due to the elegant properties of the normal distribution. The general linear mixed-effects model (Verbeke and Molenberghs 2000) has, arguably, become the most commonly used tool for analyzing continuous, normally distributed longitu-

dinal data. This model encompasses three sources of variability: (1) subject-specific effects, (2) serial correlation, which describes the phenomenon that measurements taken closer in time will exhibit a larger correlation than when they are further apart, and (3) measurement error.

When the outcome variable is discrete or categorical, techniques for correlated data are less standard, because of the lack of a discrete analogue to the normal distribution. Therefore, complete specification of the joint distribution of the response vector becomes more problematic and fully likelihood-based methods are generally awkward.

Another issue raised by this type of (non-Gaussian) outcomes is that we must distinguish between three different model families: the marginal, random-effects, or conditional model family (Molenberghs and Verbeke 2005). In a *marginal model*, the marginal expectation of a discrete or continuous response, Y, is characterized as a function of explanatory variables, X. By marginal expectation, we mean the average response over the the sub-population that shares a common value of X. The correlation among the components of Y can then be captured either by adopting a fully parametric approach or by modeling a limited number of lower-order moments only. Alternatively, in *random-effect models*, the response is assumed to be a function of explanatory variables with regression coefficients that vary from one individual to the next. This variability reflects natural heterogeneity due to unmeasured *(latent)* factors, often called random effects. Given these so-called subject-specific parameters, the responses are often assumed independent. This does not preclude that more elaborate models are possible if residual dependence is detected. Finally, a *conditional model* describes the distribution of the components of Y, conditional on X but also on (a subset of) the other components of Y. In a longitudinal context, a particular relevant class of conditional models describes a component of Y given the ones recorded earlier in time, the so-called *autoregressive* or *transition models*.

All three modeling approaches lead to the same class of linear models for Gaussian data. But in the discrete case, different (non-linear) models can lead to different interpretations for the regression coefficients. The choice of model should therefore depend on the scientific question being addressed.

Throughout the present work, we will focus mainly on flexible ways to model both linear and non-linear continuous longitudinal data. In Chapter 2, a general overview of existing relevant methodology is presented. This includes the linear mixed model mentioned earlier, but also the non-linear mixed model, fractional polynomials, splines, and finally a brief section on discrimination, classification, and clustering.

Chapter 3 covers the analysis of a pilot study on *Persistent disturbing behavior* (PDB), which refers to a chronic condition in therapy resistant psychiatric patients. Since these patients are highly unstable and difficult to maintain in their natural living environment and even in hospital wards, it is important to properly characterize this group. Here, using a score calculated from longitudinal psychiatric registration data, we characterize the difference between PDB patients and a set of control patients. These differences are studied both at a given point in time, using discriminant analysis, as well as in terms of the evolution of the score over time, using longitudinal data analysis methods. Further, using clustering techniques, the group of PDB patients is split into subgroups. Such findings are useful from a scientific as well as from an organizational point of view.

While marginal models, random-effects models, and conditional models are routinely considered to be the three main modeling families for continuous and discrete repeated measures with linear and generalized linear mean structures, respectively, it is less common to consider the taxonomy for non-linear models. In the latter situation, indeed, the focus is often exclusively on random-effects models. In Chapter 4, we consider all three families and apply them to a simple but illustrative set of data on tree circumference growth of orange trees.

In Chapter 5 we analyze data on the impact of testosterone on the dynamics of Mn^{2+} accumulation measured by magnetic resonance imaging in three songbird brain areas: the nucleus robustus arcopallii (RA), area X, and the high vocal center (HVC). Birds with and without testosterone were included in the experiment, and repeated measurements were available in both a pre and post drug administration period. We formulate a non-linear modeling strategy, allowing for the incorporation of (1) within-bird correlation, (2) the non-linearity of the profiles, and (3) the effect of treatment. For two of the outcomes (RA and area X), biological theory suggests a parametric form, while for HVC this is not the case. Since the HVC outcome bears some resemblance with the two-compartment model known from pharmacokinetics, this model was considered a sensible choice. We use a different model, based on fractional polynomials, as a sensitivity analysis for the latter. All methods used provide good fits to the data, confirm results from previous, simple analyzes undertaken in the literature, but were able to detect additional effects of treatment that had so far gone undetected. The fractional polynomial and two-compartment models provide similar substantive conclusions, the two together can be seen as a form of sensitivity analysis.

Flexible estimation of serial correlation when modeling continuous longitudinal data is the focus of Chapter 6. As mentioned earlier, four structures can be distin-

guished in the general linear mixed model: fixed effects, random effects, measurement error and serial correlation. Broadly speaking, serial correlation captures the phenomenon that the correlation structure within a subject depends on the time lag between two measurements. While the general linear mixed model is rather flexible, the need has arisen to further increase flexibility. In response, quite some work has been done to relax the model assumptions and/or to extend the model. For example, the normality assumption for the random effects has been generalized in several ways. Comparatively less work has been devoted to more flexible serial correlation structures. Therefore, we propose the use of spline-based modeling of the serial correlation function. The approach is applied to data from a pre-clinical experiment in dementia which studied the eating and drinking behavior in mice.

Chapter 7 deals with viral dynamics, a relatively new field of study that relies on mathematical models to describe the evolution of virus levels in the blood plasma, the so-called viral load, over time. Our scientific aim is to model the evolution of HIV-1 viral concentrations over time. More specifically, the goal of the analysis reported in this chapter is twofold. First, the model building steps and specification of starting values need to be systematized when fitting the bi-exponential model to viral load data with SAS procedure NLMIXED. Second, we wanted to find a suitable model for the *rebounders*, a special subgroup of patients. The data to be analyzed come from pooling three clinical trials on Prezista, a new protease inhibitor developed by the pharmaceutical company Tibotec.

Finally, in Chapter 8 we formulate concluding remarks and also indicate some topics for further research.

2

Methodology

This chapter provides a general overview of relevant existing methodology to flexibly model continuous longitudinal data. Section 2.1 introduces the linear mixed model, while the non-linear mixed model is covered in Section 2.2. Section 2.3 is devoted to fractional polynomials and Section 2.4 introduces splines. Finally, a brief discussion on discrimination, classification, and clustering is presented in Section 2.5.

2.1 The Linear Mixed Model

The linear mixed-effects model (Laird and Ware 1982, Verbeke and Molenberghs 2000) is very commonly used with continuous longitudinal data. The model will be introduced and briefly discussed.

Let $\boldsymbol{Y}_i$ denote the n_i-dimensional vector of measurements available for subject $i = 1, \ldots, N$. A general linear mixed model decomposes $\boldsymbol{Y}_i$ as:

$$\boldsymbol{Y}_i = X_i\boldsymbol{\beta} + Z_i\boldsymbol{b}_i + \boldsymbol{\varepsilon_i}, \tag{2.1}$$

in which $\boldsymbol{\beta}$ is a vector of population-average regression coefficients called fixed effects, and where $\boldsymbol{b}_i$ is a vector of subject-specific regression coefficients. The $\boldsymbol{b}_i$ describe how the evolution of the ith subject deviates from the average evolution in the population. The matrices X_i and Z_i are $(n_i \times p)$ and $(n_i \times q)$ matrices of known covariates. The random effects $\boldsymbol{b}_i$ and residual components $\boldsymbol{\varepsilon_i}$ are assumed to be independent with

distributions $N(\mathbf{0}, D)$, and $N(\mathbf{0}, \Sigma_i)$, respectively. Note that Σ_i depends on i only dimension-wise, i.e., through the number of measurements available for a particular subject. In other words, the parameters governing Σ_i are generally common to all subjects. Thus, in summary,

$$\boldsymbol{Y}_i|\boldsymbol{b}_i \sim N(X_i\boldsymbol{\beta} + Z_i\boldsymbol{b}_i, \Sigma_i), \qquad \boldsymbol{b}_i \sim N(0, D). \tag{2.2}$$

Let $f(\boldsymbol{Y}_i|\boldsymbol{b}_i)$ and $f(\boldsymbol{b}_i)$ be the density functions of $\boldsymbol{Y}_i$ conditional on $\boldsymbol{b}_i$, and of $\boldsymbol{b}_i$, respectively. The marginal density function of $\boldsymbol{Y}_i$ is then given by

$$f(\boldsymbol{Y}_i) = \int f(\boldsymbol{Y}_i|\boldsymbol{b}_i)\ f(\boldsymbol{b}_i)\ d\boldsymbol{b}_i, \tag{2.3}$$

the density of the n_i-dimensional normal distribution $N(X_i\boldsymbol{\beta}, Z_iDZ_i' + \Sigma_i)$. Further, let $\boldsymbol{\alpha}$ denote the vector of all variance and covariance parameters (usually called variance components) found in $V_i = Z_iDZ_i' + \Sigma_i$, that is, $\boldsymbol{\alpha}$ consists of the $q(q+1)/2$ different elements in D and of all parameters in Σ_i. Finally, let $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\alpha}')$ be the s-dimensional vector of all parameters in the marginal model for $\boldsymbol{Y}_i$.

Oftentimes, Σ_i in model (2.2) is chosen to be equal to $\sigma^2 I_{n_i}$ where I_{n_i} denotes the identity matrix of dimension n_i. We then call this model the conditional independence model, since it implies that the n_i responses on individual i are independent, conditional on $\boldsymbol{b}_i$ and $\boldsymbol{\beta}$. This model may imply unrealistically simple covariance structures for the response vector $\boldsymbol{Y}_i$, especially for models with few random effects. The covariance assumptions can often be relaxed by allowing an appropriate, more general, residual covariance structure Σ_i for the vector $\boldsymbol{\varepsilon_i}$ of subject-specific error components.

Diggle *et al* (2002), based on Diggle (1988), proposed such a general model. They assume that $\boldsymbol{\varepsilon_i}$ has constant variance and can be decomposed as $\boldsymbol{\varepsilon_i} = \boldsymbol{\varepsilon_{(1)i}} + \boldsymbol{\varepsilon_{(2)i}}$ in which $\boldsymbol{\varepsilon_{(2)i}}$ is a component of serial correlation, suggesting that at least part of an individual's observed profile is a response to time-varying stochastic processes operating within that individual. This type of random variation results in a correlation between serial measurements, which is usually, and quite sensibly, a decreasing function of the time separation between these measurements. Further, $\boldsymbol{\varepsilon_{(1)i}}$ is an extra component of measurement error reflecting variation added by the measurement process itself, and assumed to be independent of $\boldsymbol{\varepsilon_{(2)i}}$.

The resulting linear mixed model can now be written as

$$\begin{cases} \boldsymbol{Y}_i = X_i\boldsymbol{\beta} + Z_i\boldsymbol{b}_i + \boldsymbol{\varepsilon_{(1)i}} + \boldsymbol{\varepsilon_{(2)i}} \\ \\ \boldsymbol{b}_i \sim N(\mathbf{0}, D), \\ \\ \boldsymbol{\varepsilon_{(1)i}} \sim N(\mathbf{0}, \sigma^2 I_{n_i}), \\ \\ \boldsymbol{\varepsilon_{(2)i}} \sim N(\mathbf{0}, \tau^2 H_i), \\ \\ \boldsymbol{b_1}, \ldots, \boldsymbol{b}_{\mathrm{N}}, \boldsymbol{\varepsilon}_{(1)1}, \ldots, \boldsymbol{\varepsilon}_{(1)\mathrm{N}}, \boldsymbol{\varepsilon}_{(2)1}, \ldots, \boldsymbol{\varepsilon}_{(2)\mathrm{N}} \text{ independent}, \end{cases} \tag{2.4}$$

and the model is completed by assuming a specific structure for the $(n_i \times n_i)$ correlation matrix H_i. One usually assumes that the serial effect $\boldsymbol{\varepsilon_{(2)i}}$ is a population phenomenon, independent of the individual. The serial correlation matrix H_i then only depends on i through the number of n_i observations and the time points t_{ij} at which measurements were taken. Further, it is assumed that the $(j,\, k)$ element h_{ijk} of H_i is modeled as

$$h_{ijk} = g(|t_{ij} - t_{ik}|) \tag{2.5}$$

for some decreasing function $g(\cdot)$ with $g(0) = 1$. This means that the correlation between the measurements $\boldsymbol{\varepsilon}_{(1)ij}$ and $\boldsymbol{\varepsilon}_{(2)ik}$ only depends on the time interval between the measurements y_{ij} and y_{ik}, and decreases if the length of this interval increases.

Two frequently used $g(\cdot)$ functions are the exponential and Gaussian serial correlation functions defined as $g(u) = \exp(-\phi u)$ and $g(u) = \exp(-\phi u^2)$, respectively $(\phi > 0)$.

The marginal covariance matrix is then of the form

$$V_i = Z_i D Z_i' + \tau^2 H_i + \sigma^2 I_{n_i}. \tag{2.6}$$

A classical inferential approach is based on maximizing the marginal likelihood function

$$L_{\mathrm{ML}}(\boldsymbol{\theta}) = \prod_{i=1}^{N} \Big\{ (2\pi)^{-n_i/2} \, |V_i(\boldsymbol{\alpha})|^{-1/2} \times \exp\left(-\frac{1}{2} (\boldsymbol{Y}_i - X_i\beta)' \, V_i^{-1}(\boldsymbol{\alpha}) \, (\boldsymbol{Y}_i - X_i\beta) \right) \Big\} \tag{2.7}$$

with respect to $\boldsymbol{\theta}$. Alternatively, and with an eye on reducing small-sample likelihood-type bias, restricted maximum likelihood (REML, Harville 1974, 1977, Molenberghs and Verbeke 2000) can be used, which comes down to the maximization of the so-called REML likelihood

$$L_{\text{REML}}(\boldsymbol{\theta}) = \left| \sum_{i=1}^{N} X_i' \, V_i^{-1}(\boldsymbol{\alpha}) \, X_i \right|^{-1/2} L_{\text{ML}}(\boldsymbol{\theta}). \tag{2.8}$$

2.2 The Non-linear Mixed Model

While Model (2.4) is flexible, the limitation of parameters entering the model linearly can become too restrictive. Pinheiro and Bates (2000, Ch. 6.1) mention three main reasons why one would want to use non-linear models instead of linear models: *interpretability*, *parsimony*, and *validity beyond the range of the data*.

Linear models are *empirical* in nature, based only on the observed relationship between the response and the covariates, and do not include any theoretical considerations about the underlying mechanism producing the data.

Non-linear models, on the other hand, are often *mechanistic*, i.e., based on a model for the mechanism producing the response. As a consequence, the model parameters in a non-linear model generally have a natural physical interpretation. For example, in pharmacokinetics, the estimated absorption and elimination rate parameters are of direct interest. Furthermore, a non-linear model generally needs fewer parameters than a competitor linear model, giving a more *parsimonious* description of the data. Lastly, since non-linear models often contain parameters capturing certain typical aspects of the data, such as asymptotes and monotonicity, they usually provide more reliable predictions outside the observed range of the data.

To allow for non-linear effects in the model, we generalize (2.2) to

$$\boldsymbol{Y}_i | \boldsymbol{b}_i \sim N[f(\boldsymbol{Y}_i | X_i, Z_i, \boldsymbol{\beta}, \boldsymbol{b}_i), \Sigma_i], \qquad \boldsymbol{b}_i \sim N(0, D). \tag{2.9}$$

Here, f is a non-linear mean function. Inference usually proceeds by means of maximum likelihood (Fahrmeir and Tutz 2001, Ch. 7), unless a fully Bayesian view is adopted. To this end, the marginal likelihood needs to be calculated from (2.9). In the linear case, the marginal likelihood function has a closed-form expression, but this is no longer true in the general case. As a consequence, numerical approximations of

the likelihood function are needed, leading to more computationally intensive estimation algorithms and to less reliable inference results (Pinheiro and Bates 2000).

The most frequently used approximation methods are either based on Gaussian quadrature rules (Davidian and Gallant 1992), or Taylor series expansions. The latter consists of taking a first-order Taylor expansion of the model function f around the expected value of the random effects (Sheiner and Beal 1980, Vonesh and Carter 1992), or around the conditional mode of the random effects (Lindstrom and Bates 1990). A detailed discussion and comparison of the different approximation methods can be found in Pinheiro and Bates (1995).

Since Gaussian quadrature approximations have a higher degree of accuracy than the Taylor expansion methods, though at the cost of increased computational complexity, they are our preferred method of analysis. SAS procedure NLMIXED uses adaptive Gaussian quadrature by default and is therefore our standard analysis tool to fit non-linear mixed models. We will now summarize some of the key ideas on (adaptive) Gaussian quadrature approximation.

Gaussian Quadrature Approximation

The aim of the numerical quadrature methods is to approximate the random-effects integral itself. This is done by a weighted sum over (fixed) abscissas. Only a brief description, based on the article of Lesaffre and Spiessens (2001) of this method will be presented here.

Gaussian quadrature can be used to approximate any integral of the form

$$\int f(z)\,\phi(z)\,dz,$$

where $\phi(.)$ denotes the density function of a standard normal distribution. The approximation consists of replacing the integral by a weighted sum:

$$\int f(z)\,\phi(z)\,dz \approx \sum_{q=1}^{Q} w_q\, f(z_q),$$

where the nodes z_q are solutions of the Qth order Hermite polynomial and the w_q are suitably corresponding weights. These quadrature points z_q and weights w_q can be found in tables or can be calculated by algorithms. The higher Q the more accurate the approximation will be.

In case of univariate integration, the approximation consists of subdividing the integration region in intervals, and approximating the surface under the integrand by the sum of surfaces of the so-obtained approximating rectangles. An example is given in the left window of Figure 2.1. For this integration we have used $Q = 10$

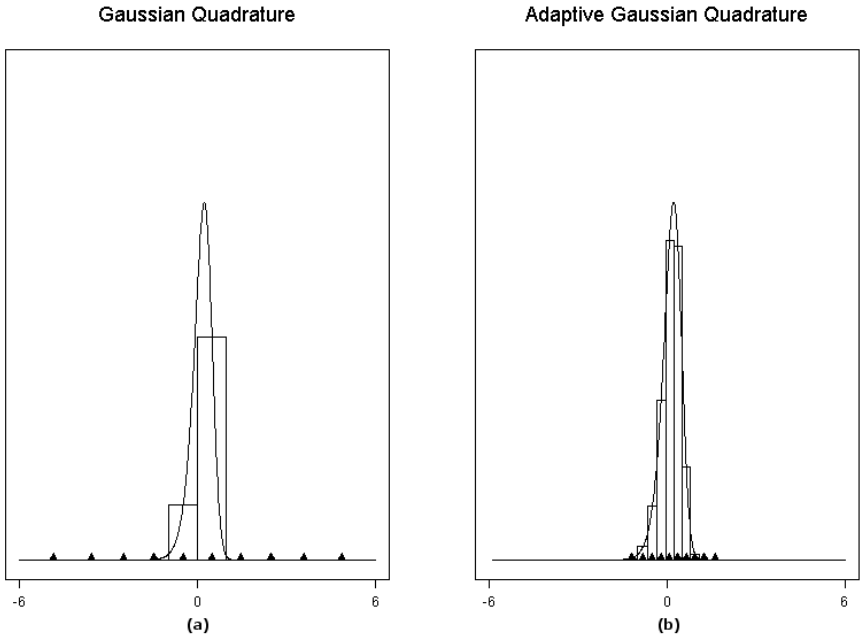

Figure 2.1: *Comparison of the positions of 10 quadrature points obtained from (a) ordinary Gaussian quadrature and (b) an adaptive Gaussian quadrature for the same integrand.*

quadrature points and these are shown as black triangles at the bottom of the figure. It is clearly seen that only 2 quadrature points out of 10 give a considerable contribution to the integral. Adaptive Gaussian quadrature solves this problem by appropriately centering and scaling the quadrature points z_q such that more quadrature points lie in the region of interest. Typically, adaptive Gaussian quadrature needs much less quadrature points than classical Gaussian quadrature. On the other hand, adaptive Gaussian quadrature is computationally much more demanding.

Empirical Bayes Estimation

Although in practice one is usually primarily interested in estimating the parameters in the marginal distribution of $\boldsymbol{Y}_i$, i.e. $\boldsymbol{\beta}$ and D, it is often useful to calculate estimates for the random effects $\boldsymbol{b}_i$ as well. They reflect between-subject variability, which makes them helpful for detecting special profiles (i.e., outlying individuals) or groups of individuals evolving differently over time. Also, estimates for the random effects are needed whenever interest is in prediction of subject-specific evolutions. In turn, these predictions can be plotted against the observed profiles to visually assess the quality of the model fit. This section is based on Molenberghs and Verbeke 2005, Section 14.2.4.

Since the random effects in the hierarchical model (2.9) are assumed to be random variables, it is most natural to estimate them using Bayesian techniques. It was also noted that the most common assumption for the marginal distribution of $\boldsymbol{b}_i$ is the multivariate normal with mean vector $\mathbf{0}$ and covariance matrix D. In the Bayesian literature, this last distribution is usually called the *prior distribution* of the parameters $\boldsymbol{b}_i$ since it does not depend on the data $\boldsymbol{Y}_i$. Once the observed values $\boldsymbol{Y}_i$ have been collected, the so-called *posterior distribution* of $\boldsymbol{b}_i$, defined as the distribution of $\boldsymbol{b}_i$, conditional on $\boldsymbol{Y}_i$, can be calculated. Using the same notation as in equation (2.3), we denote the density function of $\boldsymbol{Y}_i$ conditional on $\boldsymbol{b}_i$ for subject i by $f_i(\boldsymbol{Y}_i|\boldsymbol{b}_i)$, and the prior density function of $\boldsymbol{b}_i$ by $f(\boldsymbol{b}_i)$, the posterior density function of $\boldsymbol{b}_i$ given $\boldsymbol{Y}_i$ is given by

$$f_i(\boldsymbol{b}_i|\boldsymbol{Y}_i) = \frac{f_i(\boldsymbol{Y}_i|\boldsymbol{b}_i)\, f(\boldsymbol{b}_i)}{\int f_i(\boldsymbol{Y}_i|\boldsymbol{b}_i)\, f(\boldsymbol{b}_i)\, d\boldsymbol{b}_i}. \tag{2.10}$$

For the sake of notational convenience, we hereby suppressed the dependence of all above density functions on certain parameters in the marginal model for $\boldsymbol{Y}_i$. Unlike in the linear case, this posterior density is, in general, not a normal one. Therefore, the posterior mode, rather than the posterior mean, is used as point estimator for $\boldsymbol{b}_i$. More specifically, the estimator $\hat{\boldsymbol{b}}_i$ is the value for $\boldsymbol{b}_i$ that maximizes $f_i(\boldsymbol{b}_i|\boldsymbol{Y}_i)$, in which the unknown parameters have been replaced by their estimates obtained from maximum likelihood estimation. As in the linear case, the obtained estimates are called empirical Bayes (EB) estimates.

2.3 Fractional Polynomials

Fractional polynomials allow a wide variety of parametric shapes by considering not only integer powers of a key covariate (e.g., time), but also fractional powers. This is handy whenever no clear view on the model to be considered exists. As soon as non-integer powers are allowed for, the number of potential models is effectively endless, and it is wise to consider, a priori, a sensible model building strategy. This has been provided by Royston and Altman (1994).

Formally, Royston and Altman (1994) define a fractional polynomial as any function of the form

$$f(u) \quad = \quad \phi_0 + \sum_{k=1}^{m} \phi_k x^{(p_k)},$$

where the degree m is a positive integer, where $p_1 > \ldots > p_m$ are real-valued prespecified powers, and where ϕ_0 and $\phi_1, \ldots, \phi_m$ are real-valued unknown regression coefficients. Finally, $x^{(p_k)}$ is defined as

$$x^{(p_k)} \quad = \quad \begin{cases} x^{p_k} & \text{if } p_k \neq 0 \\ \ln(x) & \text{if } p_k = 0. \end{cases} \tag{2.11}$$

Not only the conventional powers x, x^2,... are allowable, also $\ln(x)$, $\sqrt{x}$ (for $p_k = 0.5$), $1/x$ (for $p_k = -1$), etc.

In the context of linear and logistic regression analyses, Royston and Altman (1994) have shown that the family of fractional polynomials is very flexible and that models with degree m larger than 2 are rarely required. In practice, several values for the powers $p_1, \ldots, p_m$ can be tried, and the model with the best fit is then selected.

Using a fractional polynomial within a linear or a non-linear mixed-effects model, is reasonably straightforward. One is merely required to construct the necessary covariate powers, logarithms, and interactions thereof, as a set of covariates in the data set to be analyzed. In our case, fractional polynomials will be applied to the time covariate. Of course, given the relative complexity of the non-linear mixed effects model, we propose to keep the degree m of the polynomials relatively small. One can then fit several models with a variety of powers $p_1, \ldots, p_m$.

In the context of the linear mixed model, fractional polynomials have been used both in the fixed-effects structure (Verbeke and Molenberghs 2000, Ch. 24.5) as well as in the serial correlation structure. We will return to the latter application in Section 6.3.1.

While the fractional polynomial approach is flexible, it is also empirical in nature, allowing on the one hand a wide variety of parametric shapes, useful for varied

applications, but not offering immediate biological insight into the meaning of the parameters and their estimated values. Arguably, their value lies in confirming or questioning other models, i.e., by way of sensitivity analysis, and to test treatment or other effects.

2.4 Splines

Another flexible way for obtaining a smooth fit to one's data is through *splines*, which are piecewise polynomials with components smoothly spliced together. The joining points of the polynomial pieces are called *knots*, that do not have to be evenly spaced. A spline is of degree p when the highest degree of the polynomial segments is p. Ruppert *et al* (2003) define a pth-degree spline model with knots at $\kappa_1, \ldots, \kappa_K$ as

$$f(x) = \beta_0 + \beta_1\, x + \cdots + \beta_p\, x^p + \sum_{k=1}^{K} \beta_{p+k}\, (x - \kappa_k)_+^p, \tag{2.12}$$

where $(x - \kappa_k)_+$ is the truncated power basis function, i.e., the *positive part* of the function $(x - \kappa_k)$. Other possible basis functions include the B-spline (Dierckx 1993), natural cubic spline (see, for example, Eubank 1988), and radial basis (see, for example, Green and Silverman 1994).

A simple and straightforward way to fit splines is by using ordinary least squares to estimate the (unrestricted) knot point coefficients β_{p+k}. This essentially means that the coefficient at each knot point is considered a fixed effect and this is usually referred to as a *regression spline*. However, this approach usually tends to overfit the data, leading to too coarse a regression curve, unless the number of knot points is small and their location carefully chosen (as in adaptive splines, see e.g. Zhang 2004).

Owing to the aforementioned coarseness of the parametric spline, various methods have been developed to constrain the knots' influence. Classically, the amount of smoothing is controlled by adding a term to the likelihood function, penalizing large coefficients at the knot points, which amounts to counterbalancing such coefficients' contribution to the raggedness of the curves. A candidate penalty term is $\lambda \sum_{k=1}^{K} \beta_{p+k}^2$, but there are many more. There is a vast amount of literature on the selection of the optimal smoothing parameter λ. Roughly speaking there are three (related) ways to determine the smoothing parameter λ. A first option is cross-validation, where for a grid of λ-values the squared error loss criterion is minimized in a leave-one-out cross-validation procedure. A second approach is the use of some model selection criterion, such as *Akaike's Information Criterion* (AIC, see e.g. Hurvich *et al* 1998). This method however requires a so-called equivalent number of

parameters or the effective degrees of freedom (see e.g. Ruppert *et al* 2003) to define the AIC appropriately. Thirdly, penalized splines can also be represented in mixed-model form (Verbyla *et al* 1999, Ruppert *et al* 2003), meaning that each knot point coefficient acts as a random effect. This results in a multivariate normal density entering the marginal likelihood, which then needs to be integrated out. The variance component governing these additional random effects is usually set equal for all knot points. This variance component controls and describes the degree of flexibility and smoothness. The fitted curve can be constructed by means of the empirical Bayes estimates.

The linear mixed model representation can be set up by considering the following random-spline design matrix:

$$Z_i = \begin{bmatrix} (x_1 - \kappa_1)_+ & \cdots & (x_1 - \kappa_K)_+ \\ \vdots & \ddots & \vdots \\ (x_n - \kappa_1)_+ & \cdots & (x_n - \kappa_K)_+ \end{bmatrix}.$$

Of course, such additional random effects can be combined with random effects already present in (2.1). Other modeling assumptions expressed in conjunction with (2.1) are left unaltered. Let the sole variance component governing the smoothing process be σ_u^2 and assume the residual error structure is of the conditional independence type with variance component σ_ε^2, then the smoothing parameter λ^2 can be shown to take the form $\lambda^2 = \sigma_\varepsilon^2/\sigma_u^2$ (Verbyla *et al* 1999, Ruppert *et al* 2003).

2.5 Discrimination, Classification, and Clustering

Discriminant analysis and methods for classification are obvious tools to be used when the focus is on distinguishing between groups in the data. A key tool is Fisher's linear discriminant analysis (Fisher 1936), while a large number of methodological developments have taken place, leaving us with a variety of methods.

There are two very distinct situations. The first one arises when group membership has been defined explicitly. At the same time, this does not imply that it is easy to *discriminate* between these groups, necessitating the use of formal discriminant analysis methods (Johnson and Wichern 1992). In a second situation, the researcher has a vague idea about the possibly useful split of a given group in subgroups, without precise knowledge about the number of such groups, let alone their definition. In such a case, one resorts to cluster analysis (Johnson and Wichern 1992). We will review each of these methods in the briefest way, since a vast literature on both subjects is in existence.

2.5.1 Discriminant Analysis

Discriminant function analysis is used to determine which variables discriminate between two or more naturally occurring groups. In linear discriminant analysis, we are interested in those linear features which reduce the dimensionality and simultaneously preserve class separability. Discriminant analysis can be understood as an exploratory tool to describe the dependence relations of the response variable on the given set of predictors in the observed sample of cases; the G categories of the response variable define a partition of the population Ω into G groups $(\omega_1, \omega_2, ..., \omega_G)$ and the P predictors are observed to characterize the typologies of cases within each group (Saporta 1990, McLachlan 1992). At the same time, discriminant analysis can also be used to define a decision rule for assigning a new case to one class on the basis of the observations of the given predictors in the so-called learning sample; a method such as test sample or cross-validation is considered to estimate the accuracy of the decision rule (Fisher and van Ness 1973, Celeux and Nakache 1994). Parametric, whether or not normally based, semi-parametric, and non-parametric versions have been established.

The quality of a so-obtained classification rules depends on three aspects. First, the parametric assumptions made may or may not hold, and this needs to be assessed. Second, the discrimination and classification rule obtained will be imprecise, since it is based on a finite sampling only. Third, the performance of the rule will be driven by the actual separation of the sub-populations in terms of the characteristics under study. For example, body height and weight have some discriminative power between genders in a human population, but very poorly so. It is important to note that investigators can control the first and second feature, but not the third one, unless a broader set of potential discriminators would be considered. For all of these reasons, the classification rule needs to be assessed, not so much in terms of its fit to the learning data, but rather in terms of its classification power to a test set of data.

For example, a classification rule for two normal populations, with different means but equal variance-covariance structures, takes the form:

$$(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T \boldsymbol{\Sigma}^{-1} \boldsymbol{x} - \frac{1}{2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2) \geq \ln \left[\frac{c(1|2)}{c(2|1)} \frac{p_2}{p_1} \right], \tag{2.13}$$

where $c(1|2)$ is the relative cost of misclassifying a subject that belongs to the second population into the first one and p_j $(j = 1, 2)$ is the prior group membership probability. One assigns a subject with response vector $\boldsymbol{x}$ to the first population if the inequality is satisfied and to the second one otherwise. A variety of versions exists, depending on the parametric assumptions made. An obvious extension is to

normal populations with unequal variances, leading to a quadratic rule rather than (2.13) which is of a linear type. Also extensions to general G groups rather than just two are obvious. The derivation of sample versions of population-based rules such as (2.13) is straightforward. In practice, one might want to assign a set of so-called posterior probabilities to a study subject, describing its chances of membership to each of the groups under study.

A number of formal methods have been derived to study the classification error and to assess it using datasets at hand, including cross-validation, jackknifing, and the use of learn and test samples (Johnson and Wichern 1992).

The SAS procedures DISCRIM and CANDISC can be used for discriminant analysis purposes.

2.5.2 Cluster Analysis

When interest lies in exploring whether a population could be usefully divided in subpopulations, without definite knowledge about group definition or even the number of groups, the optimal solution seems to be considering all possible partitions, together with an evaluation following a certain criterion, such as, for example, minimum variance, minimum distance, and or using a certain distance or similarity measure. However, unless the set of data at hand is extremely small, such an approach is not feasible in practice. Therefore, a number of heuristic strategies have been developed (Johnson and Wichern 1992).

One distinguishes between hierarchical and non-hierarchical clustering methods. A hierarchical method either starts from the entire datasets and then gradually and hierarchically splits it up in ever finer subsets, following a particular criterion or, in reverse, starts from the collection of singletons and groups them following a certain criterion (agglomerative procedure) until finally the entire dataset is brought together. In each of these cases, the number of clusters is then chosen following an optimality criterion. Several criteria can be used in the agglomerative case, such as so-called single linkage, complete linkage, or average linkage.

A popular non-hierarchical way is the so-called K-means method. One then partitions the subjects in K initial clusters or chooses K initial centroids. Next, the subjects are re-assigned to the cluster with the closest centroid, whereafter the centroids are computed again. This process is then repeated until convergence.

Each methods has advantages and disadvantages and these are relatively well understood. For example, single linkage tends to generate chain-like clusters which in many cases do not represent a reality one is looking for. The K-means method can

produce poorly separated clusters if two or more initial centroids happen to lie in the same physical cluster.

Overall, it is wise to consider cluster analysis an exploratory and hypothesis generating technique only. One might put more confidence in a given structure if various methods happen to produce roughly the same clustered. In any event, a cluster analysis should be followed up by a scientific discussion as to the relevance of the emerging clustering and ideally follow up studies would be undertaken to confirm the findings.

A variety of clustering procedures is available in SAS, including the procedures CLUSTER, FASTCLUS, VARCLUS, TREE, HIER, and IPFPHC.

3

Characterizing Persistent Disturbing Behavior

3.1 Introduction

Mental health care institutions in Belgium are confronted with a group of chronically therapy resistant patients. This group is problematic in the sense that no scientific definitions nor theory exists. Furthermore, there is no legislative framework in place. These patients cannot be treated satisfactorily with the latest knowledge of therapy and medication. Their behavior is disturbing in the sense that living together in their natural environment, or even in a hospital ward, is extremely difficult. Since their disease systems are unstable, and given that their behavior is persistent over time, intensive supervision over 24 hours is required. This condition is referred to as *persistent disturbing behavior* (PDB).

The current Belgian health care system is clearly not accommodating to this group. Indeed, from the two residential settings, psychiatric hospitals and psychiatric nursing homes, the former are defined as non-residential institutions for intensive specialist care. As the PDB group needs a prolonged stay in such a setting, a psychiatric hospital is not the optimal environment. In addition, a 1996 law states that a psychiatric nursing home is intended for patients with stabilized chronic psychiatric conditions. While the law does not specify the meaning of stabilized condition, it is generally

understood that PDB patients are not stable. We therefore have to conclude that mental health care does not explicitly accommodate the PDB group.

The PDB group raises four important questions. First, how can it be distinguished from related but different groups, such as patients with acute or short-term disturbing behavior. Second, since a clear definition is emerging only now, the size of the PDB group is unclear. Third, it is conceivable that the PDB group consists of a number of subgroups that can be usefully distinguished between. Finally, it is not clear where such patients should be based, even though this will in all likelihood neither be the ambulatory setting nor sheltered living. Further, psychiatric wards of general hospitals are intended for acute problems while PDB patients are clearly a chronic group.

To rectify this situation, legislative work is necessary. Before this can be done, one first needs to properly define the PDB group and undertake a quantitative analysis, formulating an answer to the aforementioned questions.

Since there is no generally accepted definition, we will use the following working definition. To be classified as PDB, a patient has to be subject to socially inadequate behavior, that is persistent and treatment resistant, disruptive for the patient's environment, and confronting the therapeutic team with unrealistic demands. The inadequate behavior can take one or several of many different forms such as multiple forms of aggression (directed to oneself or to others), sexually uninhibited behavior, agitation, loss of decorum, and suicidal behavior.

In 1998 a cross-sectional pilot study was set up in the psychiatric hospitals and the psychiatric nursing homes in the Belgian province of Limburg to (1) estimate the size of the PDB group and (2) explore factors to discriminate between PDB and non-PDB patients (Bruckers *et al* 2000). While these results are very useful, there are a number of residual issues. First, due to its cross-sectional nature, the focus is on disturbance rather than on persistence. Second, the working definition presented earlier did not exist at the time. Third, the group of patients shown for study was chosen for comparison with a non-PDB control group rather than for representativeness. This design implies that, while conclusions regarding differences between PDB and non-PDB patients, and conclusions pertaining to subgroups within the PDB group can be drawn with confidence, caution is necessary when making inferences about the magnitude of the PDB group. For the latter goal, the study should be seen as being of a pilot type. Nevertheless, it is important to know whether the group is sufficiently large so as to warrant specific components of care. Should one want to draw more refined conclusions, then a follow-up study, less prone to selection bias, would be in place.

By making use of longitudinal psychiatric registry data, we will show how the persistence aspect of the group can be studied and how insight into the PDB patients can be enhanced further. Furthermore, we will present the results of a cluster analysis, to initiate identification of subgroups within the PDB group.

The data on which our analyses are based are presented in Section 3.2, while Section 3.3 briefly discusses the implemented methods of analysis. Section 3.4 presents our findings. Precisely, after reviewing and expanding upon the cross-sectional discriminant analysis, the data are analyzed using longitudinal methodology, whereas cluster analysis provides further insight. These findings are used in Section 3.5 to formulate a perspective on the patient population with persistent disturbing behavior. Note that the content of this chapter is mainly based on the paper of Serroyen *et al* (2007a).

3.2 The Data

For every patient admitted to a residential psychiatric care setting in Belgium specific data are registered. This registration system was made mandatory in 1996 for psychiatric hospitals and in 1998 for the psychiatric nursing homes by the federal Ministry of Public Health and is called *Minimal Psychiatric Data* (MPD). The entire set of data is extensive, organized in a number of modules. A major source for the MPD instrument, but not the only one, is the so-called *Diagnostic and Statistical Manual of Mental Disorders*, known as DSM. The items, relevant for our purposes, are concisely listed in Table 3.1.

Our data set, previously used by Bruckers *et al* (2000), contains information on 611 patients from the province of Limburg about more than 200 psychiatric, physical, and sociological characteristics. The variables in this data set are mostly of a categorical or ordinal type, although some continuous variables are present as well. The four key continuous variables are the PDB score, which will be discussed in more detail in Section 3.4.1, with mean 0.62 and standard deviation 1.97, age (mean 47.48 and standard deviation 15.60), duration (mean 3365.71 and standard deviation 3143.43), and GAF score (mean 32.28 and standard deviation 80.00). The *Global Assessment of Functioning*, or GAF scale, is a numeric scale (1 through 100) used by mental health clinicians and doctors to rate the social, occupational and psychological functioning of adults. Incidentally, the GAF scale constitutes the fifth axis of the DSM-IV psychiatric classification system. It is considered a potential explanatory variable in all subsequent analyses.

To provide the reader with a perspective on the data, individual profiles of 20

Table 3.1: *Logistic regression analysis results, separately for psychiatric hospitals and psychiatric nursing homes. Odds ratios [95% confidence intervals] are reported. All items are coded such that an odds ratio greater than 1 corresponds to a less desirable score. The non-PDB group is the reference group.*

Item	Hospitals	Nursing homes
Psychiatric signs and symptoms:		
Auto-aggressive actions	5.62 [1.93; 16.42]	3.52 [1.61; 7.72]
Aggression against people	1.87 [1.21; 2.88]	
Aggression against objects		3.17 [1.26; 7.99]
Anti-social attitude		1.92 [1.05; 3.50]
Intensified supervision:		
Suicide danger	1.39 [1.19; 1.63]	
Separation/isolation		3.34 [1.12; 9.95]
Patient Functioning:		
Appearance	1.60 [1.18; 2.17]	
Respect for others	1.49 [1.03; 2.16]	1.81 [1.16; 2.82]
Socially unacceptable behavior		2.02 [1.27; 3.19]
Age	0.97 [0.95; 0.99]	
Gender	6.10 [2.89; 12.90]	
Primary diagnosis at admission:		
Mental Retardation		0.43 [0.19; 0.97]
DDAC[1]	0.22 [0.06; 0.87]	
Schizophrenia	1.75 [0.95; 3.21]	
V-codes	0.10 [0.01; 0.92]	

[1] Delirium, Dementia, Amnestic and Cognitive disorders

randomly selected subjects are presented in Figure 3.1. The average profiles and the group-specific empirical variance functions are displayed in Figure 3.2. Obviously, not all patients are observed at all times. An overview of the number of measurements available, for each of the 8 occasions and within each of the four PDB status by psychiatric sector combinations, is given in Table 3.2. In addition, the range of measurements per patient is displayed.

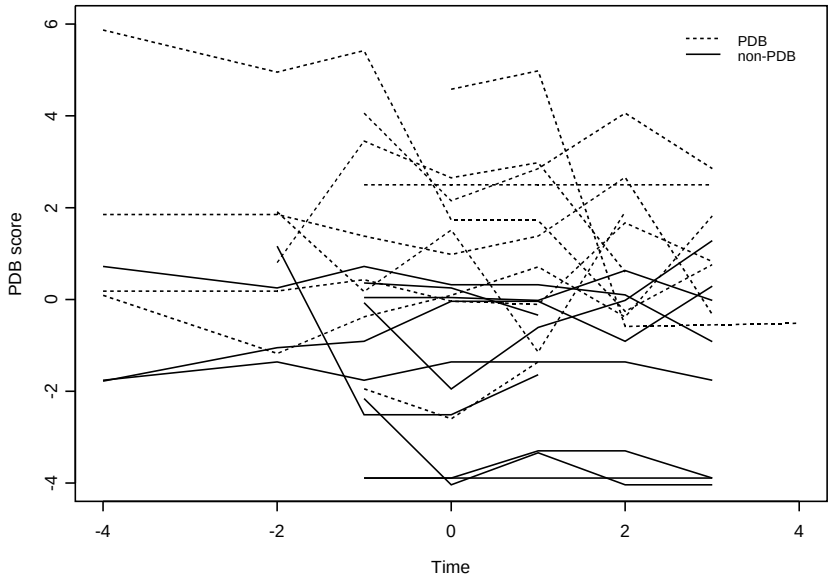

Figure 3.1: *Random sample of individual PDB score profiles for 10 PDB and 10 non-PDB patients.*

3.3 Method of Analysis

Information available in the MPD registration system was used to construct a discriminant function. Data registered in the second part of 1998 were used to develop this function. The items which make up the discriminant score have been recorded twice annually since 1996. After 2000, the legal registration framework changed, whence it is wise to restrict attention to the 1996–2000 interval.

Thus, the score was calculated at the other registration occasions as well, thus producing a longitudinal profile per patient. We employed linear mixed models to study the evolution of the mean discriminant function, for the PDB and non-PDB groups. Also, the length of stay contains very valuable information to investigate the persistence dimension. PDB patients cannot be discharged from the institution, since, due to their behavior, they are incapable of properly functioning in society.

When the focus is on distinguishing between PDB and non-PDB patients, discriminant analysis (see Section 2.5.1) is an obvious tool. Practically, two samples, one of which is of PDB type while the other consists of control patients, can then be classified based on a set of potential predictors. Linear and logistic discriminant analysis

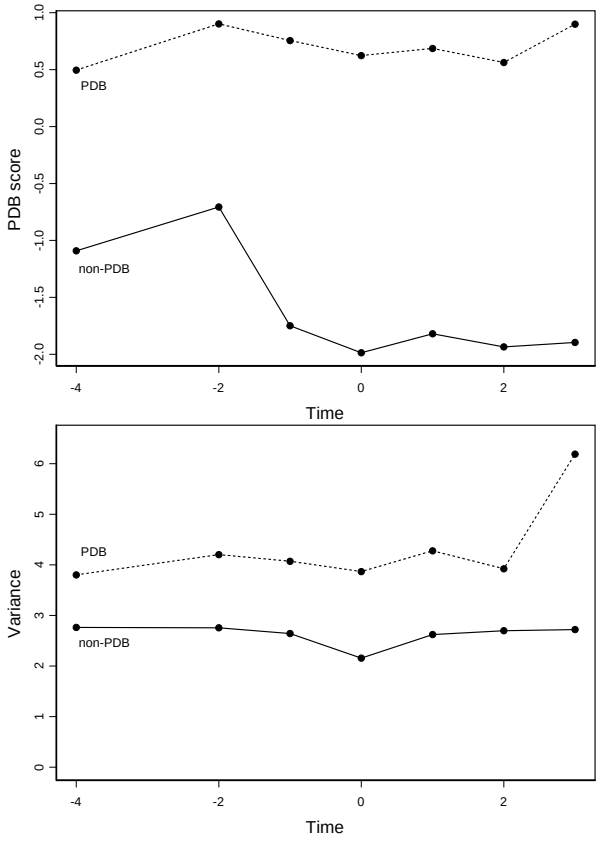

Figure 3.2: *Mean PDB score over time (top panel) and empirical variance function (bottom panel) for PDB and non-PDB group.*

are the more popular versions of the technique; the ones considered here (Johnson and Wichern 1992, Dunn and Everitt 1995, Tsuang, Tohen, and Zahner 1995, Everitt and Landau 1998, Dunn 2000).

When we further want to explore whether the PDB patient group can be divided further into subgroups, without definite knowledge about group definition or even the number of groups, cluster analysis techniques (see Section 2.5.2) can be employed. We will use the K-means method. We obviously need a distance function, and prefer the use of Gower's distance measure (Gower 1971), since it can handle all outcome types, i.e., asymmetric nominal, ordinal, interval, and ratio variables. The Gower's

Table 3.2: Number of measurements available per measurement occasion, PDB group, and sector.

		Measurement year (semester)								Range of
Group	Sector	96(2)	97(2)	98(1)	98(2)	99(1)	99(2)	00(1)	00(2)	# meas.
non-PDB	hosp.	128	180	197	202	183	158	116	115	1–8
PDB	hosp.	47	112	122	125	120	102	50	86	2–8
non-PDB	homes			214	220	211	202	189	186	2–6
PDB	homes			64	64	63	52	50	48	2–6
Total		175	292	597	611	577	514	405	435	

dissimilarity coefficient is defined as

$$d(\boldsymbol{y}_1, \boldsymbol{y}_2) = 1 - \frac{\sum_{j=1}^{n} w_j \, \delta^j_{y_1,y_2} \, d^j_{y_1,y_2}}{\sum_{j=1}^{n} w_j \, \delta^j_{y_1,y_2}},$$

where $\boldsymbol{y}_i$ $(i = 1, 2)$ is the vector of measurements on subject i, n is the number of measurements, and y_{ij} is the jth measurement on subject i. Further, w_j represents the weight for the jth variable and $w_j = 0$ when either y_{1j} or y_{2j} is missing. For symmetrically nominal, ordinal, interval, and ratio variables, $\delta^j_{y_1,y_2} = 1$, while for asymmetric nominal variables $\delta^j_{y_1,y_2} = 0$ if both y_{1j} and y_{2j} are absent and 1 otherwise. Finally, for nominal and asymmetric nominal variables, $d^j_{y_1,y_2} = 1$, if $y_{1j} = y_{2j}$, and 0 otherwise. For ordinal, interval, and ratio variables, $d^j_{y_1,y_2} = 1 - |y_{1j} - y_{2j}|$. In case of ordinal variables, the data are replaced by their corresponding rank scores. The hierarchical Ward's minimum variance method (Ward 1963) was adopted as clustering algorithm. In Ward's minimum-variance method, the distance between two clusters is the ANOVA sum of squares between the two clusters added up over all variables. At each generation, the within-cluster sum of squares is minimized over all partitions obtainable by merging two clusters from the previous generation.

3.4 Application to the Data

We will first undertake a cross-sectional study and thereafter switch to longitudinal and multivariate methods.

3.4.1 A Cross-sectional PDB Score

As mentioned in the introduction, a cross-sectional pilot study was undertaken in both the psychiatric hospitals and the psychiatric nursing homes in the province of Limburg to obtain a rough estimate of the size of the PDB group and to determine factors that can usefully distinguish between PDB and non-PDB patients. This study is in spirit of the analysis conducted by Bruckers *et al* (2000), supplemented with a number of additional analyses. Here, we focus on the 'disturbance' aspect, while the next section brings in the longitudinal 'persistence' dimension as well.

In November 1998, a number of wards were screened for PDB behavior, by an interdisciplinary team, and classified by expert opinion as PDB when the team judged that living together with the patient is hard and that s/he needed continuous supervision. The persistence dimension was approached by restricting attention to patients residing in chronic-patient wards within psychiatric hospitals or in psychiatric nursing homes. This is relevant, since patients residing in one of these wards in general already have had an intensive therapy in an acute ward and, in case of a psychiatric nursing home, also a long stay in a chronic ward.

Based on the screening, supplemented with data from the so-called *Minimal Psychiatric Data Registry* (MPD), a discriminant function was developed, producing the probability of dealing with a PDB patient, based on discriminatory MPD items. When this probability exceeds a threshold value we classify the patient as PDB. The function turned out to have good discriminative power. The screening status and the classification status agree for about 80% of the screened patients. Further details of this study are reported in Bruckers *et al* (2000).

The functional form of the discriminant function for the patients admitted in a psychiatric hospital takes the form:

$$\begin{aligned} \text{PDB}_{ij} \;=\; & -4.81 + 1.73 \cdot \text{Aggr.A}_{ij} + 0.62 \cdot \text{Aggr.P}_{ij} + 0.33 \cdot \text{Suicid}_{ij} + 0.47 \cdot \text{Appear}_{ij} \\ & +0.40 \cdot \text{Respect}_{ij} - 0.03 \cdot \text{Age}_i + 1.81 \cdot \text{Gender}_i - 1.50 \cdot \text{DDAC}_i \\ & +0.56 \cdot \text{Schizo}_i - 2.32 \cdot \text{Resid}_i + \varepsilon_{ij}, \end{aligned} \tag{3.1}$$

where 'Aggr.A' stands for aggression towards oneself (auto-aggression), 'Aggr.P' for aggression against people, 'Suicid' for suicide danger, 'Appear' for appearance, 'Re-

spect' for respect for others, 'Age' for age in 1998, 'DDAC' for the diagnostic class Delirium, Dementia, Amnestic and Cognitive disorders, 'Schizo' for the diagnostic class schizophrenia, 'Resid' for the residual diagnostic class (so-called *V-codes*) and ε_{ij} is the residual error term. The functional form for patients in psychiatric nursing homes

$$\begin{aligned} \text{PDB}_{ij} &= -6.39 + 1.26 \cdot \text{Aggr.A}_{ij} + 1.15 \cdot \text{Aggr.O}_{ij} + 0.65 \cdot \text{Asoc}_{ij} + 1.21 \cdot \text{Separ}_{ij} \\ &\quad +0.70 \cdot \text{Social}_{ij} + 0.59 \cdot \text{Respect}_{ij} - 0.85 \cdot \text{Retar}_{i} + \varepsilon_{ij}, \end{aligned}$$

with the same abbreviations as in (3.1) and in addition 'Aggr.O' standing for aggression against objects, 'Asoc' for anti-social attitude, 'Separ' for need for separation or isolation, 'Social' for socially unacceptable behavior, and 'Retar' for the diagnostic class mental retardation.

The MPD items contributing significantly to the classification of PDB *versus* non-PDB patients, as derived by Bruckers *et al* (2000), are presented in Table 3.1. Note that an odds ratio greater than one corresponds to the less desirable outcome. Given the non-PDB patients are the reference group, PDB patients fare worse on virtually all items. Note also that there are relatively large differences between the psychiatric hospitals and psychiatric nursing homes. Historically, and certainly in the late 90s when the study was conducted, the patient mix in these sectors was quite different. This situation has been in transition over the last decade, including the sectors' mission redefinition; this provides additional motivation for the currently conducted new study. These authors also quantified the amount of goodness-of-fit; they observed that the sensitivity and specificity for the psychiatric hospitals (psychiatric nursing homes) were 77.2% (71.9%) and 78.7% (85.5%), respectively. In addition to this, we calculated the ROC c statistic (Agresti 2002), which equals 0.85 for the psychiatric hospitals and 0.88 for the psychiatric nursing homes.

The use of cross-sectional information for a longitudinal goal may seem inherently contradictory. Nevertheless, it is intentional, since one wants to classify patients as early on as possible, ideally based on information at intake or right thereafter, so as to ensure the right type and level of care, as early in the treatment path as possible. As a sensitivity analysis for the fact that a cross-sectional discrimination is done with a view on longitudinal characteristics, we repeated the exercise, for one earlier follow-up occasion, 1998 (first semester), as well as for a later one, the first semester of 1999. The so-obtained results, encompassing three moments in time, are graphically represented in Figure 3.3, by way of point estimates and confidence intervals for each of the coefficients, for each one of the two sectors, and for each of the three moments in time. While there is some variation, as one could expect, the results are relatively

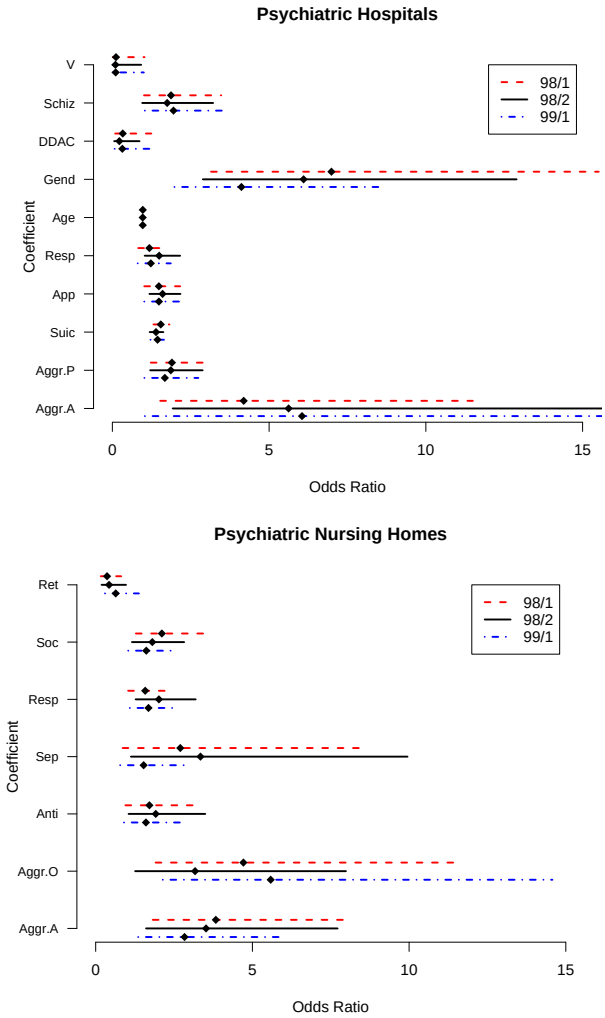

Figure 3.3: *Sensitivity analysis for the odds ratios, originally determined in 1998 (2nd semester), by re-estimating the parameters from both 1998 (1st semester) and 1999 (1st semester). For each effect, listed in Table 3.1, the odds ratios and their confidence intervals are presented, for each of the three moments and time and for each of the two sectors.*

stable, confirming that it is sensible to classify patients based on a single moment in time, even though the psychiatric condition clearly has got a longitudinal component.

A very important conclusion from Bruckers *et al* (2000) was that, following such

a discriminant rule, 35.5% of the patient population in a psychiatric hospital might belong to the PDB group, with a similar figure (32.1%) for the psychiatric nursing homes. Of course, these findings have to be taken with some caution. As stated before, the data used for analysis constituted a learning sample of PDB patients and controls, rather than a random subsample. This could be overcome by applying the rule, even when based on a learning set, to the entire population. This is likely not to resolve all issues with the data.

Therefore, as a consequence of the results derived in Bruckers *et al* (2000) as well as in this paper, a new study has been designed, with the sole purpose of refining the discrimination between PDB and non-PDB patients on the one hand, and of discerning subgroups within the PDB patients on the other hand. In this respect, it is important to note that the concept of PDB, even though the group is large, is as such relatively novel and has not received a lot of scientific interest as of yet. This new study also has got a qualitative part, primarily geared at refining the very definition of PDB.

Finally, even though the discriminant function appears to be rather stable when applied to differing moments in time, the focus remains more on the disturbance aspect than on the persistence component. Using longitudinal methods, we can do more justice to the latter, as well.

3.4.2 Longitudinal Analysis

The question arises, of course, whether or not the group considered to be PDB in 1998 indeed was chronic in their disturbing behavior. The fact that these patients are staying at long-stay wards only indicates that we are dealing with chronic disease statuses, not necessarily that the disturbing behavior is persistent. As stated before, it would be possible that the patient was going through an acute phase of disturbing behavior, something hard to disentangle based on information localized in time. This suggests the use of longitudinal methods.

We will apply the same definition of the PDB score is assumed across time, for reasons of consistency. This is practically most relevant and can be defended at least over a relatively short time span, as is the case here. However, when the time span increases and/or when other internal or external factors, such as the legal framework, would change, a careful assessment of the score's optimality across time would be in place. In principle, more complex models, such as dynamic longitudinal models, could be envisaged. However, such an approach would be essentially descriptive in nature, hard to use in practice, and less robust to idiosyncracies of the dataset. To build a

model that adequately describes the evolution of the value of the discriminant function over time we need to consider appropriate mean, variance, and covariance models. It is essential to perform an exploratory data analysis. As shown in Figures 3.1–3.2, the mean profiles for the discriminant function for the PDB and non-PDB groups are different and a non-linear structure emerges for the non-PDB group. The individual profiles follow more or less the same pattern. The figure with the individual profiles clearly shows substantial between and within variability. A key feature of the individual profiles is a vertical shift. This suggests the presence of a random intercept.

The variance functions for the PDB and non-PDB groups (Figure 3.2) display variance heterogeneity in the data. The variance is not constant over time. Moreover, the variability in the PDB group is larger than the variability in the non-PDB group.

To select a final model, describing the evolution of the discriminant function over time, we proceeded as follows. Verbeke and Molenberghs (2000) suggest selecting a variance-covariance structure based on the most complex mean structure one is prepared to consider. After selecting such a structure, the mean model can be simplified. A model including the PDB grouping indicator, time, quadratic time and pairwise interaction effects with PDB grouping was used as the most complex mean structure.

The cross-sectional analysis, based on logistic regression (Table 3.1) already indicated that the important predictors for patients from psychiatric hospitals is rather different from that of patients from psychiatric nursing homes. Therefore, it was decided to build separate models for both types of institutions.

The variance model was selected starting from the preliminary model including three PDB group-specific random effects: an intercept, a linear and a quadratic time slope. The 3×3 covariance matrix D for the random effects of each group was assumed to be unstructured. For the psychiatric hospitals, the residual error matrix was modeled using a group specific power-of-mean structure, allowing for the inclusion of covariates in the variance structure. For the psychiatric nursing homes on the other hand, a group-specific Gaussian serial correlation structure provided the best fit.

The random-effects structures of both models were simplified considering hierarchically ordered models. The significance of the effects was tested using likelihood ratio test statistics. The p-values were calculated using appropriate mixtures of χ^2 distributions as reference distribution (Verbeke and Molenberghs 2000, p. 69–72). The quadratic random slope was not significant at the 5% level of significance for both models and therefore removed from the models. However, the random intercept and linear random slope were kept in both models. These random effects and the residual matrix structures as discussed in the previous paragraph were found to be PDB group-specific.

Finally, the mean model was reduced, using the covariance structure that was just selected. For both models the mean structure for the PDB group could be simplified to a linearly increasing function with a common slope for PDB and non-PDB patients.

The reduced final model for psychiatric hospitals can be written as:

$$\begin{aligned} \text{PDB-score}_{ij} &= \beta_0 + \beta_1 \,\text{PDB}_i + \beta_2 t_{ij} \\ &\quad + (b_{1i} + b_{2i} t_{ij})\,\text{PDB}_i + (b_{3i} + b_{4i} t_{ij})\,\text{non-PDB}_i + \varepsilon_{ij}, \end{aligned} \quad (3.2)$$

where β_0 is the fixed-effects intercept, β_1 the fixed effect of the PDB group versus the non-PDB group, and β_2 is the fixed effects slope over time. The parameters b_{1i} and b_{3i} are the random intercept terms for the PDB and non-PDB groups, respectively. The subject-specific slopes are denoted by b_{2i} for the PDB group, b_{4i} for the non-PDB group, respectively, and ε_{ij} is the residual error term. The random effects have covariance matrix

$$D = \begin{pmatrix} d_{11} & d_{12} & 0 & 0 \\ d_{21} & d_{22} & 0 & 0 \\ 0 & 0 & d_{33} & d_{34} \\ 0 & 0 & d_{43} & d_{44} \end{pmatrix},$$

where the upper block refers to the PDB group and the lower block to the non-PDB group.

Parameter estimates of the model for the psychiatric hospital patients are given in Table 3.3, while Table 3.4 contains the results for the psychiatric nursing homes. The intercept is chosen to represent the mean value for the second part of 1998. The standard errors accompanying the variance components in Tables 3.3 and 3.4 should be interpreted with caution, for reasons reviewed in Verbeke and Molenberghs (2000). As Figure 3.2 already suggested, the mean value of the discriminant function for the PDB group is significantly higher than the mean value of the non-PDB group. For 1998 this is not a surprise since the function was constructed using these data. The difference between the two groups is maximal around the end of 1998 and the beginning of 1999. For the non-PDB group we note a steep decrease between 1997 and 1998. This is probably due to the effect of a successful treatment to alter the behavior of the patients. This effect is less pronounced in the PDB group, which agrees with the definition of therapy resistant patients.

Apart from an analysis using the raw PDB score as dependent variable, additional analyses were done, for both sectors, based on the log-transformed score. Parameter estimates (standard errors) are to be found in the final columns of Tables 3.3 and Table 3.4, respectively. The score is augmented by 7 prior to taking logarithms,

Table 3.3: *Parameter estimates (standard errors) for the final linear mixed-effects model for psychiatric hospitals, using restricted maximum likelihood. The model is fitted to the log-transformed PDB score as well.*

		Estimate (s.e.)	
Effect	Parameter	score	log(score+7)
Mean Structure			
Intercept	β_0	-1.36 (0.10)	1.70 (0.02)
PDB effect	β_1	2.08 (0.17)	0.32 (0.03)
Time effect	β_2	-0.10 (0.02)	-0.02 (0.003)
Random-Effects Variance Components			
Intercept (PDB)	d_{11}	2.32 (0.32)	0.05 (0.01)
Intercept (non-PDB)	d_{33}	1.96 (0.21)	0.11 (0.01)
Time (PDB)	d_{22}	0.05 (0.01)	0.001 (0.0002)
Time (non-PDB)	d_{44}	0.02 (0.01)	0.002 (0.0004)
covariance (PDB)	d_{12}	-0.05 (0.05)	-0.001 (0.001)
covariance (non-PDB)	d_{34}	0.09 (0.03)	0.01 (0.002)
Residual Variance Structure			
Power (PDB)	θ_1	0.35 (0.16)	2.19 (1.02)
Power (non-PDB)	θ_2	-1.13 (0.26)	4.29 (1.31)
Residual variance	σ^2	1.12 (0.08)	0.003 (0.002)

so as to avoid negative arguments of the logarithmic function. While parameters between these sensitivity analyses and the original ones are not directly comparable, it is important to observe that inferences made about the PDB effect would not qualitatively change when switching from the direct to the logarithmic version of the analysis.

The PDB-score profile over time is stabler in the PDB group than in the non-PDB group. Also, the variance-covariance structure contains information on the persistence dimension of the patient group under investigation. For both models it is clear that,

Table 3.4: *Parameter estimates (standard errors) for the final linear mixed-effects model for psychiatric nursing homes, using restricted maximum likelihood. The model is fitted to the log-transformed PDB score as well.*

		Estimate (s.e.)	
Effect	Parameter	score	log(score+7)
Mean Structure			
Intercept	β_0	-2.23 (0.08)	1.52 (0.02)
PDB effect	β_1	2.92 (0.27)	0.47 (0.04)
Time effect	β_2	0.03 (0.02)	0.004 (0.004)
Random Effects Variance Components			
Intercept (PDB)	d_{11}	3.74 (0.79)	0.07 (0.01)
Intercept (non-PDB)	d_{33}	1.22 (0.18)	0.05 (0.01)
Time (PDB)	d_{22}	0.00 (—)	0.00 (—)
Time (non-PDB)	d_{44}	0.01 (0.02)	0.0002 (0.001)
covariance (PDB)	d_{12}	0.07 (0.13)	0.0001 (0.003)
covariance (non-PDB)	d_{34}	0.07 (0.03)	0.002 (0.001)
Serial Structure			
Variance (PDB)	τ_1^2	1.58 (0.22)	0.02 (0.004)
Variance (non-PDB)	τ_2^2	0.14 (0.12)	0.01 (0.01)
Rate of Gaussian decrease (PDB)	$\frac{1}{\rho_1^2}$	0.26 (5.32)	0.00 (—)
Rate of Gaussian decrease (non-PDB)	$\frac{1}{\rho_2^2}$	1.63 (1.48)	2.41 (1.13)
Measurement Error Variance	σ^2	0.65 (0.08)	0.02 (0.002)

when comparing the variance of the random intercept with the measurement error and in case of the psychiatric nursing homes with the variance of the serial component, patient-specific characteristics are important. Thus, some patients intrinsically have high values while others intrinsically have low values. The variance of the random intercepts is larger in the PDB group than in the non-PDB group, while the variances for the random slopes are comparable. Furthermore, note that the variance of the

Table 3.5: *Estimated correlation matrix for PDB and non-PDB patients in psychiatric hospitals.*

	Time	**-4**	**-2**	**-1**	**0**	**1**	**2**	**3**	**4**
PDB	**-4**	1.0000	0.7250	0.6911	0.6454	0.5881	0.5208	0.4467	0.3698
	-2	0.7250	1.0000	0.7115	0.6883	0.6538	0.6087	0.5552	0.4965
	-1	0.6911	0.7115	1.0000	0.7016	0.6810	0.6495	0.6088	0.5616
	0	0.6454	0.6883	0.7016	1.0000	0.7009	0.6845	0.6581	0.6239
	1	0.5881	0.6538	0.6810	0.7009	1.0000	0.7109	0.6998	0.6797
	2	0.5208	0.6087	0.6495	0.6845	0.7109	1.0000	0.7313	0.7258
	3	0.4467	0.5552	0.6088	0.6581	0.6998	0.7313	1.0000	0.7604
	4	0.3698	0.4965	0.5616	0.6239	0.6797	0.7258	0.7604	1.0000
non-PDB	**-4**	1.0000	0.5885	0.5894	0.5855	0.5776	0.5666	0.5533	0.5385
	-2	0.5885	1.0000	0.6492	0.6577	0.6612	0.6603	0.6560	0.6491
	-1	0.5894	0.6492	1.0000	0.6845	0.6935	0.6977	0.6979	0.6950
	0	0.5855	0.6577	0.6845	1.0000	0.7191	0.7282	0.7328	0.7338
	1	0.5776	0.6612	0.6935	0.7191	1.0000	0.7521	0.7608	0.7656
	2	0.5666	0.6603	0.6977	0.7282	0.7521	1.0000	0.7826	0.7908
	3	0.5533	0.6560	0.6979	0.7328	0.7608	0.7826	1.0000	0.8101
	4	0.5385	0.6491	0.6950	0.7338	0.7656	0.7908	0.8101	1.0000

serial component for the PDB group (Table 3.4) is much larger than its counterpart for the non-PDB group. This was already observed in the exploratory data analysis. Interestingly, the rate of Gaussian decrease is much larger in the non-PDB group than in the PDB group. This indicates that stronger serial correlation exists between PDB scores in PDB patients compared to non-PDB patients in psychiatric nursing homes.

Let us also inspect the fitted correlations deriving from the estimated marginal variance-covariance matrix $V_i = Z_i D Z_i' + \Sigma_i$. This matrix is presented in Table 3.5 for the psychiatric hospitals. Table 3.6 contains the results for the psychiatric nursing homes. Considering Table 3.5, we clearly observe that the correlations between time points close in time is stronger for PDB patients than for non-PDB patients in psychiatric hospitals.

It seems logical to consider a direct comparison between the observed and fitted correlation structure, but unfortunately this is less than straightforward for two main

Table 3.6: *Estimated correlation matrix for PDB and non-PDB patients in psychiatric nursing homes.*

	Time	**-1**	**0**	**1**	**2**	**3**	**4**
PDB	**-1**	1.0000	0.6220	0.6266	0.6311	0.6356	0.6400
	0	0.6220	1.0000	0.6307	0.6351	0.6394	0.6437
	1	0.6266	0.6307	1.0000	0.6391	0.6432	0.6474
	2	0.6311	0.6351	0.6391	1.0000	0.6470	0.6510
	3	0.6356	0.6394	0.6432	0.6470	1.0000	0.6547
	4	0.6400	0.6437	0.6474	0.6510	0.6547	1.0000
non-PDB	**-1**	1.0000	0.6410	0.6143	0.6053	0.6048	0.6042
	0	0.6410	1.0000	0.6664	0.6432	0.6366	0.6377
	1	0.6143	0.6664	1.0000	0.6922	0.6720	0.6671
	2	0.6053	0.6432	0.6922	1.0000	0.7175	0.6998
	3	0.6048	0.6366	0.6720	0.7175	1.0000	0.7416
	4	0.6042	0.6377	0.6671	0.6998	0.7416	1.0000

reasons. First, linear mixed models involve three parts: (a) the fixed-effects structure; (b) the random effects; (c) the residual or serial correlation. Calculating the empirically observed correlations so as to take this layered structure into account is not without ambiguity. Even the definition of residuals, needed to calculate empirical correlations, in such a hierarchical context is a topic of some controversy. Second, data are incomplete since not all patients have a score available for all times. Unless the missing data mechanism is missing completely at random (Molenberghs and Kenward 2007), observed and expected features do not have to agree in the same way as they would if data were complete, even for a well fitting model.

In summary, a longitudinal analysis refines the perspective and enhances understanding of the PDB group, by simultaneously studying the disturbance and persistence characteristics. The analysis suggests that the group is substantial in size. Of course, given the selection of the data in the pilot study, this conclusion should be treated with caution. The aforementioned follow-up study will enable us to refine the conclusion. Nevertheless, in the meantime, it is of interest to explore whether the contingent of PDB patients can usefully be subdivided into meaningful subgroups, which could then be treated in tailor-made, high quality, wards.

3.4.3 Cluster Analysis

To further explore the group of PDB patients, we can perform a cluster analysis to suggest possible relevant therapeutic or organizational subgroups.

As stated in Section 3.3, Gower's distance measure was chosen since it can handle all outcome types, i.e., (asymmetric) nominal, ordinal, interval, and ratio variables. The hierarchical Ward's minimum-variance method was applied and two clusters retained. Since the clearest separation between these two clusters can be found in the ordinal variables, a frequency table of these variables by cluster is presented in Table 3.7. Cluster #1 appears to consist of PDB patients with higher scores on the ordinal variables compared to Cluster #2, indicating that these patients show more pathological behavior. The χ^2 tests underscore highly significant differences in distribution of scores between the two clusters. Further, it appears the mean PDB score for the first cluster (0.78) is higher than for the second one (0.47). Nevertheless, this has to be judged against the background of large variability, the standard deviations

Table 3.7: *Frequency table of the ordinal variables by cluster.*

Variable	**Cluster**	**Score**					**Total**	χ^2	*df*	**p-value**
		1	**2**	**3**	**4**	**5**				
Mobility	1	36	4	40	6	5	91	43.81	4	<.0001
	2	82	4	12	0	0	98			
Recognition	1	7	32	33	9	10	91	119.56	4	<.0001
of persons	2	85	8	4	1	0	98			
Notion of	1	11	20	25	4	31	91	119.42	4	<.0001
time	2	88	8	0	0	2	98			
Initiative	1	7	17	23	44		91	50.91	3	<.0001
	2	21	39	34	4		98			
Social	1	0	12	22	57		91	29.34	3	<.0001
	2	10	37	18	33		98			
Respect	1	2	15	24	50		91	42.37	3	<.0001
	2	8	36	43	11		98			
Conflicts	1	3	31	23	34		91	23.09	3	<.0001
	2	13	38	37	10		98			

being 1.80 and 2.11, respectively.

The identification of two clusters requires careful qualification and a number of comments are in place. First, cluster analysis is a pragmatic, exploratory method. It is therefore hard to fully formally establish that the number of clusters is equal to two, rather than three or more, or, perhaps only a single one. Therefore, our results should be taken as a mere indication that there is some room for entertaining the concept of more severe *versus* less severe PDB patients. Second, even then, one might argue it is likely for severity of PDB, as well as for other characteristics, to vary continuously across patients, rather than in a dichotomous fashion. Even then, considering a dichotomized version can be pragmatically helpful, with a view on efficiently organizing care. Arguably, these features need further study and the currently conducted follow-up study is well suited for this goal.

3.5 Perspective and Concluding Remarks

Based on discriminant analysis and longitudinal model building, the PDB score is rather different between the PDB and non-PDB groups. This is true for the mean profiles, the variance and correlation structure. Comparing PDB with non-PDB patients, the score is influenced by a different set of covariates, and for the effects in common, the magnitude of the effects is different. Also note that the different types of institutions are associated with different sets of covariates. Turning to variability, it is largest in the PDB group. This implies relatively more heterogeneity among such patients, opening perspectives for further subdivision. This can be done using cluster analysis, where discrete groups are found, or rather by considering a patient's relative position on the PDB score's scale, in case a more continuously oriented ranking is preferred.

Regarding the correlation structure, let us first turn to psychiatric hospitals. The correlation structure is subtly different between both groups. The PDB group is roughly of a first-order autoregressive type, showing relatively large correlations between adjacent measurements (around 0.75), which decreases with increasing time lag, dropping to about 0.35. Thus, the PDB group exhibits a chronic behavior from the beginning, with fluctuations happening in the long run rather than immediately. The non-PDB group correlation structure is closer to compound symmetry, amended by the fact that the correlations increase towards later times. This may suggest there is an unstable, acute phase at the beginning of the study.

Turning to psychiatric nursing homes, the picture emerging from the estimated correlation structures is different. Both are relatively close to compound-symmetry,

with a common correlation around 0.65. This is plausible from a field work point of view, because these patients are almost by definition of a chronic type.

Through the longitudinal analysis, we already established the rather heterogeneous nature of the PDB group, with a relative stability of the score within a patient. The longitudinal analysis does not allow to easily define subgroups within the PDB group, but the aforementioned heterogeneity encourages further exploration. By means of cluster analysis, we were able to suggest the presence of two clusters, characterized on the basis of the ordinal variables mobility, recognition of persons, notion of time, initiative, socially unacceptable behavior, respect for others, and conflicts. Classical contingency table analysis confirmed a significant difference between the two clusters on each of these variables. A significant difference was not found on the continuous variables.

In conclusion, the PDB patients are numerous, differ considerably from the control patients, in the sense that they exhibit a higher score. The group is also heterogeneous allowing one to further subdivide the group in clusters, based on the ordinal components of the score. Obviously, this opens perspectives for further therapeutic and/or organizational refinement. Most importantly, not only is there a need for specialized treatment entities, also further sub-specialization between such entities is to be recommended.

Further work will be directed towards refining the clustering of PDB patients by means of methods that take the longitudinal structure of the profiles into account. This might, for example, be achieved by means of latent class models (Skrondal and Rabe-Hesketh 2004).

4

Non-linear Models for Longitudinal Data

4.1 Introduction

In practice, data are often measured repeatedly over time on the same subject. Especially for Gaussian data, quite a number of approaches for analyzing longitudinal data have been developed and implemented in standard software packages. Most of the methodological work has been done in the setting of linear models and generalized linear models. While non-linear models have also been extended in various ways, the work is not as 'complete' as for the linear and generalized linear ones.

With linear models, although the *estimation* of the regression parameters must take into account the correlations in the data, their *interpretation* is essentially independent of the correlation structure. With non-linear models, including generalized linear mixed models, different assumptions about the source of correlation can lead to regression coefficients with distinct interpretations (Diggle *et al* 2002, Molenberghs and Verbeke 2005) and often widely varying magnitudes. The data analyst must therefore reflect ever so carefully on the objectives of the analysis and the source of correlation in choosing an approach.

Diggle *et al* (2002) and Molenberghs and Verbeke (2005), among others, distinguish between three different model families for longitudinal data: *marginal*, *random-*

Table 4.1: *Orange Tree Data. Measurements in mm of trunk circumference.*

	Trunk circumference				
Day	Tree 1	Tree 2	Tree 3	Tree 4	Tree 5
118	30	33	30	32	30
484	58	69	51	62	49
664	87	111	75	112	81
1004	115	156	108	167	125
1231	120	172	115	179	142
1372	142	203	139	209	174
1582	145	203	140	214	177

effects, and *conditional* models, the latter family including so-called *transition* models. While this taxonomy has been extensively used for both linear and generalized linear model settings, it is less commonly employed in a non-linear context. Arguably, it is common to make use of mixed-effects models in this area, especially in such area as pharmacokinetic and pharmacodynamic modeling (Davidian and Giltinan 1995, Molenberghs and Verbeke 2005). While the use of random-effects, or hierarchical, models is recommendable in many settings, it is sensible to also explore the possibility of using, and the relative merits of, the marginal and conditional families.

In Section 4.2, a simple set of motivating data, the orange tree data, is introduced. The three model families are surveyed and applied to the data in Section 4.3. Concluding remarks are offered in Section 4.4.

4.2 Orange Tree Data

Draper and Smith (1998, Exercise 24.N, p. 559) present data of an experiment in which trunk circumference (in mm) is measured for 5 orange trees, on 7 different occasions, over roughly a 4-year period of growth. The data are presented in Table 4.1. Profiles are plotted in Figure 4.1.

4.3 Model Families

As mentioned earlier, we can distinguish between three model families. We will now give a brief overview of these model families based on Diggle *et al* (2002) and Verbeke

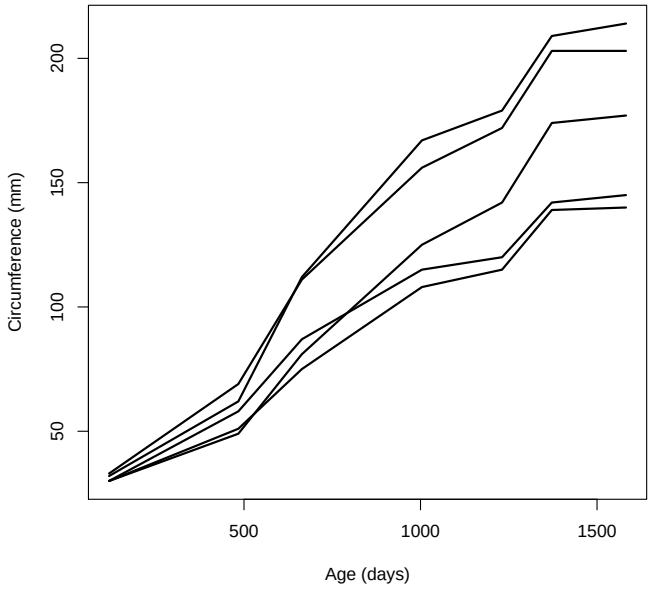

Figure 4.1: *Orange Tree Data. Growth curves of trunk circumference for each of the five trees.*

and Molenberghs (2005).

Marginal regression methods characterize the marginal expectation of a discrete or continuous response, Y, as a function of explanatory variables, X. By marginal expectation, we mean the average response over the sub-population that shares a common value of X. The methods are designed to permit separate modeling of the regression of Y on X, and the association among repeated observations of Y on each individual. Marginal models are appropriate when inferences about the population-average are the focus.

In addition to modeling the effects of covariates on the marginal expectation, we must also specify a model for the association among observations from each subject. This is to be contrasted with the random effects and transitional models where the covariate effects and the within-subject association are modeled through a single equation. All three approaches lead to the same class of linear models for Gaussian data. But in the discrete case, different (non-linear) models can lead to different interpretations for the regression coefficients. The choice of model should therefore depend on the scientific question being addressed.

In random-effects models, the response is assumed to be a function of explanatory

variables with regression coefficients that vary from one individual to the next. This variability reflects natural heterogeneity due to unmeasured *(latent)* factors. Given these so-called subject-specific parameters, the responses are often assumed independent. This does not preclude that more elaborate models are possible if residual dependence is detected. A random-effects model is a reasonable description if the set of coefficients from a population of subjects can be thought of as a sample from a distribution.

In conditional models, the sequence of repeated measures is modeled conditional upon (a subset of) the other outcomes. This could be the set of all past measurements or a subset thereof. In transition models, the subset is defined as the past measurements, often further restricted to the immediately preceding (few) measurements.

Let us now analyze the orange tree data in each of the tree model family settings. We will start with the random-effects approach, which is the most frequently used approach in the literature for analyzing the orange tree data. Then we will switch to the marginal approach, and round off with the conditional approach.

4.3.1 Random-effects Models

In a random-effects model, we focus on the expectation of the response Y_{ij} at measurement occasion j for subject i, additionally conditioning upon a random-effects vector $\boldsymbol{b}_i$:

$$E(Y_{ij}|\boldsymbol{b}_i, \boldsymbol{x}_{ij}, \boldsymbol{z}_{ij}) = h(\boldsymbol{x}_{ij}, \beta, \boldsymbol{z}_{ij}, \boldsymbol{b}_i), \tag{4.1}$$

where $\boldsymbol{z}_{ij}$ is the random-effects design matrix. Conventionally, but not always, is the distribution of the random effects assumed to be of a normal type.

The following non-linear mixed model has been proposed in the statistical literature (Pinheiro and Bates 2000):

$$\begin{aligned} Y_{ij} &= \frac{\beta_1 + b_i}{1 + \exp[-(t_{ij} - \beta_2)/\beta_3]} + \varepsilon_{ij}, \\ b_i &\sim N(0, \sigma_b^2), \\ \varepsilon_{ij} &\sim N(0, \sigma^2). \end{aligned}$$

Note that this model is non-linear in the fixed-effect parameters, but linear in the random effect b_i, simplifying the calculation of the marginal mean over the random-effects distribution. Thus, the conditional mean is

$$E(Y_{ij}|b_i) = \frac{\beta_1 + b_i}{1 + \exp[-(t_{ij} - \beta_2)/\beta_3]} \tag{4.2}$$

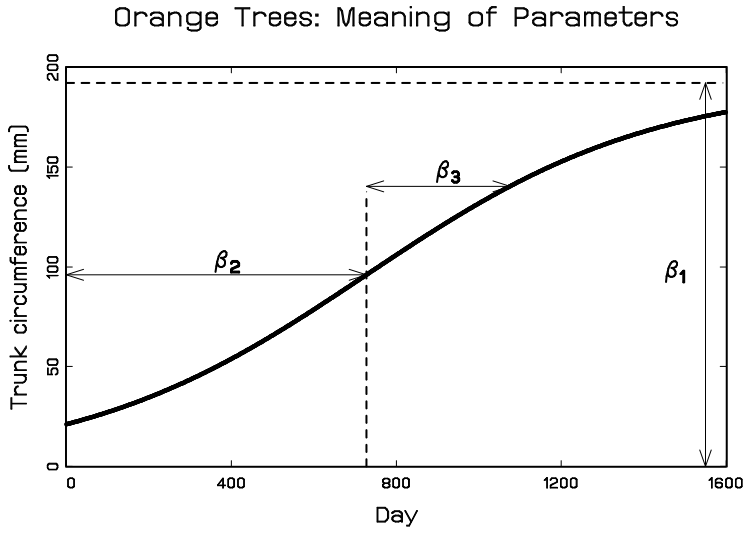

Figure 4.2: *Orange Tree Data. Interpretation of model parameters.*

while its marginal counterpart is

$$E(Y_{ij}) = \frac{\beta_1}{1 + \exp[-(t_{ij} - \beta_2)/\beta_3]}. \tag{4.3}$$

It is not to be taken for granted that this simplifying feature will occur in other applications. A graphical representation of the model, with the meaning of the parameters associated to it, is given in Figure 4.2.

To improve numerical stability while fitting the model, both the response values, trunk circumferences, and the covariate values, age, were divided by 100. The model was fitted using the SAS procedure NLMIXED.

A plot of the model fit can be found in Figure 4.3. The fitted mean profile is a smooth curve with a sigmoidal shape, although the curvature is not pronounced. The fit of the model seems acceptable. Parameter estimates and standard errors are given in Table 4.2.

Empirical Bayes predictions are graphed in Figure 4.4. The model fit seems acceptable and a set of three population-level parameters, having a clear interpretable meaning.

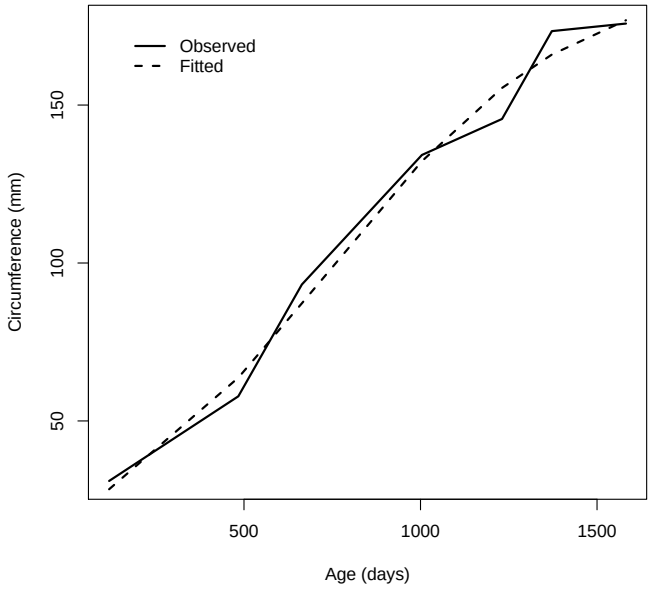

Figure 4.3: *Orange Tree Data. Plot of observed and fitted mean profiles of the random-effects model.*

4.3.2 Marginal Models

A marginal non-linear model for an outcome Y_{ij} at measurement occasion j for subject i, conditional on a vector of covariates $\boldsymbol{x}_i$, would take the form:

$$E(Y_{ij}|\boldsymbol{x}_{ij}) = h(\boldsymbol{x}_{ij}, \beta), \tag{4.4}$$

where β is the vector of regression parameters and h is the non-linear link function. This expression does not specify the full joint distribution. The association structure needs to be specified as well and this can be done, in turn, in a linear or non-linear fashion. We can then consider full likelihood approaches, which are in danger of becoming prohibitive in terms of computation, for example, when measurement sequences are of moderate to large length, or non-likelihood alternatives, such as generalized estimating equations (GEE, Liang and Zeger 1986). Such models have been given relatively little attention in the literature (Molenberghs and Verbeke 2005). One ought to be aware that this is a sensible method only when the correlation structure is not of direct interest, since it is treated merely as a nuisance characteristic in GEE.

We consider marginal model (4.3), with the errors allowed to be serially correlated, where the serial correlation function is assumed to be of an exponential type, as

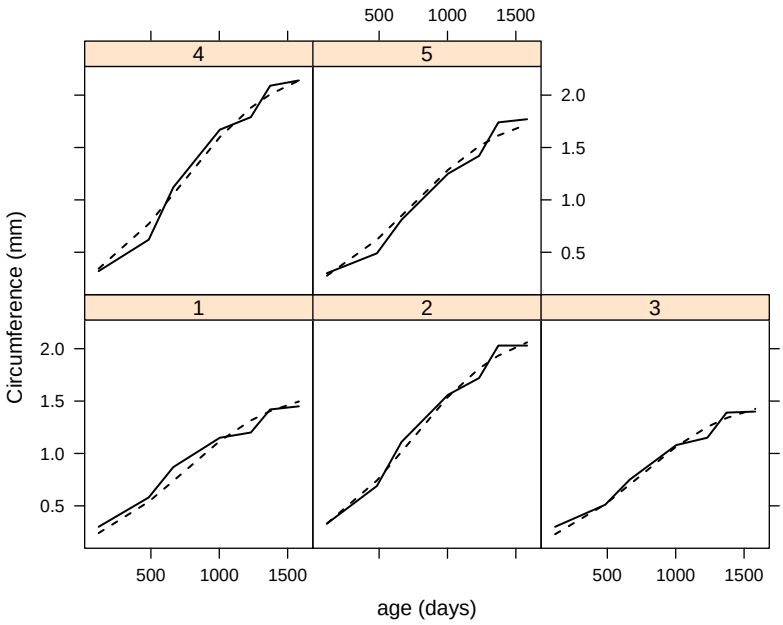

Figure 4.4: *Orange Tree Data. Plot of observed (solid line) and fitted (dashed line) individual profiles of the random-effects model.*

introduced in Section 2.1.

This marginal model was implemented in the software package R. The marginal log-likelihood function for the multivariate normal distribution was constructed and then maximized using a general purpose numerical optimizer based on a quasi-Newton method. This allowed us to directly model the serial correlation in the logistic growth model. The exponential serial correlation function was selected, since it is a simple two-parameter function and nicely converged; also, the fit is acceptable as we will see below. The orange tree data set is relatively small, forcing us to keep the modeled correlation function reasonably simple. The parameter estimates of the marginal model can be found in Table 4.3, while Figure 4.5 shows the observed and fitted profiles for the marginal model. When comparing Table 4.3 with Table 4.2, we can see that the parameter estimates are very similar.

The reason is that the random effect b_i in (4.3) enters the model in an additive way. Therefore, the random effect does not have an impact on the estimated fixed effects. The attenuation of parameter values in the marginal logistic model for binary data, as described in Diggle *et al* (2002), does not come into play, while it would,

Table 4.2: *Orange Tree Data. Parameter estimates and standard errors for random-effects model (4.2).*

Effect	Parameter	Estimate	(s.e.)
Asymptote	β_1	1.9205	(0.1566)
Half-growth	β_2	7.2791	(0.3525)
Shape parameter	β_3	3.4807	(0.2708)
$\text{Var}(b_i)$	d_{11}	0.1001	(0.0650)
$\text{Var}(\varepsilon_{ij})$	σ^2	0.0062	(0.0016)

Table 4.3: *Orange Tree Data. Parameter estimates (standard errors) for marginal model (4.3).*

Effect	Parameter	Estimate	(s.e.)
Asymptote	β_1	1.9100	(0.1440)
Half-growth	β_2	7.2463	(0.6181)
Shape parameter	β_3	3.4607	(0.4965)
Variance serial component	τ^2	0.0418	(0.0219)
Rate of exponential decrease	ϕ	24.7221	(17.6473)
Residual variance	σ^2	0.0049	(0.0023)

once the random effects entered non-linearly. In complex models, with potentially multiple random effects entering, the effect on the fixed-effects parameters would be extremely hard to predict and might require tedious computations. Such calculations would typically be of a relatively *ad-hoc* type.

4.3.3 Conditional Models

A conditional non-linear model would in addition allow $\overline{\boldsymbol{Y}}_{ij}$, the set of all outcomes except the one modeled, as an argument of h:

$$E(Y_{ij}|Y_{ik,k\neq j}, \boldsymbol{x}_{ij}) = h(\boldsymbol{x}_{ij}, \beta, \overline{\boldsymbol{Y}}_{ij}, \boldsymbol{\alpha}), \tag{4.5}$$

where $\boldsymbol{\alpha}$ is the vector of variance-component parameters. However, whereas a log-linear model produces a conditional specification, considering (4.5) in general is not

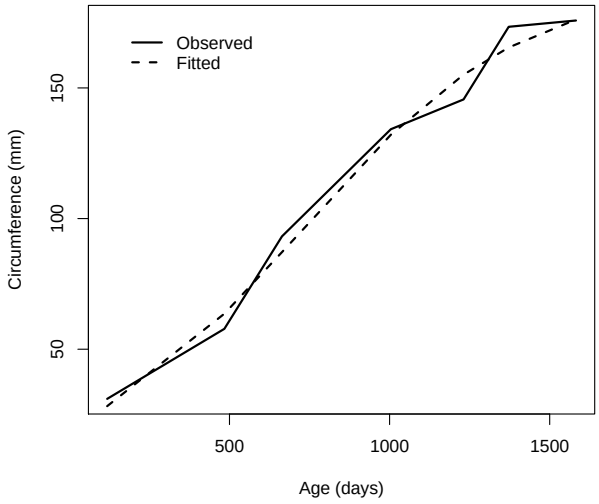

Figure 4.5: *Orange Tree Data. Plot of the observed and fitted mean profiles for the marginal model.*

guaranteed to produce a valid joint model. A non-linear version of a transition model is easier to handle and, arguably, more meaningful.

One possible conditional model for the orange tree data is a transition model where the random effect in (4.3) is replaced by the previous measurement $Y_{i,j-1}$. This model can be expressed as

$$Y_{ij} = \frac{\beta_1 + \gamma Y_{i,j-1}}{1 + \exp[-(t_{ij} - \beta_2)/\beta_3]}. \tag{4.6}$$

In Figure 4.6, we show a plot of the model fit. The fitted mean profile seems to provide an acceptable fit. This curve is not as smooth as the random effects model, but seems to follow the data closely. Individual predictions are presented in Figure 4.7. Table 4.4 provides an overview of the parameter estimates. The γ parameter estimate is positive, indicating an increment over time, or *growth*. Parameter estimates for β_1, β_2, and β_3 are clearly different from the estimates for the random-effects model. This is to be expected though, since these parameter estimates come from different model families and have a different interpretation, in the sense that they are *conditional* on the value of the prior history. This conditional interpretation makes answering the substantive question rather difficult (Diggle *et al* 2002, p. 142-144). On the other hand, σ^2 does not change as dramatically. This implies that the dependence on previous measurements re-composes the mean function, relative to the marginal model, leaving the residual component essentially untouched.

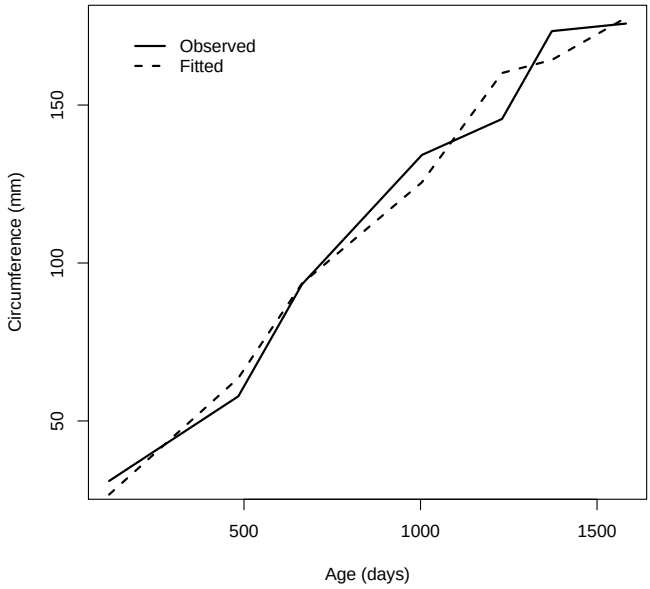

Figure 4.6: *Orange Tree Data. Plot of observed and fitted mean profiles for transition model (4.6).*

Another possibility is to add $Y_{i,j-1}$ as a covariate in the denominator of the model expression:

$$Y_{ij} = \frac{\beta_1}{1 + \exp\left(-[(t_{ij} - \beta_2)/\beta_3 + \gamma\, Y_{i,j-1}]\right)}. \tag{4.7}$$

Note that fitting such models is simple since, given the previous outcomes, measures are further assumed to be univariate. Hence, standard software for univariate non-linear models can be used. Evidently, also a tool such as the SAS procedure NLMIXED can be used, too. Figure 4.8 gives a graphical representation of the model fit. The fitted profile does not follow the observed profile as closely as in the case of Model (4.6), especially in the first half of the range in age values. Figure 4.7 shows the individual predictions. Checking the correlation matrix of the the parameter estimates also revealed signs of multicollinearity; strong correlations were observed between β_2 and β_3 (0.99), between β_3 and γ (0.96), and between β_2 and γ (0.92). Thus, issues of model stability can be raised. The parameter estimate for the asymptote (Table 4.5) is much more in line with the estimate from the random-effects model (Table 4.2) compared to the previous transition model (Table 4.4), while the opposite is true for the half-growth and shape parameter. This underscores that it is very difficult to make overall statements about the behavior of a particular modeling family in the

Table 4.4: *Orange Tree Data. Parameter estimates (standard errors) for transition model (4.6).*

Effect	Parameter	Estimate	(s.e.)
Asymptote	β_1	0.4206	(0.0819)
Half-growth	β_2	-0.4185	(1.7782)
Shape parameter	β_3	1.7820	(0.9527)
Renewal parameter	γ	0.8412	(0.0590)
Residual variance	σ^2	0.0143	(0.0034)

Table 4.5: *Orange Tree Data. Parameter estimates (standard errors) for transition model (4.7).*

Effect	Parameter	Estimate	(s.e.)
Asymptote	β_1	2.0789	(0.0891)
Half-growth	β_2	-8.3862	(3.7291)
Shape parameter	β_3	-6.6679	(2.4019)
Renewal parameter	γ	0.0336	(0.0059)
Residual variance	σ^2	0.0124	(0.0030)

non-linear case (Davidian and Giltinan 1995).

4.4 Concluding Remarks

Exactly like in the linear and generalized linear cases, one can divide non-linear models for repeated measures and otherwise hierarchical data into marginal, random-effects, and conditional models. Somehow, this taxonomy has received less consideration, since focus has been, to a large extent spear-headed up by pharmacokinetic/pharmacodynamic research, on random-effects models.

We have shown here, in a simple but instructive tree growth data example, that one can insightfully construct models within every family. For the aforementioned ubiquity of random-effects models, fitting is less of a problem and dedicated tools have been made available, such as the SAS procedure NLMIXED and the *nlme* library in S-Plus. Conditional models are usually of a transition type in longitudinal designs, in

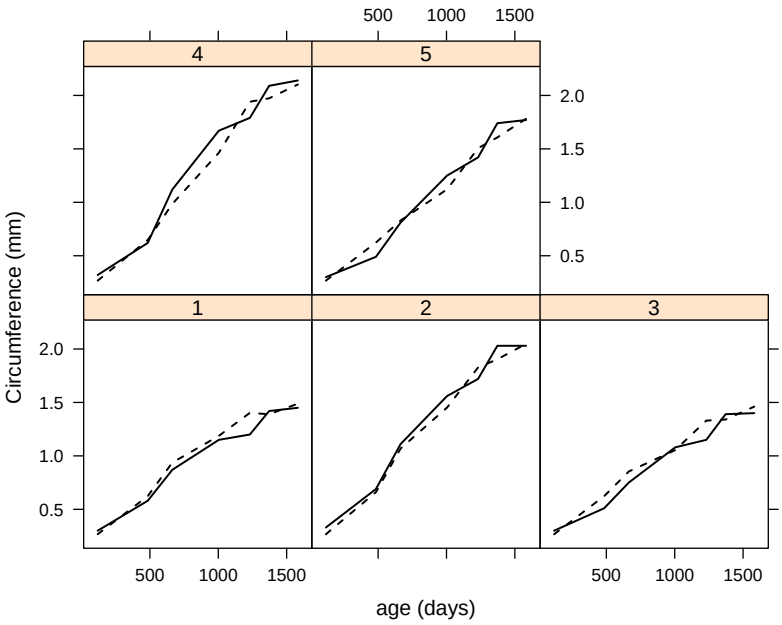

Figure 4.7: *Orange Tree Data. Plot of observed (solid line) and fitted (dashed line) individual profiles of transition model (4.6).*

the sense that measurements are independent of each other, apart from a dependence on a usually small number of previous measurements. This implies that one can fit such models using software for non-linear models, fitted to cross-sectional data.

Marginal models are a bit more challenging, since they require explicitly addressing the covariance structure. When the association is not of direct interest, generalized estimating equations type ideas can be used (Liang and Zeger 1986, Vonesh *et al* 2002). This would be possible for continuous and non-continuous data alike. At any rate, specific parametric functions for the variance and correlation structure need to be considered, such as spatial Gaussian and spatial exponential structures. Practically, the methods were implemented using a user-written function in R. In principle, it is possible to write user-defined programs in any sufficiently generic statistical programming packages with decent matrix manipulation facilities.

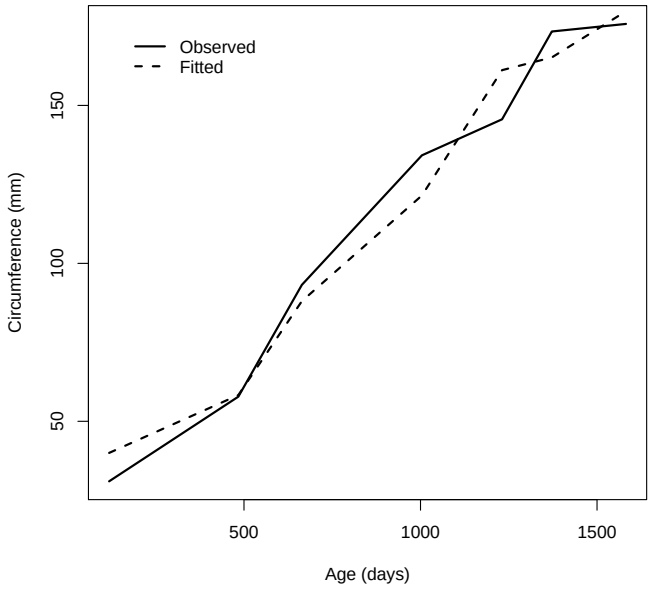

Figure 4.8: *Orange Tree Data. Plot of observed and fitted mean profiles for transition model (4.7).*

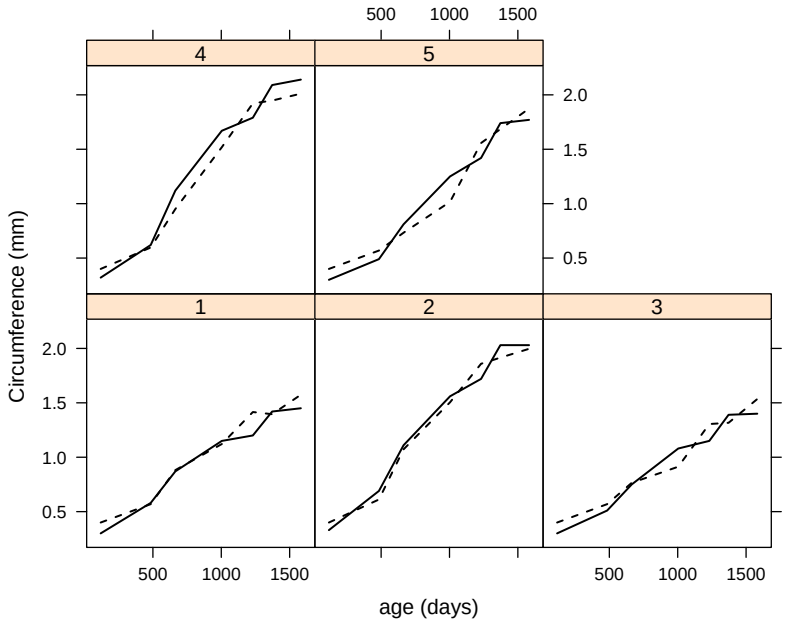

Figure 4.9: *Orange Tree Data. Plot of observed (solid line) and fitted (dashed line) individual profiles of transition model (4.7).*

5

MRI Signal Intensity Processing Based on Non-linear Mixed Modeling to Study Changes in Neuronal Activity

5.1 Introduction

This research is motivated by a study in song birds (Serroyen *et al* 2005). Earlier, Van der Linden *et al* (2002) and Van Meir *et al* (2004) established a novel in-vivo magnetic resonance imaging (MRI) approach to discern the functional characteristics of specific neuronal populations in a strongly connected brain circuitry, the so-called song control system in the songbird brain. The high vocal center (HVC), one of the major nuclei in this circuit, contains interneurons and two distinct types of neurons projecting respectively to the so-called nucleus robustus arcopallii (RA) or to area X. This is graphically represented in Figure 5.1.

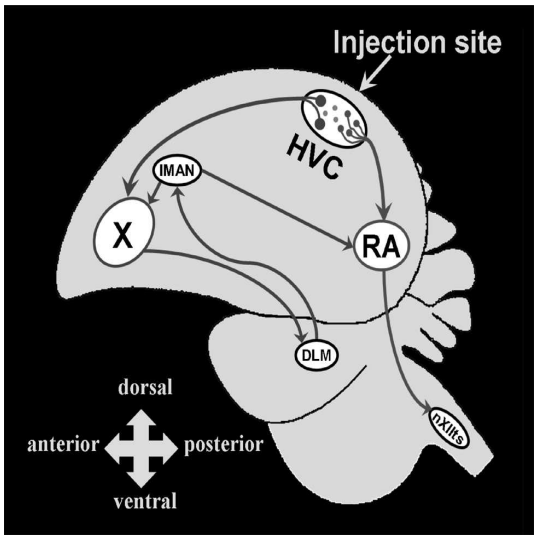

Figure 5.1: *Schematic representation of song control nuclei in the songbird brain.*

These authors analyzed the effect of testosterone on the dynamics of Mn^{2+} accumulation in RA and area X of female starling in individual birds injected with manganese in their HVC. The authors used relatively straightforward curve fitting techniques, combined with analysis of variance ideas. While simple in nature, such techniques ignore dependencies in measurements in the same bird and may be suboptimal. To improve upon this approach, we introduce a novel data processing procedure in which flexible linear and non-linear mixed-effects models are fitted to the MRI signal intensities (SI). The proposed model form allows for both bird-specific effects, as well as important parameters describing the effect of testosterone application. Using this more refined way of analysis we are able to detect testosterone effects that previously have gone unnoticed.

The outcomes analyzed are SI of RA, area X, and HVC. The former two responses could be fitted to obvious non-linear functional (sigmoidal) forms, but these have been used without taking the within-bird correlation into account. (Van der Linden *et al* 2002, Van Meir *et al* 2004). The same is not true for HVC. From graphically inspecting the data, it is clear that conventional linear models may be insufficient. We use a particular pharmacokinetic model,—the two-compartment model,—to which the HVC problem is related (Davidian and Giltinan 1995, Ch. 2) and employ fractional polynomials which extend the collection of classical polynomial shapes (Royston and

Altman 1994, Verbeke and Molenberghs 2000, Ch. 24) by way of sensitivity analysis.

The motivating study for this chapter is introduced in Section 5.2. Section 5.3 is devoted to the non-linear mixed-effects model strategy for RA and area X. Section 5.4 presents the analysis of SI of RA, while the analysis of SI at area X is discussed in Section 5.5. Further methodology, needed to tackle the HVC problem, is presented in Section 5.6. Finally, Section 5.7 is devoted to the analysis of SI of HVC.

5.2 Motivating Study

Ten first-year female starlings were caught in the wild during the winter before February and housed in two indoor cages on a stable 10–14 h light-dark light cycle, selected to maintain birds in a durable state of photosensitivity. All birds were studied by MRI for the first time between March 15 and April 30, 2001. One or two days after the first MRI measurement, the five treated birds were implanted subcutaneously in the neck region with a capsule of crystalline testosterone. The capsule was left empty for the five control birds. Birds were studied by MRI again five to six weeks after the treatment.

Previously, Van der Linden *et al* (2002) have employed the following parametric shape for a bird's profile:

$$SI_{ij}(\text{RA}) = \frac{(\phi_{0i} + \phi_{1i}G_i)T_{ij}^{\eta_{0i}+\eta_{i1}G_i}}{(\tau_{0i} + \tau_{1i}G_i)^{\eta_{0i}+\eta_{1i}G_i} + T_{ij}^{\eta_{0i}+\eta_{1i}G_i}} + \gamma_{0i} + \gamma_1 G_{1i} + \varepsilon_{ij}. \tag{5.1}$$

Here, $SI_{ij}(\text{RA})$ is the measurement of MRI signal intensity for a region of interest (RA) at occasion j for bird i, G_i is an indicator for group membership (1 for testosterone treated birds and 0 otherwise), and T_{ij} is the measurement time, referring to the before *versus* after treatment epoch. The maximal signal intensity, SI_{max}, is denoted by ϕ_{0i} for an untreated bird and $\phi_{0i} + \phi_{1i}$ for a treated one. The time required to reach 50% of this maximum (T_{50}) is τ_{0i} and $\tau_{0i} + \tau_{1i}$, respectively. The shape of the curve is governed by the parameters η_{0i} and $\eta_{0i} + \eta_{1i}$. Finally, ε_{ij} is a measurement error term, typically assumed to follow a normal distribution. Van der Linden *et al* (2002) fitted Model (5.1) to each of the birds under study, and then applied ANOVA to the estimated parameters. Such an approach rests upon the assumption that the measurements within a bird are uncorrelated.

The genesis of this model is rooted in knowledge about Mn axonal transport and changes induced in the bird's brain caused by testosterone treatment. More details

can be found in Brenowitz *et al* (1997), Van der Linden *et al* (2002), Van Meir *et al* (2004).

To properly account for such correlation, we place this model within a mixed-effects framework, where the parameters of the above model are split into fixed and random effects. This will put us in a position to analyze both the SI in RA and SI in area X outcomes. We will need additional development for SI in HVC since for this outcome no generally accepted model form exists.

5.3 A Non-linear Mixed-effects Model

Based on the idea to consider a mixed-effects approach, we will now show how (5.1) can be fitted within this framework.

Let us illustrate this modeling framework by introducing random effects into model (5.1). In this model, all parameters (ϕ_{0i}, ϕ_{1i},...) were assumed to be different from songbird to songbird, since the non-linear model was fitted to each bird separately. We now are able to analyze all data together, separating out averaged (fixed) effects from bird-specific (random) effects, using the following replacements:

$$\phi_{0i} + \phi_{1i} G_i \quad \rightarrow \quad \phi_0 + \phi_1 G_i + f_i, \tag{5.2}$$

$$\eta_{0i} + \eta_{1i} G_i \quad \rightarrow \quad \eta_0 + \eta_1 G_i + n_i, \tag{5.3}$$

$$\tau_{0i} + \tau_{1i} G_i \quad \rightarrow \quad \tau_0 + \tau_1 G_i + t_i. \tag{5.4}$$

The ϕ, η, and τ parameters are fixed effects, while the vector (f_i, n_i, t_i) is a bird-specific vector of random effects, assumed to follow a trivariate normal distribution with mean $\mathbf{0}$ and variance D. Combining model (5.1) with replacements (5.2)–(5.4), we obtain:

$$SI_{ij}(\text{RA}) = \frac{(\phi_0 + \phi_1 G_i + f_i) T_{ij}^{\eta_0 + \eta_1 G_i + n_i}}{(\tau_0 + \tau_1 G_i + t_i)^{\eta_0 + \eta_1 G_i + n_i} + T_{ij}^{\eta_0 + \eta_1 G_i + n_i}} + \gamma_0 + \gamma_1 G_i + \varepsilon_{ij}. \tag{5.5}$$

The parameters retain the meaning they had in (5.1).

Regarding the residual error terms ε_{ij}, we assume them to be mutually independent and independent from the random effects, and to be drawn from a $N(0, \sigma^2)$ distribution.

5.4 Analysis of SI at RA

We will start with a model for the second period, where treatment has been applied. Next, we will switch to the first period, where measurements are taken prior to the application of treatment and consequently no treatment effect would be expected.

Table 5.1: Parameter estimates (standard errors) for the final model, fitted to SI_{ij}(RA) at the first and second period.

		Estimate (s.e.)	
Effect	Parameter	First	Second
	ϕ_0	0.4749 (0.0451)	0.4526 (0.0478)
	η_0	2.5608 (0.1375)	2.1826 (0.0802)
	η_1		0.4285 (0.1060)
	τ_0	3.1737 (0.1658)	2.8480 (0.1761)
$\mathrm{Var}(f_i)$	d_{11}	0.0198 (0.0091)	0.0225 (0.0101)
$\mathrm{Var}(t_i)$	d_{22}	0.2438 (0.1179)	0.2881 (0.1338)
$\mathrm{Var}(n_i)$	d_{33}	0.1457 (0.0787)	
$\mathrm{Cov}(f_i, t_i)$	d_{12}	0.0587 (0.0306)	
$\mathrm{Var}(\varepsilon_{ij})$	σ^2	2.2E-04(2.0E-05)	1.9E-04(1.7E-05)

5.4.1 A Model for SI_{ij}(RA) at the Second Period

We will use model (5.5) to analyze these data. The general form of this model has 8 fixed-effects parameters, and 7 variance components (3 variances in D, 3 covariances in D, and σ^2). However, due to numerical reasons, the initial, most complex model that was fitted to this data consisted of independent random effects. The model is fitted using the SAS procedure NLMIXED, using adaptive Gaussian quadrature.

Backward selection was conducted, using likelihood ratio tests, so as to provide a model which is both parsimonious and at the same time does not exclude important effects. First, the variance d_{33} of n_i was removed. The corresponding test statistic has a $\chi^2_{0:1}$ (denoting a 50:50 mixture of a χ^2_0 and a χ^2_1 distribution, Verbeke and Molenberghs 2000, Sec. 6.3) null distribution ($p = 0.4387$). The reason for using such a mixture is that the null hypothesis lies at the boundary of the parameter space (Stram and Lee 1994). Next, fixed-effect parameters γ_0, γ_1, τ_1, and ϕ_1 are removed. The final model is:

$$SI_{ij}(\mathrm{RA}) = \frac{(\phi_0 + f_i)T_{ij}^{\eta_0+\eta_1 G_i}}{(\tau_0 + t_i)^{\eta_0+\eta_1 G_i} + T_{ij}^{\eta_0+\eta_1 G_i}} + \varepsilon_{ij}. \tag{5.6}$$

Parameter estimates and standard errors of this final model are presented in Table 5.1. Fitted curves, for each bird separately, are displayed in Figure 5.2, while

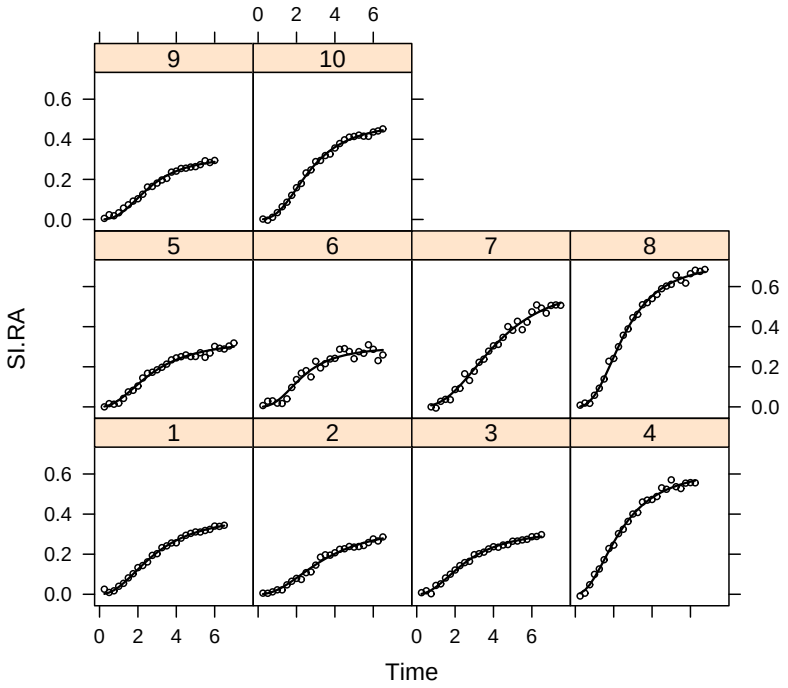

Figure 5.2: *Fitted curves for $SI_{ij}(RA)$ at the second period, for each individual bird separately.*

individual and marginal fitted curves are given in Figure 5.3. These curves support our model selection procedure and confirm the final model is a parsimonious and adequate description of the data.

It is important to note that, in contrast to previous analyzes, we do find a difference between both groups, in the sense that the shape parameter η is different between them ($\chi^2_1 = 15.44$, $p < 0.0001$). The main reason is that there is substantial between-bird variability: there is a bird-specific component in the maximum change in relative signal intensity as well as in the time required to reach 50% of the maximum. By properly accounting for this, we gain power to assess the effect of treatment, explaining why our analysis establishes a treatment effect that went undetected in previous analyzes.

5.4.2 A Model for $SI_{ij}(\mathrm{RA})$ at the First Period

Since testosterone was applied only for the second period, in principle one would not expect a difference during the first period. It is cautious to fit a model for the first

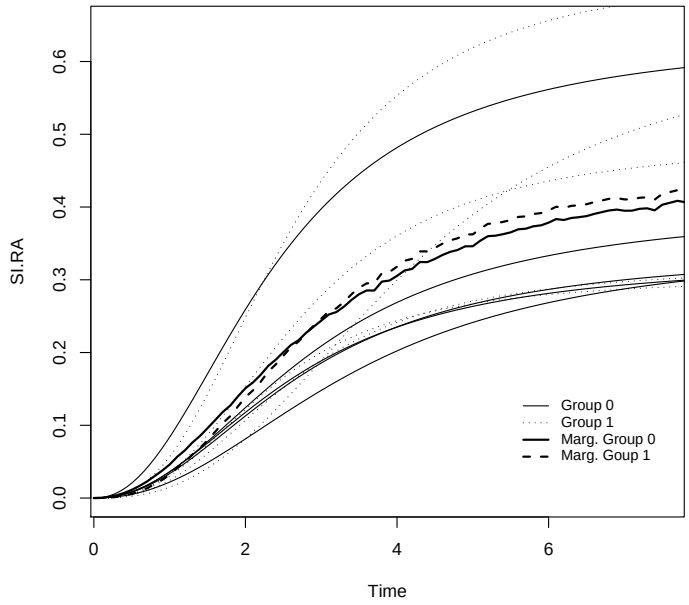

Figure 5.3: *Individual and marginal fitted curves for $SI_{ij}(RA)$ at the second period.*

period as well and to empirically verify whether this is the case indeed. Model (5.5) is used to this effect as well. The results of this model fitting are presented in Table 5.1.

Using the likelihood ratio test statistic, the initial model can be simplified. No random effect can be deleted from the model: the random effects n_i ($\chi^2_{2:3} = 17.24$, $p = 0.0004$), t_i ($\chi^2_{2:3} = 56.55$, $p < 0.0001$) and f_i ($\chi^2_{2:3} = 127.51$, $p < 0.0001$) are all highly significant. However, the covariances between the random n_i and t_i effects, and between the random n_i and f_i effects can be removed ($\chi^2_2 = 1.34$, $p = 0.5105$). Independence between random effects is not attained, because the covariance between f_i and t_i is significantly different from zero ($\chi^2_1 = 10.78$, $p = 0.0010$). In a next step, the fixed effect γ_0 as well as all group specific fixed effects γ_1, ϕ_1, τ_1 and η_1 are found to be nonsignificant and removed from the model. This indicates that there are no group differences in SI_{ij}(RA) before administration of the testosterone treatment. This leads to the following final model for SI_{ij}(RA) at the first period:

$$SI_{ij}(\text{RA}) = \frac{(\phi_0 + f_i)T_{ij}^{\eta_0+n_i}}{(\tau_0 + t_i)^{\eta_0+n_i} + T_{ij}^{\eta_0+n_i}} + \varepsilon_{ij}. \tag{5.7}$$

Note that, while some of the random effects are significant and indicate between-bird

variability, there are no group effects which is to be expected since no testosterone had been applied in the first period.

5.5 Analysis of SI at area X

In line with Section 5.4, we will first fit a model for the second period, where treatment has been applied. Next, we will switch to the first period, where measurements are taken prior to the application of treatment and consequently no treatment effect would be expected.

5.5.1 A Model for SI_{ij}(area.X) at the Second Period

Similarly, Model (5.5) can be fitted to SI_{ij}(area.X). Let us now describe subsequent model simplifications. First, the random n_i effect is removed (implying the removal of d_{33}, d_{13}, and d_{23}), using a likelihood ratio test statistic with value 4.08 and null distribution $\chi^2_{2:3}$ ($p = 0.1914$). Removal of the random t_i effect is not possible since the likelihood ratio equals 54.95 on 1:2 degrees of freedom ($p < 0.0001$). In addition, removal of the covariance between the random t_i and f_i effects is not possible ($\chi^2 = 4.35$ on 1 d.f., $p = 0.0371$). In a next step, the following fixed-effect parameters were removed: γ_0, γ_1, η_1 and τ_1. The fixed-effect ϕ_1 was found to be highly significant and therefore could not be removed from the model ($\chi^2 = 10.58$ on 1 d.f., $p = 0.0011$). This indicates a significant difference between the two groups. The resulting final model is:

$$SI_{ij}(\text{area.X}) = \frac{(\phi_0 + \phi_1 G_i + f_i) T_{ij}^{\eta_0}}{(\tau_0 + t_i)^{\eta_0} + T_{ij}^{\eta_0}} + \varepsilon_{ij}. \tag{5.8}$$

Parameter estimates and standard errors of this final model are presented in Table 5.2.

Looking at Figure 5.4, we can observe that the individually fitted sigmoidal curves describe the data quite well for all birds. Figure 5.5 plots all individual as well as marginal average fitted curves per group. This marginal effect was obtained using sampling based methods. It is clear that SI_{ij}(area.X) is higher for most treated birds (group 1) compared to the untreated birds (group 0), which is confirmed by the significance of the ϕ_1 parameter, the group component of the maximum change in the relative signal intensity.

Table 5.2: Parameter estimates (standard errors) for the final model, fitted to SI_{ij}(area.X) at the first and second period.

		Estimate (s.e.)	
Effect	Parameter	First	Second
	ϕ_0	0.1864 (0.0346)	0.1035 (0.0261)
	ϕ_1		0.1331 (0.0312)
	η_0	2.2167 (0.2578)	2.3462 (0.1498)
	η_1	-0.2925 (0.1566)	
	τ_0	5.5328 (0.8420)	3.7264 (0.3262)
$\text{Var}(f_i)$	d_{11}	0.0039 (0.0021)	0.0043 (0.0022)
$\text{Var}(t_i)$	d_{22}		0.5054 (0.2881)
$\text{Cov}(f_i, t_i)$	d_{12}		0.0344 (0.0229)
$\text{Var}(\varepsilon_{ij})$	σ^2	2.0E-04 (1.8E-05)	1.6E-04 (1.4E-05)

5.5.2 A Model for SI_{ij}(area.X) at the First Period

Model (5.5) can also be fitted to SI_{ij}(area.X) at the first period. The initial, most complex model that was fitted to this data consisted of independent random effects. This means that the covariances in the D matrix were set equal to zero initial due to numerical reasons.

Some effects in the initial model may not be necessary. This leads us to model simplifications, which will be discussed next. First, the random n_i effect is removed ($\chi^2_{2:3} = 0.005$, $p = 0.997$). Secondly, also the random t_i effect can be removed from the model ($\chi^2_{1:2} = 0.44$, $p = 0.656$). Removal of the random f_i effect is not possible ($\chi^2_{0:1} = 247.86$, $p < 0.0001$). In a next step, the following fixed-effect parameters are removed: γ_0, γ_1, ϕ_1 and τ_1. The fixed-effect η_1 was found to be significant and therefore was not removed from the model ($\chi^2 = 4.54$ on 1 d.f., $p = 0.0332$). This indicates a significant difference between the two groups before the testosterone treatment was administered. The resulting final model is:

$$SI_{ij}(\text{area.X}) = \frac{(\phi_0 + f_i)T_{ij}^{\eta_0+\eta_1 G_i}}{(\tau_0)^{\eta_0+\eta_1 G_i} + T_{ij}^{\eta_0+\eta_1 G_i}} + \varepsilon_{ij}. \tag{5.9}$$

Parameter estimates and standard errors of this final model can be found in Table 5.2.

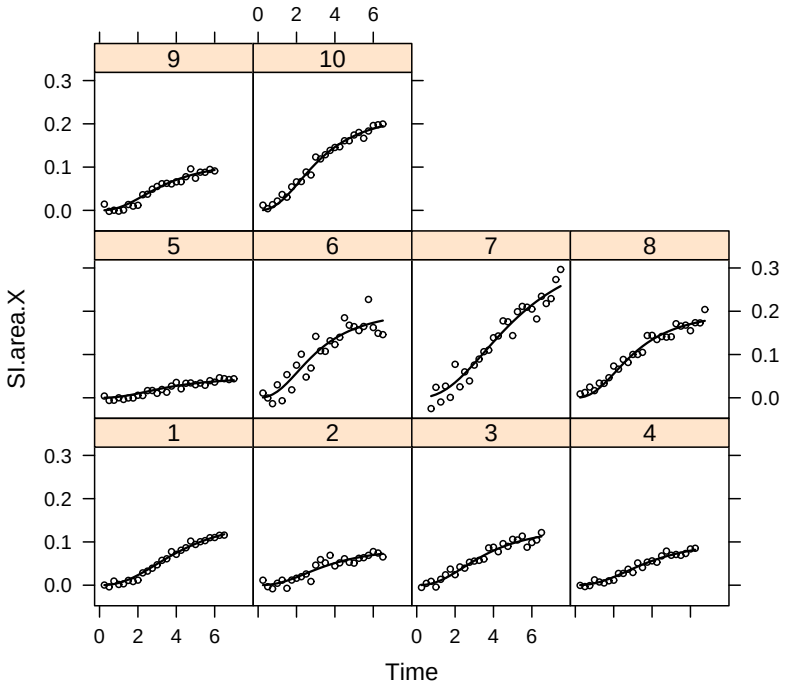

Figure 5.4: *Fitted curves for $SI_{ij}(area.X)$ at the second period, for each individual bird separately.*

In contrast to Model (5.7) for S_{ij}(RA) at the first period, Model (5.9) does contain a significant treatment difference in the shape parameter. While a priori perhaps not expected, one should notice that the corresponding p value is sufficiently borderline and hence the effect is likely due to the occurrence of multiple comparisons.

5.6 Model Strategies for HVC

As stated before, there is a clear view on the hierarchical model needed to analyze RA and area X, rooted in the non-linear model used in the literature and previewed in the previous section. This is less the case for HVC. Therefore, it seems prudent to consider at least two different modeling strategies: (1) fractional polynomials, a pragmatic approach and introduced in Section 2.3, and (2) a plausible, scientifically based modeling strategy, namely a bi-exponential model, which we will introduce now.

The HVC problem is strongly connected to pharmacokinetic theory, which studies

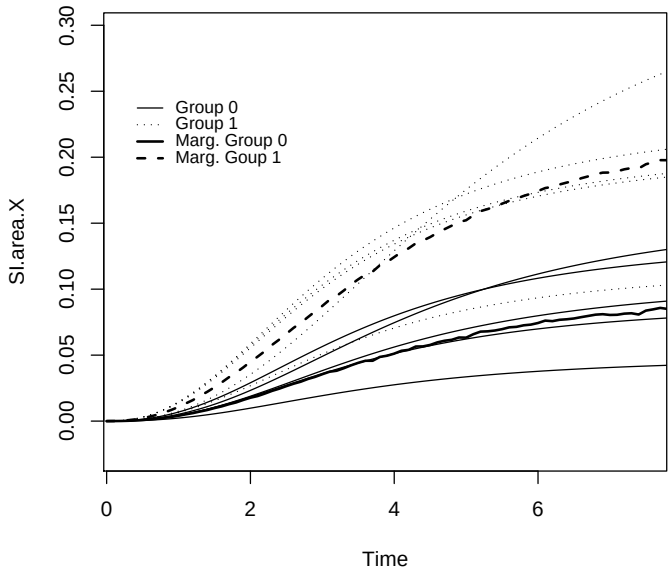

Figure 5.5: *Individual and Marginal fitted curves for $SI_{ij}(area.X)$ at the second period.*

the dispersion of a compound through a living organism. Since the HVC region can be regarded as the *central compartment* from which manganese is dispersed to area X and RA, a two-compartment model seems a reasonable choice. Such model is described by a differential equation with unknown parameters, solution of which yield the so-called two-compartment or *bi-exponential* model (Davidian and Giltinan 1995, Ch. 2):

$$Y_{ij} = \beta_{i1} \exp(-\beta_{i2}\, t_{ij}) - \beta_{i3} \exp(-\beta_{i4}\, t_{ij}) + \varepsilon_{ij}, \tag{5.10}$$

with bird-specific parameters $\beta_{i1}, \beta_{i2}, \beta_{i3}, \beta_{i4} > 0$ and where Y_{ij} is the response (SI) at HVC for bird i at time point j. The β's are the parameters to be estimated and, in what follows, they can be split into fixed and random effects. The parameters β_{i2} and β_{i4} describe the rate of exponential decay in concentration, while β_{i1} and β_{i3} are intercept parameters.

An at first sight odd feature of this model is the functional equality of both systematic terms. The model is identified by virtue of its non-linearity, and problems occur only in very specific and data-analytically unrealistic settings (e.g., when $\beta_2 \equiv \beta_4$).

More details can be found in Davidian and Giltinan (1995, Ch. 2).

To guarantee positive β's, the following reparametrization can be used:

$$Y_{ij} = \exp(\widetilde{\beta}_{i1}) \exp[-\exp(-\widetilde{\beta}_{i2}\, t_{ij})] - \exp(\widetilde{\beta}_{i3}) \exp[-\exp(-\widetilde{\beta}_{i4}\, t_{ij})] + \varepsilon_{ij}. \quad (5.11)$$

In the next section, we will implement this model, together with the fractional polynomial model.

5.7 Analysis of SI at HVC

As mentioned in the previous section, there is no clear view on the parametric shape for a model for the high vocal center (HVC). In this section, we will implement the two proposed forms for such a model. First, the two-compartment model will be applied. Second, we will make use of fractional polynomials, which is a general and flexible family of parametric shapes.

5.7.1 A Model for SI_{ij}(HVC) at the Second Period

Let us first turn to the pharmacokinetic two-compartment model (5.11). Decomposing the $\widetilde{\beta}$'s into fixed and random effects, thereby including a group effect as well, we obtain:

$$\begin{aligned} Y_{ij} \;=\; & e^{(\beta_1+\gamma_1 G_i+b_{1i})} \exp[-e^{(-\beta_2+\gamma_2 G_i+b_{2i})\, t_{ij}}] \\ & -e^{(\beta_3+\gamma_3 G_i+b_{3i})} \exp[-e^{(-\beta_4+\gamma_4 G_i+b_{4i})\, t_{ij}}] + \varepsilon_{ij}. \end{aligned} \quad (5.12)$$

The initial, most complex model consisted of independent random effects. This means that the covariances in the D matrix were assumed to be zero. Parameter estimates and standard errors of this model are presented in Table 5.3. All random effects are needed: the random b_{3i} effect is associated with a $\chi^2_{0:1}$ of 7.83 ($p = 0.0026$), while the other random effects b_{1i}, b_{2i} and b_{4i} are all highly significant ($p < 0.0001$). Deleting all four treatment parameters γ from the model does not lead to a significant decrease in likelihood ($\chi^2_4 = 5.54$, $p = 0.2362$). The resulting final model equals:

$$Y_{ij} = e^{(\beta_1+b_{1i})} \exp[-e^{(-\beta_2+b_{2i})\, t_{ij}}] - e^{(\beta_3+b_{3i})} \exp[-e^{(-\beta_4+b_{4i})\, t_{ij}}] + \varepsilon_{ij}. \quad (5.13)$$

A plot of the individual fitted curves (not shown) indicates that the two-compartment model fits the data adequately. The bottom right graph in Figure 5.6 plots all individual as well as marginal average fitted curves per group based on the initial model. As before, this marginal effect was obtained using sampling based methods. Although the marginal SI_{ij}(HVC) profile of the treated birds (group 1) appears to be positioned

Table 5.3: Parameter estimates (standard errors) for the final two-compartment model, fitted to SI_{ij}(HVC) at the first and second periods.

		Estimate (s.e.)	
Effect	Parameter	First	Second
	β_1	0.8306 (0.0921)	0.7964 (0.1330)
	γ_1	-0.3596 (0.1305)	
	β_2	-2.7425 (0.1974)	-2.7088 (0.0627)
	γ_2	0.2500 (0.2879)	
	β_3	1.2516 (0.6802)	-0.5711 (0.6436)
	γ_3	-2.5839 (1.0243)	
	β_4	0.9094 (0.1846)	1.2311 (0.4906)
	γ_4	0.0051 (0.4076)	
$\mathrm{Var}(b_{1i})$	d_{11}	0.0416 (0.0191)	0.1744 (0.0792)
$\mathrm{Var}(b_{2i})$	d_{22}	0.1810 (0.1045)	0.0241 (0.0157)
$\mathrm{Var}(b_{3i})$	d_{33}	2.0659 (1.2933)	1.5106 (1.1534)
$\mathrm{Var}(b_{4i})$	d_{44}	0.0690 (0.1086)	1.3865 (1.1047)
$\mathrm{Var}(\varepsilon_{ij})$	σ^2	0.0050 (4.7E-04)	1.4E-03 (1.4E-04)

higher than the untreated birds (group 0), the model simplification showed that this difference is not significant.

Let us now apply the fractional polynomial approach to the HVC data. These were fitted with a range of power combinations. The combination associated with the highest likelihood value consists of ln(time) and $\sqrt{\text{time}}$. This leads to the following initial model:

$$\begin{aligned} SI_{ij}(\text{HVC}) &= (\alpha_0 + \alpha_1 G_i + a_i) + (\lambda_0 + \lambda_1 G_i + l_i)\ln(t_{ij}) \\ &\quad + (\delta_0 + \delta_1 G_i + d_i)\, t_{ij}^{0.5} + \varepsilon_{ij}. \end{aligned} \tag{5.14}$$

Parameter estimates and standard errors for the final model are presented in Table 5.4. No significant group differences were found: λ_1 (χ^2_1=1.03, $p = 0.3093$), α_1 (χ^2_1=1.78, $p = 0.1816$), δ_1 (χ^2_1=2.22, $p = 0.1362$). All random effects are needed (all $p < 0.0001$), but the covariance between a_i and l_i can be removed from the model

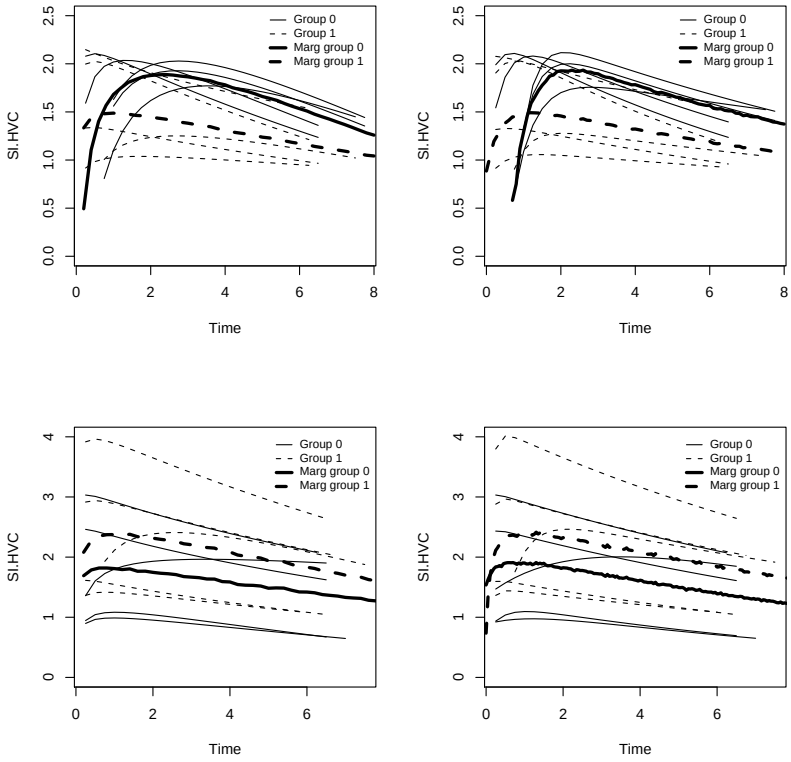

Figure 5.6: *Individual and marginal fitted curves for $SI_{ij}(HVC)$. The first row corresponds to the first period and the second row to the second period. The fractional polynomial model can be found on the left-hand side, the two-compartment model on the right-hand side.*

(χ^2_1=1.14, $p = 0.2857$). The final model equals:

$$SI_{ij}(\text{HVC}) = (\alpha_0 + a_i) + (\lambda_0 + l_i)\ln(t_{ij}) + (\delta_0 + d_i)t_{ij}^{0.5} + \varepsilon_{ij}. \tag{5.15}$$

We note from inspecting the fitted curves (figure not shown), that the individually fitted curves describe the data reasonably well for all birds. The bottom left graph in Figure 5.6 plots all individual as well as marginal average fitted curves per group based on the initial model. As before, this marginal effect was obtained using sampling based methods. Although the marginal SI_{ij}(HVC) profile of the treated birds (group 1) appears to be positioned higher than the untreated birds (group 0), the model simplification showed that this difference is not significant.

Table 5.4: Parameter estimates (standard errors) for the final fractional polynomial model, fitted to SI_{ij}(HVC) at the first and second period.

		Estimate (s.e.)	
Effect	Parameter	First	Second
	α_0	3.4277 (0.2190)	2.9382 (0.3732)
	α_1	-1.3765 (0.3080)	
	λ_0	1.3278 (0.2354)	0.4140 (0.1277)
	λ_1	-1.0455 (0.3286)	
	δ_0	-1.7436 (0.1897)	-0.8395 (0.1247)
	δ_1	1.1783 (0.2631)	
$\text{Var}(a_i)$	d_{11}	0.2091 (0.1052)	1.3857 (0.6227)
$\text{Var}(l_i)$	d_{22}	0.2502 (0.1205)	0.1597 (0.0748)
$\text{Var}(d_i)$	d_{33}	0.1432 (0.0633)	0.1493 (0.0532)
$\text{Cov}(a_i, d_i)$	d_{13}	-0.0632 (0.0351)	-0.2675 (0.1271)
$\text{Cov}(l_i, d_i)$	d_{23}	-0.1743 (0.0858)	-0.1192 (0.0583)
$\text{Var}(\varepsilon_{ij})$	σ^2	0.0076 (0.0007)	0.0022 (0.0002)

It is important to note that, while the two-compartment and fractional polynomial modeling strategies are rather different in nature, both coincide in the conclusions that (1) there is no treatment effect at the second period and (2) there is a strong indication, by way of random effects, for between-bird variability.

5.7.2 A Model for SI_{ij}(HVC) at the First Period

Table 5.3 shows the parameter estimates and standard errors for the pharmacokinetic two-compartment model (5.11). Due to numerical reasons the covariances in the D matrix were assumed to be zero. The random effects b_{1i} ($\chi^2_{0:1} = 4.35$, $p = 0.0185$), b_{2i} ($\chi^2_{0:1} = 32.47$, $p < 0.0001$), b_{3i} ($\chi^2_{0:1} = 26.50$, $p < 0.0001$), b_{4i} ($\chi^2_{0:1} = 3.64$, $p = 0.0282$) are needed in the model. Since no convergence could be achieved when removing either γ_2 or γ_4 from the model, the overall treatment effect was tested by removing all four γ parameters simultaneously. This test indicated a significant difference between the two treatment groups ($\chi^2_4 = 11.70$, $p = 0.0197$). Since no model simplification was possible, the final two-compartment model is identical to

the initial model (5.12). A graph (not shown) shows that this model fits the data adequately for all birds. The marginalized and individually fitted curves can be found in top right graph in Figure 5.6.

As for the second period, fractional polynomials were fitted with the same range of power combinations for the first period. The power combination associated with the highest likelihood is exactly the same as for the second period. Parameter estimates and standard errors of this model are presented in Table 5.4. All random effects are needed in the model: the random effects a_i ($\chi^2_{2:3} = 61.14$, $p < 0.0001$), l_i ($\chi^2_{2:3} = 94.03$, $p < 0.0001$) and d_i ($\chi^2_{2:3} = 47.90$, $p < 0.0001$) are all highly significant. However, the covariance between the random a_i and l_i effects can be removed ($\chi^2_1 = 1.20$, $p = 0.2726$). The fixed-effects α_1 ($\chi^2_1 = 11.06$, $p = 0.0009$), λ_1 ($\chi^2_1 = 7.25$, $p = 0.0071$) and δ_1 ($\chi^2_1 = 12.34$, $p = 0.0004$) all appear highly statistically significant. The model formulation for the final fractional polynomial model for SI_{ij}(HVC) at the first period is equal to the initial model (5.14). The individually fitted curves for this model are not shown but the fit is excellent. The marginalized fitted curves can be found in the top left graph in Figure 5.6. Although no treatment was administered, it is clear that the untreated birds (group 0) have higher response values then the treated birds (group 1).

5.8 Concluding Remarks

In this chapter, we have analyzed data to study the impact of testosterone in songbird brains. Precisely, the impact of Manganese enhanced MRI deduced signal intensity is studied in three areas within the brain: the nucleus robustus arcopallii (RA), area X, and the high vocal center (HVC). Birds with and without testosterone were included in the experiment, and repeated measurements were available in both a pre and post drug administration period. Since the birds' profiles are obviously highly non-linear, an appropriate modeling strategy had to be formulated, taking three aspects into account: (1) within-bird correlation, due to the repeated measures nature of the experiment, (2) the non-linearity of the profiles, and (3) the effect of treatment.

We successfully utilized the non-linear mixed-effects modeling framework to this task. It allows to elegantly incorporate these three aspects within a single modeling paradigm. The framework was able to cope with the different situation arising for RA and area X on the one hand and HVC on the other hand. Indeed, from analogous biological processes, one already deduced a particular parametric form for the profiles in the RA and area X cases. This functional form was incorporated within the modeling framework. We were able to confirm previously established treatment effects,

while additional effects were found with our model which had not been detected with the simpler modes of analyzes, underscoring the strength of the non-linear modeling framework. While Van der Linden *et al* (2002) and Van Meir *et al* (2004) did not find significant differences due to testosterone for SI at RA during the second period, we were able to establish the nature of the bird-specific effects: these are due to strong bird-specific components in the maximum change in relative signal intensity and in the time required to reach 50% of the maximum.

For HVC, the *a priori* situation was less clear and a more exploratory route had to be adopted. First, a fully pragmatic route was considered in the form of fractional polynomials. Second, the similarity between the problem at hand and the two-compartment model in pharmacokinetics was exploited to formulate an alternative model. While the former method is entirely flexible, the second one typically leads to parameters with a clearer interpretation. Further, since no formal theory was available, it is cautious to use two rather different approaches, and one method can be seen as a sensitivity analysis for the other. Having said this, the two-compartment model arguably has a plausible basis as a model generating mechanism. At any rate, for HVC, the results from both fits were perfectly in agreement with each other; they provided a good fit to the empirically observed curves, found no treatment effects at the second period together with strong evidence for the presence of random effects, pointing to considerable bird-specific components, i.e., considerable between-bird variability.

6

Flexible Estimation of Serial Correlation in Linear Mixed Models

6.1 Introduction

Arguably, the linear mixed effects model, as introduced in Section 2.1, has become the most commonly used tool for analyzing continuous, normally distributed longitudinal data. In its general formulation, based on Diggle's (1988) model, four structures can be distinguished: fixed effects, random effects, measurement error and serial correlation. Broadly speaking, serial correlation captures the phenomenon that the correlation structure within a subject depends on the time lag between two measurements. Often, indeed, measurements taken closer in time will exhibit a larger correlation than when they are further apart. Diggle (1988) proposed the *semi-variogram* as a convenient graphical tool to study the overall variance-covariance structure and to separate it into its three constituents. For this tool to be applicable, one has to assume a constant variance over time and restrict the random-effects structure to a random intercept only.

While the above model is oftentimes sufficiently flexible, the need has arisen for

further flexibility. In response, quite some work has been done to relax the model assumptions and/or to extend the model. One strand of research is directed towards flexible covariance-structure modeling (Pan and Mackenzie 2003), while another strand of research has considered spline-based formulations for the random-effects structure (Verbyla *et al* 1999, Ruppert *et al* 2003). Such spline-based models can be implemented, on a routine basis, in the SAS procedure GLIMMIX. Ruppert *et al* (2003) present the necessary S-Plus code to fit their model.

Comparatively less work has been devoted to more flexible serial correlation structures. Diggle and Verbyla (1998) proposed kernel smoothing to provide a nonparametric estimator for the covariance structure without assuming stationarity. As mentioned by these authors, in practice it often becomes essential to impose structural restrictions on the covariance matrix either by smoothing or by fitting a parametric model. In their approach they clearly opt for (nonparametric) smoothing, while we believe that the spline-based approach, which we will introduce further on, strikes a good balance between the parametric and smoothing ideas. This methods combines flexibility with a broad range of potential uses, such as estimation, inference, and prediction over ranges of covariate values; the latter is less straightforward with local methods.

Verbeke, Lesaffre, and Brant (1998) presented an extension of the semi-variogram, allowing for random effects other than merely a random intercept. While elegant in concept, the method is not invariant to the choice of transformation on which it is based. Lesaffre *et al* (2000) used *fractional polynomials* (Royston and Altman 1994) to obtain a flexible yet still fully parametric description of the serial correlation function. This is an appealing idea, worth of further refinement. Consequently, it is taken up in this paper. Next to this, we also propose the use of spline-based modeling of the serial correlation function.

When one is not directly interested in the correlation structure as such, but merely needs to correct for it, the generalized estimating equations (GEE) approach of Liang and Zeger (1986) can be adopted. Even in this situation, however, there are reasons to prefer a mixed model approach. First, this is the case when subject-specific predictions are needed. Second, the full likelihood-based mixed models are preferable when one is confronted with missing data and the assumption of missing completely at random (MCAR, Little and Rubin 2002) is considered too restrictive and one needs to revert to missing at random (MAR).

Section 6.2 introduces the motivating case study for this chapter, of which the analysis is taken up in Section 6.4. Our proposals for flexible serial correlation methodology are described in Section 6.3. Note that the content of this chapter is mainly

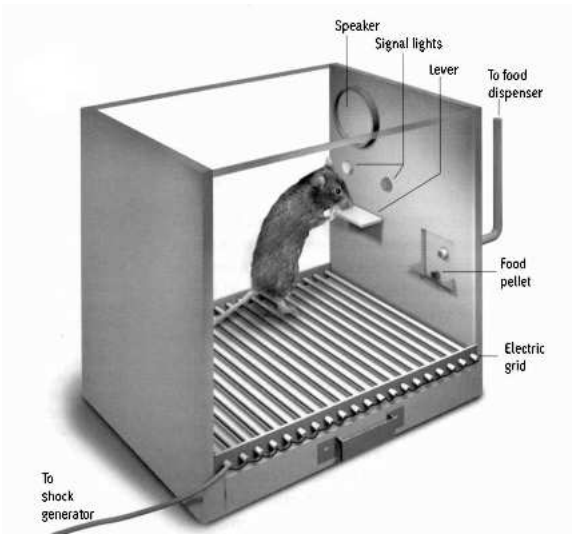

Figure 6.1: *Schematic representation of a Skinner box.*

based on the paper of Serroyen *et al* (2007b).

6.2 Motivating Case Study

Alzheimer's disease (AD) and other dementias have been defined by cognitive and non-cognitive symptomatology. These neuropsychological characteristics are referred to as *Behavioral and Psychological Signs and symptoms of Dementia (BPSD)*. Besides these behavioral disturbances and psychological symptoms described by Reisberg *et al* (1987), demented patients develop changes in eating and drinking behavior. The data introduced in this section were obtained from a study which was set up to investigate behavioral changes in genetically modified mice. These so-called transgenic APP23 mice were genetically engineered based on an animal model for dementia (Vloeberghs *et al* 2004). The specific aim of the study was to investigate whether this valuable mouse model develops eating and drinking disturbances. The APP23 mice were compared with wild-type (WT) control littermates. The total sample size was 85, of which 44 were transgenic mice and 41 were controls.

Eating and drinking behavior were simultaneously recorded for one week by employing so-called Skinner boxes (see Figure 6.1) placed inside ventilated isolation compartments. Each mouse cubicle was equipped with a pellet feeder and a water bottle (optical lickometer) to provide 20 mg dustless precision pellets of the rodent

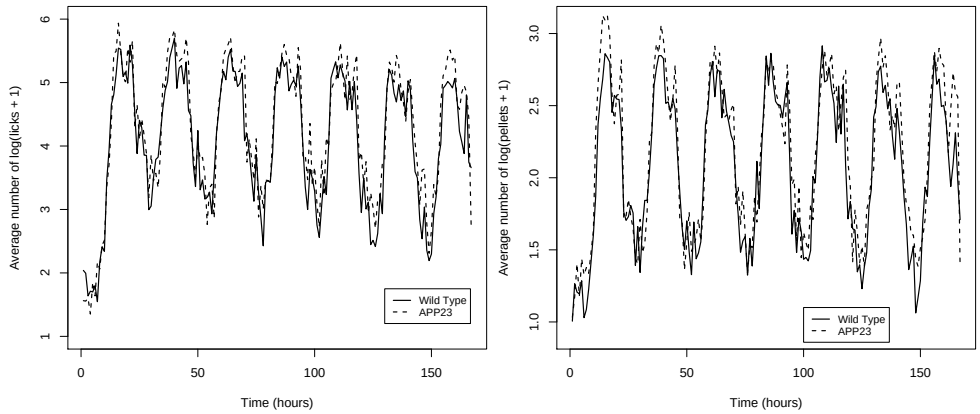

Figure 6.2: *Average evolutions for log-transformed number of licks and pellets over time.*

grain-based formula and tap water. Photocell sensors were used to detect pellet removal, i.e., the number of pellets taken, and the number of licks at the drinking tube. Registration periods typically started Wednesday at 10 am and ended exactly 167 hours later on Wednesday at 9 am. During this 1-week recording period, the *12-hour light—12-hour dark* cycle was continued in the same way as in the facility where mice were previously housed (i.e., lights off at 8 pm).

The response variables were defined as the total number of licks and pellets per hour. Since these responses showed severe right-tailed skewness, they were transformed to log(response + 1). Figure 6.2 presents the average evolutions in the log-transformed number of licks and pellets over time for the WT and APP23 group. A circadian pattern can clearly be observed: the mice show more activity at night (e.g., after 12 hours) compared to during the day (e.g., after 24 hours). Let $\log(l_{ij} + 1)$ be the log-transformed number of licks for mouse i at time point j. The observed individual profiles for $\log(l_{ij} + 1)$ of 5 randomly selected mice are shown in Figure 6.3, while the corresponding observed variance function is portrayed in the left hand panel of Figure 6.4. The variability is not constant and the circadian pattern also seems to be present in the variance structure. The circadian rhythm thus appears to be a dominant biological factor in this experiment. The right hand panel of Figure 6.4 presents the semi-variogram for $\log(l_{ij} + 1)$, based on a random-intercept model with an unstructured model for the mean. However, we should be careful in interpreting this semi-variogram, since the non-constant variance is a clear indication of non-stationarity, thereby rendering the semi-variogram less than trustworthy. It is

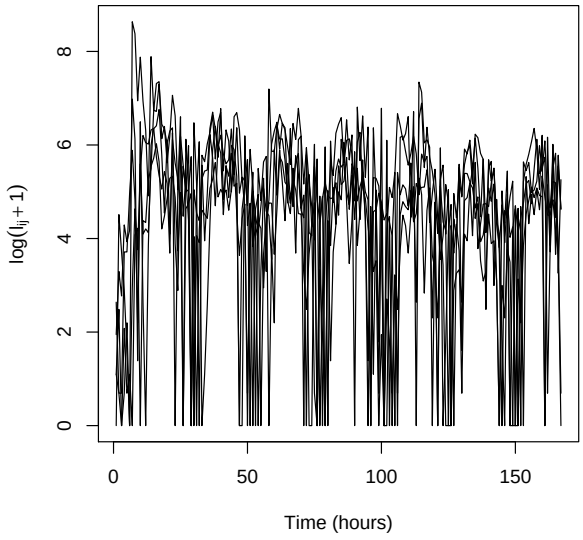

Figure 6.3: *Observed profiles for* $\log(l_{ij}+1)$ *of 5 randomly selected mice.*

therefore prudent to consider the semi-variogram for exploratory purposes only. Nevertheless, there appears to be signs of a circadian trend in this graph as well. This was also the direct motivation for studying if the circadian pattern was also present on the level of serial correlation.

6.3 Flexible Serial Correlation Structures

In analogy with choosing flexible functions and modeling concepts for the fixed and random effects structures, it would also be desirable to dispose of flexible tools for the serial structure. Lesaffre *et al* (2000) proposed fractional polynomials to flexibly model this structure, which, together with some issues surrounding it, will be reviewed briefly in Section 6.3.1. The subsequent section deals with using penalized spline ideas, the concept of which was introduced in Section 2.4, when describing the serial association. All of these methods are rooted in studying the function $g(\cdot)$ in (2.5).

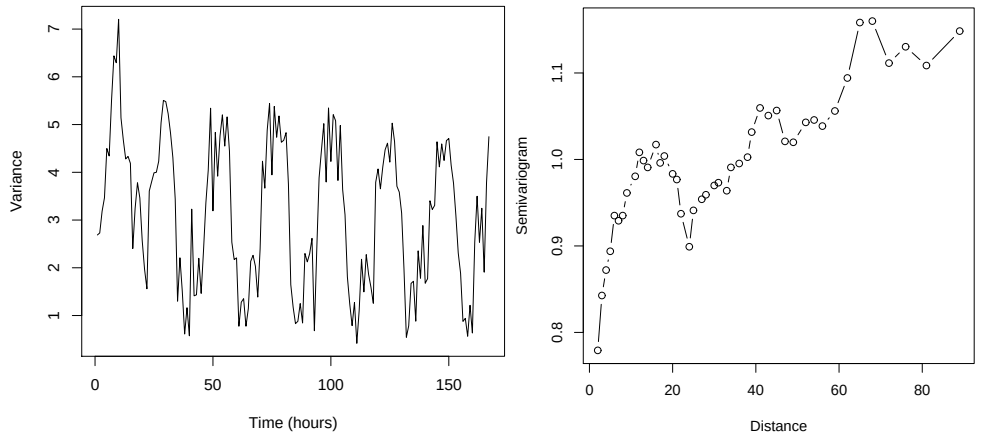

Figure 6.4: *Left hand panel: observed variance function for* $\log(l_{ij}+1)$. *Right hand panel: Semi-variogram for* $\log(l_{ij}+1)$.

6.3.1 Fractional Polynomials

Lesaffre *et al* (2000) applied fractional polynomials to model the serial correlation function $g(\cdot)$. Their model is of the form

$$\tau^2 g(u) = \exp\left\{\phi_0 + \sum_{j=1}^{m} \phi_j \, u^{(p_j)}\right\}. \tag{6.1}$$

This parametrization does not *a priori* ensure that $g(\cdot)$ is a decreasing function, nor that it is has an upper bound of 1. Furthermore, these authors construct a fractional polynomial of degree 4, using the power set $\{0, 0.5, 1, 2\}$. This is in clear contrast with the recommended degree of $m = 2$. As a result, concerns of multicollinearity and model stability can be raised.

The fractional polynomial approach, using the recommended degree of $m = 2$, will be applied to the case study data and the results will be presented in Section 6.4.

6.3.2 Splines

Returning to the smoothing spline ideas laid out in Section 2.4, we are now in a position to formulate a spline model for the serial process:

$$\tau^2 g(u) = \frac{\exp(\phi_0)}{1 + \exp\{\phi_1 + \phi_2 \, \log(u) + \sum_{k=1}^{K} \phi_{k+2}[\log(u) - \log(\kappa_k)]_+\}}. \tag{6.2}$$

This means that ϕ_0 acts as a (strictly positive) intercept, capturing the variance of the serial correlation component, τ^2. Further, ϕ_1 acts as an intercept, ϕ_2 is the

linear slope and $\phi_3, \ldots, \phi_{K+2}$ are the spline coefficients associated with the serial correlation function $g(\cdot)$. The logistic link ensures that the estimated $g(\cdot)$ function stays within the $[0, 1]$ interval. Defining the spline on the log-scale (using argument $\log(u)$) and constraining ϕ_2 to be strictly positive, the serial correlation function satisfies the natural assumption that $\lim_{u \to 0} g(u) = 1$. Additionally, the use of the log-scale improved convergence considerably.

For a particular rich enough set of knots, a penalty term is added to (2.7) to obtain a smooth fit, leading to the following marginal likelihood function:

$$l(\boldsymbol{\theta}) = L_{\text{ML}}(\boldsymbol{\theta}) + \lambda \sum_{k=1}^{K} \phi_{k+2}^2, \tag{6.3}$$

where again the smoothing parameter λ controls the amount of smoothing. In principle, it is conceivable to develop methods for an optimal, data-driven selection of λ, as those briefly discussed in Section 2.4. The main difficulty however is that criteria such as cross-validation and the effective degrees of freedom (being the trace of the so-called smoother matrix in the classical setting) are defined on the scale of the data, and it is no means clear how to translate these concepts to serial correlation. The same holds for the mixed model representation. The aforementioned integration can be done analytically in case of normal random spline effects, but this is no longer true for spline effects in the serial correlation function. In this case the integration can be carried out using conventional numerical integration (e.g., Gaussian quadrature, Laplace approximation) or sampling based (e.g., Monte Carlo Markov chain) methods.

A data-driven optimal selection of the smoothing parameter falls outside of the scope of this paper. Instead, we limit the number of (well-chosen) knots and try out some fixes values for the smoothing parameter. The number and position of knot points for the spline function has to be chosen wisely. This is generally the case in smoothing when using only a limited number of knots, but maybe even more so when modeling the covariance matrix, where parsimonious modeling is very important. Covariance modeling can in general be considered as computationally heavy, and this is particularly true for our case study, where we have 167 measurement occasions. Further, the semi-variogram might help in choosing the location of knot points when the data is balanced and stationarity can be assumed. In practice, splines with different sets of knot points can be fitted to the data. This then gradually gives an idea of the shape of the serial correlation function. Since most pairs of data can be formed for shorter time lags, it is also most sensible to focus at least some knot points in this range.

All analyzes were performed using the statistical software package R (R Development Core Team 2007). The marginal log-likelihood expression for (6.3) was constructed and then maximized using a general purpose numerical optimizer based on a quasi-Newton method. The R code used for fitting the final model can be found in the Appendix at the end of this chapter.

6.4 Analysis of Case Study

First, a model for the mean number of licks will be presented. Then, the approaches as described above in Section 6.3 will be applied to the case study data introduced in Section 6.2.

The mean structure was modeled using the six-parameter function:

$$\log(l_{ij}+1) = \beta_0 + \beta_1\, t_{ij} + \beta_2\, t_{ij}^2 + \beta_3\, \log(t_{ij})^2 + \beta_4\, \sin(\beta_5\, t_{ij}). \tag{6.4}$$

Note that this model includes one non-linear parameter, i.e., the sine frequency parameter β_5. However, since we opted for a fully marginal approach, capturing the within-subject association through the serial correlation function, no random effect enters the likelihood in a non-linear way. This means that no random effect needs to be numerically or approximately integrated out from the likelihood function and therefore, we can still use the marginal likelihood expression (2.7) to fit (6.4). The resulting model fit for the mean number of licks is shown in Figure 6.5.

Let us now turn to the main topic of interest, namely the modeling of the serial correlation function. After fitting models over a range of smoothing parameter values, a small value ($\lambda = 0.01$) was chosen since this improved convergency, while it only had a small impact on the actual fit. A classical exponential function, a fractional polynomial fit, together with spline fits at two different sets of knot points of the serial correlation function $g(.)$ for number of licks is shown in Figure 6.6. The combination of fractional powers that provided the best fit was $\{0.5, 1\}$, i.e. $\sqrt{u}$ and u. The choice of knot point locations did not seem to have a large impact on convergence, as long as they were not positioned too close to each other. For this reason, we chose to set the knot points at 6 hourly intervals. The results obtained with different sets of knot points were all consistent.

The spline fits indicate that the serial correlation function is non-monotone. The (point-wise) 95% confidence bands for this model fit is presented in the left hand graph in Figure 6.7. The non-monotone trend is indeed confirmed by this graph. This fact would go entirely unnoticed with a conventional serial correlation approach

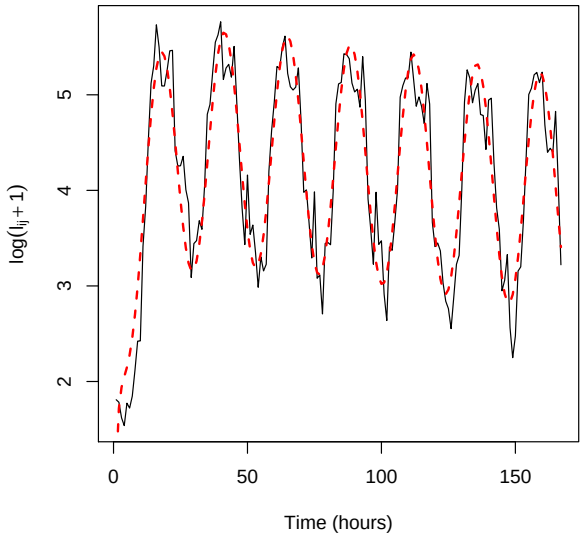

Figure 6.5: *Observed mean profile (solid) and fitted mean model (dashed) for* $\log(l_{ij}+1)$.

and is also missed by the fractional polynomial fit. For example, we now see that the serial correlation is substantially lower for a 12-hour time lag than for one of 24 hours. Very likely, this can be ascribed to the circadian rhythm. This 24-hours pattern could also be observed in Figure 6.2 and Figure 6.3. It therefore plays a role in the mean structure, variance structure and the correlation structure simultaneously. Since a classical serial correlation model would not allow for this, it is conceivable that in such a model the mean structure fit would be distorted, rendering associated inferences less reliable. To check if the estimation of the serial correlation function was not distorted by the non-constant variance, the variance itself was modeled with a sine function. The right hand graph in Figure 6.7 shows the observed and fitted variance function. Although modeling the variance function improved the likelihood considerably, the estimated serial correlation function remained virtually unchanged.

The fact that the one-parameter exponential function cannot detect this type of serial correlation also shows through the difference in log-likelihood between the exponential, and the spline model with knots located at $u = 6, 12, 18, 24, 36$. Precisely, the test statistic equals $2(27395.0 - 27306.9) = 176.2$, which under the null follows a χ^2_6. This represents a considerable improvement in fit of the spline model compared to the exponential model.

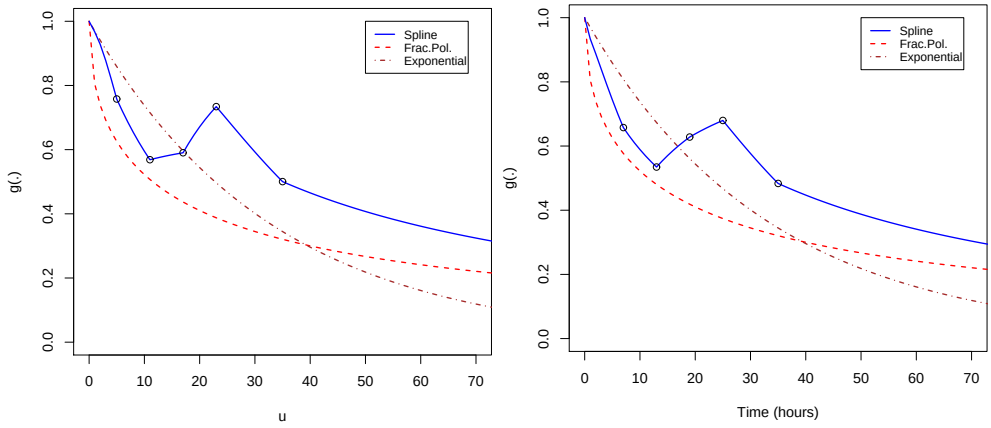

Figure 6.6: *Exponential, fractional polynomial and spline fit of the serial correlation function $g(.)$ for number of licks. Left hand graph: knot points at $u = 6, 12, 18, 24, 36$. Right hand graph: knot points at $u = 8, 14, 20, 26, 36$.*

Note that all knot points were positioned for time lags below 36 hours, confirming that this is the time frame were quite a lot is happening, in contrast to larger lags. Of course, as mentioned before, there is more information in a set of data about shorter lags, since relatively more pairs corresponding to such lags can be formed.

6.5 Concluding Remarks

Flexible serial correlation structures, in agreement with flexible random-effects modeling, are necessary when modeling complex longitudinal profiles, especially with a long period of follow up and/or a large number of measurements within subjects. To this end, we have proposed a spline-based approach. Such a parametric spline approach works acceptably well, as long as the number of knot points is chosen to be relatively small compared to the number of time points. In our case study, we essentially used 5 knot points for 167 follow-up occasions. The choice of the knot points' position, too, is important, both for the quality of the fit as well as for convergence of the updating algorithm.

Convergency can be problematic when fitting an elaborate covariance structure. However, in the analysis of the case study the proposed spline approach actually performed better than some of the simpler serial correlation based models, such as one featuring Gaussian serial correlation. Arguably, the specific parameterization in combination with the added flexibility allows for a better fit and ultimately therefore

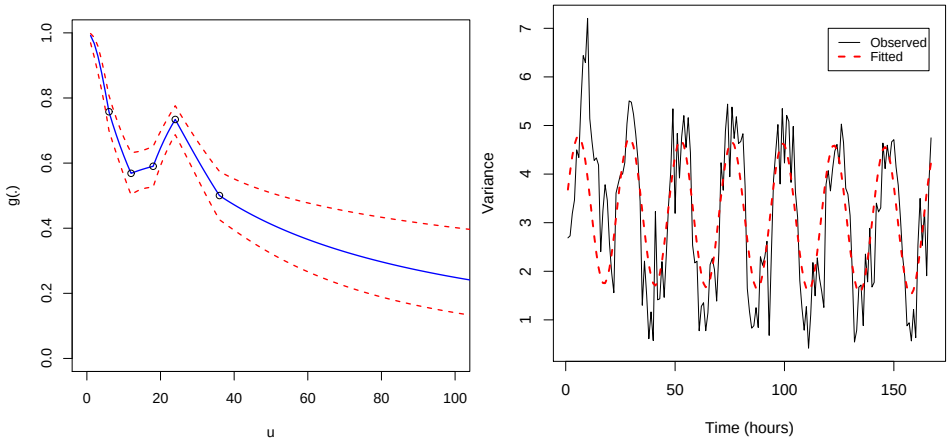

Figure 6.7: *Spline model with knot points at* $u = 6, 12, 18, 24, 36$. *Left hand graph: spline fit with 95% confidence bands. Right hand graph: observed and fitted variance function.*

better convergence.

Since model (6.4) includes a non-linear parameter, i.e., frequency of the sine wave, we opted for a fully marginal approach, thus omitting random effects. Although it is possible to include both random effects and serial correlation in a linear mixed-effects model, this is usually not advisable. Diggle *et al* (1994) and Verbeke and Molenberghs (2000) argue that, in applications, the effect of serial correlation is often dominated by the combination of random effects and measurement error. In practice, this is often reflected by way of estimation problems. Therefore, restricting the model's covariance structure to serial correlation and measurement error will often be a sensible choice.

The presented spline approach provides a flexible alternative to the simple classical models (e.g., the exponential model), which is useful when interest lies directly in the shape of the serial correlation function. The approach can also be used for sensitivity analysis purposes, i.e., checking if an assumed simple serial correlation function is adequate.

For the case study data, the standard errors of the fixed-effects parameter estimates did not change substantially when comparing the relatively simple exponential serial correlation model with the more elaborate spline model. Thus, this particular analysis could not provide evidence for a gain in efficiency related to inference about the fixed-effects parameters. However, a simulation study might shed some light on the fact if this conclusion is true in general.

The choice of the smoothing parameter λ is rather subjective by nature. In our analysis, we chose a small value since this improved convergence, without smoothing out the non-monotone trend in the fitted serial correlation function and without adversely impacting the model's fit.

Admittedly, observations made in a case study are always a bit ad hoc. Therefore, we performed a small simulation study (details not reported), which largely confirmed the finding that adding knot points can improve the fit, while at the same time causing rapid variance increases and having a detrimental influence on convergence.

Appendix

The R code used for fitting model (6.4):

```
#########################
#   Data Manipulation   #
#########################
vloeb <- read.table("c:/temp/data/serial/vloeberghs/pred_licks.txt",header=T)
yi <- as.data.frame(matrix(NA,nrow=length(vloeb[,1]),ncol=3))
schaal <- 100                   # scaling factor
yi[,2] <- vloeb$time / schaal   # time point indicator
yi[,1] <- vloeb$subject         # subject indicator

N <- length(unique(yi[,1]))     # number of subjects
ni <- length(unique(yi[,2]))    # number of repeated measurements
ti <- matrix(yi[,2][yi[,1]==1],ncol=ni,nrow=ni)/schaal
tij <- ti[,1]

######################
#  Adding Gradient   #
######################
 # this function computes the central difference approximation for 'f'
cd <- function (x, f, ..., eps = 1e-04) {
    n <- length(x)
    res <- numeric(n)
    ex <- pmax(abs(x), 1)
    for (i in 1:n) {
        x1 <- x2 <- x
        x1[i] <- x[i] + eps * ex[i]
        x2[i] <- x[i] - eps * ex[i]
        diff.f <- c(f(x1, ...) - f(x2, ...))
        diff.x <- 2 * max(abs(c(x1[i] - x[i], x2[i] - x[i])))
        res[i] <- diff.f / diff.x
    }
    res
 }
```

```
##########################
# Loglikelihood Function #
##########################
loglik71 <- function(theta) {
  beta <- theta[1:6]
  sigma <- theta[7:10]
  alpha <- theta[11:(13+nkn)]

  # Spline-based serial correlation function:
  U <- rep(list(matrix(NA,nrow=ni,ncol=ni)),nkn)
    for (i in 1:nkn) U[[i]] <- (log(abs(ti-t(ti))) - log(kn[i]))*
                                  (log(abs(ti-t(ti))) - log(kn[i])>0)
  cv <- alpha[3]*log(abs(ti-t(ti)))
    for (j in 1:nkn) cv <- cv+alpha[3+j]*U[[j]]
    diag(cv) <- -Inf
  covi <- (exp(alpha[1]))/(1+exp(-(alpha[2]+cv)))

  # Mean function:
   Mu <- beta[1] + beta[2]*tij + beta[3]*tij^2 + beta[4]*log(tij)^2
          + beta[5]*sin(beta[6]*tij)

  # Variance function:
   sigma2 <- sigma[1]^2 + sigma[2]*tij + sigma[3]*sin(sigma[4]*tij)

  # Inverted marginal covariance matrix:
   Wi <- solve(covi + sigma2*diag(ni))

   ll <- 0
  # Marginal LogLikelihood function for multivariate normal distr.
   lli <- function(x) -(log((2*pi)^(-ni/2)) + log(sqrt(det(Wi)))
           + (diag(-(t(x-Mu)%*%Wi%*%(x-Mu)/2)))
           + lambda*sum(alpha[4:(3+nkn)]^2) )
   ll <- sum(tapply(vloeb$loglicks, yi[,1], FUN=lli))
   ll
}

grad.loglik71 <- function(thetas) cd(thetas, loglik71)

# Knot point locations:
kn <- c(6,12,18,24,36)/schaal
nkn=length(kn)
```

```
# Smoothing parameter:
lambda <- 0.01

# Starting values:
Start71 <- c(4.55,-0.042,-0.196,-0.144,-1.24,26.6,
             1.877,-0.172,1.51,26.9,
             -0.5,-7,-4, rep(0,nkn))
loglik71(Start71)

# Using numerical optimizer to minimize negative loglikelihood:
fit71 <- optim(Start71, loglik71, grad.loglik71, hessian=T,method="L-BFGS-B",
               lower=c(0,rep(-Inf,4),0, 0, rep(-Inf,nkn+6)),
               control=list(trace=1,maxit=750,REPORT=5))
```

7

Flexible Modeling of Viral Dynamics in HIV-1-infected Patients

7.1 Introduction

The human immunodeficiency virus (HIV) is a retrovirus that can lead to acquired immunodeficiency syndrome (AIDS), a condition in humans in which the immune system begins to fail, leading to life-threatening opportunistic infections. HIV infection in humans is now considered a global pandemic with infection rates as high as 25% in southern and eastern Africa. As of January 2006, the Joint United Nations Programme on HIV/AIDS (UNAIDS) and the World Health Organization (WHO) estimate that AIDS has killed more than 25 million people since it was first recognized on December 1, 1981, making it one of the most destructive pandemics in recorded history (UNAIDS 2006).

The high mutation rate of HIV is the main reason why the human immune system cannot gain control over an HIV infection and why it is so difficult to develop a cure or vaccine against HIV.

There are two species of HIV that infect humans: HIV-1 and HIV-2. HIV-1 is

more virulent and it is the cause of the majority of HIV infections globally. HIV-2 is less transmittable and is largely confined to West Africa.

Current treatment for HIV infection consists of highly active antiretroviral therapy (HAART), consisting of a combination of protease and reverse transcriptase inhibitors. HAART allows the stabilization of the patients symptoms and viremia, but it neither cures the patient, nor alleviates the symptoms, and high levels of HIV-1, often HAART resistant, return once treatment is stopped.

Our scientific aim is to model the evolution of HIV-1 viral concentrations over time. More specifically, the goal of the analysis reported in this chapter is twofold. First, the model building steps and specification of starting values need to be systematized when fitting the bi-exponential model to viral load data with SAS procedure NLMIXED. Second, we wanted to find a suitable model for the *rebounders*, a special subgroup of patients that will be described later.

The motivating case study is presented in Section 7.2, while some fundamental concepts from the field of viral dynamics are reviewed in Section 7.3. A model building strategy to fit the bi-exponential model on viral load data will be introduced in Section 7.4. Finally, the results of applying the model building strategy to the case study data are shown in Section 7.5.

7.2 Motivating Case Study

The data to be analyzed come from pooling three clinical trials on Prezista, also known as *darunavir*, previously TMC114, a new protease inhibitor (PI) developed by Tibotec (Spinosa-Guzman *et al* 2007). The first two trials, TMC114-C202 and TMC114-C213, are randomized, controlled, partially-blinded Phase II trials aimed at determining the dosing, antiviral activity, safety, and tolerability of *darunavir*, formulated as an oral tablet, and administered with a low dose of *ritonavir* (RTV). Participants were required to have a baseline HIV RNA (viral load) of greater than 1000 copies/ml, had previous treatment with PIs, non-nucleoside reverse transcriptase inhibitors (NNRTIs), and nucleoside reverse transcriptase inhibitors (NRTIs), and have at least one primary PI mutation at screening, and to be currently taking a stable PI-containing regimen at screening for at least 8 weeks prior to study entry. After completion and approval of the 24 week dose finding analysis, all subjects on darunavir were instructed to switch to the recommended dose of 600 mg BID[1]. The subjects in the control group continued their therapy unchanged until the end of the

[1] twice daily

trial.

Subjects who did still benefit from their assigned antiretroviral regimen beyond 96 weeks of treatment were offered the possibility to extend their treatment with an additional 48 weeks. Subjects had to provide their explicit informed consent for this extension. Original analyzes included 319 patients in TMC114-C202 and 318 patients in TMC114-C213.

The third trial (TMC114-C215) is an open-label trial in HIV-1 infected subjects who failed trial treatment in the TMC114-C202 or TMC114-C213 trial and who may derive benefit from TMC114 therapy, as judged by the investigator. The trial evaluates the safety and tolerability of TMC114/RTV, in addition to an individually optimized background antiretroviral therapy. Additionally, the antiviral activity is evaluated. 452 subjects were accrued in trial TMC114-C215.

As mentioned before, a large number of patients underwent a dose switch during the study and switched to the recommended dose of 600 mg BID. Therefore, to reduce the heterogeneity in the data, only the patients who received the recommended dose from the start and control patients were retained for analysis. The control patients received an investigator selected PI on top of an optimised background regimen. This brings the final sample size to 591 patients. The number of available measurements per subject ranged from 1 to 24, with a median of 14.

Before presenting the general model building strategy for analyzing viral load data, we will now review some fundamental concepts from the field of viral dynamics.

7.3 Viral Dynamics

Viral dynamics is a relatively new field of study that relies on mathematical models to describe the evolution of virus levels in the blood plasma, the so-called viral load, over time (Nowak and May 2000). Basic models of viral dynamics describe the interaction between (1) cells susceptible to infection (target cells), (2) different classes of infected cells, and (3) virus.

Within a few months of HIV infection, patients typically attain a constant or set-point viral load, which is roughly maintained for years. The patients enter a steady state of chronic infection (Di Mascio *et al* 2004). When this steady state is perturbed with potent HAART therapies, the concentration of HIV-1 in blood plasma shows an initial rapid exponential decay, usually followed by a slower second phase of exponential decay (Perelson *et al* 1997). The first phase stems from the emptying of the pool of actively virus-producing infected T-cells, the replenishment of which

is downsized by the treatment, while the second phase can be attributed to gradual release of the virus from *secondary* sources that have a longer half-life, such as infected tissue macrophages, dendritic cells or activation of latently infected lymphocytes.

The biphasic decay in viral load can be described by a bi-exponential model:

$$V_{ij} = V_0 \left[A\, e^{-\alpha\, t_{ij}} + B\, e^{-\beta\, t_{ij}}\right] + \varepsilon_{ij}, \tag{7.1}$$

where V_{ij} is the viral load for subject i at measurement occasion j, V_0 is the baseline viral load and t_{ij} is the time indicator. The parameters α and β describe the rate of exponential decay, while A and B are intercept parameters. Finally, ε_{ij} is the residual error term. This model is very similar to the two-compartment model used in Section 5.6, only with the addition of V_0, which is basically a scaling factor. Because the data are of longitudinal form, we will embed this model in a mixed-effects setting.

After the biphasic decrease in viral load levels, some subjects show a *rebound*, i.e., a sudden rise in viral load levels during treatment. This rebound is generally caused by the emergence of a drug-resistant virus strain. Non-adherence to the therapy may also come into play, with partial adherence a possible factor of the faster accumulation of drug-resistant mutations (Sethi *et al* 2003). We propose to extend the bi-exponential model by adding a third phase, allowing us to capture the rebound: a logistic growth process (Pinheiro and Bates 2000, p. 274). The three-parameter logistic growth model can be formulated as

$$y(x) = \frac{\theta_1}{1 + e^{\theta_2 + \theta_3\, x}}, \tag{7.2}$$

where θ_1 is the upper asymptote, θ_2 plays the role of an intercept and θ_3 is the slope parameter. Combining the bi-exponential and logistic growth model leads to the following rebound model:

$$V_{ij} = V_0 \left[A\, e^{-\alpha\, t_{ij}} + B\, e^{-\beta\, t_{ij}} + \frac{\theta_1}{1 + e^{\theta_2 + \theta_3\, t_{ij}}}\right] + \varepsilon_{ij}. \tag{7.3}$$

All analyses were performed using the NLMIXED procedure in SAS 9.1.

7.4 Model Building Strategy

The first goal of the reported analysis is to develop a general model building strategy to fit the bi-exponential model on viral load data, including the selection of proper starting values.

We will now present this model building strategy, based on a two-stage model building approach. The developments are made so as to facilitate implementation in the SAS procedure NLMIXED. Precisely, the following steps are taken.

1. Individual bi-exponential model fit per patient (with grid search for starting values)
2. Summarize parameter estimates from individual fits:
   (a) Fixed effects: median
   (b) Random effects (standard deviation): interquartile range
   (c) Treatment effect: summarize individual parameters by treatment group
3. Overall bi-exponential model fit (ignoring longitudinal nature of the data):
   (a) Use the median of individual parameter estimates as starting values
   (b) Inflate the residual variance, to accommodate the impact resulting from between-subject variability
4. Introduce random effects:
   (a) Use interquartile distance of individual parameter estimates as starting values
   (b) Add random effects one-by-one in decreasing order of interquartile distance value
   (c) Likelihood ratio tests to check whether random effects are needed
   (d) If the model becomes unstable or otherwise does not converge:
      i. Switch the numerical integration method to non-adaptive Gaussian quadrature, as opposed to SAS' default method, which is adaptive Gaussian quadrature
      ii. Switch the numerical integration method to *first order optimization* (FIRO) and use the estimates obtained as starting values for Gaussian quadrature
      iii. Reparameterize the model or rescale the parameters to ensure that all estimates are approximately of same order of magnitude
      iv. Use a grid of starting values for covariances and random effect variances
      v. Omit the last added random effect from model and try adding random effect elsewhere in the model (for example, random β instead of B)
5. Introduce (fixed) treatment effects: use likelihood ratio tests to check if treatment effects are significant

Let us now apply this strategy to the case study introduced in Section 7.2.

7.5 Application to TMC114 data

Since viral loads generally evolve exponentially over time, they are most commonly expressed on a logarithmic scale. The observed mean log-transformed (base 10) viral load profile and the individual profiles of 20 patients randomly selected from the pooled trials are presented in Figure 7.1. The mean profile exhibits the biphasic pattern, as discussed in the previous section. During the first few weeks of treatment, there is a clear, sharp decrease in mean log(viral load), followed by a slower phase. The individual profiles plot indicates that there is substantial variability, both within and between subjects. The viral load of most patients drops below the detection limit within the first few weeks of treatment. Some patients have highly irregular profiles and do not show the typical biphasic pattern. However, this is not uncommon for the type of patients selected for these trials, i.e., patients who did not respond well to previous treatments (salvage patients). Another important aspect of the data is that there is a detection limit at 50 HIV RNA copies/ml, corresponding to a log(viral load) value of 1.7. This means that the data are left censored. Nevertheless, in our analysis we will ignore this aspect, since it is less than straightforward to extend the bi-exponential mixed-effects model to account for censoring.

7.5.1 Bi-exponential Model

The model building strategy of Section 7.4 will now be applied to the TMC114 data. The patients who show a rebound are excluded in this part of the analysis, since this behavior is believed to be caused by another underlying biological mechanism. Section 7.5.2 deals with finding a suitable model for these patients.

Before the model building strategy was put into practice, the bi-exponential model was reparameterized. To start, both sides of the model equation in (7.1) are transformed to a logarithmic scale with base 10, to improve numerical stability. In the classical bi-exponential model, the parameters A, B, α, and β are considered to be strictly positive (Davidian and Giltinan 1995). However, since a number of patients have irregular profiles and do not show a monotonically decreasing biphasic pattern, the α and β are not parameterized to be strictly positive, ensuring sufficient model flexibility. Finally, the residual variance is estimated using the parameterization $Var(\varepsilon_{ij}) \equiv \sigma_\varepsilon^2 = \exp(2 \cdot \xi)$, where ξ represents the log-transformed standard deviation of the residual error term.

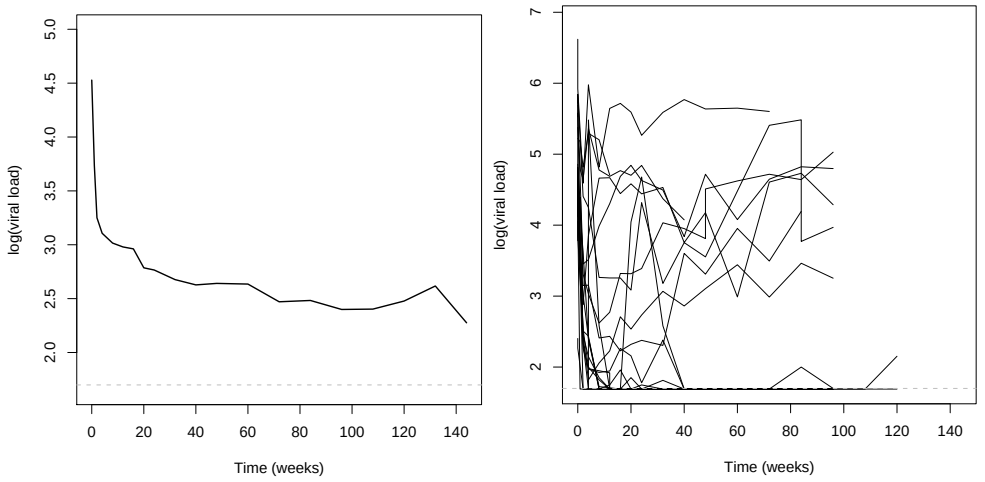

Figure 7.1: *Left hand panel: observed mean log(viral load) profile. Right hand panel: individual profiles of 20 patients randomly selected from the pooled trials. Note that about one-third of the patients do not show the typical bi-phasic pattern. Dashed grey line represents detection limit.*

The first model building step is to fit a bi-exponential model to each patient separately (*by subject approach*). The corresponding model is given by

$$\log_{10}(V_{ij}) = \log_{10}\left(V_0\,[e^{A_i}\,e^{-\alpha_i\,t_{ij}} + e^{B_i}\,e^{-\beta_i\,t_{ij}}]\right) + \varepsilon_{ij}. \tag{7.4}$$

Table 7.1 provides an overview of the summarized parameter estimates for model (7.4). To facilitate interpretation, this table also includes back-transformed median estimates, consistent with the parameterization of model (7.1). After fitting the overall bi-exponential regression model, the next step in the model building process is to introduce random effects into the model. Adding four random effects leads to the following model expression:

$$\log_{10}(V_{ij}) = \log_{10}\left(V_0\,[e^{(A+a_{1i})}\,e^{-(\alpha+a_{2i})\,t_{ij}} + e^{(B+b_{1i})}\,e^{-(\beta+b_{2i})\,t_{ij}}]\right) + \varepsilon_{ij}. \tag{7.5}$$

To further improve numerical stability, the random-effect covariance parameters are expressed in terms of inverse Fisher's Z transformed correlations. For example, the covariance between the first and second random effect is parameterized as

$$d_{12} = r_{12} \cdot (\sigma_{11} \cdot \sigma_{22}),$$

Table 7.1: Bi-exponential model. Summary statistics for estimates obtained in by subject approach. Medians are also back-transformed.

Parameter	Median	IQR[1]	Mean	Std[2]	Transformation	Transformed Estimate
A	-0.1649	0.4939	-0.5035	1.7482	e^A	0.8480
α	1.6193	1.7730	2.3500	3.0524	α	1.6193
B	-5.0229	4.1073	-6.7810	13.9849	e^B	0.0066
β	0.0008	0.0208	-0.1627	0.7823	β	0.0008
σ_ε^2	0.0284	0.0510	0.0610	0.1034		

[1] Interquartile Range [2] Standard Deviation

where

$$r_{12} = \frac{\exp(c_{12}) - 1}{\exp(c_{12}) + 1}.$$

Also, σ_{11} and σ_{22} are the random-effect standard deviation parameters, and c_{12} is the transformed covariance parameter to be optimized. Parameter estimates and standard errors for model (7.5) are provided in Table 7.2. The correlations between the random-effect parameters are relatively weak. The strongest correlation, i.e., $r_{13} = -0.8189$, is negative and occurs between the intercepts of the two exponential phases. This means that a high (low) intercept in one phase is compensated by a low (high) intercept in the other phase.

Comparing Table 7.1 and Table 7.2, it can be concluded that the median and interquartile range can be adequately used as starting values for fixed- and random-effect parameters of model (7.5), respectively. The estimated random effect standard deviations are consistently smaller than the interquartile range, which is probably due to *shrinkage* (Verbeke and Molenberghs 2000, p. 80). This phenomenon is most apparent for standard deviations (Table 7.1), which are even considerably larger than the corresponding interquartile ranges. The standard deviations are inflated, owing to an unstable model fit for some individual patients, resulting in outlying parameter estimates for these patients.

The last model building step involves introducing treatment effects into the bi-exponential model. As mentioned earlier, only patients who did not undergo a dose switch were retained in the analysis. This means that two treatment groups can be identified: patients receiving the recommended dose from the start (treatment group)

Table 7.2: Bi-exponential model. Parameter estimates and standard errors for random effects model. Fixed-effect parameter estimates are also back-transformed.

Effect	Parameter	Estimate	(s.e.)	Transformation	Transformed Estimate
	A	-0.3836	(0.0366)	e^A	0.6814
	α	1.7037	(0.0546)	α	1.7037
	B	-4.0841	(0.0328)	e^B	0.0168
	β	0.0030	(0.0007)	β	0.0030
Std(b_{1i})	σ_{11}	1.7304	(0.0228)		
Std(b_{2i})	σ_{22}	0.0112	(0.0006)		
Std(a_{1i})	σ_{33}	0.1426	(0.0309)		
Std(a_{2i})	σ_{44}	0.7906	(0.0267)		
Cor(b_{1i}, b_{2i})	r_{12}	-0.2898	(0.0402)		
Cor(b_{1i}, a_{1i})	r_{13}	-0.8189	(0.1459)		
Cor(b_{1i}, a_{2i})	r_{14}	0.0233	(0.0057)		
Cor(b_{2i}, a_{1i})	r_{23}	0.2212	(0.2651)		
Cor(b_{2i}, a_{2i})	r_{24}	0.0350	(0.0030)		
Cor(a_{1i}, a_{2i})	r_{34}	-0.1086	(0.0284)		
Var(ε_{ij})	σ^2_ε	0.1006	(0.0021)		

and the patients receiving an investigator selected PI (control group). Therefore, the model was extended by adding a treatment indicator to all four fixed effects, which takes on the value 0 for the control group and 1 for the treatment group. The random a_{1i} effect was estimated close to zero and therefore deleted from the model. The estimates and standard errors of the resulting model fit are presented in Table 7.3. The parameters A_t, α_t, B_t, and β_t represent the treatment effect, i.e., the difference in parameter estimate for the treatment group compared to the control group.

7.5.2 Rebound Model

As mentioned in Section 7.3, some subjects show a *rebound* after the biphasic decrease in viral load levels. To capture this rebound, the bi-exponential model is extended by

Table 7.3: Bi-exponential model. Parameter estimates and standard errors for random effects model including treatment effects. Fixed-effect parameter estimates are also back-transformed.

Effect	Parameter	Estimate	(s.e.)	Transformation	Transformed Estimate
	A	-0.5198	(0.0890)	e^{A}	0.5946
	A_t	0.3340	(0.0929)	e^{A+A_t}	0.8304
	α	1.7712	(0.0490)	α	1.7712
	α_t	-0.2028	(0.0150)	$\alpha + \alpha_t$	1.5684
	B	-2.9034	(0.0578)	e^{B}	0.0548
	B_t	-1.6907	(0.0706)	e^{B+B_t}	0.0101
	β	-0.0079	(0.0013)	β	-0.0079
	β_t	0.0093	(0.0015)	$\beta + \beta_t$	0.0014
Std(b_{1i})	σ_{11}	1.4069	(0.0195)		
Std(b_{2i})	σ_{22}	0.0108	(0.0005)		
Std(a_{2i})	σ_{44}	0.7719	(0.0209)		
Cor(b_{1i}, b_{2i})	r_{12}	-0.3709	(0.0354)		
Cor(b_{1i}, a_{2i})	r_{14}	0.0269	(0.0019)		
Cor(b_{2i}, a_{2i})	r_{24}	0.1132	(0.0045)		
Var(ε_{ij})	σ^2_ε	0.0950	(0.0020)		

adding a logistic growth phase.

The case study data contained a variable indicating whether a patient is a rebounder or not, based on certain clinical criteria. After visual inspection of the individual viral load profiles, some additional patients were identified as being a rebounder. This brings the total number of rebounders to 228.

In line with Section 7.5.1, both sides of the model equation in (7.3) were transformed to a logarithmic scale with base 10. Opposed to the implementation of the bi-exponential model, β was exponentiated in the rebound model, forcing the second phase to decrease over time. Not imposing this restriction can lead to an unstable model, since the logistic part of the model and the second phase can both try to capture the observed rebound. This happens mainly when the patient rebounds early

Table 7.4: Rebound model. Summary statistics for estimates obtained in by subject approach. Medians are also back-transformed.

Parameter	Median	IQR[1]	Mean	Std[2]	Transformation	Transformed Estimate
A	-0.0757	0.3449	-0.2971	1.5569	e^A	0.9271
α	2.1388	1.8559	3.1153	3.1414	α	2.1388
B	-3.8732	2.6865	-5.2699	5.4575	e^B	0.0208
β	-2.7069	2.2474	-4.2403	5.3995	e^β	0.0667
θ_1	4.4987	1.2285	5.3890	9.9694		
θ_2	25.0648	25.6249	36.4454	37.4573		
θ_3	-1.4920	4.1498	-3.8585	6.4588		
σ^2_ε	0.0558	0.1158	0.1418	0.2440		

[1] Interquartile Range [2] Standard Deviation

during therapy, meaning before or shortly after the start of the second phase, implying a lack of data on this second phase. Furthermore, θ_1 was parameterized as 10^{θ_1} to obtain estimates for the asymptote on the log(viral load) scale and to increase numerical stability. This leads to the following model parametrization:

$$\log_{10}(V_{ij}) = \log_{10}\left(V_0 \left[e^A\, e^{-\alpha\, t_{ij}} + e^B\, e^{-\exp(\beta)\, t_{ij}} + \frac{10^{\theta_{1i}}}{1 + e^{\theta_{2i} + \theta_{3i}\, t_{ij}}}\right]\right) + \varepsilon_{ij}. \quad (7.6)$$

Using the model building strategy described earlier, the rebound model is first fitted separately for each rebounding patient. Table 7.4 provides an overview of estimates from this step. A typical rebound profile and predicted curve for one selected patient is displayed in the left hand panel of Figure 7.2. The right hand panel shows the predicted curve with underlying bi-exponential and logistic curves for the same patient. The model prediction captures the rebound pattern reasonably well.

The mixed-model approach, using random effects to capture the between-subject variability in the logistic growth curve parameters (θ_1, θ_2, and θ_3) did not lead to acceptable model fits on the level of the individual subjects. This can probably be attributed to the restrictive normality assumption on the empirical Bayes (EB) estimates and the resulting shrinkage effect. Therefore, it was decided to treat the θ-parameters as covariates in the final model, fixing their values at the estimates obtained in the individual fit per patient. The repeated nature of the data was taken

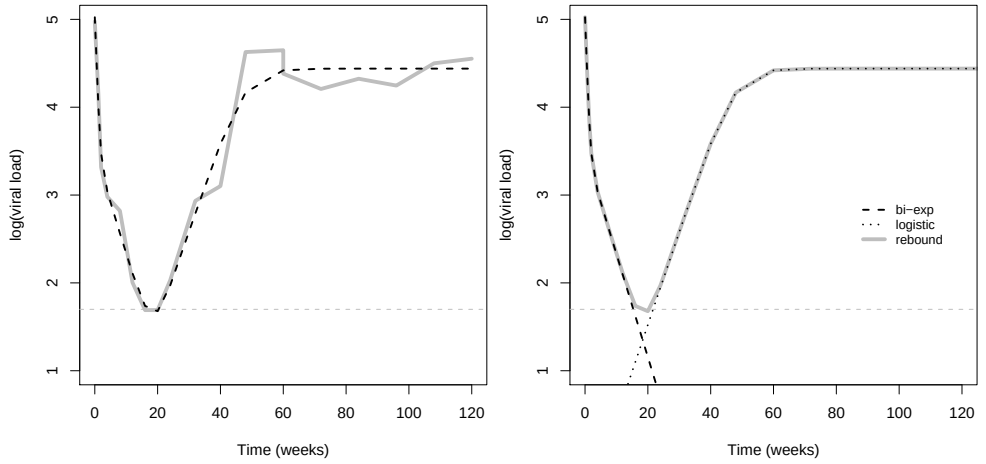

Figure 7.2: *Left hand panel: Observed profile (solid) and predicted curve (dashed) for one selected patient. Right hand panel: Predicted curve, and underlying bi-exponential and logistic curves. Dashed grey line represents detection limit.*

into account by including random effects in the bi-exponential part of the model.

Three random effects were found to be statistically significant, i.e., a random intercept, b_{1i}, and slope, b_{2i}, in the second phase of the bi-exponential part of the model, and a random slope, a_{2i}, in the first phase. The final model thus becomes

$$\log_{10}(V_{ij}) = \log_{10}\Big(V_0\ \Big[e^{A}\ e^{-(\alpha+a_{2i})\ t_{ij}}\ +\ e^{(B+b_{1i})}\ e^{-(\exp(\beta+b_{2i}))\ t_{ij}} + \frac{10^{\theta_{1i}}}{1+e^{\theta_{2i}+\theta_{3i}\ t_{ij}}}\Big]\Big) + \varepsilon_{ij}. \tag{7.7}$$

The inclusion of the random effects b_{1i} and b_{2i} also appeared to have a stabilizing effect on the standard error estimation of the B and β parameters. The parameter estimates and standard errors for model (7.7) can be found in Table 7.5.

Of clinical importance is also the time point at which the rebound phase starts, the so-called point of rebound. This information cannot be derived directly from a model parameter. Therefore, it was decided to define the point of rebound as the time point where the logistic growth curve exceeds the bi-exponential curve. By applying this rule to model (7.7), a subject-specific estimated point of rebound can be obtained. For the viral load profile in Figure 7.2, this would be around week 20.

Table 7.5: Rebound model. Parameter estimates and standard errors for random effects model. Fixed-effect parameter estimates are also back-transformed.

Effect	Parameter	Estimate	(s.e.)	Transformation	Transformed Estimate
	A	-0.1349	(0.0482)	e^A	0.8738
	α	2.1614	(0.1044)	α	2.1614
	B	-4.1652	(0.1787)	e^B	0.0155
	β	-2.8413	(0.1891)	e^β	0.0583
Std(b_{1i})	σ_{11}	1.7342	(0.1266)		
Std(b_{2i})	σ_{22}	1.0750	(0.1196)		
Std(a_{2i})	σ_{44}	0.6500	(0.0535)		
Cor(b_{1i}, b_{2i})	r_{12}	0.2801	(0.1097)		
Cor(b_{1i}, a_{2i})	r_{14}	-0.4058	(0.1197)		
Cor(b_{2i}, a_{2i})	r_{24}	0.1870	(0.1064)		
Var(ε_{ij})	σ^2_ε	0.1033	(0.0034)		

7.5.3 Biological interpretation

The essential viral dynamics models are typically a set of ordinary differential equations (ODE) describing a predator-prey like behavior between virus and cells. When therapy is started, the solution to these ODE's has the form of equation (7.1). The parameters A and B can be interpreted as proportions of different infected cell types that contribute to the production of new virions. The slopes α and β represent a combination of the effectiveness of the used therapy and the respective half-lives of the different types of infected cells. Assuming a 100% inhibition of the viral replication, these half-lives would correspond to $\ln(2)/\alpha$ and $\ln(2)/\beta$.

Conform the literature, we find half-lives between 2 and 3 days for infected and activated CD4+ T-cells (the pool that contributes to the typical high levels of viral load). However, the results for the half-lives (and proportion) of longer lived cells are influenced by the detection limit of the assay.

7.6 Concluding Remarks

The model building strategy gave good results for the TMC114 dataset and also for another Tibotec dataset on viral loads (results not shown here). Manual tweaking of the starting values and the numerical integration methods was only needed for the most complex models. However, complete automation of this type of NLME models is probably unrealistic. Pinheiro and Bates (2000) even described the determination of reasonable starting values for the parameters in a non-linear model as *somewhat of an art*. Hand tweaking will most likely be a necessity for the most complex models.

The proposed model building strategy can be extended even further by, for example, modeling of the variance function, random effects by treatment group, etc. However, this is beyond the scope of this chapter, since convergence can be problematic for these type of complex models, even when carefully fitted by hand. Additionally, the TMC114 study comprised so-called salvage patients, which almost by definition have more irregular viral load profiles.

These types of mechanistic models are generally sensitive to the quality of the data. This means that the data should roughly show the pattern assumed by the fitted model. In contrast, linear models are much more stable. This is also the reason why a new model was developed specifically for rebounding patients, since the rebound is assumed to be caused by a fundamentally different biological mechanism.

8

Concluding Remarks and Further Research

In this dissertation we have focussed on methods for modeling continuous, i.e., Gaussian, longitudinal data. We have shown that a flexible, rich set of tools is available for analyzing this type of data.

In the repeated measures setting, each of the three model families, which were compared in detail in Chapter 4, model both the dependence of the response on the explanatory variables and the autocorrelation among the responses. Ignoring this correlation leads to incorrect inferences about the fixed-effect regression coefficients, and to a loss of efficiency, that is, less precise estimates. This point was illustrated in Chapter 5, where we were able to establish an additional treatment effect that had gone undetected in previous, simpler analyzes. By properly accounting for bird-specific effects, we gained power to assess the effect of treatment, underscoring the strength of the non-linear mixed modeling framework.

In addition to a gain in efficiency, the modeling of within-subject correlation can also be of direct scientific interest. In Chapter 3, the correlation structure was examined to describe the persistence dimension of patients exhibiting persistent disturbing behavior (PDB). In Chapter 6, we focused on serial correlation and we proposed a spline-based approach to flexibly model the serial correlation function. Applying this method to data from a pre-clinical experiment in dementia, enabled us to show that a

circadian pattern played a role in the mean structure, variance structure and the correlation structure simultaneously. However, as Davidian and Giltinan (1995, p. 330) mentioned, second moment behavior is inherently difficult to characterize, and this is especially true for correlation parameters. This also means that a substantial amount of information, i.e., a large dataset, is needed when drawing conclusions about the nature of the correlation structure.

In contrast, when one is not directly interested in the correlation structure as such, but merely needs to correct for it, the generalized estimating equations (GEE) approach of Liang and Zeger (1986) can be adopted. However, this method is not really appealing when one is confronted with missing data, since it requires the missingness mechanism to be missing completely at random (MCAR, Little and Rubin 2002), which is often unrealistic. Therefore, alternatives have been proposed such as weighted generalized estimating equations (WGEE, Robins, Rotnitzky, and Zhao 1994) and multiple imputation based generalized estimating equations (MI-GEE, Schafer 2003), to obtain valid inferences under the missing at random (MAR) assumption.

In Chapter 3, we took up the issue of discriminating between PDB and non-PDB patients, using longitudinal data analysis techniques. Conventional cluster analysis methods were also employed to study the important issue of subgroups within the data and hence in the PDB group. However, further research will be directed towards refining the clustering of PDB patients by means of methods that take the longitudinal structure of the profiles into account. In this respect, a first attempt was made by Bruckers, Serroyen, and Molenberghs (2007), who performed an analysis based on conventional linear mixed models, and on so-called growth-mixture and latent class growth models (Nagin 1999, Nagin and Tremblay 2001, Erosheva, Fienberg, and Lafferty 2004).

Non-linear models, especially in the field of pharmacokinetics, are often mechanistic in nature, and therefore are rooted in specific theoretical considerations about the underlying mechanism producing the data. As a consequence, these models usually do not leave much freedom for model adaptations, compared to linear models, since every parameter generally has a natural physical interpretation, which we do not want to distort. This notwithstanding, in Chapter 7 we were able to extend the basic bi-exponential viral dynamic model to incorporate a possible rebound in viral load levels. Another important aspect of the data analyzed in this chapter, is the detection limit at 50 HIV RNA copies/ml, which means that the data are left censored. However, it is less than straightforward to extend the bi-exponential mixed-effects

model to account for censoring. This is therefore a possible topic for further research.

We were able to perform all analyzes within the likelihood framework by relying on the elegant properties of the (multivariate) normal distribution. This assumption of normality, however, brings us to an important step in the process of data modeling, namely checking the various features of the fitted model. This usually involves verifying goodness-of-fit of the model, checking model assumptions, and detecting possibly influential observations. More than a decade ago, Davidian and Giltinan (1995, p. 328) already indicated that there is an urgent need for diagnostics to complement all of the methods discussed in their book. Surprisingly, little work has been done on model checking and model diagnostics in non-linear mixed models and this topic alone would definitely deserve further research. One of the most common approaches to assessing the quality of the fit for these types of models are based on informal, graphical techniques, e.g., plotting the observed versus the predicted values per subject. However, even in the linear mixed model setting a seemingly simple concept such as residuals can be defined in several ways (Verbeke and Molenberghs, Ch. 11). For example, the marginal residual $\boldsymbol{Y}_i - X_i\widehat{\beta}$ reflects how a specific profile deviates from the overall population means and can therefore be interpreted as a residual. Alternatively, the subject-specific residual $\boldsymbol{Y}_i - X_i\widehat{\beta} - Z_i\widehat{\boldsymbol{b}}_i$ measures how much the observed values deviate from the subject's own predicted line. Finally, the estimated random effects $\widehat{\boldsymbol{b}}_i$ can also be seen as residuals since they reflect how much specific profiles deviate from the population average. In non-linear case, there is an additional level of complication in fitting, and this complexity may obscure key features of the data (Davidian and Giltinan 1995).

Another model diagnostic approach, besides residual analysis, is to use influence measures. In general, influence measures aim at determining whether some observations have undue influence on the estimates of the model parameters and hence how sensitive is the fitting of the model to such observations. With longitudinal data we actually need to distinguish between influential subjects (influence of the observations from a particular subject) and influential observations (a particular observation from a particular subject). One approach to detecting influential observations is local influence (Cook 1986). Using a case-weight perturbation scheme where it is investigated how much the parameter estimates are affected by changes in the weights of the log-likelihood contributions of specific subjects, Lesaffre and Verbeke (1998) derive local influence measures in linear mixed models. Ouwens, Tan, and Berger (2001) extended these local influence measures in generalized linear mixed models, while Lee and Xu (2004) adapted them to the setting of non-linear mixed models. Although it

is clear that basic methodological work has been done, more research needs to take place before these methods can be applied on a routine basis and implemented in standard statistical software packages.

This dissertation has demonstrated that a rich toolkit is available in the setting of continuous longitudinal data, enabling us to answer a broad range of questions. However, to round off with a quote from Davidian and Giltinan (1995, p. 331), 'modeling of nonlinear repeated measures data, although useful, is not an enterprise that should be undertaken lightly.'

References

Agresti, A. (2002). *Categorical Data Analysis* (2nd ed.). New York: Wiley.

Brenowitz, E.A., Margoliash, D., and Nordeen, K.W. (1997). An introduction to birdsong and the avian song system. *Journal of Neurobiology*, **33**, 495–500.

Bruckers, L., Molenberghs, G., Poncelet, J., Brouns, K., Cuypers, W., Slaets, H., and Vanheyst, I. (2000). Identificatie en inschatting van de omvang van de groep patiënten met persisterend storend gedrag. *Acta Hospitalia*, **40**, 21–30.

Bruckers, L., Serroyen, J., and Molenberghs, G. (2007) Cluster analysis of persistent disturbing behavior patients using longitudinal profiles. *Submitted for publication.*

Cook, R.D. (1986). Assessment of local influence, *Journal of the Royal Statistical Society - Series B*, **48,** 133–169.

Davidian, M. and Gallant, A.R. (1992). Smooth nonparametric maximum likelihood estimation for population pharmacokinetics, with application to quinidine, *Journal of Pharmacokinetics and Biopharmaceutics* , **20**, 529–556.

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

Dierckx, P. (1993). *Curve and Surface Fitting with Splines.* Clarendon, Oxford.

Diggle, P.J. (1988). An approach to the analysis of repeated measures. *Biometrics*, **44**, 959–971.

Diggle, P.J., Liang, K.-Y., and Zeger, S.L. (1994). *Analysis of Longitudinal Data.* Oxford Science Publications. Oxford: Clarendon Press.

Diggle, P.J., Heagerty, P.J., Liang, K.-Y., and Zeger, S.L. (2002). *Analysis of Longitudinal Data (2nd ed.).* Oxford Science Publications. Oxford: Clarendon Press.

Diggle, P.J. and Verbyla, A.P. (1998). Nonparametric estimation of covariance structure in longitudinal data. *Biometrics*, **54**, 401–415.

Di Mascio, M., Ribeiro, R.M., Markowitz, M., Ho, D.D., and Perelson, A.S. (2004). Modeling the long-term control of viremia in HIV-1 infected patients treated with antiretroviral therapy, *Mathematical Biosciences*, **188**, 47–62.

Draper, N.R., and Smith, H. (1998). *Applied Regression Analysis (3rd ed.)*. New York: John Wiley & Sons.

Dunn, G. and Everitt, B. (1995). *Clinical Biostatistics*. London: Arnold.

Dunn, G. (2000). *Statistics in Psychiatry*. London: Arnold.

Erosheva, E., Fienberg, S., and Lafferty, J. (2004). Mixed-membership models of scientific publications. *Proceedings of the National Academy of Sciences*, **101S**, 5220–5227.

Eubank, R.L. (1988). *Spline Smoothing and Nonparametric Regression*. New York: Marcel Dekker.

Everitt, B.S. and Landau, S. (1998). The use of multivariate statistical methods in psychiatry. *Statistical Methods in Medical Research*, **7**, 253–277.

Fahrmeir, L. and Tutz, G. (2001). *Multivariate Statistical Modelling Based on Generalized Linear Models*. Heidelberg: Springer-Verlag.

Fisher, R.A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, **7**, 179–188.

Gower, J. (1971). A General Coefficient of Similarity and Some of its Properties. *Biometrics*, **27**, 857–874.

Green, P.J., and Silverman, B.W. (1994). *Nonparametric Regression and Generalized Linear Models*. London: Chapman & Hall.

Harville, D.A. (1974). Bayesian inference for variance components using only error contrasts. *Biometrika*, **61**, 383–385.

Harville, D.A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association*, **72**, 320–340.

Hurvich, C.M, Simonoff, J.S., and Tsai, C.-L. (1998). Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion. *Journal of the Royal Statistical Society - Series B*, **60**, 271–293.

Johnson, R.A. and Wichern, D.W. (1992). *Applied Multivariate Statistical Analysis.* 3rd ed. Englewood Cliffs: Prentice-Hall.

Laird, N.M. and Ware, J.H. (1982). Random effects models for longitudinal data. *Biometrics*, **38**, 963–974.

Lee, S.-Y., and Xu, L. (2004). Influence analyses of nonlinear mixed-effects models, *Computational Statistics & Data Analysis*, **45**, 321–341.

Lesaffre E. and Spiessens B. (2001). On the effect of the number of quadrature points in a logistic random-effects model: an example, *Applied Statistics*, **50**, 325–335.

Lesaffre, E., Todem, D., Verbeke, G. and Kenward, M. (2000). Flexible modelling of the covariance matrix in a linear random effects model. *Biometrical Journal*, **42**, 807–822.

Lesaffre, E. and Verbeke, G. (1998). Local influence in linear mixed models. *Biometrics*, **54**, 570–582.

Liang, K.-Y. and Zeger, S.L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, **73**, 13–22.

Lindstrom, M.J. and Bates, D.M. (1990). Nonlinear mixed-effects models for repeated measures data, *Biometrics*, **46**, 673–687.

Little, R.J.A. and Rubin, D.B. (2002). *Statistical Analysis with Missing Data* (2nd ed.). New York: John Wiley & Sons.

McLachlan, G.J. (1992). *Discriminant Analysis and Statistical Pattern Recognition.* New York: John Wiley & Sons.

Molenberghs, G. and Verbeke, G. (2005). *Models for Discrete Longitudinal Data*, Springer Series in Statistics, Springer-Verlag, New York.

Nagin, D.S. (1999). Analyzing developmental trajectories: a semiparametric, group-bases approach. *Psychological Methods*, **4**, 139–157.

Nagin, D.S. and Tremblay, R.E. (2001). Analyzing developmental trajectories of distinct but related behaviors: a group-based method. *Psychological Methods*, **6**, 18–34.

Nowak, M.A., and May, R.M. (2000). *Virus Dynamics: Mathematical Principles of Immunology and Virology*, Oxford University, Oxford.

Ouwens, M., Tan, F., and Berger, M. (2001). Local influence to detect influential data structures for generalized linear mixed models. *Biometrics*, **57**, 1166–1172.

Pan, J. and Mackenzie, G. (2003). On modelling mean-covariance structures in longitudinal studies. *Biometrika*, **90**, 239–244.

Perelson, A.S., Essunger, P., Cao, Y., Vesanen, M., Hurley, A., Saksela, K., Markowitz, M., and Ho, D.D. (1997) Decay characteristics of HIV-1-infected compartments during combination therapy, *Nature*, **387**, 188–191.

Pinheiro, J.C. and Bates, D.M. (1995). Approximations to the log-likelihood function in the nonlinear mixed-effects model. *Journal of Computational and Graphical Statistics*, **4**, 12–35.

Pinheiro, J.C. and Bates, D.M. (2000). *Mixed-Effects models in S and S-Plus*, Statistics and Computing, Springer-Verlag, New York.

R Development Core Team (2007). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Reisberg, B., Borenstein, J., Salob, S.P., Ferris, S.H., Franssen, E., Georgotas, A. (1987). Behavioral symptoms in Alzheimer's disease: phenomenology and treatment. *Journal of Clinical Psychiatry*, **48**, 9–15.

Robins, J.M., Rotnitzky, A., and Zhao, L.P. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association*, **89**, 846–866.

Royston, P. and Altman, D.G. (1994). Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling. *Applied Statistics*, **43**, 429–468.

Ruppert, D., Wand, M.P. and Carroll, R.J. (2003). *Semiparametric Regression*. New York: Cambridge University Press.

Saporta, G. (1990). *Probabilités Analyse des Données et Statistique*, Paris, Editions Technip.

Schafer, J.L. (2003). Multiple imputation in multivariate problems when the imputation and analysis models differ. *Statistica Neerlandica*, **57**, 19–35.

Serroyen, J., Molenberghs, G., Verhoye, M., Van Meir, V., and Van der Linden, A. (2005). Dynamic Manganese-Enhanced MRI Signal Intensity Processing Based on Nonlinear Mixed Modeling to Study Changes in Neuronal Activity, *Journal of Agricultural, Biological, and Environmental Statistics*, **10**, 170–183.

Serroyen, J., Bruckers, L., Rogiers, G., and Molenberghs, G. (2007a). Characterizing persistent disturbing behavior using longitudinal and multivariate techniques. *Submitted for publication.*

Serroyen, J., Molenberghs, G., Aerts, M., Vloeberghs, E., De Deyn, P., and Verbeke, G. (2007b). Flexible Estimation of Serial Correlation in Linear Mixed Models. *Submitted for publication.*

Sethi, A.K., Celentano, D.D., Gange, S.J., Moore, R.D., and Gallant, J.E. (2003) Association between Adherence to Antiretroviral Therapy and Human Immunodeficiency Virus Drug Resistance, *Clinical Infectious Diseases*, **37**, 1112–1118.

Sheiner, L.B. and Beal, S.L. (1980). Evaluation of methods for estimating population pharmacokinetic parameters. I. Michaelis-Menten model: Routine clinical pharmacokinetic data. *Journal of Pharmacokinetics and Biopharmaceutics*, **8**, 553–571.

Skrondal, A. and Rabe-Hesketh, S. (2004). *Generalized latent variable modeling.* London: Chapman & Hall/CRC.

Spinosa-Guzman, S., Vangeneugden, T., Hoetelmans, R., De Meyer, S. (2007). *A Phase II, randomized, controlled, partially blinded trial to investigate the efficacy, safety, and dose-response relationship of TMC114/ritonavir in 3-class-experienced HIV-1 infected subjects, followed by an open-label period on the recommended dose of TMC114/ritonavir. Analysis with cut-off date of 18 December 2006, at which time all subjects had reached Week 96 or discontinued before.* Tibotec Pharmaceuticals Ltd., TMC114-C202, Clinical Research Report.

Stram, D.O. and Lee, J.W. (1994). Variance components testing in the longitudinal mixed effects model. *Biometrics*, **50**, 1171–1177.

Tsuang, M.T., Tohen, M., and Zahner, G.P. (1995). *Textbook in Psychiatric Epidemiology.* New York: Wiley-Liss.

UNAIDS, Joint United Nations Programme on HIV/AIDS (2006). *Overview of the global AIDS epidemic*, 2006 Report on the global AIDS epidemic.

Van der Linden, A., Verhoye, A., Van Meir, V., Tindemans, I., Eens, M., Absil, P., and Balthazart, J. (2002). In vivo manganese-enhanced magnetic resonance imaging reveals connections and functional properties of the songbird vocal control system. *Neuroscience*, **112**, 467–474.

Van Meir, V., Verhoye, M., Absil, P., Eens, M., Balthazart, J., and Van der Linden, A. (2004). Differential effects of testosterone on neuronal populations and their connections in a sensorimotor brain nucleus controlling song production in songbirds: a manganese enhanced-magnetic resonance imaging study. *Neuroimage*, **21**, 914–923.

Verbeke, G. (1995). *The linear mixed model. A critical investigation in the context of longitudinal data analysis.* Ph.D. thesis, Catholic University of Leuven, Faculty of Science, Department of Mathematics.

Verbeke, G., Lesaffre, E. and Brant, L.J. (1998). The detection of residual serial correlation in linear mixed models. *Statistics in Medicine*, **17**, 1391–1402.

Verbeke, G. and Molenberghs, G. (2000). *Linear mixed models for longitudinal data*, Springer Series in Statistics, Springer-Verlag, New-York.

Verbyla, A.P., Cullis, B.R., Kenward, M.G. and Welham, S.J. (1999). The analysis of designed experiments and longitudinal data by using smoothing splines. *Applied Statistics*, **48**, 269–311.

Vloeberghs, E., Van Dam, D., Engelborghs, S., Nagels, G., Staufenbiel, M. and De Deyn, P.P. (2004). Altered circadian locomotor activity in APP23 mice: A model for BPSD disturbances. *European Journal of Neuroscience*, **20**, 2757–2766.

Vonesh, E.F. and Carter, R.L. (1992). Mixed-effects nonlinear regression for unbalanced repeated measures, *Biometrics*, **48**, 1–18.

Vonesh, E.F., Wang, H., Nie, L., and Majumdar, D. (2002). Conditional Second-Order Generalized Estimating Equations for Generalized Linear and Nonlinear Mixed-Effects Models, *Journal of the American Statistical Association*, **97**, 271–283.

Ward, J.H. (1963). Hierarchical Grouping to Optimize an Objective Function. *Journal of American Statistical Association*, **58**, 236–244.

Zhang, H. (2004). Mixed effects multivariate adaptive splines model for the analysis of longitudinal and growth curve data. *Statistical Methods in Medical Research*, **13**, 63–82.