SAPIENZA
UNIVERSITÀ DI ROMA

Dottorato di Ricerca in Statistica Metodologica
Tesi di Dottorato XXIV Ciclo
Dipartimento di Statistica, Probabilità e Statistiche Applicate

Mixed effect joint models for longitudinal responses with dropout: estimation and sensitivity issues

Sara Viviani

Contents

1 Introduction **1**

2 Mixed Models for longitudinal responses **5**
2.1 Longitudinal Data 5
2.2 General Linear models 7
2.3 Generalized Linear models 7
2.3.1 Marginal models 8
2.3.2 Transition models 10
2.3.3 Random effect models 11
2.4 Maximum Likelihood Estimation for GLMMs 12
2.4.1 Maximum Likelihood 12
2.4.2 Gaussian Quadrature 13

3 Missing Data **17**
3.1 Missing data issues 17
3.2 Rubin's taxonomy 20
3.3 Model frameworks 23
3.4 Shared parameter models 25

4 Local sensitivity to nonignorability in shared parameter models **29**
4.1 What is a sensitivity analysis? 29
4.1.1 Sensitivity tools 30
4.2 Local sensitivity to non-ignorability 32
4.3 Local sensitivity in shared parameter models 35
4.4 Shared Parameter Models for Gaussian Random Variables 37
4.4.1 ISNI in Shared Parameter Models 41
4.4.2 Assessing ISNI variability 43
4.5 AIDS data 44
4.6 Primary Biliary Cirrhosis data 46
4.7 Chronic Kidney Disease data 47
4.8 Simulation study 48
4.8.1 Study design 48
4.8.2 Results 50

4.9 Discussion 52

5 Joint modeling for discrete longitudinal responses and time to dropout 59
5.1 The Bayesian multivariate joint model 59
5.2 The generalized linear mixed joint model 61
5.2.1 The EM algorithm 62
5.2.2 The Poisson case 62
5.2.3 The Binomial case 63
5.3 Simulation study 64
5.3.1 Simulation design 64
5.3.2 Simulation results 65
5.4 The AIDS data set 66
5.5 MMT Data 68
5.6 Discussion 69

6 Concluding remarks 71

A Calculations for Chapter 4 73
A.1 Calculating the ISNI for the Shared Parameter Models 73

Bibliography 76

Acknowledgements 83

To Alessio

Chapter 1

Introduction

This thesis deals with the issue of parameter estimation when a mixed effect model for longitudinal data with drop-out is entailed. In longitudinal studies, commonly two kind of information are recorded: repeated measurements of a response of interest, and realizations of a survival time which describes the individual participation to the study. In the literature, see for instance Diggle *et al.* (1994), one possible model framework to jointly consider these information is represented by *shared parameter models* (SPM); in this framework we assume that the two processes, longitudinal and survival, are dependent and that this dependence is due to sharing a set of coefficients (fixed and random). One particular case of this class of models are the *joint models* (JM), where the expected value of the response at time t is assumed to influence the hazard of the event, see Wulfsohn and Tsiatis (1997).

The structure of this manuscript is as follows. In the context of maximum likelihood (ML) estimation, one has to deserve attention to some crucial points; first, since some modelling assumptions are untestable, one has to bear in mind that sensitivity of model parameter estimates to these assumptions should be carefully analysed. Second, while ML estimation for joint models when the longitudinal response is assumed to be Gaussian are quite well studied from a theoretical and computational point of view (see for instance the `JM` library in R), appropriate modelling and computational tools are lacking when the response is still distributed according to a member of the exponential family but it is not necessary Gaussian. This thesis aims at answering to both questions by extending previous literature approaches in two directions: sensitivity in SPM models and ML estimation in JM when the response is discrete-valued.

Chapter 2 describes the main model frameworks introduced in literature to describe the variability of a longitudinal outcome with respect to a set of covariates/factors. Particular attention is on linear mixed models, where random coefficients are introduced to account for dependence between repeated measurements corresponding to the same subject over time. These coefficients represent the influence of unobservable heterogeneity on the adopted parametric structure.

On the other hand, during a longitudinal study, some individuals may drop-out

prematurely due to different reasons. A missing data analysis is therefore of interest. Chapter 3 is devoted to the definition of missing data, to the explanation of the standard taxonomy by Rubin (1976) and to the discussion of model frameworks to deal with longitudinal response with drop-out.

Chapters 4 and 5 describe the main contributions of the thesis.

As far as standard shared parameter models and joint models are concerned (Chapter 4), and the longitudinal response is assumed to be a conditional Gaussian random variable, the sensitivity of parameter estimates to the assumptions upon ignorability of the survival process is explored. In the field of missing data, the latter represents a relevant hypothesis which is however untestable, since it implies the dependence of the drop-out mechanism on the unobserved longitudinal outcome values. A useful screening tool in this context is the *Index of local sensitivity to ignorability* (ISNI), proposed by Troxel *et al.* (2004) and extended by Ma *et al.* (2005). This index is based on a Taylor expansion of the likelihood function (with respect to the longitudinal parameter vector) of a non-ignorable model around the value which defines the ignorability of the missing data. The sensitivity to non-ignorability has been evaluated through the ISNI in different model frameworks, but not for SPMs. In Viviani *et al.* (2011) the extension of the index to the shared parameter model framework is described. The objectives of this analysis are several: to obtain an analytical formulation of the index in the case of shared parameter models (see Appendix A), to compare the sensitivity of SPMs and JMs to non-ignorability, to give solution to some interpretative issues through the formulation of a relative index of non-ignorability.

The main results highlight a higher sensitivity for JMs with respect to SPMs, mainly due to the fact that, in the former, the interpretation of parameter estimates for the longitudinal sub-model changes whether one considers a non-ignorable or an ignorable drop-out mechanism. On the other hand, the SPM is seen to be more sensitive as far as the intercept estimate is concerned, leading to unbiased estimates of the covariates effects when one moves from the ignorability to the non-ignorability assumptions. These remarks are relevant in missing data models, since they allow to understand how parameter estimates are influenced by the assumption of non-ignorability of the drop-out mechanism and, therefore, how inference could change when an ignorable procedure rather than a joint model is adopted.

As far as the index interpretation is concerned, a relevant issue is that the ISNI is an *absolute* measure of the change in parameter estimates, and this may cause difficulties in assessing sensitivity of the parameter estimates. Ma *et al.* (2005) have dealt with this issue by defining a relative index as the ratio between the ISNI and the standard error of the estimate under ignorability. This formulation, though, can be computationally unstable (the standard error could be very close to zero and it is not intended to give an estimate of the ISNI variability); both a simulation study and applications to benchmark datasets deal with this issues in details. In Viviani *et al.* (2011), two alternative tools are proposed: the ratio between the ISNI for a parameter and the corresponding estimate under ignorability, and the ratio between the ISNI and an estimate of the

corresponding sampling variability. The first formulation follows the principle that the ISNI is a measure of parameter estimate changes when moving from ignorability to non-ignorability; hence, the comparison to the value of the estimate under ignorability may lead to a relative measure of change. The second formulation is based on the evaluation of the index variability, which is estimated by considering two different methods: a Monte Carlo approach and a regression based approach. The latter is based on an approximation of the ISNI as the slope of the linear curve which describes the variability of the longitudinal ML estimate as a function of the ignorability parameter. While the ISNI is not a formal model parameter and therefore the notion of sampling variability is somewhat questionable, we must notice that when moving from non-ignorability to ignorability parameter estimates for JMs change interpretation; therefore, the sampling variability observed under ignorability accounts mainly for this aspect. Empirical and simulation results highlight an easier interpretation of these relative formulations with respect to the existing relative formulation based on the ratio to the standard error of the estimate under ignorability.

The second part of the manuscript deals with the extension of the model of Wulfsohn and Tsiatis (1997) and of the R library JM, see Rizopoulos (2010), to the case of longitudinal responses with distribution belonging to the exponential family (Chapter 5). This part answers to a twofold question: first, the need of a general model for describing informative drop-out where discrete-valued responses and a discrete/continuous time the drop-out event are observed. Second, to give a formal and a computational basis for ML estimation to informative drop-out models when a discrete longitudinal outcome is at hand. In the case when individuals may drop-out prematurely from the study and one would adopt a joint modelling approach, not so much has been done in the literature and the available methods are mainly based on Bayesian approaches, usually fully parametric and quite complex from estimation and interpretation perspectives. Moreover, there are no implemented libraries for parameter estimation in this context. Our proposal is a new formulation of a joint model (for discrete responses) in the context of maximum likelihood, referred to as *Generalized Linear Mixed Joint Models* (GLMJM). The key idea is basically the same of the standard JM, i.e. that the expected value of the outcome of interest at time t influences the hazard of the event (drop-out) at that time, with the difference that the hazard function depends on a transformation of the linear predictor instead of the linear predictor itself. The ignorability parameter is then associated to the expected value of the response evaluated at time t and to the corresponding effect on the hazard of the drop-out event at that time point. The model is implemented for responses with Poisson and Binomial distributions, and parameter estimation is conducted via an EM algorithm followed by a Quasi-Newton algorithm for higher speed of convergence. The analytical formulation of the model is discussed in Chapter 5, where ML estimation is also considered with particular emphasis on the iterations of the EM algorithm, which has been structured in a set of functions to be inserted in a further R library. Quasi-Newton steps are performed by using the `optim` function in R. Several simulation studies and applications to benchmark datasets are

presented to study the behaviour of the proposed approach under both mild and extreme departures from the ignorable setting. Since the JM properly nests an ignorable drop-out model, attention has been devoted to study the distribution of parameter estimates when the "true" drop-out process is based on a ignorable structure as well as when we move by the hypothesis of ignorability.
The main results of the simulation study are that maximum likelihood estimates obtained by fitting the GLMJM present a better behaviour than the corresponding estimates under the hypothesis of ignorability when the ignorability parameter is different from zero. On the other hand, when the latter is assumed to be null, the estimates obtained from model based on ignorability and non-ignorability assumptions are qualitatively and quantitatively equivalent.
Applications to benchmark datasets lead to intuitive results. The first application is a HIV study, see Goldman *et al.* (1996) and Carlin and Louis (2009), with the objective of comparing the efficacy/safety of two antiretroviral drugs by recording a longitudinal response (the CD4 cell count). The parameter estimates for the standard Joint Model, where the CD4 cell counts are assumed to be conditionally Gaussian random variables, are compared to the estimates obtained through the Poisson Joint Model, where the response variable is assumed to follow a Poisson distribution. Both models suggest a non-ignorable drop-out mechanism with a negative estimate for the ignorability parameter. The second dataset focuses on a MMT program, see Alfó and Aitkin (2000), where heroin users are observed for 26 weeks and the longitudinal response is positivity to morphine, a marker of recent heroin use. The GLMJM is compared to an autoregressive model which does not account for the non-ignorability of the drop-out process. The significant time effect in the autoregressive model agrees with the significance of the non-ignorability parameter in the GLMJM, and their effect seems to represent a time selection for patients who do not respond to the methadone treatment.

Finally, Chapter 6 give some concluding remarks and some suggestions for future research developments in this field.

Chapter 2

Mixed Models for longitudinal responses

In this Chapter, a review of standard models for longitudinal data is presented. In Section (2.1), we introduce basic concepts and structures for longitudinal data and related issues. In Section (2.2) we briefly describe the General Linear Model with general covariance structure, while in Section (2.3) we illustrate the main model framework for repeated measurements when the response in not Gaussian. We give particular emphasis on generalized linear mixed models and related computational issues.

2.1 Longitudinal Data

Longitudinal data are a common product of several kind of studies, where measurements of a response variable are taken on the same individuals over several occasions. The main advantage with longitudinal data is that one can distinguish between changes over time within individuals from baseline differences among subjects, see Diggle *et al.* (1994). Hence, while some individuals could begin at a higher or a lower level of the variable of interest, the evolution through time could follow a different pattern.

To deal with longitudinal observations, specific statistical methods are needed. In fact, the set of observations corresponding to one subject measured at different time points are usually associated; this correlation must be taken into account to draw valid scientific inferences.

A nice field of application for longitudinal models is growth data, where biological indexes are recorded for children at different ages and growth curves are drawn. When each child is followed through time, this design is useful to make the distinction between *cohort-effects*, i.e. cross-sectional differences *between* children of the same age, and *age-effects*, i.e. the evolution of biological markers *within* each child. This phenomenon is visible in Figure 2.1, where the plot of height (cm) for subjects followed from 0 to 25 years of age is shown. The children are born in early gestational age and do not

show catch-up growth; they are randomized between two levels of growth hormone, see Willemsen *et al.* (2011).

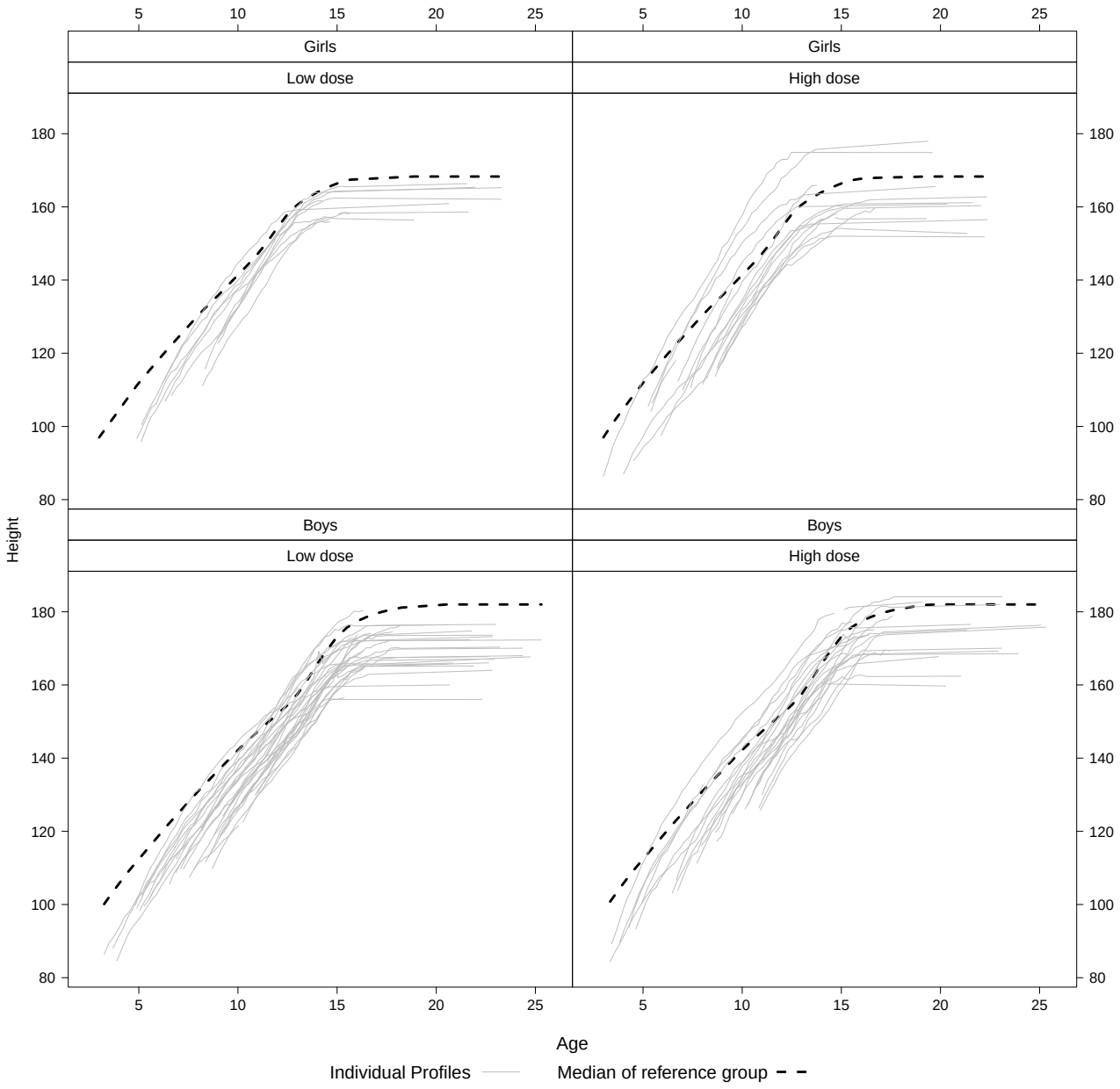

Figure 2.1: Growth curve of heights.

Let us now introduce some notation. Random variables are denoted by capital letters, while observations are indicated by small letters; bold capital and small characters represent matrices and vectors, respectively.

Let $Y_i(t)$ be the longitudinal outcome and $\mathbf{x}_i(t)$ the p-dimensional vector of explanatory variables corresponding to subject $i = 1, \ldots, n$ observed at time $t = 1, \ldots, T_i$. The

expected value and the variance of the response variable are given by $\mathbb{E}[Y_i(t)] = \mu_i(t)$ and $\mathbb{V}[Y_i(t)] = v_i(t)$.
The observations for a given individual i are collected into a T_i-vector $\mathbf{y}_i(t) = [y_i(1), \ldots, y_i(T_i)]$, where $\text{Cov}[y_i(j), y_i(k)] = v_i(j,k)$ and $\mathbf{R}_i$ is the corresponding $T_i \times T_i$ correlation matrix.
The basic model for longitudinal analyses is the following linear model:

$$y_i(t) = \beta_0 x_{i0}(t) + \beta_1 x_{i1}(t) + \ldots + \beta_p x_{ip}(t) + \varepsilon_i(t), \tag{2.1}$$

that is, in vector formulation:

$$\mathbf{y}_i = \beta^\mathsf{T} \mathbf{x}_i + \varepsilon_i, \tag{2.2}$$

where $\beta = (\beta_0, \ldots, \beta_p)$ is a p-dimensional vector of unknown but fixed regression parameters, $\varepsilon_i = (\varepsilon_i(1), \ldots, \varepsilon_i(T_i))$ is the vector of zero-mean errors and $\mathbf{x}_i$ the row vector of the design matrix. We will further discuss models for longitudinal data in Section (2.2). As we will point out, model choice is not straightforward.

2.2 General Linear models

As described in Section 2.1, the basic linear model for repeated measurements is expressed by equation (2.1). In this paragraph, we will further deepen assumptions upon the general linear model for longitudinal data and highlight structural and interpretation issues.
Let us indicate with $\mathbf{Y} = (\mathbf{y}_1, \ldots, \mathbf{y}_n)$ the matrix containing the longitudinal response for subject $i = 1, \ldots, n$, and assume to have a balanced design with $t = 1, \ldots, T$ measures for each individual, the generic individual vector being $\mathbf{y}_i = (y_i(1), \ldots, y_i(T))$. The general linear model for longitudinal data assumes $\mathbf{Y}$ to be a multivariate normal variable, i.e.

$$\mathbf{Y} \sim MVN(\beta^\mathsf{T}\mathbf{X}, \sigma^2\mathbf{V}), \tag{2.3}$$

where $\sigma^2\mathbf{V}$ is a block-diagonal covariance matrix, with non null $T \times T$ blocks $\sigma^2\mathbf{V}_0$. The block-diagonal covariance structure allows us to estimate the variation in the repeated measurements among different occasions. Nevertheless, parametric assumptions could sometimes represent a good approximation of the variance structure and may lead to simplifications.
For instance, in some situations, it is realistic to assume a *uniform correlation* structure between any two occasions relative to the same subject, or either an *exponential correlation* form.

2.3 Generalized Linear models

Different model frameworks are available to analyse longitudinal data. We will discuss them in this Section, trying to critically revise model choice and estimation issues. In

fact, before applying any specific model, one should conduct a preliminary analysis to explore data structure.
Standard tools for exploring longitudinal data could be either analytical and graphical. For instance, a basic display is the scatterplot of the response variable against time, for all the individuals in the sample or alternatively for a representative subset. Figure 2.2 shows individual profiles for reading ability scores on 221 children and 4 occasions (data avaible at **http://www.duke.edu/curran/**).

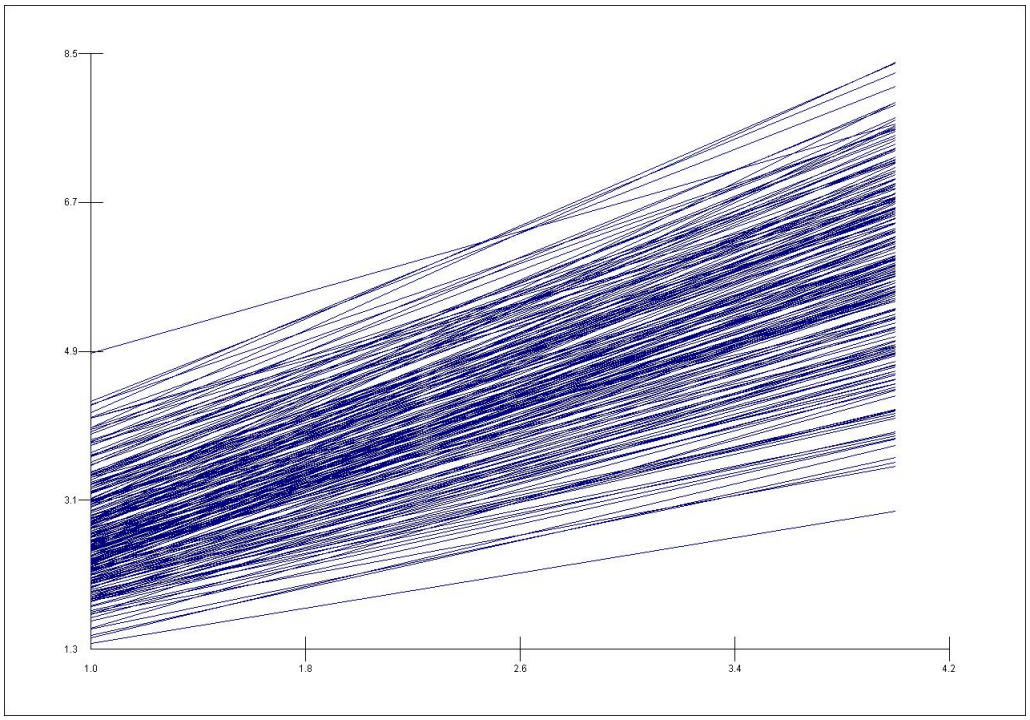

Figure 2.2: Individual profiles for reading ability.

This simple graph shows a number of important patterns; nevertheless, the huge number of subjects makes it complicated from an interpretational perspective. To avoid confusion, one could fit a mean curve, computed using different kind of smoothers (such as lowess, splines or kernel) to highlight a common pattern. This simple solution involves many relevant considerations about model choice.

2.3.1 Marginal models

The strategy of considering the *mean* behaviour is typical of **marginal models**, see for instance Prentice (1988). This could be considered the model specification closer to a cross-sectional study, see for instance Fitzmaurice *et al.* (2004). The focus is, in fact, on finding a proper functional specification for the expected value of the response $\mu_i(t)$ when considered separately from the variance $v_i(t)$ and the within-subject association

expressed by the correlation matrix $\mathbf{R}_i$.
This is a three part model that can be summarized by the following steps.

1. The expected value of the response depends on covariates through a link function: $g(\mu_i(t)) = \beta^{\mathsf{T}}\mathbf{x}_i(t)$;

2. The variance of each outcome is a function of the mean and a scale parameter: $v_i(t) = \Phi f(\mu_i(t))$;

3. The correlation matrix is a function of an association parameter α and the mean: $\mathbf{R}_i = \{\rho_i(j,k) = a(\alpha, \mu_i(t,j,k))\}$.

The separate modelling of mean and covariance structure leads to *population-averaged* interpretation of regression parameters. This means that conclusions about parameters estimates can be related to the whole sample, not being influenced by the within subjects correlation. Moreover, marginal models are valid under any distributional assumption, by following a likelihood or a quasi-likelihood approach.
In this last respect, a convenient estimation procedure is GEE (*Generalized Estimating Equations*), see Liang and Zeger (1986) and Zeger and Liang (1996) . The key idea of this approach is to incorporate the covariance matrix into the usual score equations for a generalized linear model. More specifically, as conditions $1-2$ for the marginal model are specified, the third is made by assuming the following Cholesky decomposition:

$$\mathbf{V}_i = \mathbf{A}_i^{\frac{1}{2}}\mathbf{R}_i\mathbf{A}_i^{\frac{1}{2}}, \tag{2.4}$$

where $\mathbf{A}_i^{\frac{1}{2}}$ is a diagonal matrix containing the standard deviation of the response variable $\sqrt{\Phi f(\mu_i(t))}$. $\mathbf{V}_i$ is referred to as as 'working' covariance matrix.
GEE estimation moves along the lines of Generalized least squares (GLS) for the linear model, which leads to the solution

$$\widehat{\beta} = \left(\sum_{i=1}^{n}\mathbf{x}_i^{\mathsf{T}}\mathbf{\Sigma}_i^{-1}\mathbf{x}_i^{\mathsf{T}}\right)^{-1}\sum_{i=1}^{n}\mathbf{x}_i^{\mathsf{T}}\mathbf{\Sigma}_i^{-1}\mathbf{y}_i,$$

where $\mathbf{\Sigma}$ is the *true* covariance matrix of the response.
The corresponding GEE estimator is given by solving the generalized estimating equations

$$\sum_{i=1}^{n} D_i^{\mathsf{T}}V_i^{-1}(y_i - \mu_i) = 0, \tag{2.5}$$

where $D_i = \frac{\partial\mu_i}{\partial\eta_i}\frac{\partial\eta_i}{\partial\beta}$.
Because GEE depends on both β and α, a two-stages procedure is required.

1. V_i is estimated, given starting values for α and Φ, and the corresponding estimate for β is obtained from (2.5);

2. The current estimate $\widehat{\beta}_{\text{GEE}}$ is used to update α and Φ on the basis of studentized residuals:

$$e_i(t) = (y_i(t) - \widehat{\mu}_i(t))/\sqrt{v_i(t)} \tag{2.6}$$

2.3.2 Transition models

As we have pointed out in the previous Section, marginal models and GLMs for cross-sectional studies lead to similar conclusions in terms of parameter estimates. This leads to simplificated interpretation, but, on the other hand, it causes a loss of information, deriving from the individual evolution of the response over time.
An attempt to take into account the trend of the outcome of interest for each observed subject are *transition models*, see Korn and Whittermore (1979), Wong (1986) and Ware *et al.* (1988). The key idea of this group of models is to consider the influence of the *past history* $\mathcal{H}_i(t)$ on the current value of the response. Hence, the mean $\mu_i(t)^C = \mathbb{E}(Y_i(t) \mid \mathcal{H}_i(t))$ and the variance $v_i(t)^C = \mathbb{V}(Y_i(t) \mid \mathcal{H}_i(t))$ of the response are conditional on the past responses $y_{t-1}, \ldots, y_{t-q}$ up to a lag of order q, and a set of covariates.
A particular case of transition models is the class of first order Markov chains, where mean and variance at a given time are conditional on the value corresponding to the previous occasion of the response, i.e. $\mu_i(t)^C = \mathbb{E}(Y_i(t)|\dagger_i(t-1))$ and $v_i(t)^C = \gtrapprox(Y_i(t)|\dagger_i(t-1))$.
Given a specific link function g, the generic transition model can be expressed by

$$\begin{cases} g(\mu_i(t)^C) = \beta^{\mathsf{T}}\mathbf{x}_i(t) + \sum_{r=1}^{s} f_r(\mathcal{H}_i(t), \alpha) \\ v_i(t)^C = \Phi f(\mu_i(t)^C), \end{cases} \tag{2.7}$$

i.e. the past outcomes are treated as additional covariates after the proper transformation through the function f_r. If the model for the conditional mean is correctely specified, the past responses can be considered as independent events.
Transition models can be fitted using a likelihood approach. The joint distribution of the responses for subject i is

$$f(y_i(1), \ldots, y_i(T_i); \beta, \alpha) = f(y_i(T_i) \mid y_i(T_i{-}1), \ldots, y_i(1); \beta, \alpha) \cdots f(y_i(2) \mid y_i(1); \beta, \alpha) f(y_i(1); \beta, \alpha),$$

that can be simplified to

$$f(y_i(1), \ldots, y_i(T_i); \beta, \alpha) = f(y_i(T_i) \mid y_i(T_i{-}1); \beta, \alpha) \cdots f(y_i(2) \mid y_i(1); \beta, \alpha) f(y_i(1); \beta, \alpha),$$

for a first order Markov model.
For a GLM with a given link function, it is not straightforward to derive the distribution $f(y_i(t) \mid y_i(t{-}1))$ without additional assumptions. An appealing approach is to consider the *conditional likelihood* of $Y_i(2), \ldots, Y_i(T_i)$ given $Y_i(1)$, which is obtained by omitting $f(y_i(1))$ from the previous equation. The resulting estimates are less efficient than MLEs. If the likelihood is still intractable, GEE are a valid alternative (see previous paragraph for more details).

2.3.3 Random effect models

The linear regression model in equation (2.3) defines the association between repeated observations on the same subject through the covariance matrix, and could be seen as an element of a wide class of models containing the specific parametrization of *random effects model*, see Laird and Ware (1982). In fact, these models introduce correlation as due to random variation in unobserved, subject-specific quantities called *random effects* or *random coefficients*. The logical implications of this parametrization is that there exists a natural heterogeneity among observed subjects, deriving either from unobserved characteristics or different effects of measured covariates. This variability could include genetic or environmental factors that are thought to be represented by a zero mean random variable.
The basic assumption of a random effect model is that unobserved individual-specific heterogeneity among individuals could be represented by the random variability in the regression coefficients, where the following regression model holds:

$$y_i(t) = (\beta_0 + b_{i0}) + \beta_1 \mathbf{x}_{i1}(t) + \ldots + \beta_p \mathbf{x}_{ip}(t) + \varepsilon_i(t),$$

and $b_{i0} \sim \mathrm{g}(0, \alpha^2)$. In this case, random variability influences the intercept only, that is a baseline variability among subjects is assumed.
The linear formulation of the random effects model could be extended to generalized linear models with a given link function, leading to the following:

$$g(\mu_i(t)) = (\beta_0 + b_{i0}) + \beta_1 \mathbf{x}_{i1}(t) + \ldots + \beta_p \mathbf{x}_{ip}(t),$$

where $\mu_i(t) = \mathbb{E}[Y_i(t)|b_{i0}, \mathbf{x}_i)]$. When a set of random regression coefficients is used, $\mathbf{b}_i \sim h(\mathbf{0}, \mathbf{D})$, and the covariance matrix $\mathbf{D}$ needs to be estimated. It is worth noting that, since in most cases the random coefficients correspond to some explanatory variables only, random effects models can be referred to as *mixed models*, to highlight the difference between random and fixed coefficients.
Formally, the basic assumptions for a random effect GLM are:

1. *Conditional independence*: given $\mathbf{b}_i$, the responses $Y_i(1), \ldots, Y_i(T_i)$ are independent and follow an exponential family distribution with density $f(y_i(t)|\mathbf{b}_i)$, with expected value $\mu_i(t) = \mathbb{E}(Y_i(t|\mathbf{b}_i, \mathbf{X}_i))$, where the model $g(\mu_i(t)) = \mathbf{x}_i(t)^\mathsf{T}\beta + \mathbf{d}_i(t)^\mathsf{T}\mathbf{b}_i$ hold, and $\mathbf{d}_i(t)$ is a subset of $\mathbf{x}_i(t)$.

2. $\mathbf{b}_i$, $i = 1, \ldots, n$ are i.i.d. zero-mean random variables with multivariate density function $\mathrm{h}(\cdot)$ and covariance matrix $\mathbf{D}$.

3. $\mathbf{b}_i$ and $\mathbf{x}_i(t)$ are mutually independent.

The philosophy underlying a random effect model is opposite to marginal models. The objective is, in fact, to make inference on individual behaviour rather than on population average.

2.4 Maximum Likelihood Estimation for GLMMs

In this Section we give some details on parameter estimation for GLMMs. In this field, two main approaches have been proposed:

1. *Conditional likelihood*, where random effects are treated as fixed parameters, and, conditionally, the response follows a standard multivariate exponential distribution, see Andersen (1973);

2. *Maximum likelihood*, where random effects are treated as random variables with a multivariate distribution and must be integrated out of the likelihood. In this case, $\mathrm{h}(\cdot)$ is considered a zero mean gaussian distribution with variance-covariance matrix $\mathbf{D}$.

We will deepen the latter approach, leaving further details for the conditional likelihood method to literature, see for instance McCullagh and Nelder (1989).

2.4.1 Maximum Likelihood

As it has been outlined before, in the likelihood framework the vector of random coefficients $\mathbf{b}_i$ is assumed to follow a given probability distribution $\mathrm{h}(\cdot)$. This apparently simple assumption leads to one relevant consequence: the possibility to estimate individual trajectories over time. Thus, if heterogeneity in the sample is high, individual estimation of model coefficient is more reliable; on the other hand, if variability between subjects is slight, a population based approach could represent a good method for parameter estimation.
The likelihood as a function of both the fixed effect vector β and the random effects covariance matrix $\mathbf{D}$ is defined by integrating out the random effects:

$$L(\beta, \mathbf{D}; \mathbf{Y}) = \prod_{i=1}^{n} \int_{\mathbf{b}_i} \prod_{j=1}^{T_i} f(\mathbf{y}_i \mid \beta) f(\mathbf{b}_i \mid \mathbf{D}) d\mathbf{b}_i, \tag{2.8}$$

In some cases, i.e. for Gaussian linear model with Gaussian random effects, (2.8) may have a closed solution, but in most situations one must apply numerical integration methods such as Gaussian quadrature (see Section 2.4.2).
To find maximum likelihood estimates, a common procedure is to use the EM algorithm, see Dempster *et al.* (1977). This algorithm is based on the evaluation of the score function through two steps:

1. The **E-step**, evaluates the expectation of the score function given observed data and current values for model parameters;

2. The **M-step** solves the score functions updating parameters estimates.

The observed data score functions to be evaluated are

$$\begin{cases} \mathbf{s}_\beta(\delta \mid \mathbf{Y}) = \sum_{i=1}^{n} \sum_{j=1}^{T_i} x_{ij} \left[y_{ij} - \mathbb{E}\left\{ \mu_{ij}(\mathbf{b}_i) \mid \mathbf{y}_i \right\} \right] \\ \mathbf{s}_{\mathbf{D}}(\delta \mid \mathbf{Y}) = \frac{1}{2}\mathbf{D}^{-1} \left\{ \sum_{i=1}^{n} \mathbb{E}\left\{ \mu_{ij}(\mathbf{b}_i \mathbf{b}_i^{\mathsf{T}}) \mid \mathbf{y}_i \right\} \mathbf{D}^{-1} - \frac{m}{2}\mathbf{D}^{-1} \right\}, \end{cases} \tag{2.9}$$

with respect to β and $\mathbf{D}$, respectively.

2.4.2 Gaussian Quadrature

When $\mathbf{b}_i$ has a limited dimension, a potential technique to solve the integral in (2.8) is the Gauss-Hermite quadrature, introduced by Naylor and Smith (1982) in the Bayesian framework and further extended to mixed models with binary data by Anderson and Aitkin (1985). Liu and Pierce (1994) presented an interesting modification of the Gaussian quadrature for transformed variables.
The term *Gaussian* quadrature is due to numerical approximation of the univariate integral

$$\int_{-\infty}^{+\infty} e^{-\theta^2} f(\theta) d\theta$$

with a Gaussian-type polynomial (Gauss-Hermite formula)

$$\sum_{i=1}^{K} \omega_i f(\theta_i)$$

where

$$\omega_i = \frac{2^{K-1} K! \sqrt{\pi}}{K^2 \left[H_{K-1}(\theta_i) \right]^2}$$

and θ_i is the ith zero of the Hermite polynomial $H_K(\theta)$, see Davis and Rabinowitz (1967). For a suitably regular function $h(\theta)$, and

$$g(\theta) = h(\theta) \left(2\pi\sigma^2 \right)^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} \left(\frac{\theta - \mu}{\sigma} \right)^2 \right\}$$

it can be proved that

$$\int_{-\infty}^{+\infty} g(\theta) d\theta \approx \sum_{i=1}^{K} w_i g(z_i)$$

where $w_i = w_k \exp(\theta_i) \sqrt{2\sigma^2}$ and $z_i = \mu + \sqrt{2\sigma^2 \theta_i}$ are the weights and the locations of the Gaussian-Hermite quadrature. Table 2.1 shows the first 5 quadrature locations and weights, see Salzer *et al.* (1952):

k	z_i	w_i
2	± 0.707107	0.886227
3	0	1.18164
	± 1.22474	0.295409
4	± 0.524648	0.804914
	± 1.65068	0.0813128
5	0	0.945309
	± 0.958572	0.3936190
	± 2.020180	0.0199532

Table 2.1: Quadrature locations and weights for the Gauss-Hermite quadrature, $k = 1, \ldots, 5$

By applying the previous concepts to the likelihood in (2.8), the integral is approximated as follows:

$$L(\beta, \mathbf{D}; \mathbf{Y}) \approx \prod_{i=1}^{n} \sum_{k=1}^{K} \left[\prod_{j=1}^{T_i} f(\mathbf{y}_i \mid \beta_k) \right] f(\mathbf{b}_k \mid \mathbf{D}) \pi_k. \tag{2.10}$$

This means that the likelihood is approximately equal to the likelihood of a finite mixture of component specific densities with known proportions π_k and locations $\mathbf{b}_i$.
When the random coefficient vector is unidimensional, i.e. when random intercept model are considered, b_k represents one of the quadrature points and π_k the probability mass associated to it. On the other hand, when $\mathbf{b}_k$ has dimension $q > 1$, i.e. with random coefficients, $\mathbf{b}_k = (b_{k1}, \ldots, b_{kq})$ is a vector of quadrature points with associated probability mass $\pi_k = \prod_{j=1}^{q} \pi_{kj}$. In this case, a useful transformation of a multivariate Gaussian random variable is

$$\mathbf{b}_i = \mathbf{D}^{\frac{1}{2}} \mathbf{b}_i^*,$$

where $\mathbf{b}_i \sim MVN(\mathbf{0}, \mathbf{D})$ and $\mathbf{b}_i^* \sim MVN(\mathbf{0}, \mathbf{I})$. This allows us to work with the density $f(\mathbf{b}_{i1}^*, \ldots, \mathbf{b}_{iq}^*) = \prod_{j=1}^{q} f(\mathbf{b}_{iq}^*) \sim MVN(\mathbf{0}, \mathbf{I})$. Hence, the quadrature dimension is K^q. For istance, for $q = 2$, at each pair (b_{ij1}, b_{ij2}) is associated a fixed weight $\pi_{j1} \cdot \pi_{j2}$.
To derive the score function expression, we rewrite the density function as follows:

$$\frac{\partial f(\mathbf{y}_i \mid \mathbf{b}_k)}{\partial \delta} = f(\mathbf{y}_i \mid \mathbf{b}_k) \frac{\partial \log f(\mathbf{y}_i \mid \mathbf{b}_k)}{\partial \delta},$$

obtaining

$$\mathbf{s}(\delta) \approx \sum_{i=1}^{n} \sum_{k=1}^{K} w_{ik}(\delta) \frac{\partial \log f(\mathbf{y}_i \mid \mathbf{b}_k)}{\partial \delta},$$

where

$$w_{ik}(\delta) = \frac{\pi_k f(\mathbf{y}_i \mid \mathbf{b}_k)}{\sum_{l=1}^{K} \pi_l f(\mathbf{y}_i \mid \mathbf{b}_{li})}$$

and $\sum_{k=1}^{K} w_{ik}(\delta) = 1$. The weights $w_{ik}(\delta)$ represent the posterior probabilities that the ith unit belongs to the kth component.
An efficient way of computing the posterior weights is via the EM algorithm (see the previous Section):

1. E-step: the expected value score function in (2.11) is computed conditionally based on the current values of parameter estimates and observed data;

2. M-step: for given weights the likelihood is maximized through a *Fisher scoring* algorithm;

It is interesting to note that for $n, K \to \infty$, the estimates of δ are consistent and asymptotically normal the under usual regularity conditions.

Numerical issues may hold for Gaussian quadrature since locations and corresponding weights are fixed, and this may cause an improper approximation when the integrand function is not suitably regular. An alternative approach is the *adaptive Gaussian quadrature*, where the integration points are resampled in subintervals of the integration domain. For instance, the Gauss-Kronrod quadrature, see Kronrod (1964), is an adaptive Gaussian quadrature method where error is estimated by evaluating special points, see *Kronrod points*. By suitably picking these points, abscissas from the previous iteration can be reused as part of the new set of points, whereas usual Gaussian quadrature would require recomputation of all abscissas at each iteration. This is particularly important when some specified degree of accuracy is needed but the number of points needed to achieve this accuracy is not known ahead of time, Calvetti *et al.* (2000).

When the integral in (2.8) has large dimension, a Monte Carlo approach could be computationally more efficient , given that the complexity of the algorithm depends only linearly on the number of dimensions q. For a nice review of Monte Carlo methods, see for instance James (1980). In this case the likelihood assumes the form:

$$L(\delta) \approx \frac{1}{K} \sum_{k=1}^{K} f(\mathbf{y}_i \mid \mathbf{b}_{ik}),$$

$\mathbf{b}_{ik}$ are K realizations of the multidimensional random variable $\mathbf{b}_i^*$, a zero-mean random variable with covariance matrix I and density $\mathrm{h}^*(\cdot, I)$. Assuming a known density $\mathrm{h}^*(\cdot, I)$, the integral is computed as the arithmetic mean of $g^*(\cdot, I)$ corresponding to the simulated values of the random coefficients $\mathbf{b}_i^*$. In this case, the score function can be written as follows:

$$\mathbf{s}(\delta) = \sum_{i=1}^{n} \sum_{k=1}^{K} w_{ik}(\delta) \frac{\partial \log \left[f(\mathbf{y}_i \mid \mathbf{b}_i) \right]}{\partial \delta}, \tag{2.11}$$

where $w_{ik}(\delta) = f(\mathbf{y}_i \mid \mathbf{b}_{ik}) / \left(\sum_{l=1}^{K} f(\mathbf{y}_i \mid \mathbf{b}_{il}) \right)$. Hence the derivatives $\partial \log \left[f(\mathbf{y}_i \mid \mathbf{b}_i) \right] / \partial \delta$ can be computed following the same procedure as for the Gaussian quadrature, but substituting $\mathbf{b}_k$ with their realizations $\mathbf{b}_{ik}$ and applying the same EM algorithm.

Chapter 3

Missing Data

Longitudinal data may suffer from missingness. This means that subjects may not be measured in some of the planned occasions, or exit the study at a given time before the end of the study. In this Chapter, we will deal with both these situations and describe models to treat data with some unobserved entries.

3.1 Missing data issues

Let us suppose to observe a balanced longitudinal study, where $i = 1, \ldots, n$ subjects are observed at $t = 1, \ldots, T$ occasions. The response is usually organized in a matrix form as follows:

$$\mathbf{Y} = \begin{pmatrix} y_1(1) & \ldots & y_1(T) \\ . & . & . \\ y_n(1) & \ldots & y_n(T) \end{pmatrix}, \tag{3.1}$$

i.e. in a rectangular matrix. We will deal with statistical analysis when the $\mathbf{Y}$ matrix includes some unobserved entries. We will refer to these issues as statistical analysis with missing data, see Little and Rubin (2002). The basic idea of missing data models is that in several contexts, a *missing* information could represent an *additional* information. To make this point clearer, we illustrate the following example.

Example 3.1.1. *Let us consider the longitudinal study (analysed by Rizopoulos and Ghosh (2011)) on chronic kidney disease, where patients underwent transplantation with a graft. The longitudinal outcome is the Glomerular Filtration Rate (GFR), measured during a 10-years follow-up. From Figure 3.1 it is evident that subjects who complete the study present different response patterns, when compared to subjects who drop out from the study. More precisely, dropped-out patients present a mean profile which decreases more quickly than the profiles corresponding to the whole sample and to those who complete the study.*

Different ideas may be derived from the previous example. First of all, the data matrix in (3.1) can be subject to *dropout*, and subjects may exit the study and do not

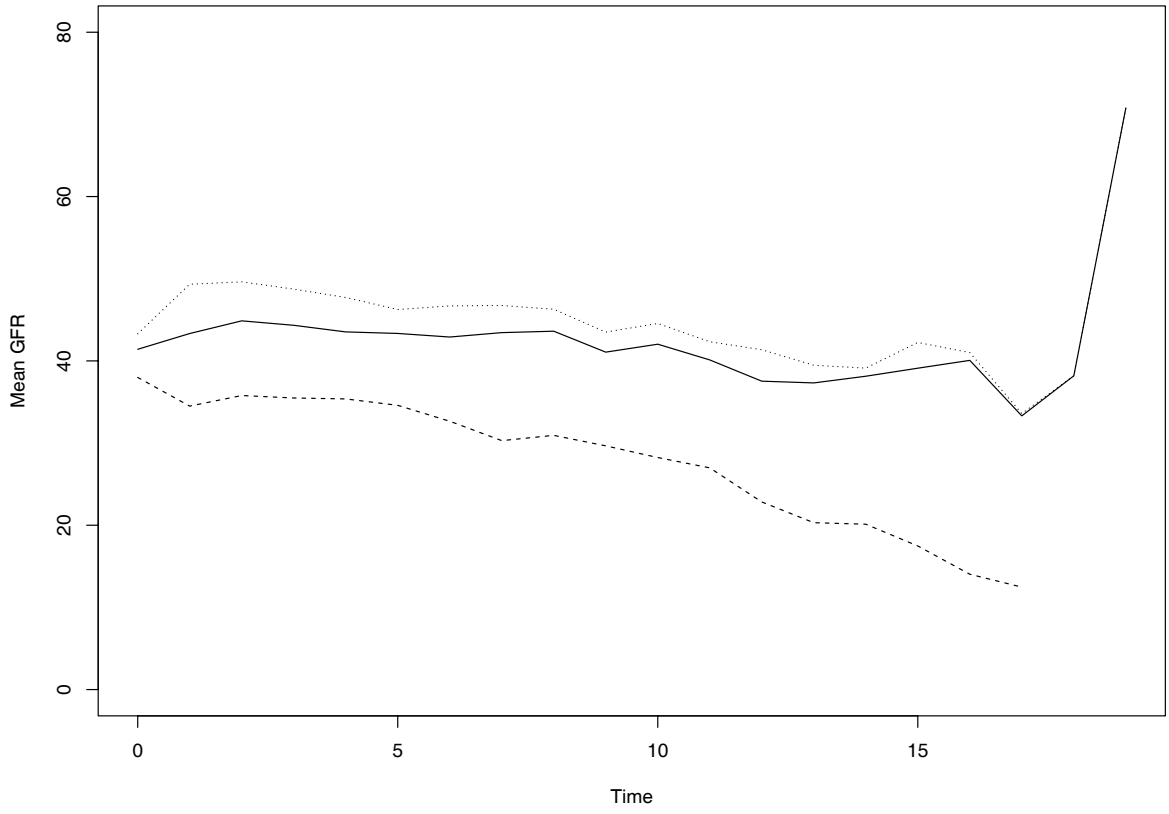

Figure 3.1: Mean profiles of GRF for the overall sample (solid line), patients who complete the study (dotted line) and patients who drop out from the study (dashed line).

re-enter. For a set of n_d drop-outs, the occasions $T_{d1}, \ldots, T_{dn_d}$ are ordered by rows such that $T_{d1} \leq T_{d2} \leq \ldots \leq T_{dn_d} \leq T$. The data matrix looks like the following:

$$\mathbf{Y}_{\text{drop}} = \begin{pmatrix} y_1(1) & \ldots & y_1(T_{d1}) & NA & NA & \ldots & NA \\ & & & \vdots & & & \\ y_{n_d}(1) & \ldots & y_{n_d}(T_{dn_d}-1) & y_{n_d}(T_{dn_d}) & NA & \ldots & NA \\ y_{n_d+1}(1) & \ldots & y_{n_d+1}(T_{dn_d}-1) & y_{n_d+1}(T_{dn_d}) & y_{n_d+1}(T_{dn_d+1}) & \ldots & y_{n_d+1}(T) \\ & & & & \ddots & & \\ y_n(1) & & & \ldots & & & y_n(T) \end{pmatrix}. \tag{3.2}$$

In this case, the missing-data mechanism is referred to as *attrition* (or monotone drop-out), and it is quite common in clinical trials, where patients drop out from the study for side effects that can be related to the drug, to lack of efficacy or to healing. Molemberghs

and Kenward (2007) defined these subjects as *lost to follow-up*. The pattern of attrition is called *monotone*, because data can be arranged as in matrix (3.2) such that n_d subjects are potentially missing and the remaining $n - n_d$ are completers. As Little and Rubin (2002) point out, the missing pattern is rarely monotone, but often it can be close enough to.
On the other hand, the missing pattern could be *intermittent*, i.e. subjects may not participate only to some non contiguous occasions, leading to *non-monotone* missing endpoints. In this case, the missingness may be due to patients skipping a visit for practical or administrative reasons or to measurement equipment failure. It is than evident that the reasons behind these pattern are more complicated to be modelled. Just to give an example, the majority of clinical trials present dropout, and a small fraction of them non-monotone patterns. For these reasons, in this work we will put the emphasis on issues related to dropout.
Another important assumption we make is that the missigness process hides true values that could be meaningful for the analysis. In this context, we focus on three common procedures applied to handle missingness:

1. **The complete cases analysis (CC)**: it simply eliminates the dropped observations, by including in the analysis only the completers. Even if this approach is of immediate application, it leads to severe inference complications. First of all, it causes a relevant *loss of information*, which may lead to inefficient estimators. Moreover, in some cases (these will be discussed in the following paragraph), inferences based on CC may suffer from bias.

2. **The imputation method**: it substitutes the missing values with other values derived through some procedures. As Dempster and Rubin (1987) pointed out, methods from this class are both seductive and dangerous. In fact, they may create the illusion that the dataset is complete afterall, making the implicit assumption that imputed and complete observations contain the same kind of information.

3. **The last observation carried forward (LOCF) analysis**: it represents a special imputation method that, in presence of drop-out, carries the last observed value for all the remaining occasions until the end of the follow-up. The assumption that patients remain with the same response level since drop-out all along the study is quite unrealistic, expecially for clinical trials, where it is reasonable to expect that individual profiles change after leaving the current treatment. Further, like for imputation methods, it could be problematic to treat imputed subjects in the same way as completers.

These methods can be defined as *ad-hoc* practical procedures rather than as methods based on statistical principles, and the underlying hypotheses are strong and often unrealistic. This is of course against the previous assumption of meaningflulness of missing values, and we will not further deepen these approaches.

3.2 Rubin's taxonomy

In this Section we give some definitions regarding the missing data process. First, to clarify what we mean for missing data mechanism (or process) we define the missing indicator as

$$\delta_i(t) = \begin{cases} 1 \text{ if } & y_i(t) \text{ is missing} \\ 0 \text{ if } & y_i(t) \text{ is present.} \end{cases} \tag{3.3}$$

Hence, it is clear that the dropout event can be treated as a random variable describing a stochastic process varying through time for each individual. As we will further deepen in the next paragraphs, this concept can be extended also to include survival analysis model when continuous time.
Literature on missing data has so far been based on specific definition of the missing process. By indicating with $\mathbf{y}_i^o$ the observed and $\mathbf{y}_i^m$ the unobserved outcome, Rubin (1976) defined missing data as:

- **Missing Completely At Random** (MCAR), when the distribution of the missing data indicator do not depend either on observed or unobserved data, i.e.

$$P(\delta_i(t)|\mathbf{y}_i^o, \mathbf{y}_i^m, \mathbf{x}_i) = P(\delta_i|\mathbf{x}_i)$$

- **Missing At Random** (MAR), when the event depends only on the observed information (response, covariates or both), that is

$$P(\delta_i|\mathbf{y}_i^o, \mathbf{y}_i^m, \mathbf{x}_i) = P(\delta_i|\mathbf{y}_i^o, \mathbf{x}_i)$$

- **Missing Not At Random** (MNAR), when the event is assumed to be related both with the observed and unobserved response.

Here the focus is on understanding whether the dropout process is related to the measurement process, see Diggle and Kenward (1994). In case of a mechanism generating MCAR missing data, a CC analysis may lead to less efficient but still valid estimators, since the dropout is assumed to be random and it does not add any information suitable for inference. On the other hand, the MAR hypothesis, when adopting a likelihood based approach based on the *observed* information only, is equivalent to the assumption of MCAR. When the dropout is assumed to be MNAR, standard inference procedures, lead other than to inefficiency, to potential bias in parameter estimates. While the former definition entails the mechanism generating missing data, a further taxonomy could be of interest when the influence of such missingness is inspected with regards to parameter estimates. In this case, a dropout process is defined to be:

- **Ignorable** when a combination of MAR and separability in model parameters between the measurement and the dropout processes hold;

- **Non-ignorable** (or **informative**) when either parameter separability or a MNAR missing data mechanism hold.

According to Shafer (1997), by parameter separability we mean that the joint parameter space (η, ξ) is the Cartesian product of the individual parameter spaces for η and ξ, where η represents the parameter vector for the longitudinal process and ξ the parameter vector for the missing data process.

A nice review of the objectives and assumptions of models to treat longitudinal data with drop-out is given by Diggle and Henderson (2007). In particular, they consider the simple case where a quantitative response $\mathbf{Y}$ should potentially be measured at two occasions, $t = 1, 2$, but could not be recorded at $t = 2$ for subjects who drop-out. By focusing on the estimation of $\eta(t) = \mathbb{E}[\mathbf{Y}(t)]$, parametrizations so far can be described by the following model:

$$\begin{cases} \mathbf{Y}(1) = \mu(1) + \varepsilon(1) \\ \mathbf{Y}(2) = \mu(2) + \varepsilon(2) \\ \mathbb{E}(\varepsilon(1)) = \mathbb{E}(\varepsilon(2)) = 0, \end{cases} \tag{3.4}$$

where $\mathbf{Y}(t)$ is the response at time t. To be more specific, at $t = 2$ we can distinguish between the unobserved part of the response $Y_{\text{miss}}(2)$ and the outcome corresponding to subjects who complete the study $Y_{\text{obs}}(2)$. Then, a model for the outcome can be formulated as

$$\begin{cases} \mathbf{Y}(1) = \mu(1) + \varepsilon(1) \\ \mathbf{Y}_{\text{obs}}(2) = \mu_{\text{obs}}(2) + \varepsilon_{\text{obs}}(2) \\ \mathbf{Y}_{\text{miss}}(2) = \mu_{\text{miss}}(2) + \varepsilon_{\text{miss}}(2) \\ p(\delta_i(2) = 0 \mid \mathbf{X}) = \pi(\mathbf{X}) \\ \mathbb{E}(\varepsilon(1)) = \mathbb{E}(\varepsilon_{\text{obs}}(2)) = \mathbb{E}(\varepsilon_{\text{miss}}(2)) = 0, \end{cases} \tag{3.5}$$

where $\delta_i(2)$ is the drop-out indicator at time $t = 2$, $\mu(1)$, $\mu_{\text{miss}}(2)$ and $\mu_{\text{obs}}(2)$ are the marginal expectations of $\mathbf{Y}(1)$, $\mathbf{Y}_{\text{miss}}(2)$ and $\mathbf{Y}_{\text{obs}}(2)$, respectively. Given model structure (3.4), the different objectives in missing data models can be summarized as follows:

- Model the *realized* second response, which is given by

$$\mathbf{Y}(2) = \delta_i(2)\mathbf{Y}_{\text{obs}}(2) + (1 - \delta_i(2))\mathbf{Y}_{\text{miss}}(2). \tag{3.6}$$

  Most of published studies are implicitly based on the strong and untestable assumption that $\mathbf{Y}_{\text{miss}}(2) = \mathbf{Y}_{\text{obs}}(2)$. This hypothesis may hold when the drop-out does not affect the measurement process other than making the response unobserved. Otherwise, it could lead to misleading inference about $\mathbf{Y}(2)$. For istance, when dropout is caused by death, the missing response at $t = 2$ can not be hypothetically measured after that time, and model 3.6 is meaningless. Thus, one

should pay attention to the nature of drop-out, to choose the proper parametrization.

- Model the *conditional* second response, i.e. make inference on the response at $t = 2$ conditional on not dropping out:

$$\mathbf{Y}(2) = \begin{cases} \mathbf{Y}_{\text{obs}}(2) & \text{if } \delta_i(2) = 0 \\ \text{undefined} & \text{if } \delta_i(2) = 1. \end{cases} \tag{3.7}$$

  Hence, only completers contribute to inference. This objective is perfectly proper, when the goal is to study the response for completers.

- Model the *hypothetical* second response, i.e. assume

$$\mathbf{Y}(2) = \mathbf{Y}_{\text{obs}}(2). \tag{3.8}$$

  This approach differs from the one proposed in (3.7) because the latter is based on the *conditional* distribution of the response at $t = 2$, whereas model (3.8) focuses on the *marginal* one. Thus, if we assume $\mathbf{Y}_{\text{obs}}(2) = \mathbf{Y}_{\text{miss}}(2)$ the realized and hypothetical response simply coincide.

In practice, a combinations of objectives may also be appropriate. For instance, Kurland (2005) discuss an application where two causes of drop-out are considered: death and possibly informative loss to follow-up. Inference is conducted on the *hypothetical* distribution of the response in absence of loss due to follow-up, by *conditioning* on not being died (combination of models (3.7) and (3.8)).

Most methods proposed in the literature, see for istance Hogan and Laird (1997), Hogan *et al.* (2004) and Davidian *et al.* (2005), can be summarized in the following categories:

1. *Procedures based on Complete Cases or Imputation analysis*, already described in Section 3.1. In the case of MCAR these procedures do not lead to biased estimates.

2. *Weighting procedures*, common in population based surveys and based on the modification of the Horvitz and Thompson (1952) estimator for the population mean as

$$\sum_{i=1}^{n} (\pi_i \widehat{p}_i)^{-1} y_i / \sum_{i=1}^{n} (\pi_i \widehat{p}_i)^{-1},$$

   where π_i is the (known) probability of inclusion in the sample for the i-th unit, while $\widehat{p}_i$ is the estimate of the probability of the response being observed for the i-th unit, usually the proportion of responding units in a subclass of the sample. This approach can be placed in the framework of a MAR drop-out model, given that the goal is to estimate $\widehat{p}_i$, leaving unspecified the relationship between $\mathbf{Y}_{\text{obs}}$ and $\mathbf{Y}_{\text{miss}}$.

3. *Model-based procedures*, aiming at model the joint distribution of the measurement process $\mathbf{y}_i$ and the dropout process (T_i, δ_i), where T_i is either the follow-up duration when $\delta_i = 0$ or the time-to-drop-out (due to several reasons) when $\delta_i = 1$. Different models correspond to different factorizations of the joint distribution. It is straightforward that these procedures can be adequate to handle MNAR mechanisms. We will further deepen this approach in Section 3.3.

In this context, a relevant issue entails hypothesis testing. In fact, it is potentially infeasible to verify whether the missing mechanism is dependent on *unobserved* data. Diggle (1990) and Ridout (1994) developed an hypothesis testing considering a MCAR based null hypothesis against a MAR based alternative hypothesis. More recently, sensitivity analysis approaches has been developed, see for instance Verbeke *et al.* (2001), Troxel *et al.* (2004), Ma *et al.* (2005) and Creemers *et al.* (2010). In Chapter 4 we will explore this approach on a theoretical point of view and will focus on how the assumptions on the drop-out process affect parameter estimates of the longitudinal process. This is of interest if one is not interested in establishing the realism of the hypotheses on the missing process, but, more practically, their potential effects on inference.

3.3 Model frameworks

While the taxonomy in Section (3.2) entails the classification of missing data mechanisms and their effect on corresponding inferences, in this Section we will discuss the procedures to joint model the distribution for the primary (longitudinal) and the drop-out processes.

Lee and Nelder (2009) notice that:

> "In the statistical literature unobservables appear with various names such as random effects, latent processes, factor, missing data, unobserved future observations, potential outcomes, and so on."

These names often correspond to different model structures, that sometimes represent the final product of a *philosophical* point of view rather than an appropriate step-to-step procedure. As pointed out by Henderson *et al.* (2000), an optimal model choice in the missing data framework is cumbersome to be obtained, given that the underlying assumptions are frequently untestable.

A general scheme for existing models is shown in Table 3.1, where the joint distribution of the longitudinal and the drop-out processes, $P(\mathbf{y}_i, \delta_i|\mu, \phi)$, is specified.

More precisely, one can distinguish between:

- *Pattern-Mixture Models*, that formulate separate sub-models for $(\mathbf{Y}_{\text{obs}} \mid \delta_i(t) = 0)$ and $(\mathbf{Y}_{\text{obs}}, \mathbf{Y}_{\text{miss}} \mid \delta_i(t) = 1)$. Hence, valid inference for $\mathbf{Y}_{\text{obs}}$, and *conditional* (on $\delta_i(t) = 1$) inference for $\mathbf{Y}_{\text{miss}}$. This is appealing for studies where the main objective is to compare the response distribution in subgroups with possibly different drop-out times. It can be less immediate when one would allow drop-out to

Models	$P(\mathbf{y}_i, \delta_i \mid \eta, \phi)$	References
Pattern Mixture	$P(\mathbf{y}_i \mid \delta_i, \eta) P(\delta_i \mid \phi)$	Little (1993)
Selection	$P(\mathbf{y}_i \mid \eta) P(\delta_i \mid \mathbf{y}_i, \phi)$	Diggle *et al.* (1994)
Shared Parameter	$\int_{b_i} P(y_i \mid b_i, \eta) P(r_i \mid b_i, \phi) P(b_i) db_i$	Wu and Carrol (1988)

Table 3.1: Avaliable models for longitudinal data with dropout.

depend on the history of subjects only, since it allows dependence on the future also.

- *Selection Models*, see Diggle *et al.* (1994), based on an explicit model to handle the distribution of the drop-out process given the measurement mechanism. In the specific case of model (3.4) with two time occasions, by assuming $(\mathbf{Y}(1), \mathbf{Y}(2)) \sim N(\mathbf{0}, \sigma^2 \mathbf{V})$, where V represent the corresponding correlation matrix, the drop-out process distribution is given by:

$$P[\delta_i(2) = 1] = \frac{\exp(\beta_0 + \beta_1 \mathbf{Y}(1) + \alpha \mathbf{Y}(2))}{1 + \exp(\beta_0 + \beta_1 \mathbf{Y}(1) + \alpha \mathbf{Y}(2))}, \tag{3.9}$$

  with the assumption $\mathbf{Y}(2) = \mathbf{Y}_{\text{obs}}(2) = \mathbf{Y}_{\text{miss}}(2)$. Hence, drop-out is MAR when $\alpha = 0$. The inference on $Y_{miss}(2)$ is based on estimating $P[\mathbf{Y}(1), \mathbf{Y}(2)]$ with its conditional expectation, derived from the conditional distribution of $(\mathbf{Y}(2) \mid \mathbf{Y}(1))$. On the other hand, correct inference depends on untestable assumptions of normality for $(\mathbf{Y}(1), \mathbf{Y}(2))$ and on the use of a logistic model for the drop-out.

- *Shared Parameter Models.* If the omitted covariates substantially contribute to the response distribution, a mixed model for the longitudinal process could be appealing. When the unobservable heterogeneity is assumed to be shared by the measurement and the drop-out processes, a *shared parameter model* holds. By following the model in (3.9), i.e. assuming a logistic model for the probability of drop-out, in the simple case of two measurement occasions the model can be written as follows:

$$\begin{cases} \mathbf{Y}(1) = \mu(1) + \mathbf{b} + \varepsilon_1 \\ \mathbf{Y}(2) = \mu(2) + \mathbf{b} + \varepsilon_2 \\ \mathbf{b} \sim N(0, \mathbf{D}) \\ (\varepsilon_1, \varepsilon_2) \sim N(0, \sigma^2) \\ P[\delta_i(2) = 1] = \dfrac{\exp\{\beta_0 + \alpha b\}}{1 + \exp\{\beta_0 + \alpha \mathbf{b}\}}, \end{cases} \tag{3.10}$$

  where independence between $\mathbf{b}$, ε_1 and ε_2 and conditional independence between $\mathbf{Y}(1)$ and $\mathbf{Y}(2)$ given the random effects hold. The basic idea is to define a more

general model than (3.9), where sources of dependence between the longitudinal and the drop-out processes can be expressed also through unobsered or omitted covariates, accounting for quite general, at least in theory, missing data process. Even if these models are directly identifiable, the distributional assumptions are generally untestable. We will explore this approach in detail in Section (3.4).

3.4 Shared parameter models

In this Section, we give a detailed review of the wide class of models referred to as shared parameter models. The first formulation was probably introduced by Wu and Carrol (1988), where the focus was on longitudinal data subject to right censoring due to participants' death or withdrawal. They referred to this phenomenon as *primary right censoring process*, and propose to model this kind of event through a probit model. In formulas, they assume the response follow a random effects model of the form

$$\begin{cases} \mathbf{y}_i = \mathbf{x}_i\mathbf{b}_i + \varepsilon_i \\ P(\delta_i, T; \xi, \mathbf{b}_i) = \Phi(\xi^{\mathsf{T}}\mathbf{b}_i) \end{cases} \tag{3.11}$$

where $\mathbf{b}_i$ are normally distributed random effects, T is the predefined length of the study, ξ the set of regression parameters for the dropout process and Φ the normal cumulative density function of a standard Gaussian random variable. Hence the longitudinal and the right-censoring primary processes are assumed to share some sources of unobservable individual-specific variability; in other words, the individual *biological* variability is assumed to depend *only* on the fact that different subjects have different propensity to drop-out, or, more correctly, that the underlying heterogeneity influence both the measurement and the censoring mechanisms.

The shared parameter model has been further developed, among others, by Follmann and Wu (1995) and Wulfsohn and Tsiatis (1997). Follmann and Wu (1995) assume that the influence of missing data on the random effect distribution can be summarized by a location change which appear to influence the primary response, while Wulfsohn and Tsiatis (1997) focus mainly on how the parameters of the drop-out process, modelled through a Cox regression model, are influenced by missing values in time dependent covariates.

An nice overview of existing methods is given by Tsiatis and Davidian (2004), where joint modelling of a longitudinal continuous response and a measure of possibly censored time-to-event are discussed. The longitudinal process may be considered as a time dependent covariate in the time-to-event process, leading for instance to a proportional hazard model of the form

$$\lambda_i(t) = \lambda_0(t)\exp\{\xi^{\mathsf{T}}\mathbf{W}_i + \alpha\mathbf{Y}_i(1,t)\}$$

where $\mathbf{W}_i$ are baseline covariates (for instance the treatment) and $\mathbf{Y}_i(1,t)$ is the history of the longitudinal process up to time t. $\mathbf{Y}_i(1,t)$ can be considered as a *surrogate marker*,

see Prentice (1989), i.e. satisfying the following conditions: i) treatments must have an effect on the time-to-event, ii) treatment must have an effect on the marker, and iii) the effect of the treatments should manifest through the marker, i.e. the risk of an event given a specific marker trajectory should be independent of treatment. Inference on ξ and α is pursued by maximizing the partial likelihood

$$\prod_{i=1}^{n}\left[\frac{\exp\{\xi^{\mathsf{T}}\mathbf{W}_i+\alpha\mathbf{Y}_i(t)\}}{\sum_{i=1}^{n}\exp\{\xi^{\mathsf{T}}\mathbf{W}_i+\alpha\mathbf{Y}_i(t)}\right].$$

As stressed before, however, the longitudinal marker $\mathbf{Y}_i(t)$ may include missing values due to drop-out or withdraw from the study. To consider this mechanism, a shared parameter model is postulated by defining a random effect model for the longitudinal outcome:

$$\mathbb{E}\mathbf{Y}_i(t)=f(t)^{\mathsf{T}}\mathbf{b}_i,$$

where $f(t)$ is a vector of functions of time, and $\mathbf{b}_i$ are independent Gaussian random effects representing between-subjects variation in the features of the *true* longitudinal trajectories. The corresponding model for the time-to-event is therefore

$$\lambda_i(t)=\lambda_0(t)\exp\{\xi^{\mathsf{T}}\mathbf{W}_i+\alpha f(t)^{\mathsf{T}}\mathbf{b}_i\}. \tag{3.12}$$

This is the most used parametrization in the context of shared parameter models. Self and Pawitan (1992) propose a longitudinal model of the form (3.12) with the term $\exp\{\xi^{\mathsf{T}}\mathbf{W}_i\}$ replaced by $1+\xi^{\mathsf{T}}\mathbf{W}_i$ to make the hazard depend linearly on $\mathbf{b}_i$, while Pawitan and Self (1993) consider a parametric model for the hazard in (3.12).

A limiting aspect of shared parameter models is that they assume a perfect correlation between the random effects in the longitudinal and in the drop-out processes. To solve this problem and propose a more general and flexible approach, Rizopoulos *et al.* (2008b) introduce dependence between the dropout and the longitudinal random effects through copula functions, see Nelsen (1999). This parametrization is similar to the one in Henderson *et al.* (2000), where a bivariate, correlated Gaussian process is introduced to account for dependence. Hence, the bivariate (joint) random effect density takes the form:

$$p(b_{yi},b_{ti}\mid\delta_i)=\begin{cases}p(b_{ti};\xi), & \text{if } \delta_i(t_i)=0\\ \mathcal{C}(p(b_{yi};\eta),p(b_{ti};\xi)) & \text{if } \delta_i(t_i)=1,\end{cases}$$

where $\mathcal{C}(\cdot,\cdot)$ is the density of the copula $C(\cdot)$, b_{yi} and b_{ti} are the (correlated) random effects for the longitudinal and the time-to-event processes.

Among others, Huang *et al.* (2009) discuss diagnostic methods to check random effect model misspecification.

We now present the likelihood formulation for a shared parameter model. We define

$$\begin{cases} p(\delta_i(t) \mid \mathbf{b}_i, \xi) & \text{as the density of the dropout process,} \\ p(y_i(t) \mid \mathbf{b}_i, \eta) & \text{as the density of the longitudinal process,} \\ p(\mathbf{b}_i \mid \mathbf{D}) & \text{as the density of the random effects,} \\ & \text{usually a zero centered multivariate Gaussian.} \end{cases}$$

The likelihood function can be written as

$$L(\mathbf{y}_i, \delta_i, \mathbf{b}_i; \theta) = \prod_{i=1}^{n} \int_{\mathbf{b}_i} p(\delta_i(t) \mid \mathbf{b}_i, \xi) p(y_i(t) \mid \mathbf{b}_i, \eta) p(\mathbf{b}_i \mid \mathbf{D}) d\mathbf{b}_i, \tag{3.13}$$

where θ is the vector of all model parameters. Numerical approximations of the integral in (3.13), such as Gaussian Quadrature described in Section 2.4.2, are usually needed. On the other hand, some authors stress that parametric assumptions (usually normality) on $\mathbf{b}_i$ may be restrictive and sometimes incorrect. For instance, Tsiatis and Davidian (2001) propose a *conditional score* approach, that does not make any distributional assumptions upon the underlying random effect distribution, leading to unbiased estimates of ξ and η treating $\mathbf{b}_i$ as a nuisance parameter and conditioning on a sufficient statistic. Song *et al.* (2002a) follow a semi-parametric likelihood approach where $\mathbf{b}_i$ is assumed to have a conditional density in a class $\mathcal{H}$ that account for skewness, and includes multi-modal distributions (the Gaussian distribution is a specific case of this class).

Example 3.4.1. *In the framework of shared parameter models,* ***Joint Models*** *(JM), introduced by Wulfsohn and Tsiatis (1997), may represent a useful parametrization. This model assumes, for a continuous longitudinal response $y_i(t)$, a linear mixed model of the form*

$$y_i(t) = \beta^{\mathsf{T}} X_i(t) + \mathbf{b}_i^{\mathsf{T}} Z_i(t) + \varepsilon_i(t).$$

Let us indicate by $m_i(t) = \beta^{\mathsf{T}} X_i(t) + \mathbf{b}_i^{\mathsf{T}} Z_i(t)$ the expected value of the longitudinal process; the corresponding hazard for the event at time t is given by

$$\lambda_i(t) = \lambda_0(t) \exp\{\gamma^{\mathsf{T}} \mathbf{w}_i + \alpha m_i(t)\}.$$

In this sense, the (error-free) "true" pattern of the longitudinal outcome rather than the observed value is assumed to influence the hazard of an event. By "true" pattern we mean the value of the predicted process at time t, as postulated by the underlying model (which may also include with splines or other functional forms).
Interesting enough, in this parametrization both the fixed and the random effects are shared by the two processes of interest. This means that the maximum likelihood estimate for β has a contribution from both the longitudinal and the drop-out processes. Thus, parameter separability holds only when $\alpha = 0$; the same is true for the ignorability of the missing data process.

Chapter 4

Local sensitivity to nonignorability in shared parameter models

In this Chapter we study the sensitivity of shared parameter models to assumptions about nonignorability of the dropout process. In Section 4.1, we give a general description of what we mean by sensitivity analysis, and why such analysis should be performed in presence of missing data. In Section 4.1.1 we give an overview of the main sensitivity tools proposed in literature so far. We introduce the Index of Local Sensitivity to ingnorability (ISNI) in Section 4.2, while in Section 4.3 we describe a new proposal concerning local sensitivity which is defined to deal with shared parameter models. The adopted parameterizations are described in Section 4.4, where we further derive our proposed sensitivity approach for this class of models. Sections 4.5, 4.6 and 4.7 discuss the use and properties of the approach in a series of benchmark datasets, while 4.8 deals with a large scale simulation study designed to analyse proposed indexes behaviour in a variety of empirical situations. Section 4.9 gives concluding remarks and outlines potential developments.

4.1 What is a sensitivity analysis?

Sensitivity analysis aims at assessing to what extent the conclusions that can be drawn by adopting a particular modelling structure are dependent on the (explicit or implicit) assumptions one makes on it. By definition, it is clear that several approaches are plausible, depending on *which modeling assumptions* the researcher is interested in, and which procedures he/she choices. In this paragraph, we will mention the more interesting sensitivity approaches proposed in the literature where dependence between the longitudinal outcome and the missing data mechanism is concerned, with a particular emphasis on those defined to check for assumptions about the dropout process in shared parameter models, see Section 3.4.

4.1.1 Sensitivity tools

Several approaches have been proposed to investigate the sensitivity of model parameter estimates to assumptions regarding the drop-out mechanism, but the majority of those has been focused on Selection and Pattern Mixture Models (SeM and PMM), see e.g. Troxel *et al.* (2004), Ma *et al.* (2005), Molenberghs and Verbeke (2005) and Molemberghs and Kenward (2007), with minor contributions to the framework of Shared Parameter Models. In this Section, we will discuss the main sensitivity tools proposed in the literature, focusing on Shared Parameter Models (SPMM in the following).
As we have previously mentioned, sensitivity means exploring the robustness of a given model to one or more model assumptions. From this perspective, one should decide *which* assumptions are to be considered; this depends on either the aim of the analysis or the prior knowledge of the researcher about which assumptions are weaker or more questionable from a theoretical and/or empirical point of view.
For instance, it is known that shared parameter models are based on the longitudinal and drop-out process sharing one or more random effects. Rizopoulos *et al.* (2008a) explore the sensitivity of shared parameter models to assumptions upon the random effect distribution when parameter estimates and corresponding standard errors are considered. The authors find out that, as the number of repeated measures per individual in the longitudinal process increases, the maximum likelihood estimator of model parameters $\widehat{\eta}$ under any distributional assumptions upon the random effects tends to converge to the maximum likelihood estimator under the correct model for $\mathbf{b}_i^*$. The reason for this result is that, as the number of repeated measures increases, the longitudinal measumerement process becomes the dominating part of the posterior distribution $p(\mathbf{b}_i \mid \mathbf{y}_i, \delta_i; \theta)$, implying that the choice of the prior distribution has a minor role. On the other hand, the effect on estimated standard errors of this distributional assumption could be more prominent.
Another important issue in SPM regards the definition of the dependence structure between the longitudinal and the dropout process. As it has been pointed out in Section 3.4, the standard SPM postulates the existence of a *perfect* correlation between the two mechanisms, given that they share the *same* random coefficients. It is clear that this hypothesis may not appropriately describe the data structure. It may happen that, for instance, the correlation between the random coefficients of the two processes exists but it is different from 1. Sensitivity of shared parameter models to the association between the missing and the longitudinal processes has been conducted by Rizopoulos *et al.* (2008b). The authors describe the dependence structure between the random effects through the application of *copulas*, see Section 3.4. The results show that different kind of copulas, chosen to describe the association structure, can significantly alter parameter estimates.
In the context of missing data, a relevant but usually untestable assumption is the ignorability of the drop-out process. It is known that SeM, PMM and SPM assume that the drop-out mechanism is MNAR and model potential dependence through the introduc-

tion of non-ignorable parameters (SeM and SPM) or through conditioning the observed response on the observed pattern of participation to the study (PMM). In this context, one may wonder whether the parameter estimates obtained through the corresponding MAR model are different from the parameter corresponding to the MNAR assumption, i.e. if the parameter estimates are *sensitive* to the assumption of non-ignorability.
Alfó *et al.* (2010) propose an informal approach to check sensitivity of parameter estimates from mixed logistic model when missing data are supposed to be non-ignorable and evaluate the results in a simulation study. They consider binary longitudinal data and focus on how much significance of parameter estimates changes, when a pattern mixture model (with interaction between the length of the participation to the study and all the other covariates in the model) is fitted as preliminary sensitivity tool. The following example will clarify this concept.

Example 4.1.1. *Let us consider a simulation study where the binary longitudinal response is defined according to*

$$\begin{aligned} y_i(t) \mid \mathbf{x}_i(t), \mathbf{b}_i &\sim Bin(1, \pi_{it}), \\ \text{logit}(\pi_{it}) &= (\beta_0 + b_{i1}) + x_i(t)\beta_1 + z_i(t)(\beta_2 + b_{i2}) \quad i = 1, \ldots, n \\ & \quad t = 1, \ldots, T \end{aligned}$$

where $X_{it} \sim N(0, 3)$ *and* $Z_{it} \sim N(0, 1.5)$ *are the design matrices for the fixed and random effects, respectively,* $\mathbf{b}_i \sim MVN\left(\mathbf{0}, \begin{bmatrix} 1 & 0 \\ 0 & 1.2 \end{bmatrix}\right)$ *are the random coefficients and* $n = \{100, 200, 500\}$ *the sample sizes. Moreover the measurement occasions are* $T = \{10, 15\}$ *and the fixed effects are* $\beta_0 = 0.5$, $\beta_1 = 0.5$ *and* $\beta_2 = -0.7$. *The drop-out process is described by an exponential random variable* $t_i \sim \text{Exp}(\lambda_i)$ *where three MNAR mechanisms are considered:*

1. $\lambda_i = \exp(\sum_{t=1}^{T} y_{it}) = \exp(\sum y_{it}^0 + y_{it}^m)$
2. $\lambda_i = \exp(\rho W + b_{i2})$, $W \sim N(0, 1)$, $\rho = 2$
3. $\lambda_i = \exp(\rho W + b_{i1})$ ",

while ignorable assumptions are expressed by the following scenarios:

- *MCAR:* $\lambda_i = \lambda$
- *MAR:* $\lambda_i = \exp(\sum_{t=1}^{S_i} y_{it}) = \exp(\sum y_{it}^o)$

The fitted pattern mixture model of the form

$$y = (\beta_0 + b_1) + \beta_1 X + (\beta_2 + b_2)Z + \gamma_0 S + \gamma_1 X * S + \gamma_2 Z * S$$

, where $S = \sum_{t=1}^{T} 1 - \delta_i(t)$, *where* $\delta_i(t) = 0$ *if the individual is observed at time* t, $\delta_i(t) = 1$ *otherwise. The variable* S *plays a central role to understand how the propensity to stay*

in the study influences parameter estimates and their significance. The authors guess that a significant estimate for γ_0, γ_1 or γ_2 implies that the missing data mechanism is non ignorable as far as parameter estimates in the "main" model, i.e. β_0, β_1 and β_2 are concerned. Tables 4.1 and 4.2 represent the PMM parameter estimates when MCAR, MAR and MNAR assumtions hold.

Setup			Parameters			perc. sign.		
	n	T	$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{\beta}_2$	γ_0	γ_1	γ_2
MCAR	100	10	0.447	0.480	-0.633	0.070	0.040	0.050
	100	15	0.572	0.523	-0.650	0.060	0.100	0.110
	200	10	0.540	0.488	-0.694	0.079	0.050	0.109
	200	15	0.556	0.483	-0.654	0.094	0.059	0.084
	500	10	0.483	0.499	-0.651	0.062	0.047	0.016
	500	15	0.432	0.461	-0.621	0.134	0.045	0.149
MAR	100	10	**0.883**	0.466	-0.734	0.08	0.04	0.100
	100	15	**0.636**	0.483	-0.703	0.14	0.02	0.190
	200	10	**0.798**	0.529	-0.634	0.07	0.09	0.050
	200	15	**0.712**	0.478	-0.716	0.1	0.07	0.070
	500	10	**0.739**	0.493	-0.592	0.125	0.028	0.069
	500	15	**0.635**	0.487	-0.657	0.122	0.068	0.041

Table 4.1: PMM for the longitudinal process. MCAR and MAR case

The proportion of simulated samples where the γs, representing the interaction between covariates (intercept included) and the length of the observed individual sequence, S, are significant at a prespecified α level, say $\alpha = 0.05$, are substantially higher, when the dropout is MNAR. To sum up, PMM can be viewed as a sensitivity analysis tool to gain information on the dependence link between the two processes, as well as on the potential effects of this dependence on parameter estimates for the longitudinal process. When a random intercept model is used for the longitudinal process, MNAR data lead, at most, to biased intercept estimates, as shown by Table 4.2, case MNAR3. In this case, we do not really need shared intercept parameter models.

In the following Sections, we will consider another approach based on the concept of *local sensitivity*, that focuses on the changes of maximum likelihood estimates as one moves from the MAR to the MNAR assumption.

4.2 Local sensitivity to non-ignorability

In this Section, we review the general formulation of the *Index of Local Sensitivity to Nonignorability* (ISNI), proposed by Troxel *et al.* (2004) and extended by Ma *et al.* (2005). Other interesting extensions of this approach to Selection Models are dealt with

Setup			Parameters			perc. sign.		
	n	T	$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{\beta}_2$	γ_0	γ_1	γ_2
MNAR1	100	10	**1.615**	0.481	-0.735	**0.550**	0.06	0.070
	100	15	**1.266**	0.511	-0.678	**0.570**	0.06	0.090
	200	10	**1.318**	0.508	-0.669	**0.851**	0.104	0.119
	200	15	**1.366**	0.478	-0.712	**0.845**	0.068	0.078
	500	10	**1.461**	0.492	-0.654	**1.000**	0.029	0.059
	500	15	**1.256**	0.483	-0.64	**0.990**	0.04	0.070
MNAR2	100	10	0.617	0.537	**0.302**	0.090	0.060	**0.710**
	100	15	0.448	0.512	**0.259**	0.090	0.060	**0.820**
	200	10	0.590	0.475	**0.299**	0.020	0.030	**0.900**
	200	15	0.444	0.496	**0.247**	0.110	0.070	**0.970**
	500	10	0.499	0.506	**0.404**	0.045	0.036	**0.991**
	500	15	0.516	0.499	**0.306**	0.087	0.058	**1.000**
MNAR3	100	10	**1.904**	0.532	-0.712	**0.830**	0.07	0.070
	100	15	**1.813**	0.562	-0.867	**0.870**	0.05	0.130
	200	10	**1.573**	0.497	-0.677	**0.995**	0.035	0.139
	200	15	**1.281**	0.471	-0.647	**1.000**	0.06	0.104
	500	10	**1.902**	0.511	-0.659	**1.000**	0.078	0.097
	500	15	**1.645**	0.507	-0.698	**1.000**	0.047	0.047

Table 4.2: Average parameter estimates and proportion of samples where parameter estimates are significant at level 0.05. MNAR case.

by Xie (2008), Xie (2009) and Qian and Xie (2010). The ISNI is defined to measure the local sensitivity of ML parameter estimates to departures from the MAR assumption, i.e. to investigate how much maximum likelihood parameter estimates for the longitudinal process are influenced by the hypothesis about ignorability of the drop-out mechanism. We will start by adopting the modelling structure given by Wulfsohn and Tsiatis (1997). In this Chapter, the interest will lie only on the parameters for the longitudinal process, described by a linear mixed model.

Let $\widehat{\beta}(\alpha)$ and $\widehat{\gamma}(\alpha)$ denote the ML estimates for the longitudinal and the survival process parameters in a general shared parameter model, when a non null value of α is kept fixed. On the other hand, let $\widehat{\beta}_0$ and $\widehat{\gamma}_0$ be the corresponding ML estimates under the MCAR model obtained by setting $\alpha = 0$. The ISNI measures the rate of change of $\widehat{\beta}(\alpha)$ from $\widehat{\beta}_0$, for a unit displacement of α from 0; it is based on the derivative of $\widehat{\beta}(\alpha)$ with respect to α, evaluated at $\widehat{\beta}_0$, $\widehat{\gamma}_0$ and $\alpha = 0$. To derive the index, the likelihood function is expanded around $(\widehat{\beta}_0, \widehat{\gamma}_0, \alpha = 0)$; by writing $\theta = (\theta_1, \theta_2, \theta_3) = (\beta, \gamma, \alpha)$, we

have:

$$\begin{aligned} L(\beta,\gamma,\alpha) &\approx L(\beta_0,\gamma_0,\alpha) + \left[(\beta-\widehat{\beta}_0)',(\gamma-\widehat{\gamma}_0)',\alpha\right]\nabla L \\ &+ \frac{1}{2}\left[(\beta-\widehat{\beta}_0)',(\gamma-\widehat{\gamma}_0)',\alpha\right]\nabla^2 L\left[(\beta-\widehat{\beta}_0)',(\gamma-\widehat{\gamma}_0)',\alpha\right]', \end{aligned}$$

where $\nabla L = \{\nabla L_i\}_{i=1,2,3}$ and $\nabla^2 L = \{\nabla^2 L_{ij}\}_{i,j=1,2,3}$ represent the score vector and the Hessian matrix for the parameter vector. The index of sensitivity to nonignorability is defined as follows:

$$\text{ISNI} = \nabla\widehat{\beta}(\alpha)\big|_{\alpha=0} = -(\nabla^2 L_{11})^{-1}\nabla^2 L_{13}\big|_{\alpha=0}. \tag{4.1}$$

The main advantage of this approach is that it does not actually require to fit the MNAR model, but only to compute the corresponding Hessian matrix. The previous expansion allows us to write the ML estimate of β as a function of α; according to Xie (2008) we may write:

$$\beta(\alpha) \cong \beta_0 + \frac{\partial\beta(\alpha)}{\partial\alpha}\alpha. \tag{4.2}$$

yielding to the following approximation:

$$\text{ISNI} \cong \frac{1}{\alpha}\left(\beta(\alpha) - \beta_0\right), \tag{4.3}$$

which represents the ratio of the difference between the MNAR and the MAR estimates to the value of the nonignorability parameter. Xie (2008) extend the ISNI methodology to handle longitudinal non-Gaussian data subject to non-ignorable dropout, by assuming a Selection Model. The author considers the approximation $\beta_1(\alpha) = \beta_1(0) + \text{ISNI}\alpha$ for the MNAR parameter estimate and observes that this approximation seems to be related to the drop-out proportion. However, as it can be easily noticed, the proposed ISNI represents an absolute measure and therefore can be hard to interpret. In fact, we need to assess whether the observed ISNI value could be due to a substantial departure of the MNAR estimates from the MAR ones, or rather to sampling variability in the index, since when a SPM is fitted, $\alpha = 0$ means that the β estimates change interpretation. For this purpose, many relative formulations have been proposed; for instance, Troxel *et al.* (2004) consider the ratio of the ISNI to the standard error of the corresponding MAR estimate, stating sensitivity when this ratio is greater than 1. However, as it can be evinced by looking at expression (4.3), the standard error of the MAR estimate clearly understates the index variability, since it neglects the variability of the ISNI, which could be present also when the adopted model is the true one and $\alpha = 0$.

An interesting alternative for a relative formulation of the ISNI is the relative ISNI (*isni*), based on the ratio of the absolute ISNI to the corresponding MAR estimate; for example, if the j-th element of β is considered, the *isni* is defined as follows:

$$isni_{\beta_j} = \frac{ISNI_{\beta_j}}{\widehat{\beta}_{0j}}. \tag{4.4}$$

Since the ISNI is the rate of change of a parameter estimate, it could be reasonable to compare this change to the corresponding MAR estimate; this ratio would help in assessing the effective weight of the displacement, due to α moving away from zero. We suggest to consider a value of $|isni| \geq 0.5$ as indicating some sensitivity; this means that, for a unit change of the association parameter from zero, the ML parameter estimate varies by 50% of its actual value. The *isni* represents a direct comparison between the index and the corresponding parameter estimate calculated under MAR assumption and leads to a direct and straightforward interpretation. However, some concerns could raise when the MAR estimate is near to zero; in this case, an interesting alternative could be represented by the ratio of the ISNI to a measure of its sampling variability. As it can easily observed by looking at (4.2), this variability can not be consistently estimated by the standard error of the MAR estimate. Rather, we propose to estimate the ISNI as the slope in the regression model (4.2) by using, as a formal response, the parameter estimates $\widehat{\beta}(\alpha)$ calculated for the data at hand (generated from the MAR hypothesis, with true $\alpha = 0$), by fixing α to a set of predetermined values. These values serve as a covariate in the regression model, where $\widehat{\beta}_0$ represents the intercept, the ISNI is the corresponding slope and the index variability could be approximated by the slope standard error. This may lead to an estimate of the sampling variability for the index, since, for $\alpha = 0$, the only source of variability is the one around the ISNI mean value.

4.3 Local sensitivity in shared parameter models

In this Section, we study the sensitivity of inferences to assumptions regarding the drop-out mechanism when a shared parameter model is considered. The majority of literature on sensitivity to the assumptions upon ignorability of drop-out process has focused on Selection and Pattern Mixture Models, see e.g. Troxel *et al.* (2004), Ma *et al.* (2005), Molenberghs and Verbeke (2005) and Xie (2008). Not too much has been done so far for the Shared Parameter Model, with the only exception of the recent work by Creemers *et al.* (2010). In the latter, the authors consider various specifications for a SPM, obtained by varying the random effect structure. In these parameterizations, they introduce a *scale* parameter which is not identifiable, representing the effect of the dependence between the missingness process and the missing observations given the observed ones. This parameter plays the role of *sensitivity parameter*; only by fixing the sensitivity parameter the model can be fitted. A grid in the sensitivity parameter space is defined, and for each of the values, the model is fitted and the missing response is imputed using conventional multiple imputation procedures. Re-fitting the model using these imputations, and summarizing the different inferences resulting from different imputations into a single set of inference, purports to sensitivity analysis. However, the obtained results represent a function not only of the random effect structure but also of the adopted imputation procedure. Therefore, we cannot distinguish between sensitivity to model structure and/or missing data imputation procedure.

For this reason, we focus on local sensitivity and extend the Index of Local Sensitivity (ISNI), see Section 4.2, to shared parameter models framework; our goal is at measuring how much the maximum likelihood estimates are influenced by assumptions regarding the dependence between the longitudinal outcome and the drop-out mechanism. The main advantage of this approach is that it does not require a complete shared parameter model to be fitted but, rather, it is based on quantities that can be calculated by fitting a missing at random (MAR) model.
As far as the model structure is entailed, we will discuss two different parameterizations. First, the standard Shared Parameter Model, where subject-specific random effects induce association between the longitudinal and the survival process. Second, the so called Joint Model, where the expected value of the primary (longitudinal) response influences the current risk of the event; in this model structure, fixed and random effects are shared by the longitudinal and the survival model. Since the ISNI is an absolute measure of changes in parameter estimates induced by departures from the MAR assumption, we propose a relative formulation based on the ratio between the ISNI and the corresponding MAR estimate, highlighting potential interpretation and drawbacks of this (relative) index. Next, we focus on the sampling variability of the index, providing different estimates for the corresponding standard error; this will help us define a further relative index, defined as the ratio of the ISNI to the sampling variability when the true model is MAR. We will discuss three longitudinal studies to highlight different conclusions about sensitivity and ISNI behaviour. The first one is a clinical trial concerning AIDS, see Goldman *et al.* (1996) and Carlin and Louis (2009), with the objective of comparing the efficacy/safety of two antiretroviral drugs by recording a longitudinal response (the CD4 cell count) and the time to death of 467 HIV infected patients. We are particularly interested in investigating how much the longitudinal evolution of the CD4 cells count and the occurrence of the event are related. As a second example, we consider the primary biliary cirrhosis (PBC) data, see Murtaugh *et al.* (2002), where 312 patients have been considered to test for treatment effect on survival after adjusting for the longitudinal bilirubin levels. Third, we consider a longitudinal study on chronic kidney disease, see Rizopoulos *et al.* (2008a), where 407 patients undergoing a primary renal transplantation with a graft in the University Hospital of Leuven (Belgium) between 1983 and 2000 have been considered. Further, we discuss the performance of abolute and relative indices through a simulation study where several scenarios for the number of subjects, the random effect covariance structure, the association structure between the longitudinal and the survival process are considered. This would help us study indices behaviour when the model is correctly specified and when it is partially/globally misspecified.

4.4 Shared Parameter Models for Gaussian Random Variables

In this Section we describe two particular parametrizations of the Linear Shared Parameter Models we are interested in.

Longitudinal studies often record two types of outcomes: repeated measurements of a response of interest, and realizations of a time-to-event process representing drop-out. An example can be derived from AIDS studies, where interest lies in the longitudinal evolution of markers such as the CD4 cells count. In this setting, the longitudinal response is directly related to the event process, where the event could be represented by sieroconvertion, dropout due to several reasons, or death; after an event has occurred, longitudinal measurements are no longer collected or considered nonrelevant, therefore inducing dropout. In many cases, the dropout process may depend on the unobserved values of the longitudinal response, thus corresponding to a missing not at random framework. Literature has so far concentrated on three modeling frameworks for the joint analysis of a longitudinal outcome and a (nonrandom) dropout process, these have been have reviewed in Chapter 3: Selection Models, Pattern Mixture Models, and Shared Parameter Models. Even though these models can be quite flexible in the specification of the dropout mechanism, not too much trust must placed on corresponding inferences due to potential sensitivity of parameter estimates to (unverifiable) modelling assumptions. For istance, Little (1995) suggest sensitivity analyses to assess the effect on inferences of alternative assumptions about the drop-out process about target quantities. In particular, as it has been pointed out, among others, by Molemberghs and Kenward (2007), the MNAR model is not fully verifiable from the data, since it is based on the assumption that the dropout mechanism depends on *unobservable* variables. This phenomenon is complicated by the fact that, for every MNAR model fitted to a set of data, there is a MAR counterpart providing exactly the same fit. This can be proved for shared parameter model through the following steps:

1. The likelihood for the MNAR model is

$$L(\cdot) = \prod_{i=1}^{n} \int_{\mathbf{b}_i} p(\mathbf{y}_i \mid \eta, \mathbf{b}_i) p(\delta_i \mid \phi, \mathbf{b}_i) p(\mathbf{b}_i \mid \mathbf{D}) d\mathbf{b}_i.$$

   This leads to the maximum likelihood estimates $\widehat{\eta}$, $\widehat{\phi}$ and $\widehat{\mathbf{D}}$ for the longitudinal and dropout model parameters and the random effects covariance matrix.

2. The hypothetical fit to the *fully observed* data is

$$p(\mathbf{y}_i, \delta_i \mid \widehat{\eta}, \widehat{\phi}, \widehat{\mathbf{D}} = \int_{\mathbf{b}_i} p(\mathbf{y}_i \mid \eta, \mathbf{b}_i) p(\delta_i \mid \phi, \mathbf{b}_i) p(\mathbf{b}_i \mid \mathbf{D}) d\mathbf{b}_i.$$

3. The dropout mechanism is MAR if

$$p(\delta_i \mid \phi, \mathbf{b}_i) = p(\delta_i \mid \phi).$$

This means that the random effects are not shared by the dropout and the longitudinal processes, but represent only between subjects variation in the longitudinal process due to repeated measurements taken at different time points on the same individuals.

4. The fit obtained from a MNAR model is exactly riproducible from a MAR model given that

$$L(\cdot) = \prod_{i=1}^{n} \int_{\mathbf{b}_i} p(\mathbf{y}_i \mid \eta, \mathbf{b}_i) p(\delta_i \mid \phi, \mathbf{b}_i) p(\mathbf{b}_i \mid \mathbf{D}) d\mathbf{b}_i \tag{4.5}$$

$$= \prod_{i=1}^{n} p(\mathbf{y}_i \mid \eta) p(\delta_i \mid \phi), \tag{4.6}$$

which means that it is possible to find η^* and ϕ^* such that the second equivalence in (4.5) is verified.

The basic assumption of shared parameter models is that of *conditional independence*, expressed by the following equations:

$$\begin{aligned} p(T_i, \delta_i, y_i \mid b_i; \theta) &= p(T_i, \delta_i \mid b_i; \theta) p(y_i \mid b_i; \theta) \\ p(y_i \mid b_i; \theta) &= \prod_j p\{y_i(t_{ij}) \mid b_i; \theta\}. \end{aligned}$$

That is, conditionally on the random effects, the repeated measurements for the generic individual are independent, and the same is true for the longitudinal and the survival processes. The random coefficients allow for within-individual dependence in the longitudinal process, and for dependence between the longitudinal and the survival processes.

Let $Y_i(t)$ represent a longitudinal continuous response recorded for the ith subject $(i = 1, \ldots, n)$ at time t and let $y_i(t)$ be the corresponding observed outcome. First, we will assume that $Y_i(t)$ follows a linear mixed model of the form:

$$Y_i(t) = \beta^{\mathsf{T}} \mathbf{x}_i(t) + \mathbf{b}_i^{\mathsf{T}} \mathbf{z}_i(t) + \varepsilon_i(t), \tag{4.7}$$

where $\varepsilon_i(t) \sim N(0, \sigma^2)$ is the measurement error, $\mathbf{b}_i \sim MVN(0, D)$ is a set of random coefficients, while $\mathbf{x}_i(t)$ and $\mathbf{z}_i(t)$ are the design vectors corresponding to fixed and random effects, respectively. The random terms $\mathbf{b}_i$ and ε_i are assumed to be independent; the first is time-invariant and shared by the response for the i-th subject i responses, while the latter represents unstructured, time-varying, random variation from the *true* individual signal. Often, longitudinal outcomes are not observed for the whole study period since some individuals may leave the study before its designed end, potentially due to a secondary event. That is, the occurrence of the event at time t may induce dropout in the longitudinal outcome since no longitudinal measurements are collected at time t or afterwards. The dropout is said to be non random if the probability of dropout, conditional on the observed unit characteristics, still depends, either directly

or indirectly, on the unobserved longitudinal responses. In this case, standard estimation procedures may lead to inconsistent parameter estimates. Formally, let y_i^o and y_i^m denote the observed and the missing longitudinal responses for the i-th unit, $i = 1, \ldots, n$. Let us denote by $T_i = \min(T_i^*, C_i)$ the observed failure time for the ith individual, taken as the minimum between the true event time T_i^* and the censoring time C_i, in most empirical cases the study completion time; δ_i is the corresponding event indicator defined by $\delta_i = I(T_i^* \leq C_i)$. The longitudinal response y_i is therefore observed before T_i and is missing at and after T_i, and the observed failure time T_i may represent either a true event time or a censoring time. Typically, we assume that the longitudinal process is associated with T_i^*, i.e with the true event time, but is independent of C_i. According to this framework, dropout could be due to the occurrence of a particular event (with $T_i^* \leq C_i$ and $\delta_i = 1$) or to censoring due to potentially noninformative events (with $T_i^* > C_i$ and $\delta_i = 0$). In this sense, a clear distinction should be made between dropouts due to events where longitudinal responses are no longer available (eg death events which may represent censoring events) and dropouts where longitudinal responses could have been registered should the subject have been participating to the study. Our aim is to account for potentially informative dropouts and investigate their effect on model parameter estimates for the longitudinal outcome; shared parameter models (SPMs) represent an appealing framework for joint modeling of longitudinal and survival processes, since the repeated measurements and the time to dropout are assumed to share a set of time-invariant, subject-specific random effects, which induce dependence in the univariate profiles as well as between the two processes.

To be more general, the association between the longitudinal (primary) and the dropout (secondary) process is defined by adopting two different parametrizations: the first one, referred to as the *joint model*, introduces the expected value of the longitudinal process in the model describing the hazard function for the survival process. The second one, referred to as the standard *shared parameter model*, postulates that a set of random effects is shared by the longitudinal and the survival model. In both cases, the survival process is assumed to follow a proportional hazard model, Cox (1972). Let us first consider the joint model as formulated by Fawcett and Thomas (1996) and Wulfsohn and Tsiatis (1997); here, the risk of experiencing the event at time t depends on the expected value of the longitudinal outcome at the same time, where random biological is not considered. The model can be expressed by a set of two equations, one for the longitudinal and the other one for the survival part:

$$\begin{cases} Y_i(t) = m_i(t) + \varepsilon_i(t) \\ h_i(t \mid M_i(t), \mathbf{w}_i) = h_0(t) \exp\{\gamma^\mathsf{T} \mathbf{w}_i + \alpha m_i(t)\}, \quad i = 1, \ldots, n \end{cases} \tag{4.8}$$

where $m_i(t) = \beta^\mathsf{T} \mathbf{x}_i(t) + b_i^\mathsf{T} \mathbf{z}_i(t)$ is the expected value of the longitudinal process at time t for the ith individual, $M_i(t) = \{m_i(u), 0 \leq u < t\}$ represents the history of $m_i(t)$ until time t, W_i is a row vector of baseline covariates and $h_0(t)$ denotes the baseline risk function. The degree of dependence between the longitudinal and the survival processes

is measured by the association parameter α, which is introduced to assess potential nonignorability of the missing data mechanism. As it can be easily noticed, in model (4.8), the survival process share the set of fixed and random effects defining the linear predictor for the longitudinal response. Thus, we have two potential sources of non-ignorability: when $\alpha \neq 0$ the two processes are not independent, since they share the same set of random coefficients; furthermore, fixed effects β appear in both submodels and parameter distinctiveness does not hold. Adopting a different perspective, we may look at standard SPMs, where the two submodels share only a set of random coefficients; the aim here is to study how unobserved individual-specific variability in the longitudinal process influences the time to the event. In this case, the hazard function can be defined as follows:

$$\begin{aligned} h_i(t \mid b_i, W_i) &= h_0(t) \exp\{\gamma^\mathsf{T} W_i + \alpha^\mathsf{T} b_i\} \\ &= h_0(t) \exp\{\gamma^\mathsf{T} W_i + \alpha_1 b_{1i} + \ldots + \alpha_k b_{ki}\}. \end{aligned} \tag{4.9}$$

This model structure can be of interest should the focus be on identifying the influence on time-to-event of subject-specific unobservable characteristics, when distinguished from the observed ones. In this case, α represents a vector of k elements, and each of its components is the nonignorability parameter for the corresponding random effect. A relevant issue concerns the meaning of a (near) null estimate for α. In both models (4.8) and (4.9), this leads to a missing completely at random (MCAR) model, that is, to the assumption that the dropout mechanism, once conditioned on available covariates, does not depend on the longitudinal response, either observed or missing. In fact when $\alpha = 0$ model parameters in the two submodels are distinct, the joint probability of the dropout and the longitudinal processes can be factorized as follows:

$$p\left(T_i, \delta_i, Y_i(t)\right) = p(T_i, \delta_i) p(Y_i(t)).$$

The same can be done with the log-likelihood function with respect to the fixed effects in the longitudinal process. As we adopt a maximum likelihood approach to parameter estimation, we know that parameters estimates derived from maximizing the likelihood of the longitudinal process, that is $p(Y_i(t))$, yield maximum likelihood estimates that are valid under both MCAR and MAR assumptions, i.e. under the hypothesis that the dropout mechanism depends on the observed responses only. Thus, while $\alpha = 0$ implies a MCAR mechanism, it leads to parameter estimates that are still valid under MAR hypotheses.

To illustrate that a shared parameter model corresponds to a MNAR mechanism, let us suppose that T_i^* denote the event (e.g. drop-out) time for the i-th individual, i.e. $T_i = \min(T_i^*, C_i)$, and let us indicate with $\mathbf{y}_i^o$ and $\mathbf{y}_i^m$ the set of observed (before time T_i) and missing (at and after time T_i) longitudinal measurements for the i-th subject. The missing data mechanism, i.e. the conditional distribution of the dropout process

given the complete longitudinal data vector $(\mathbf{y}_i^o, \mathbf{y}_i^m)$, is given by:

$$\begin{aligned} p(T_i \mid \mathbf{y}_i^o, \mathbf{y}_i^m) &= \frac{\int p(T_i \mid \mathbf{b}_i) p(\mathbf{y}_i^o, \mathbf{y}_i^m \mid \mathbf{b}_i) p(\mathbf{b}_i) d\mathbf{b}_i}{\int p(\mathbf{y}_i^o, \mathbf{y}_i^m \mid \mathbf{b}_i) p(\mathbf{b}_i) d\mathbf{b}_i} \\ &= \int p(T_i \mid \mathbf{b}_i) p(\mathbf{b}_i \mid \mathbf{y}_i^o, \mathbf{y}_i^m) d\mathbf{b}_i, \end{aligned} \tag{4.10}$$

which depends on $\mathbf{y}_i^m$ through the posterior distribution of the random effects.

4.4.1 ISNI in Shared Parameter Models

In this section, we present the general formulation of the ISNI for shared parameter models, when association structures (4.8) and (4.9) are considered. A more detailed derivation can be found in Appendix A. The log-likelihood function for a general SPM can be written as follows:

$$\ell(\theta) = \ell(\theta \mid T_i, \delta_i, \mathbf{y}_i) = \sum_i \log \int_{\mathbf{b}_i} p(T_i, \delta_i \mid \mathbf{b}_i; \theta) p(\mathbf{y}_i \mid \mathbf{b}_i; \theta) p(\mathbf{b}_i; \theta) d\mathbf{b}_i, \tag{4.11}$$

where

$$\begin{aligned} p(T_i, \delta_i \mid \mathbf{b}_i; \theta) &= p(T_i \mid \mathbf{b}_i; \theta)^{\delta_i} S(T_i \mid M_i(t), W_i; \theta)^{1-\delta_i} \\ &= h(T_i \mid M_i(t), W_i; \theta)^{\delta_i} S(T_i \mid M_i(t), W_i; \theta), \end{aligned}$$

and θ is the overall model parameter vector. If a joint model parameterization is adopted, the survival function at time t is given by

$$S(t \mid M_i(t), W_i; \theta) = \exp\left\{ - \int_0^t h_0(s) \exp\left\{ \gamma^\mathsf{T} W_i + \alpha m_i(s) \right\} ds \right\}, \tag{4.12}$$

while, if the association between the longitudinal response and the time to dropout is based on sharing random coefficients, the survival function can be written as follows

$$S(t \mid \mathbf{b_i}, W_i; \theta) = \exp\left\{ - \int_0^t h_0(s) \exp\left\{ \gamma^\mathsf{T} W_i + \alpha^\mathsf{T} \mathbf{b}_i \right\} ds \right\}. \tag{4.13}$$

The integral in (4.12) does not have an analytical solution, while the one in (4.13) has a closed form. To calculate the ISNI, we need the second order derivatives of the log-likelihood. For a general shared parameter model, see e.g. Rizopoulos *et al.* (2009),

the score function takes the form :

$$\begin{aligned} s(\theta) = \frac{\partial \ell(\theta)}{\partial \theta} &= \sum_i s_i(\theta) = \sum_i \frac{\partial}{\partial \theta^{\mathsf{T}}} \log \int p(T_i, \delta_i \mid \mathbf{b}_i; \theta) p(y_i \mid \mathbf{b}_i; \theta) p(\mathbf{b}_i; \theta) d\mathbf{b}_i \\ &= \sum_i \frac{1}{p(T_i, \delta_i, y_i; \theta)} \frac{\partial}{\partial \theta^{\mathsf{T}}} \int p(T_i, \delta_i \mid \mathbf{b}_i; \theta) p(y_i \mid \mathbf{b}_i; \theta) p(\mathbf{b}_i; \theta) d\mathbf{b}_i \\ &= \sum_i \frac{1}{p(T_i, \delta_i, y_i; \theta)} \int \frac{\partial}{\partial \theta^{\mathsf{T}}} p(T_i, \delta_i \mid \mathbf{b}_i; \theta) p(y_i \mid \mathbf{b}_i; \theta) p(\mathbf{b}_i; \theta) d\mathbf{b}_i \\ &= \sum_i \int \left[\frac{\partial}{\partial \theta^{\mathsf{T}}} \log \left\{ p(T_i, \delta_i \mid \mathbf{b}_i; \theta) p(y_i \mid \mathbf{b}_i; \theta) p(\mathbf{b}_i; \theta) \right\} \right] \\ &\qquad \times \frac{p(T_i, \delta_i \mid \mathbf{b}_i; \theta) p(y_i \mid \mathbf{b}_i; \theta) p(\mathbf{b}_i; \theta)}{p(T_i, \delta_i, y_i; \theta)} d\mathbf{b}_i \qquad (4.14) \\ &= \sum_i \int q(\theta, \mathbf{b}_i) p(\mathbf{b}_i \mid T_i, \delta_i, y_i; \theta) d\mathbf{b}_i, \end{aligned}$$

where

$$\begin{aligned} q(\theta, b_i) &= \frac{\partial}{\partial \theta^{\mathsf{T}}} \log p(T_i, \delta_i, y_i, b_i; \theta) \\ &= \frac{\partial}{\partial \theta^{\mathsf{T}}} \left\{ \log p(T_i, \delta_i \mid b_i; \theta) + \log p(y_i \mid b_i; \theta) + \log p(b_i; \theta) \right\}. \end{aligned}$$

For the index pair (u, v) and the i-th individual, the generic element in the Hessian matrix takes the following form:

$$\begin{aligned} \frac{\partial s_i(\theta_u)}{\partial \theta_v} &= \frac{\partial}{\partial \theta_v} \int q(\theta_u, \mathbf{b}_i) p(\mathbf{b}_i \mid T_i, \delta_i, y_i; \theta_u) d\mathbf{b}_i \\ &= \int \frac{\partial q(\theta_u, \mathbf{b}_i)}{\partial \theta_v} p(\mathbf{b}_i \mid T_i, \delta_i, y_i; \theta) d\mathbf{b}_i + I_1, \end{aligned}$$

where

$$\begin{aligned} I_1 &= \int q(\theta_u, \mathbf{b}_i) \left\{ \frac{\partial \log p(\mathbf{b}_i \mid T_i, \delta_i, y_i; \theta_u)}{\partial \theta_v} \right\}^{\mathsf{T}} p(\mathbf{b}_i \mid T_i, \delta_i, y_i; \theta_u) d\mathbf{b}_i \\ &= \int q(\theta_u; \mathbf{b}_i) \Bigg\{ \frac{\partial \log p(T_i, \delta_i \mid \mathbf{b}_i; \theta_u) + \log p(y_i \mid \mathbf{b}_i; \theta_u) + \log p(\mathbf{b}_i; \theta_u)}{\partial \theta_v} - \\ &\quad \frac{\partial \log p(T_i, \delta_i, y_i; \theta_u)}{\partial \theta_v} \Bigg\}^{\mathsf{T}} p(\mathbf{b}_i \mid T_i, \delta_i, y_i; \theta_u) d\mathbf{b}_i \\ &= \int q(\theta_u, \mathbf{b}_i) \left\{ q(\theta_v, \mathbf{b}_i) - s_i(\theta_v) \right\}^{\mathsf{T}} p(\mathbf{b}_i \mid T_i, \delta_i, y_i; \theta_u) d\mathbf{b}_i. \end{aligned}$$

and, as outlined before, $s_i(\theta)$ is the score vector for the ith subject. To calculate the ISNI, we have to consider the matrix with (u, v)-th element given by

$$H_{\theta_u \theta_v} = \frac{\partial s_i(\theta_u)}{\partial \theta_v} \Big|_{\alpha=0} \qquad (4.15)$$

Here, α is a constant or a vector depending on the adopted model structure, i.e. on the sources of variability that induce dependence between the two processes.
When we adopt a joint model, α is a constant and the ISNI is a vector, while when we adopt a shared parameter model, α is a vector and the corresponding ISNI is a matrix with $j = 1, \ldots, K$ columns, each representing the ISNI with respect to a generic element α_j in α (see Appendix A for further details). From a computational point of view, we have to analytically compute the score vector with respect to the parameters of interest and then proceed to numerical derivation. To calculate the Hessian, we apply Gauss–Hermite quadrature since the integral with respect to $\mathbf{b}_i$ in the score function (4.14) does not have a closed form.

4.4.2 Assessing ISNI variability

A further key question concerns how to assess the sampling variability of the ISNI, when the MAR assumption it true. This could be of interest for different reasons: first, to assess the precision of the ISNI estimate; second, it could lead to a valid alternative for a relative index of sensitivity, rather than comparing the absolute ISNI to the standard error of the MAR estimate, as in Troxel *et al.* (2004).
Moreover, Shared Parameter Models are based on the so called *non-separability* of parameter estimates. This means that, for any value of α, there could be sensitivity to non-ignorability even if the proportion of dropout is null. In this case, we have $\delta_i = 0$, and

$$p(T_i, \delta_i \mid \mathbf{b}_i) = h_i(t)^{\delta_i} S_i(t) = S_i(t),$$

where $S_i(t) = \exp\{-\int_0^t h_0(t) \exp\{\gamma^\mathsf{T}\mathbf{W}_i + \alpha m_i(t)\}\}$ for the joint model parametrization and $S_i(t) = \exp\{-\int_0^t h_0(t) \exp\{\gamma^\mathsf{T}\mathbf{W}_i + \alpha \mathbf{b}_i)\}\}$ for the standard shared parameter model. Hence, the dependence between the two processes still holds, and α could assume a non-null estimate also in case of no subjects dropping out.
Previous considerations allow us to suggest that an estimate of ISNI variability calculated under the MAR assumption could be an appealing measure of the behaviour of the index, and the relative formulation ISNI/se(ISNI) may account for the index variability and for interpretational issues due to parameter non-separability. As it can be noticed by looking at the simulation study we will illustrate at the end of this paragraph (see Table 4.3), the standard error of the ISNI when α is assumed to be null is different from zero, meaning that the ISNI is not exactly null under non-ignorability.
To find a measure of the ISNI variability, we may approximate the ISNI as the slope of the regression model in equation (4.2), where $\beta(\alpha)$ is computed from the joint model by fixing a priori different values for the nonignorability parameter, and β_0 is fixed at the MAR estimate. By definition, the ISNI represents the slope of the tangent curve to $\beta(\alpha)$ at $\alpha = 0$.
The approximation in equation (4.2) gives an interesting framework to obtain an estimate of the ISNI standard error. To show that, we performed the following small

simulation study. We have drawn samples from the joint model (4.8) with $\alpha = 0$ following the simulation design in Section 4.8. We define α as a vector containing equally spaced values in the range $[-2, 2]$, compute β_0 as the MAR estimate from the simulated data and $\beta(\alpha)$ as the MNAR estimates corresponding to fixed values for α. We estimate the ISNI from model (4.2) and calculate the corresponding standard error with respect to the time variable. This procedure has been repeated for $B = 100$ samples. The resulting standard error estimate, the Monte Carlo estimate of the standard error obtained from the simulation study discussed in section 4.8 when data are generated according a *true* MAR model, i.e. with $\alpha = 0$, and the standard error of the MAR estimate are shown in Table 4.3 with respect to the time variable effect.

Regression	0.041
Monte Carlo	0.072
MAR	0.004

Table 4.3: Simulation results. Estimates of the standard error of the ISNI under a MAR assumption: regression model estimate, Monte Carlo estimate, standar error of the MAR estimate for the time effect.

It is evident that the proposed approximation to the ISNI standard error and the Monte Carlo estimate of the same quantity seem to agree; therefore, this could be interpreted as the *sampling* variability of the ISNI around its mean value when $\alpha = 0$, i.e. under a MAR model, due to parameter non-separability. Whereas, the standard error of the MAR estimate is much smaller.
Finally, Figure (4.1) represents the regression approximation of the ISNI and corresponding current values of $\beta(\alpha)$ with varying α.

It is clear that the linear approximation is better as α approaches 0.

4.5 AIDS data

We look back at the motivating example briefly mentioned in Section 4.3, see Goldman *et al.* (1996) and Carlin and Louis (2009). The longitudinal study on 467 HIV infected patients with the aim at comparing the efficacy and safety of two randomly assigned antiretroviral drugs: didanosine (ddI) and zalcitabine (ddC). The longitudinal response is the CD4 cell count, recorded at the randomization time and after 2, 6, 12 and 18 months. By the end of the study, 188 patients have died, corresponding to 60% censoring.

For the longitudinal process, we consider a linear mixed model of the form:

$$y_{it} = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + b_{0i} + b_{1i} x_{i1} + \varepsilon_{it}, \tag{4.16}$$

where x_{i1} and x_{i2} represent time and interaction between treatment and time, respectively. For the survival part, Weibull proportional hazard model is adopted,

$$h_{it} = \xi t^{\xi-1} \exp\{\gamma_0 + \gamma_1 w_i + \alpha m_i(t)\}.$$

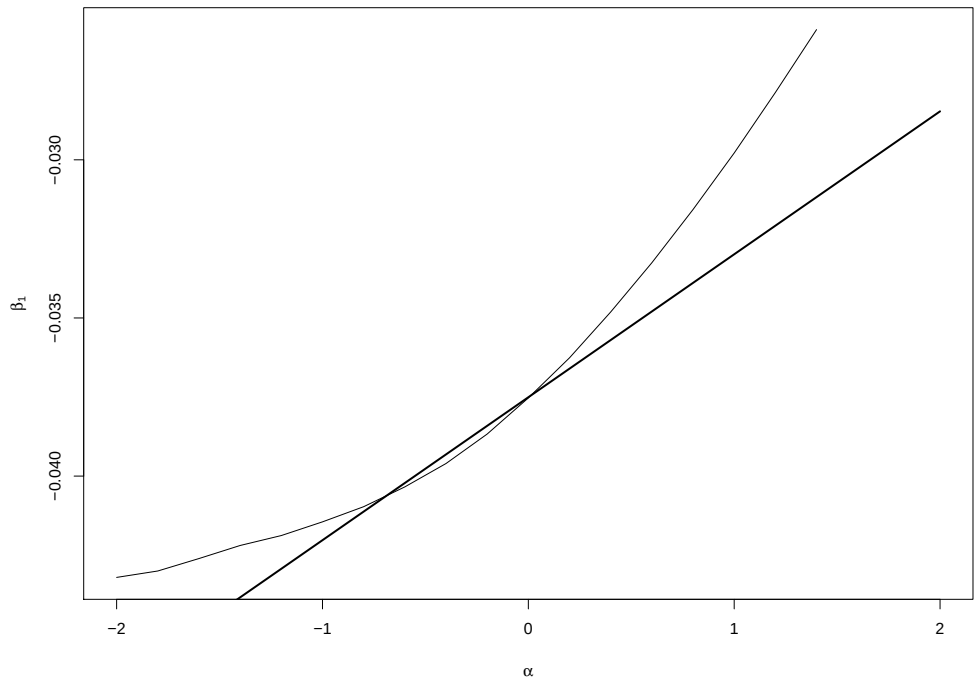

Figure 4.1: Linear approximation of the ISNI.

For this and the following analysed dataser, we will focus only on sensitivity of JM parameters estimates, since the objective will be the evaluation of the performance of the ISNI rather than the comparison of the sensitivity between the two parametrizations.

Table 4.4 shows the values of the absolute and the relative ISNI for the AIDS dataset.

	$\widehat{\beta}_0$	se$(\widehat{\beta}_0)$	$\widehat{\beta}$	ISNI	ISNI$/\widehat{\beta}_0$	ISNI/se$(\widehat{\beta}_0)$
Intercept	2.512	0.042	2.554	-0.022	-0.009	-0.524
Time	-0.037	0.004	-0.041	0.036	-0.972	9.000
Treat*Time	0.008	0.006	0.005	-0.003	-0.375	-0.500

Table 4.4: Absolute and relative ISNI for the AIDS data set. $\alpha = -0.844$, standard deviation of the longitudinal response $\widehat{\sigma} = 0.368$ and random effects covariance matrix $D_1 = (0.7594, -0.0005, -0.0005, 0.0013.$

The joint model seems to be quite robust to misspecification of the nonignorability parameter, suggesting that small departures from $\alpha = 0$ slightly affect parameters estimates. If one looks at the parameters estimates under the MAR and the MNAR joint model, it is clear that they do not experience a wide change. This aspect is visible

also in the ISNI and ISNI/$\widehat{\theta}_0$ formulations, whereas the ISNI/se($\widehat{\beta}_0$) may be misleading with respect to the time covariate. For this reason, in Table 4.5 we show the alternative relative version of the ISNI, i.e. ISNI/se($ISNI$) and the regression based estimates of the standard error of the index, se(ISNI).

	ISNI/se(ISNI)	se(ISNI)
Intercept	-2.200	0.010
Time	0.400	0.090
Treat*Time	-0.428	0.007

Table 4.5: Regression based relative ISNI for the AIDS data set. $\alpha = -0.844$.

Comparing the absolute and the regression based relative version of the ISNI, we can conclude that, for this dataset, fitting a MAR model would not result in large changes in the parameter estimates for the longitudinal process.

4.6 Primary Biliary Cirrhosis data

A second example comes from the primary biliary cirrhosis (PBC) data collected by the Mayo Clinic from 1974 to 1984, see Murtaugh *et al.* (2002). PBC is a fatal liver desease characterized by inflammatory destruction of the small bile ducts within the liver, which may lead to cirrhosis of the liver. In this study, 312 patients are considered; 158 were randomly assigned to recieve D-penicillamine and 154 placebo. By the end of the study, 140 patients (45%) died, 143 (46%) were alive and 9% were transplanted. We are interested in testing for treatment effect on survival after adjusting for the longitudinal bilirubin levels. Given that the distribution of the observed bilirubin serum shows a certain skeweness, we consider its natural logarithm. Here, we model the longitudinal dependence of the the bilirubin serum and the survival of enrolled patients by employing the following model for the bivariate (longitudinal and survival) process:

$$\begin{cases} \log(y_{it}) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5} + b_{0i} + b_{2i} x_{i2} + \varepsilon_{it} \\ h_{it} = h_0(t) \exp\{\gamma_0 + \gamma_{1i} w_{i1} + \alpha m_i(t)\} \end{cases}, \quad (4.17)$$

where $x_{i1}, \ldots, x_{i5}$ represent treatment, time, interaction between treatment and time, age and gender covariates, while w_{i1} is the treatment. The aim is at exploring sensitivity of parameter estimates in the longitudinal model to the assumption about the missing (survival) data process. The results of the sensitivity analysis are shown in Table 4.6.

This dataset presents a similar situation to the previous one. Even though a quite significant change in parameter estimate is experienced by the gender covariate effect, the ISNI/$\widehat{\beta}_0$ and the ISNI/se($\widehat{\beta}_0$) present high values for the interaction between time and treatment. On the other hand, the other parameters do not experience relatively

	$\widehat{\beta}_0$	se($\widehat{\beta}_0$)	$\widehat{\beta}$	ISNI	ISNI/$\widehat{\beta}_0$	ISNI/se($\widehat{\beta}_0$)
Intercept	0.739	0.349	0.893	-0.021	-0.028	-0.060
Treatment	**-0.141**	0.117	**0.118**	0.023	-0.163	0.196
Time	0.179	0.018	0.174	0.003	0.017	0.166
Age	0.001	0.006	-0.006	1e-04	0.100	0.016
Gender	**-0.202**	0.180	**-0.102**	0.014	-0.069	0.078
Treat*Time	-0.004	0.025	-0.007	-0.049	12.271	-0.960

Table 4.6: Absolute and relative ISNI for the PBC data set. $\alpha = 1.261$

wide changes. At this point, we show in Table 4.7 the regression based standard errors for the ISNI and the relative formulation of the index proposed in this Section.

	ISNI/se(ISNI)	se(ISNI)
Intercept	-0.697	0.0301
Treatment	**1.074**	0.0214
Time	0.214	0.014
Age	0.002	0.0474
Gender	**1.272**	0.0110
Treat*Time	-0.379	0.0129

Table 4.7: Regression based relative ISNI for the PBC data set. $\alpha = 1.261$.

This relative formulation allows us to highlight the change in the gender parameter estimate change and thus it seems a more appropriate sensitivity tool.

4.7 Chronic Kidney Disease data

Last, we turn back to the third example introduced in Section 4.3. This longitudinal study entails 407 patients suffering from chronic kidney disease who underwent, between 1/21/1983 and 8/16/2000, a primary renal transplantation with a graft from a deceased or living donor at the University Hospital of the Catholic University of Leuven (Belgium), see Rizopoulos *et al.* (2008a). Chronic kidney disease, also known as chronic renal disease, is a progressive loss of renal function which can be described by using five stages; each stage is a progression through an abnormally low and progressively worse glomerular filtration rate. The clinical interest is focused on the long term performance of the new graft, and, in particular, on analyzing the time to graft failure, if any. For these purposes, during the follow-up period, patients were periodically tested for the condition and performance of their kidneys; for this purpose, the Glomerular Filtration Rate (GFR), that measures the filtration rate of the kidneys, is considered as a longi-

tudinal response. By the end of the study, 126 patients have suffered for a graft failure, corresponding to 31% of patients exiting the study.

We consider the following formulation for the longitudinal process:

$$y_{it} = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + b_{0i} + b_{1i} x_{i1} + \varepsilon_{it}, \tag{4.18}$$

where x_{i1} and x_{i2} represent time since transplantation and gender, respectively. Furthermore, we postulate a Weibull proportional hazard model for the survival process, with hazard function:

$$h_i(t) = \xi t^{\xi-1} \exp\{\gamma_0 + \gamma_1 w_i + \alpha m_i(t)\},$$

where the association is defined according to model structure (4.8), where w_i represents patients' age. Table 4.8 shows MAR ($\alpha = 0$) and MNAR ($\alpha \neq 0$) estimates as well as the values of the absolute and relative ISNI computed for the kidney dataset.

	MAR		MNAR JM	
Longit. Proc.	$\widehat{\beta}_0$	se($\widehat{\beta}_0$)	$\widehat{\beta}$	se($\widehat{\beta}$)
Intercept	7.169	0.161	7.144	0.028
Time	**-0.101**	**0.015**	**-0.012**	**0.002**
Gender	**-0.417**	**0.104**	**-0.676**	**0.033**
	ISNI	ISNI/($\widehat{\beta}_0$)	ISNI/se($\widehat{\beta}_0$)	ISNI/se(ISNI)
Intercept	-0.078	-0.011	-0.484	0.490
Time	0.007	-0.071	0.467	**1.207**
Gender	0.012	-0.029	0.115	**0.722**

Table 4.8: MAR and MNAR parameter estimates for the kidney data set. $\widehat{\alpha} = -1.395$.

Also in this example, the relative ISNI calculated by the ratio of the absolute index and the regression standard error leads to a clearer interpretation of the effective change in parameter estimates, where one moves from MAR to MNAR assumptions, i.e. from $\alpha = 0$ to $\alpha \neq 0$.

4.8 Simulation study

To investigate the empirical behaviour of the ISNI when a shared parameter model is considered, we performed the following simulation study.

4.8.1 Study design

We simulate the longitudinal response $y_i(t)$ from a Gaussian distribution with mean $m_i(t) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + b_{i0} + b_{i1} x_{1i}$ and standard deviation $\sigma = 0.37$, for a sample

of $n = 467$ individuals observed for $T = 5$ or $T = 15$ occasions, respectively. For the drop-out process, we have adopted the hazard function

$$h_i(t) = h_0(t) \exp\{\gamma_0 + \gamma_1 w_i + \alpha m_i(t)\}, \tag{4.19}$$

for the joint model, and

$$h_i(t) = h_0(t) \exp\{\gamma_0 + \gamma_1 w_i + \alpha_0 b_{0i} + \alpha_1 b_{1i}\}, \tag{4.20}$$

for the standard shared parameter model. In this study, x_{1i} represents the time and x_{2i} the interaction between time and treatment, w_i is randomly drawn from a standard Gaussian distribution, while the design matrices have been simulated according to the AIDS dataset, discussed in Section 4.5. A Weibull baseline hazard function is adopted, i.e. $h_0(t) = \xi t^{\xi-1}$. Fixed effects for the longitudinal process are equal to $\beta_0 = 2.51$, $\beta_1 = -0.37$, $\beta_2 = 0.82$, $\sigma = 0.37$, $\gamma_0 = -3.31$, $\gamma_1 = 0.15$. Individual censoring times have been randomly drawn from an exponential distribution with mean chosen to result in about 50% censoring. To investigate the effect of the random effect covariance structure on parameter estimates, we have considered two different covariance matrices:

$$D_1 = \begin{pmatrix} 0.7594 & -0.0005 \\ -0.0005 & 0.0013 \end{pmatrix},$$

and

$$D_2 = \begin{pmatrix} 0.5 & 0.01 \\ 0.01 & 0.5 \end{pmatrix}.$$

While D_1 is the estimated random effect covariance matrix for the AIDS dataset, D_2 describes a homoscedastic random effect covariance structure with a positive association between b_{i1} and b_{i2} and a higher variability in b_{i2} (comparable to the variability in b_{i1}). To account for different degrees of dependence between the longitudinal and the dropout processes, we have simulated data according to different values for the nonignorability parameter; namely, we used $\alpha = 0$ (MAR model), $\alpha = -0.5$, $\alpha = -1$ and $\alpha = -1.5$ for model (4.19) and $\alpha = (0, 0)$, $\alpha = (0, -1)$, $\alpha = (0, -1.5)$, $\alpha = (-1, 0)$, $\alpha = (-1.5, 0)$, $\alpha = (-1, -1)$, $\alpha = (-1.5, -1.5)$ for model (4.20). The median number of observed measurements per individual is $\overline{n}_i = 3$ (when $T = 5$) and $\overline{n}_i = 7$ (with $T = 15$).

On each simulated dataset, we have computed the ISNI, the *isni*, the ISNI/se(θ_0); when parametrization (4.19) was used, the ISNI is calculated through the R package `JM`, Rizopoulos (2010), while R code written by the author (available on request) has been used for model (4.20). Further, we performed a sensitivity analysis to model misspecification, simulating observed data from model (4.19) and computing absolute and relative indexes of parameter estimates for model (4.20), and viceversa. The aim is to assess robustness of parameter estimates to model specification; this is quite a major point, since the ISNI is known to measure departures from the MAR estimates only when the fitted model is the true one.

4.8.2 Results

Figure 4.2 shows 2.5%, 50% and 97.5% percentiles over $N = 1000$ simulated samples of the three indices ISNI, $ISNI/\theta_0$ and $ISNI/se(\theta_0)$, corresponding to the parameters of the longitudinal process when the joint model parameterization is adopted. We may observe a higher sensitivity in the results as α deviates from 0; we may also notice that the length of the observed individual sequence, denoted by the median number of available responses, $\overline{n}_i$, has a substantial effects on the sensitivity of parameter estimates, all other parameters kept fixed. In particular, for $T = 15$, ie $\overline{n}_i = 7$, the joint model is considerably more robust when compared to the setting where $T = 5$ and $\overline{n}_i = 3$. Higher the absolute value of the nonignorability parameter, higher the variability in the ISNI values. On the contrary, the random effect covariance structure does not seem to produce substantial effect on the ISNI distribution; the results corresponding to covariance matrix D_2 are further explored in Table 4.10.

Results for the shared parameter model are illustrated in Figures 4.3 and 4.4, and in Figure 1 in the supplementary material.

Table 4.9 shows the mean and the median for the MAR parameters estimates and the corresponding standard errors and quantiles, while Table 4.10 contains the simulation results for each value of α in the JM.

The ISNI is computed with respect to the association parameter vector $\alpha = (\alpha_1, \alpha_2)$, where the two elements correspond to the intercept (α_1) and the time effect (α_2). As a first point, it may be observed that the results seem to be slightly affected by the structure of the random effect covariance matrix.

For what concerns the interpretation of the three indexes, the ratio between the ISNI and the standard error of the corresponding MAR estimate assumes widely large values, since the standard error is, in most analysed cases, close to zero, or, to be more precise, on another scale than the ISNI; also, the absolute index might be misleading, as the corresponding values can not be directly interpreted. On the other hand, in this simulation study, the proposed relative index (*isni*) leads to a clearer interpretation, since it provides a direct comparison of potential changes in parameter estimates to the corresponding MAR estimates. As can be evinced by looking at Figures 4.3 and 4.4 when compared to Figure 4.2, the observed sensitivity of β_2 estimates is reduced with ISNI values approaching to zero. This empirical evidence could suggest that fixed effect estimates are not influenced by the presence of MNAR mechanisms if the SPM is the *true* model. Furthermore, these findings should help reconsider the wide use of SPMs to recover potential bias due to misspecified missing data mechanism when fixed effects are the main focus of the analysis. On the contrary, when estimates of β_1 are considered, the corresponding sensitivity is quite high in both the joint and the shared parameter model structures, even if it is decreasing with increasing length of the individual sequence. The sensitivity is often negligible but when $\alpha_2 = -1$ in the SPM structure with covariance matrix D_1. In all other cases, with a higher number of time occasions, only the intercept shows some sensitivity to wrong assumptions about the missing data mechanism.

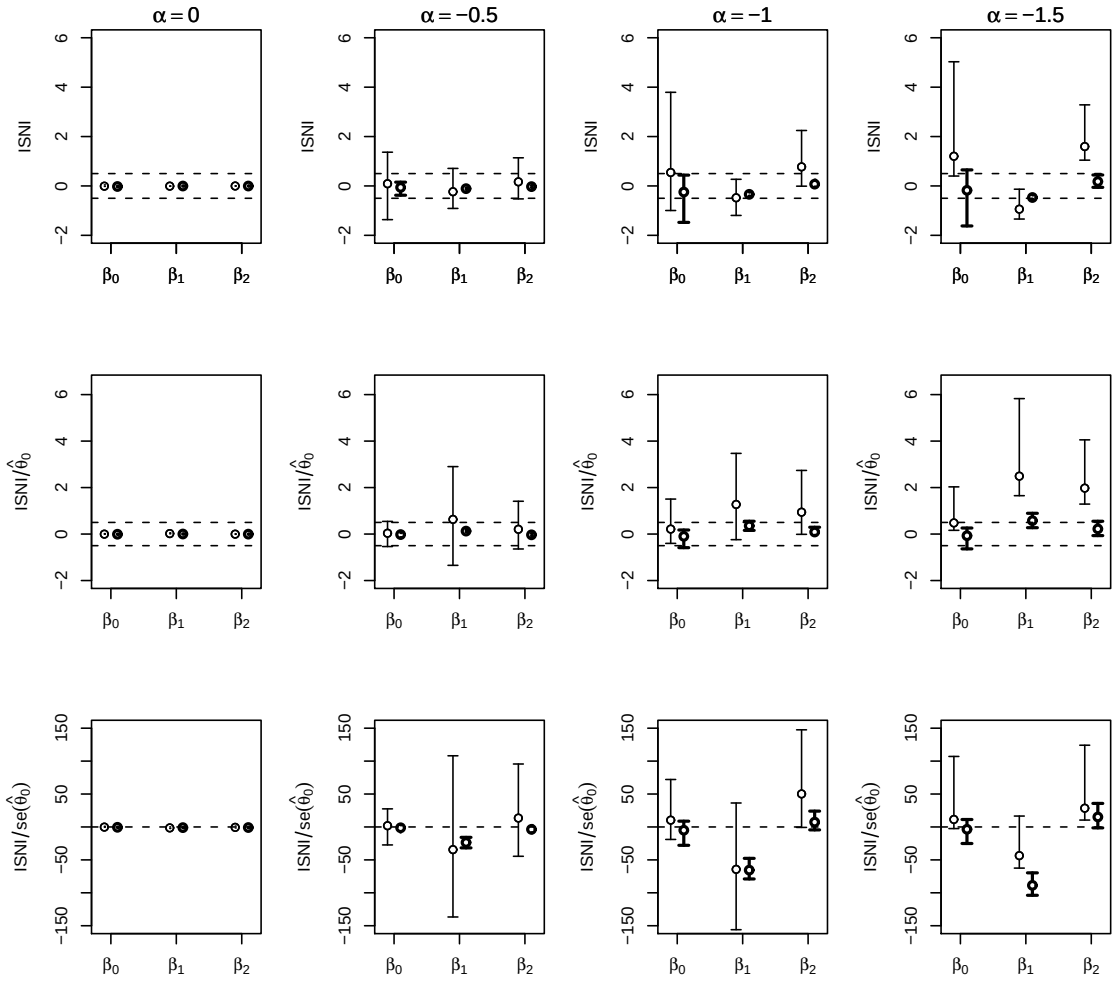

Figure 4.2: Simulation study, results for the joint model on 1000 samples. The circle denotes the median and the edges the 2.5% and 97.5% percentiles of absolute and relative ISNI for $\overline{n}_i = 3$ (thin line) and $\overline{n}_i = 7$ (bold line). The dashed line corresponds to $y = \pm 0.5$ for ISNI and $ISNI/\theta_0$ and $y = 0$ for the $ISNI/se(\theta_0)$. Covariance matrix D_1.

When sensitivity to model misspecification is considered, the results in Figure 4.6 show a higher sensitivity when the joint model is the true one, when compared to the standard shared parameter model; the sensitivity is found to be decreasing with $\alpha_1 = \alpha_2 = \alpha$, where α_1 and α_2 are the association parameters for the SPM formulation, while α represents the association parameter for the JM structure, with the intercept showing the highest sensitivity. However, the median of absolute and relative ISNIs is close to zero; therefore, we may conclude that both proposed modelling structures are quite robust to model misspecification. This represents an appealing result given the

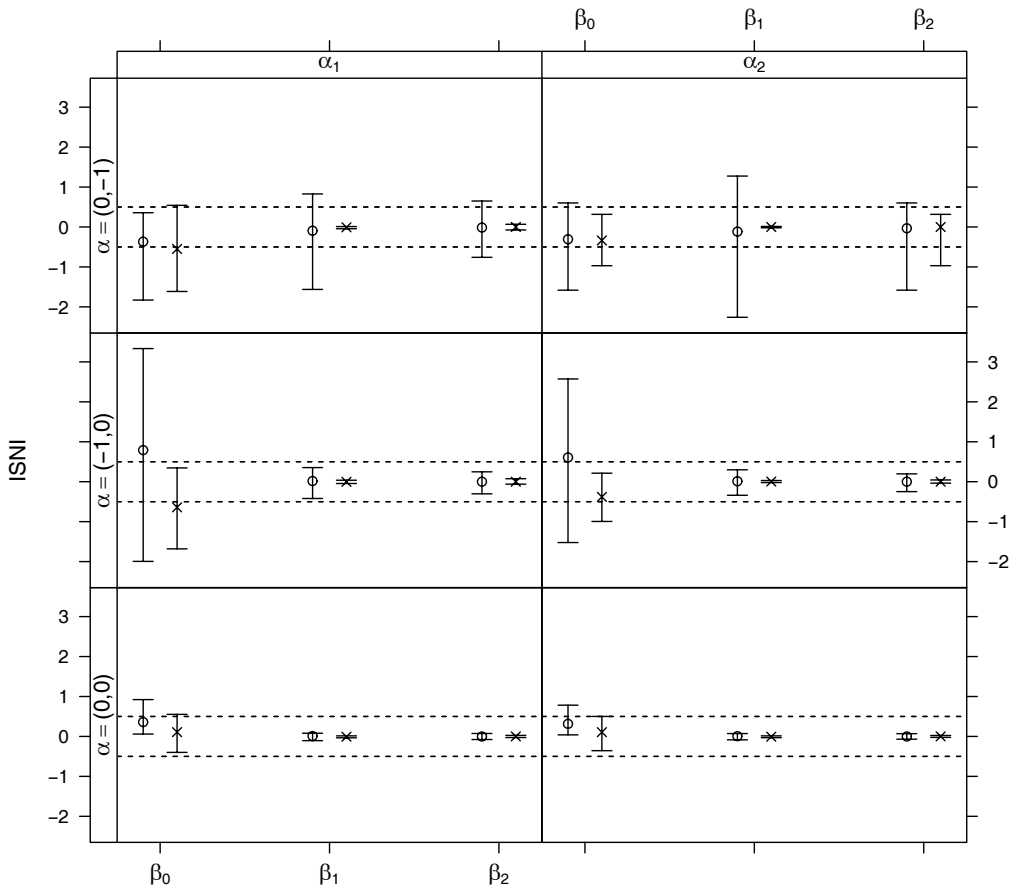

Figure 4.3: Simulation study, absolute ISNI for shared parameter model on 1000 samples, covariance matrix D_1. The points denotes the median and the edges the 2.5% and 97.5% percentiles of the ISNI for $\overline{n}_i = 3$ (circle) and $\overline{n}_i = 7$ (cross). The dashed lines corresponds to $y = \pm 0.5$.

wide use of SPM in empirical and theoretical contexts, see Follmann and Wu (1995), Pulkstenis and Landis (1998) and Ten Have *et al.* (1998).

4.9 Discussion

In this Chapter, we have discussed a local sensitivity analysis based on deriving the ISNI for the general class of shared parameter models, considering several structures to account for dependence between the longitudinal outcome and the time to dropout processes. We have focus on changes registered in model parameters for the longitudinal process due to potential misspecification in the missing data mechanism and in the dependence structure. The data application and the simulation study have shown a slight sensitivity of model parameter estimates for the joint model when the departure

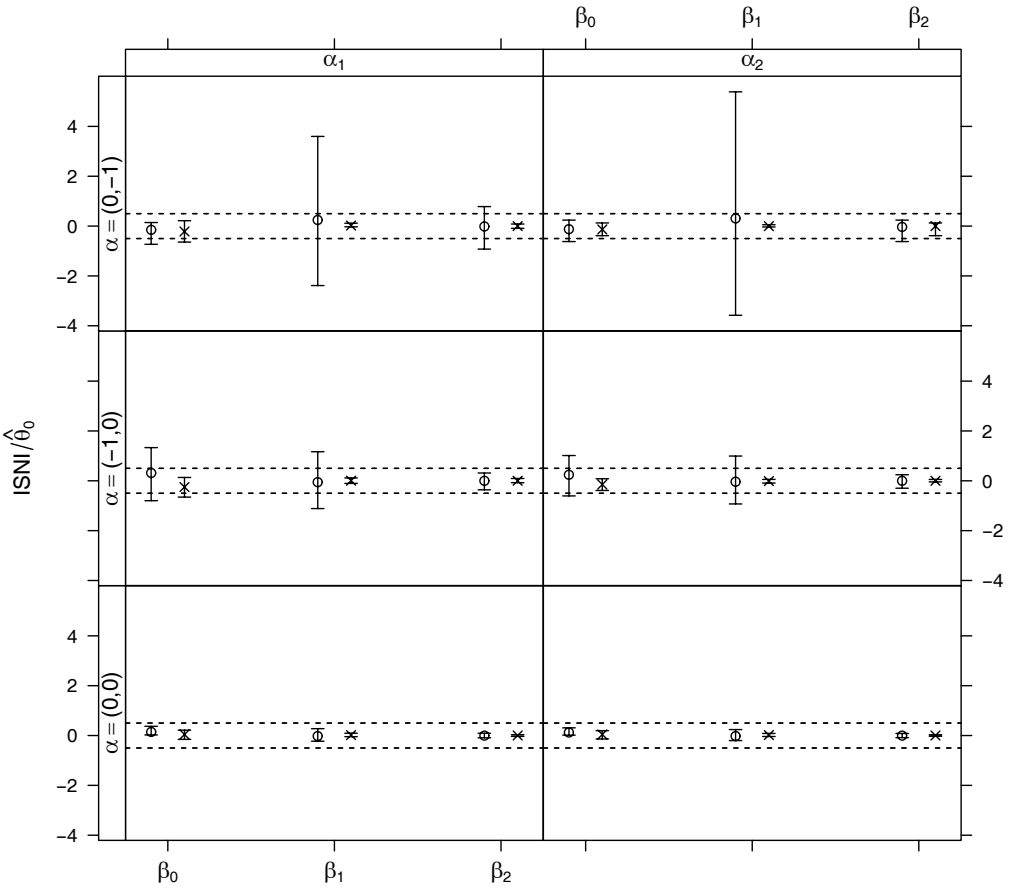

Figure 4.4: Simulation study, ISNI/$\widehat{\theta}_0$ for shared parameter model on 1000 samples, covariance matrix D_1. The points denotes the median and the edges the 2.5% and 97.5% percentiles of the index for $\overline{n}_i = 3$ (circle) and $\overline{n}_i = 7$ (cross). The dashed lines corresponds to $y = \pm 0.5$.

from MAR is not too large, and a decreasing effect of the assumptions of ignorability for the missing data mechanism when the length of the individual sequences increases. The random effect covariance structure does not seem to play a substantial role. On the other hand, the standard shared parameter model has performed well with respect to sensitivity, but it has experienced a slight dependence on the random effects covariance structure. In addition, the sensitivity with respect to model misspecification seems to be larger when the true model is the joint model, compared to the case when the true model is the standard shared parameter model.

As a further contribution, we have proposed a relative index of local sensitivity, given by the ratio of the ISNI values to the corresponding parameter estimates under the MAR model, that seems to lead to a clearer interpretation of parameters sensitivity, while the classical ISNI and ISNI/se($\widehat{\theta}_0$) may be misleading, at least in some cases when a shared

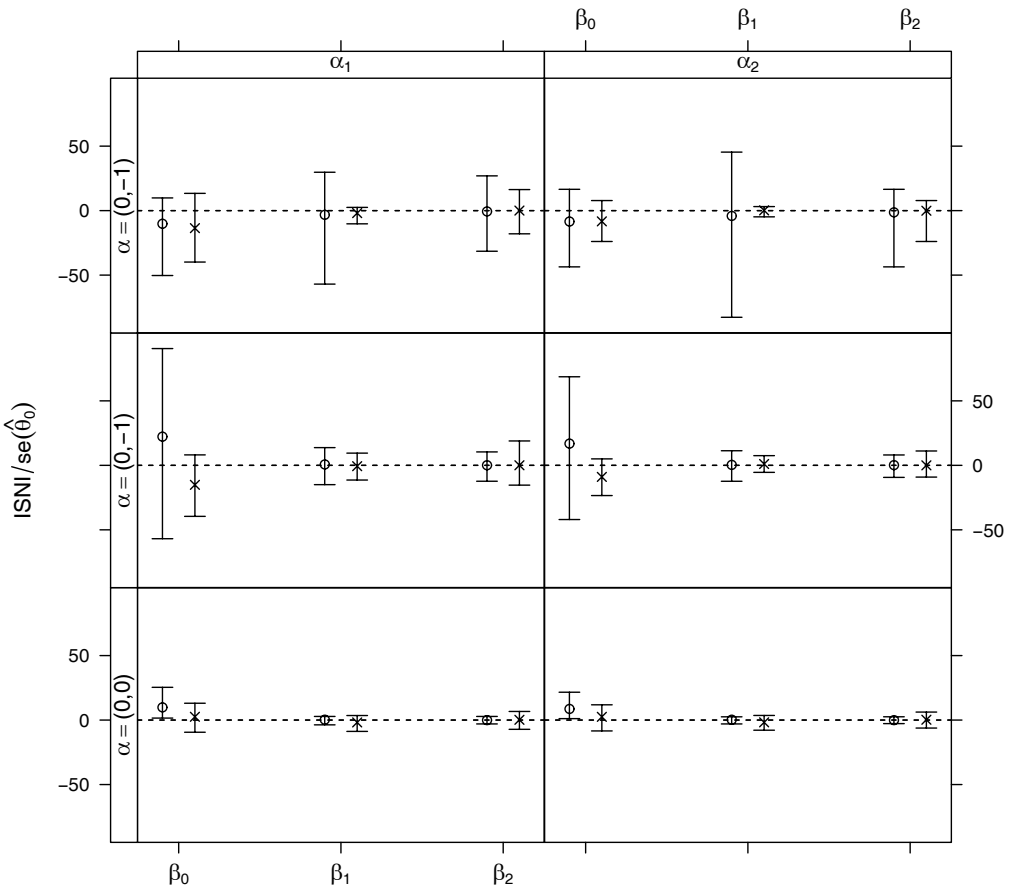

Figure 4.5: Simulation study, ISNI/se($\widehat{\theta_0}$) for shared parameter model on 1000 samples, covariance matrix D_1.The points denotes the median and the edges the 2.5% and 97.5% percentiles of the index for $\overline{n}_i = 3$ (circle) and $\overline{n}_i = 7$ (cross). The dashed lines corresponds to $y = 0$.

parameter model is adopted. By using an approximation developed by looking at the ISNI definition, we have provided an estimate of the sampling variability of the index when the MAR hypothesis is true, which could help define a further relative measure of parameter sensitivity to departures from the MAR assumption. This approach is based on the empirical evidence that, also when the MAR hypothesis is true, the ISNI may take non null values; therefore, this absolute measure of displacement should be compared to a measure of its sampling variability, which may be linked to parameter non-separability. In conclusion, as illustrated in the three benchmark data examples, the sensitivity analysis based on the ISNI relative formulations can be helpful to avoid inefficient uses of shared parameter models.

	True	Mean	Median	se	2.5% Quantile	75% Quantile
			$\alpha = 0$			
Intercept	2.512	2.511	2.483	0.036	2.484	2.535
Time	-0.375	-0.375	-0.376	0.009	-0.382	-0.369
Time * Treatment	0.821	0.814	0.822	0.009	0.815	0.827
			$\alpha = -0.5$			
Intercept	2.512	2.489	2.487	0.042	2.463	2.519
Time	-0.375	-0.380	-0.380	0.010	-0.387	-0.374
Time * Treatment	0.821	0.799	0.799	0.017	0.788	0.813
			$\alpha = -1$			
Intercept	2.512	2.493	2.493	0.052	2.462	2.525
Time	-0.375	-0.381	-0.380	0.008	-0.388	-0.372
Time * Treatment	0.821	0.793	0.793	0.016	0.778	0.807
			$\alpha = -1.5$			
Intercept	2.512	2.492	2.491	0.048	2.462	2.524
Time	-0.375	-0.381	-0.380	0.012	-0.389	-0.374
Time * Treatment	0.821	0.783	0.783	0.021	0.767	0.800
			$\alpha = 0.5$			
Intercept	2.512	2.571	2.487	0.042	2.542	2.600
Time	-0.375	-0.367	-0.366	0.021	-0.379	-0.354
Time * Treatment	0.821	0.782	0.799	0.029	0.762	0.802
			$\alpha = 1$			
Intercept	2.512	2.612	2.614	0.044	2.584	2.641
Time	-0.375	-0.402	-0.403	0.038	-0.423	-0.377
Time * Treatment	0.821	0.707	0.702	0.065	0.662	0.754
			$\alpha = 1.5$			
Intercept	2.512	2.629	2.630	0.043	2.602	2.658
Time	-0.375	-0.508	-0.509	0.097	-0.571	-0.454
Time * Treatment	0.821	0.618	0.621	0.160	0.532	0.703

Table 4.9: Simulation study: descriptive statistics for the MAR parameter estimates

	β true	β MAR	se(β)	ISNI	se(ISNI)	$min(\|ISNI\|)$	$max(\|ISNI\|)$
$\alpha = 0$, $\omega = 0.56$							
Intercept	2.510	2.474	0.035	-0.027	0.016	7.7e-04	0.085
Time	-0.370	-0.384	0.026	0.027	0.015	1.4e-04	0.052
Time*Group	0.820	0.850	0.022	0.032	0.012	1.7e-05	0.041
$\alpha = -0.5$, $\omega = 0.50$							
Intercept	2.510	2.425	0.042	0.828	0.130	2.0e-05	0.582
Time	-0.370	-0.534	0.031	0.439	0.125	1.4e-03	0.560
Time*Group	0.820	0.628	0.035	-1.433	0.224	4.9e-04	1.199
$\alpha = -1$, $\omega = 0.50$							
Intercept	2.510	2.284	0.043	1.409	0.141	8.8e-05	0.608
Time	-0.370	-0.605	0.034	0.380	0.152	3.7e-04	0.695
Time*Group	0.820	0.607	0.039	-0.998	0.261	1.4e-03	0.975
$\alpha = -1.5$, $\omega = 0.48$							
Intercept	2.510	2.423	0.044	-0.353	0.187	6.8e-03	0.809
Time	-0.370	-0.585	0.027	-0.404	0.213	0.001	1.114
Time*Group	0.820	0.519	0.034	-0.198	0.335	3.2e-04	1.776
$\alpha = 0.5$, $\omega = 0.257$							
Intercept	2.510	2.551	0.045	-0.003	0.023	5.6e-06	0.118
Time	-0.370	-0.378	0.023	-0.02	0.051	2.5e-05	0.242
Time*Group	0.820	0.778	0.031	0.05	0.040	2.2e-04	0.183
$\alpha = 1$, $\omega = 0.124$							
Intercept	2.510	2.605	0.045	-0.015	0.033	5.4e-05	0.134
Time	-0.370	-0.398	0.032	0.077	0.094	4.6e-06	0.339
Time*Group	0.820	0.745	0.056	0.131	0.129	5.7e-05	0.567
$\alpha = 1.5$, $\omega = 0.100$							
Intercept	2.510	2.645	0.045	-0.061	0.043	5.3e-07	0.157
Time	-0.370	-0.483	0.032	-0.002	0.173	3.5e-07	0.889
Time*Group	0.820	0.802	0.056	0.274	0.307	1.2e-07	1.482

Table 4.10: Simulations results corresponding to JM for each value of α, covariance matrix D_2 and proportion of dropout ω.

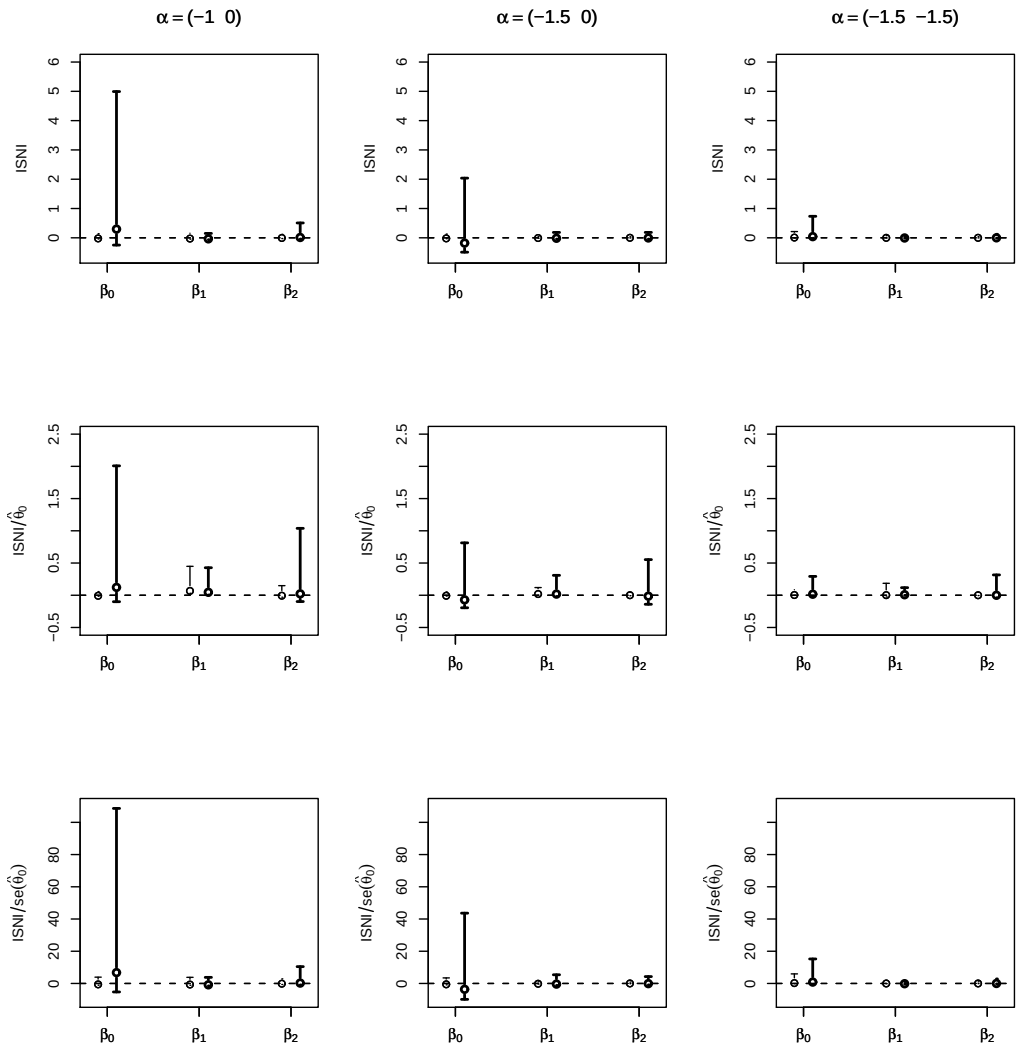

Figure 4.6: Simulation study, results for model misspecification on 1000 datasets and covariance matrix D_1. The circle denotes the median and the edges the 2.5% and 97.5% percentiles of absolute and relative ISNIs with respect to α_1 when the true model is the SPM and the ISNI is computed for the JM (thin line) and vice versa (bold line). The dashed line corresponds to $y = 0$.

Chapter 5

Joint modeling for discrete longitudinal responses and time to drop-out

Several studies in different disciplines collect longitudinal non-Gaussian data, such as binary or counted outcomes; an interesting review in this context is given by Molenberghs and Verbeke (2005). As in studies with Gaussian responses, it is common for some subjects to drop-out prematurely; the occurrence of such drop-outs leads to missing data and poses an additional challenge to draw correct statistical inference. One appealing approach to treat this statistical issue is by extending the joint models described in Chapter 4 to non-Gaussian responses. In this context, although some proposals have been introduced, see for instance Rizopoulos and Ghosh (2011) in the Bayesian framework, very limited statistical software is available. In this Chapter, we illustrate one proposal of Joint Model for non-Gaussian data and show how parameter estimation can be performed with ad hoc *R* code.
The Chapter is organized as follows. Section 5.1 illustrates the Bayesian approach to multivariate Joint modelling proposed by Rizopoulos and Ghosh (2011), while Section 5.2 describes the proposed Generalized Linear Mixed Joint Model from an analytical point of view. The computational issues are dealt in Section 5.2.1. Sections 5.2.2 and 5.2.3 introduces the GLMJM in the cases of Poisson and Binomial longitudinal outcomes. To investigate the behaviour of parameter estimates a simulation study is adopted in Section 5.3, while applications to benchmark datasets are available in Sections 5.5 and 5.4. Finally, Section 5.6 gives some concluding remarks and possible future developments.

5.1 The Bayesian multivariate joint model

When a discrete longitudinal outcome is recorded together with a survival time describing the participation to the study, joint modelling could represent properly the

association between the two processes. However, not many proposals can be found in literature in this field. A Bayesian approach has been proposed by Rizopoulos and Ghosh (2011). In their paper, the authors postulate a joint model for a multivariate longitudinal outcome (with both discrete and continuous) and a time-to-event. By indicating with $\mathbf{Y}_i = (\mathbf{y}_{i1}^{\mathsf{T}}, \ldots, \mathbf{y}_{iK}^{\mathsf{T}})$ the K-variate response matrix for the ith subject, $i = 1, \ldots, n$, they allowed each longitudinal response to be recorded at different time points $t_{ij,k}$. The longitudinal response is assumed to follow a generalized linear mixed effects model, where the conditional distribution of $\mathbf{Y}_{ik}$ given a vector of random effects b_{ik} is assumed to be a member of the exponential family, with linear predictor $f_{ik} = g_k[\mathbb{E}(y_{ik} \mid b_{ik})]$, where $g_k(\cdot)$ is a known one-to-one link function, $y_{ik}(t)$ is the value of the longitudinal outcome for the ith subject at time t, and f_{ik} is an unknown function which is assumed to describe the true, possibly non linear, longitudinal profile for the kth outcome. This last function is approximated using a spline-based approach. Let $\lambda_k = \{\lambda_{lk}; l = 1, \ldots, L_k\}$ represent an increasing sequence of knot positions, then

$$f_{ik} = B_{ik}(\beta_k^{(1)}, b_{ik}^{(1)}) + H_{ik}(t; \beta_k^{(2)}, b_{ik}^{(2)}, \lambda_k).$$

Hence, f_{ik} is approximated by the sum of two parts: a time independent part, $B_{ik}(\cdot)$, which contains a set of baseline covariates with corresponding vector of fixed effects $\beta_k^{(1)}$ and random effects $b_{ik}^{(1)}$; and a time dependent part, $H_{ik}(t)$, approximated by a natural cubic spline function with knots at λ_{lk}, while fixed and random coefficients $\beta_k^{(2)}$ and $b_{ik}^{(2)}$ are used to include possible interactions of baseline covariates with time-dependent terms.

The interaction between the longitudinal outcome and the survival time is captured via a relative risk model of the form

$$h_i(t \mid \mathcal{F}_i^H(t), \mathbf{w}_i) = h_0(t) \exp\left\{ \mathbf{w}_i \gamma^{\mathsf{T}} + \sum_{k=1}^{K} m_{ik}\left\{ f_{ik}(t), \alpha_k \right\} \right\},$$

where $\mathcal{F}_i^H(t) = \{f_{ik}(s), 0 \leq \ s \leq t, 1 \leq k \leq K\}$ is the history of the true and unobserved longitudinal process up to time t, $\mathbf{w}_i$ denotes the vector of baseline covariates with regression coefficients γ and $m_{ik}(\cdot)$ specifies which components of the longitudinal process for the k-th outcome is related to the survival time, and is assumed to follow different parameterizations. Moreover, α_k represents the effect of the longitudinal outcome on the risk function.

The authors adopt a Bayesian formulation for the proposed semiparametric multivariate joint model, since the random effect dimensions is large and a classical maximum likelihood approach could lead to cumbersome expressions. The posterior distribution of the parameters, conditional on the observed data is derived using an MCMC algorithm which can be written as follows:

$$p(\theta, \mathbf{b}_i \mid \mathbf{y}_i, T_i, \delta_i) \propto \left[\prod_{k=1}^{K} \prod_{j=1}^{n_{ik}} p(y_{ij,k} \mid b_{ik} : \theta_y) \right] p(T_i, \delta_i \mid \mathbf{m}_i(\cdot); \theta_t) p(\mathbf{b}_i; \theta_b) p(\theta_y, \theta_t, \theta_b).$$

This approach is particularly useful when the longitudinal response is high-dimensioned and has seen to lead to interesting results both in simulation and application studies.

5.2 The generalized linear mixed joint model

In this Section, our aim is at proposing an approach called generalized linear mixed model, which represents an extension of the linear joint model of Wulfsohn and Tsiatis (1997) to non-Gaussian responses. This approach is simple and takes into account one longitudinal response which is joint modelled with a survival time, but in a likelihood based context. In the proposed parametrization, the *mean* value of the longitudinal outcome at time t is assumed to influence the survival process. Let $Y \sim EF(\theta_i(t))$, i.e. let Y be a random variable with distribution belonging to the exponential family; then, the following generalized linear mixed joint model (GLMJM) may be defined:

$$\begin{cases} g(m_i(t)) = \beta^{\mathsf{T}}\mathbf{X}_i(t) + \mathbf{b}_i^{\mathsf{T}}\mathbf{Z}_i(t) \\ h_i(t \mid M_i(t), \mathbf{W}_i) = h_0(t)\exp\{\gamma^{\mathsf{T}}\mathbf{W}_i + \alpha m_i(t)\}, \end{cases} \tag{5.1}$$

where $m_i(t) = g^{-1}(\beta^{\mathsf{T}}\mathbf{X}_i(t) + b_i^{\mathsf{T}}\mathbf{Z}_i(t) = m(\theta_i(t))$ and $\theta_i(t)$ is the canonical parameter of the distribution and $g(\cdot)$ is a given link function. For canonic links, $g(m_i(t)) = \theta_i(t)$, where $\theta_i(t)$. In general, model (5.1) can be used when one would study how the expected value of the longitudinal process influences the risk of the drop-out event. Following the theory described in Chapter 4, the log-likelihood for model (5.1) is given by

$$\ell(\theta) = \ell(\theta \mid T_i, \delta_i, \mathbf{y}_i) = \sum_i \log \int_{\mathbf{b}_i} p(T_i, \delta_i \mid \mathbf{b}_i; \theta) p(\mathbf{y}_i \mid \mathbf{b}_i; \theta) p(\mathbf{b}_i; \theta) d\mathbf{b}_i, \tag{5.2}$$

where

$$p(\mathbf{y}_i \mid \mathbf{b}_i; \theta) = \exp\left\{B(\theta)\mathbf{y}_i - A(\theta) + C(\mathbf{y}_i)\right\},$$

i.e. the density belongs to the exponential distribution family, and the time-to-event process is defined as

$$p(T_i, \delta_i \mid \mathbf{b}_i; \theta) = h(T_i \mid M_i(t), \mathbf{W}_i; \theta)^{\delta_i} S(T_i \mid M_i(t), \mathbf{W}_i; \theta) \tag{5.3}$$

where the random effects $\mathbf{b}_i$ are assumed to follow a Gaussian distribution, i.e. they account for dependence among the repeated measurements over time corresponding to the same individual, see Molemberghs *et al.* (2011).
The corresponding score vector can be written as follows:

$$s(\theta) = \sum_i \int q(\theta, \mathbf{b}_i) p(\mathbf{b}_i \mid T_i, \delta_i, y_i; \theta) d\mathbf{b}_i, \tag{5.4}$$

where

$$q(\theta, b_i) = \frac{\partial}{\partial \theta^{\mathsf{T}}} \left\{\log p(T_i, \delta_i \mid b_i; \theta) + \log p(y_i \mid b_i; \theta) + \log p(b_i; \theta)\right\}.$$

5.2.1 The EM algorithm

In this section we focus on the estimation of $\theta = (\theta_y, \theta_t, \theta_b)$, i.e. the parameter for the longitudinal, the survival and the latent processes, respectively. The maximum likelihood estimates in the joint modeling are typically obtained using standard maximization algorithms such as the EM and the Newton-Rapson. The key component to apply these two algorithms is the score vector in (5.4). It can be noted that the observed data score vector is expressed as the expected value of the complete data score vector with respect to the posterior distribution of the random effects. From a computational point of view, this implies that when the score equations are solved with respect to θ with $p(\mathbf{b}_i \mid T_i, \delta_i, y_i; \theta)$ calculated on the basis of the value of θ derived at the previous iteration, leading to the EM algorithm.

More specifically, the algorithm can be summarized as follows.

for $i \in 1 : iter.EM$ **do**

E-STEP: Compute the posterior random effects distribution $p(\mathbf{b}_i \mid T_i, \delta_i, y_i; \theta)$, through the conditional distributions $p(T_i, \delta_i \mid \mathbf{b}_i; \theta)$, $p(T_i, \delta_i \mid \mathbf{b}_i; \theta)$ and $p(\mathbf{b}_i; \theta)$.

M-STEP: Compute the maximum likelihood estimates for the random effect covariance matrix D, the longitudinal parameters β and the survival parameters θ_t as follows:

$$\widehat{D} = n^{-1} \sum_i Cov(b_{i1}, b_{i2} \mid T_i, \delta_i, \theta)$$

$$\widehat{\beta}_{i+1} = \widehat{\beta}_i - \left\{ \frac{\partial}{\partial \beta^\intercal} S(\widehat{\beta}_i) \right\}^{-1} S(\widehat{\beta}_i)$$

$$\widehat{\theta}_{t,i+1} = \widehat{\theta}_{i,t} - \left\{ \frac{\partial}{\partial \theta_t^\intercal} S(\widehat{\theta}_{t,i}) \right\}^{-1} S(\widehat{\theta}_{t,i})$$

if convergence **then**

break.

end if

end for

while !convergence **do**

QUASI-NEWTON STEP $\widehat{\theta} = \arg\max_\theta \ell(\theta)$

(`optim` function in R)

end while

The joint model can be specialized by specifying a particular member of the exponential family.

5.2.2 The Poisson case

Let us assume $y_i(t)$ values have been recorded for subjects $i = 1, \ldots, n$ at time $t = 1, \ldots, T$; if $y_i(t)$ are the observed values of a Poisson random variable, the density

function for the response is

$$p(y_i(t) \mid b_i) = \frac{\exp(-\lambda_i)\lambda_i^{y_i(t)}}{y_i(t)!},$$

where $\lambda_i = \exp(\beta^\mathsf{T}\mathbf{X}_i(t) + \mathbf{b}_i^\mathsf{T}\mathbf{Z}_i(t)) = \mathbb{E}(Y_i(t)) = \mathbb{V}(Y_i(t))$. Hence,

$$\log p(y_i(t) \mid \mathbf{b}_i; \theta) = -\exp(\beta^\mathsf{T}\mathbf{X}_i(t) + \mathbf{b}_i^\mathsf{T}\mathbf{Z}_i(t)) + y_i(t)(\beta^\mathsf{T}\mathbf{X}_i(t) + \mathbf{b}_i^\mathsf{T}\mathbf{Z}_i(t)) - \log[y_i(t)!].$$

The survival model, assuming a Weibull distribution is given by (5.3), where the hazard function is

$$h_i(t \mid M_i(t), \mathbf{W}_i; \theta) = \xi t^{\xi-1} \exp\left\{\gamma^\mathsf{T}\mathbf{W}_i + \alpha \exp(\beta^\mathsf{T}\mathbf{X}_i(t) + \mathbf{b}_i^\mathsf{T}\mathbf{Z}_i(t))\right\} \tag{5.5}$$

and the survival function is

$$S(t \mid M_i(t), \mathbf{W}_i; \theta) = \exp\left\{-\int_0^t h_i(s)ds\right\}.$$

The score vector for the fixed effects in the longitudinal models can be written as

$$\begin{aligned} s(\beta) &= \sum_{i=1}^{n} \int_{\mathbf{b}_i} -\mathbf{x}_i \exp\left\{\beta^\mathsf{T}\mathbf{x} + \mathbf{b}_i^\mathsf{T}\mathbf{z}_i)\right\} + \mathbf{x}_i\mathbf{y}_i - \int_0^t h_0(s)\alpha\mathbf{x}_i \exp\left\{\beta^\mathsf{T}\mathbf{x}_i + \mathbf{b}_i^\mathsf{T}\mathbf{z}_i\right\} \\ &\quad \exp\left\{\gamma^\mathsf{T}\mathbf{w}_i + \alpha\exp(\beta^\mathsf{T}\mathbf{x}_i + \mathbf{b}_i^\mathsf{T}\mathbf{z}_i)\right\} \mathrm{d}s\mathrm{d}\mathbf{b}_i, \end{aligned}$$

while the Hessian matrix assumes the form

$$\begin{aligned} \frac{\partial s(\beta)}{\partial \beta} &= \sum_{i=1}^{n} \int_{\mathbf{b}_i} -\mathbf{x}_i^\mathsf{T}\mathbf{x}_i \exp\left\{\beta^\mathsf{T}\mathbf{x}_i + \mathbf{b}_i^\mathsf{T}\mathbf{z}_i\right\} + \delta_i\alpha\mathbf{x}_i^\mathsf{T}\mathbf{x}_i \exp\left\{\beta^\mathsf{T}\mathbf{x}_i + \mathbf{b}_i^\mathsf{T}\mathbf{z}_i\right\} \\ &- \int_0^t h_0(s)\alpha\mathbf{x}_i \exp\left\{\beta^\mathsf{T}\mathbf{x}_i + \mathbf{b}_i^\mathsf{T}\mathbf{z}_i\right\} \exp\left\{\gamma^\mathsf{T}\mathbf{w}_i + \alpha\exp(\beta^\mathsf{T}\mathbf{x}_i + \mathbf{b}_i^\mathsf{T}\mathbf{z}_i\right\} \mathrm{d}s \\ &- \int_0^t h_0(s)\alpha^2\mathbf{x}_i \exp\left\{\beta^\mathsf{T}\mathbf{x}_i + \mathbf{b}_i^\mathsf{T}\mathbf{z}_i\right\} \mathbf{x}_i^\mathsf{T} \exp\left\{\beta^\mathsf{T}\mathbf{x}_i + \mathbf{b}_i^\mathsf{T}\mathbf{z}_i\right\}^\mathsf{T} \\ &\quad \exp\left\{\gamma^\mathsf{T}\mathbf{w}_i + \alpha\exp(\beta^\mathsf{T}\mathbf{x}_i + \mathbf{b}_i^\mathsf{T}\mathbf{z}_i)\right\} \mathrm{d}s\mathrm{d}\mathbf{b}_i. \end{aligned}$$

5.2.3 The Binomial case

When the primary outcome is distributed as a binomial random variable of size n, $y_i(y)$ successes and probability of success $p_i(t)$, the corresponding density function on logaritmic scale is

$$p(y_i(t) \mid \mathbf{b}_i) = \log\binom{n}{y_i(t)} + y_i(t)\log p_i(t) + (n - y_i(t))\log(1 - p_i(t)),$$

where $p_i(t) = \dfrac{\exp(\beta^\mathsf{T}\mathbf{X}_i(t) + \mathbf{b}_i^\mathsf{T}\mathbf{Z}_i(t))}{1 + \exp(\beta^\mathsf{T}\mathbf{X}_i(t) + \mathbf{b}_i^\mathsf{T}\mathbf{Z}_i(t))}$. The survival time follows a Weibull distributed hazard of the following form:

$$h_i(T_i \mid \mathbf{b}_i) = h_0(t)\exp\left\{\gamma^\mathsf{T}\mathbf{W}_i + \alpha\frac{\exp(\beta^\mathsf{T}\mathbf{X}_i(t) + \mathbf{b}_i^\mathsf{T}\mathbf{Z}_i(t))}{1 + \exp(\beta^\mathsf{T}\mathbf{X}_i(t) + \mathbf{b}_i^\mathsf{T}\mathbf{Z}_i(t))}\right\}$$

is the hazard function, and

$$S(T_i \mid \mathbf{b}_i) = \exp\left\{-\int_0^{T_i} h_0(s)\exp\left\{\gamma^\mathsf{T}\mathbf{W}_i + \alpha\frac{\exp(\beta^\mathsf{T}\mathbf{X}_i(t) + \mathbf{b}_i^\mathsf{T}\mathbf{Z}_i(t))}{1+\exp(\beta^\mathsf{T}\mathbf{X}_i(t) + \mathbf{b}_i^\mathsf{T}\mathbf{Z}_i(t))}\right\}\right\}$$

is the survival function.
The score vector and the Hessian matrix are given by, respectively:

$$\begin{aligned} s(\beta) &= y_i(t)\mathbf{x}_i - n\mathbf{x}_i^\mathsf{T}\frac{\exp\left\{\beta^\mathsf{T}\mathbf{x}_i + \mathbf{b}_i^\mathsf{T}\mathbf{z}_i\right\}}{1+\exp\left\{\beta^\mathsf{T}\mathbf{x}_i + \mathbf{b}_i^\mathsf{T}\mathbf{z}_i\right\}} + \delta_i\alpha n\mathbf{x}_i^\mathsf{T}\frac{\exp\left\{\beta^\mathsf{T}\mathbf{x}_i + \mathbf{b}_i^\mathsf{T}\mathbf{z}_i\right\}}{\left[1+\exp\left\{\beta^\mathsf{T}\mathbf{x}_i + \mathbf{b}_i^\mathsf{T}\mathbf{z}_i\right\}\right]^2} \\ &- \int_0^t h_0(s)\mathbf{x}_i^\mathsf{T}\frac{\exp\left\{\beta^\mathsf{T}\mathbf{x}_i + \mathbf{b}_i^\mathsf{T}\mathbf{z}_i\right\}}{\left[1+\exp\left\{\beta^\mathsf{T}\mathbf{x}_i + \mathbf{b}_i^\mathsf{T}\mathbf{z}_i\right\}\right]^2}\alpha n\exp\left\{\gamma^\mathsf{T}\mathbf{w}_i + \frac{\exp\left\{\beta^\mathsf{T}\mathbf{x}_i + \mathbf{b}_i^\mathsf{T}\mathbf{z}_i\right\}}{1+\exp\left\{\beta^\mathsf{T}\mathbf{x}_i + \mathbf{b}_i^\mathsf{T}\mathbf{z}_i\right\}}\right\}\mathrm{d}s\mathrm{d}\mathbf{b}_i, \end{aligned}$$

and

$$\begin{aligned} \frac{\partial s(\beta)}{\partial\beta} &= -n\mathbf{x}_i^\mathsf{T}\mathbf{x}_i\mathbf{B} + \delta_i\alpha n\mathbf{x}_i\mathbf{B} \\ &- \int_0^t h_0(s)\mathbf{x}_i\mathbf{A}\alpha^2 n\exp\left\{\gamma^\mathsf{T}\mathbf{w}_i + \frac{\exp\left\{\beta^\mathsf{T}\mathbf{x}_i + \mathbf{b}_i^\mathsf{T}\mathbf{z}_i\right\}}{1+\exp\left\{\beta^\mathsf{T}\mathbf{x}_i + \mathbf{b}_i^\mathsf{T}\mathbf{z}_i\right\}}\right\}\mathbf{x}_i^\mathsf{T}\mathbf{B}\mathrm{d}s\mathrm{d}\mathbf{b}_i, \end{aligned}$$

where

$$\mathbf{A} = \frac{\left[2\mathbf{x}_i\exp\left\{\beta^\mathsf{T}\mathbf{x}_i + \mathbf{b}_i^\mathsf{T}\mathbf{z}_i\right\}\right]\left[1 - 2\mathbf{x}_i\exp\left\{\beta^\mathsf{T}\mathbf{x}_i + \mathbf{b}_i^\mathsf{T}\mathbf{z}_i\right\} - 1/2\mathbf{x}_i\exp\left\{\beta^\mathsf{T}\mathbf{x}_i + \mathbf{b}_i^\mathsf{T}\mathbf{z}_i\right\}^2\right]}{\left[1+\exp\left\{\beta^\mathsf{T}\mathbf{x}_i + \mathbf{b}_i^\mathsf{T}\mathbf{z}_i\right\}\right]^4},$$

and

$$\mathbf{B} = \frac{\exp\left\{\beta^\mathsf{T}\mathbf{x}_i + \mathbf{b}_i^\mathsf{T}\mathbf{z}_i\right\}}{\left[1+\exp\left\{\beta^\mathsf{T}\mathbf{x}_i + \mathbf{b}_i^\mathsf{T}\mathbf{z}_i\right\}\right]^2}.$$

5.3 Simulation study

To study the behaviour of parameter estimates for model (5.1) we have drawn a simulation study. In Section 5.3.1 we describe the simulation settings while in Section 5.3.2 we illustrate the main results.

5.3.1 Simulation design

For the Poisson case, we simulate $N = 100$ samples of the longitudinal response through the canonical model function:

$$\log(m_i(t)) = (\beta_0 + b_{i0}) + (\beta_1 + b_{i1})x_{i1}(t) + \beta_2 x_{i2}(t), \tag{5.6}$$

where $\beta_0 = 0.65$, $\beta_1 = 0.04$, $\beta_2 = -0.69$. Moreover, x_{i1} is a sequence from 0 to $t.max_i$, where $t.max_i$ is the maximum follow up time for subject i and x_{i2} is a binary random variable with parameter $p = 0.5$, representing a treatment variable. The random effects

are drawn form a multivariate Gaussian random variable with triangular covariance matrix $D = (0.640, 0.006, 0.005)$.
The simulated survival times are drawn from the hazard function:

$$h_i(t) = \xi t^{\xi-1} \exp\left\{\gamma_0 + \gamma_1 x_{i2} + \alpha m_i(t)\right\},$$

where x_{i2} contains the baseline values of $x_{i2}(t)$ in (5.6), $\xi = 1.8$, $\gamma_0 = -3.2$ and $\gamma_1 = -0.5$.

In the case of a binomial longitudinal outcome, we simulate $N = 100$ samples of 550 units for a binomial response with size $n = 11$ with the same covariates as in the Poisson simulation design, and the expected value of the longitudinal process is given by

$$m_i(t) = \frac{\exp\left\{\theta_i(t)\right\}}{1 + \exp\left\{\theta_i(t)\right\}}, \tag{5.7}$$

and

$$h(T_i \mid \mathbf{b}_i) = h_0(t) \exp \gamma^{\mathsf{T}} \mathbf{W}_i + \alpha m_i(t).$$

5.3.2 Simulation results

The mean of the parameter estimates for the Poisson joint model and the corresponding MAR estimates are shown in Table 5.1.

	β_{TRUE}	$\widehat{\beta}_0$	$\widehat{\beta}$	SE	$Q_{0.25}$	$Q_{0.75}$
$\alpha = 0$, $\widehat{\alpha} = -0.036$ and $\omega = 0.985$						
Intercept	0.65	0.655	0.655	0.076	0.601	0.713
Time	0.04	0.041	0.042	0.012	0.034	0.052
Group	-0.69	-0.698	-0.708	0.123	-0.789	-0.645
$\alpha = 1.2$, $\widehat{\alpha} = 0.805$ and $\omega = 0.695$						
Intercept	0.65	0.419	0.580	0.391	0.423	0.583
Time	0.04	-0.018	0.031	0.106	0.001	0.032
Group	-0.69	-0.468	-0.463	0.307	-0.561	-0.351
$\alpha = -1.2$, $\widehat{\alpha} = -0.911$ and $\omega = 0.987$						
Intercept	0.65	0.378	0.475	0.108	0.315	0.535
Time	0.04	0.059	0.032	0.032	0.026	0.055
Group	-0.69	-0.429	-0.507	0.106	-0.589	-0.448

Table 5.1: Mean and quantiles of the longitudinal parameter estimates for the Poisson joint model and the corresponding MAR model for different values of α, the ML estimate $\widehat{\alpha}$ and proportion of dropout ω.

It can be noted that, as α is fixed to zero, i.e. in the ignorability case, the parameter estimates under MAR and MNAR models are close to each other and to the true

values. On the other hand, as α moves from zero, the MAR model becomes less precise. This phenomenon can be recognized also in the empirical distribution of the parameter estimates shown in Figure 5.1.

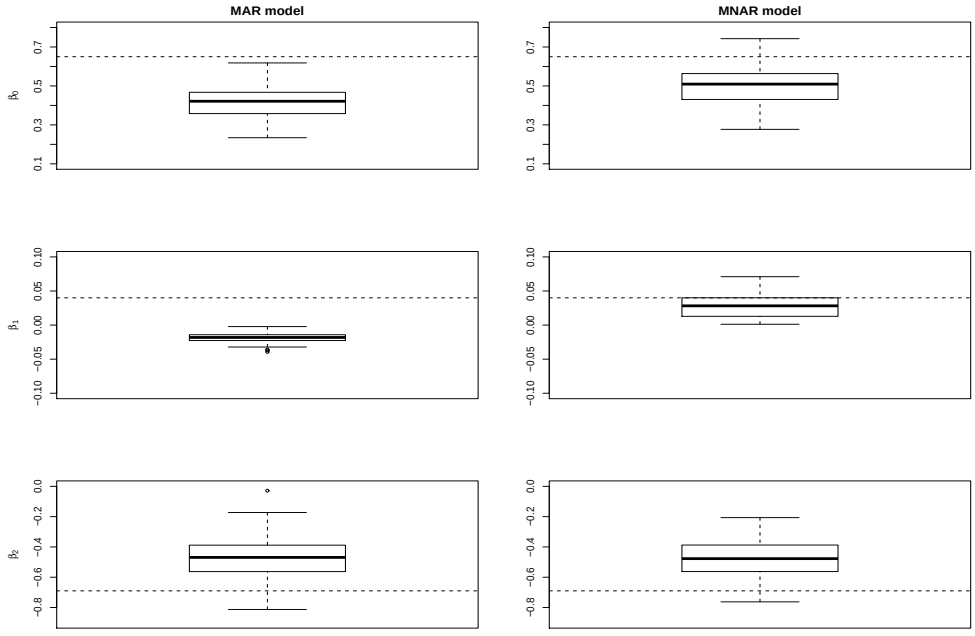

Figure 5.1: Simulation results: Poisson longitudinal response, empirical distribution for ML estimates under MAR (left side) and MNAR (right side) assumptions.

While the MAR model does not include the true parameter value in the interquartile range either for the intercept and the time effect, the joint model leads to more precise parameter estimates for these effects. Moreover, it can be noticed that the group effect (β_2) is estimated well both from the MAR and the MNAR model. This may happen because the variable is not time dependent.

Parameter estimates for the Binomial model corresponding to the MAR and the MNAR assumptions are shown in Table 5.2.

In this case, α is fixed to a value close to zero, but the empirical distribution of the parameter estimates is still more precise than the ones corresponding to MAR estimates, as it can be seen in Figure 5.2.

5.4 The AIDS data set

In this Section, we compare the parameter estimates for the standard Joint Model applied to the AIDS dataset (see Section 4.5), where the CD4 cell counts are assumed

	β_{TRUE}	$\widehat{\beta}_0$	$\widehat{\beta}$	SE	$Q_{0.25}$	$Q_{0.75}$
$\alpha = 0$, $\widehat{\alpha} = 0.001$ and $\omega = 0.678$						
Intercept	1.05	1.040	1.061	0.080	0.996	1.120
Time	-0.15	-0.156	-0.159	0.007	-0.166	-0.153
Group	-0.39	-0.381	-0.356	0.116	-0.431	-0.280
$\alpha = 0.25$, $\widehat{\alpha} = 0.6560$ and $\omega = 0.352$						
Intercept	1.05	1.051	1.084	0.132	0.993	1.174
Time	-0.15	-0.149	-0.164	0.077	-0.159	-0.148
Group	-0.39	-0.401	-0.411	0.159	-0.531	0.327
$\alpha = -0.25$, $\widehat{\alpha} = -0.519$ and $\omega = 0.132$						
Intercept	1.05	1.010	1.047	0.136	0.956	1.173
Time	0.15	0.142	0.143	0.203	0.137	0.149
Group	-0.39	-0.387	-0.409	0.009	-0.556	-0.228

Table 5.2: Mean and quantiles of the longitudinal parameter estimates for the Binomial joint model and the corresponding MAR model for different values of α, the ML estimate $\widehat{\alpha}$ and proportion of dropout ω.

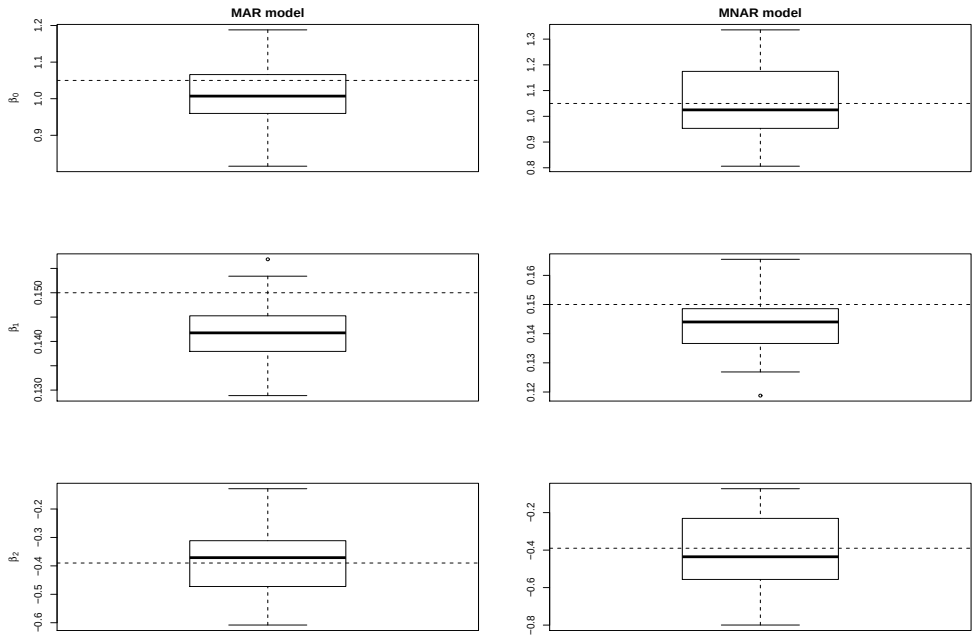

Figure 5.2: Simulation results: Binomial longitudinal response, empirical distribution for ML estimates under MAR (left side) and MNAR (right side) assumptions.

to come from a Gaussian distribution, to the Poisson Joint Model, where this variable is assumed to follow a Poisson distribution.

	$\widehat{\beta}_{\text{MAR}}$	$\widehat{\beta}_{\text{JM}}$	SE	z-value	p-value
PJM, $\widehat{\alpha} = -0.16$ and $\omega = 0.6$					
Intercept	1.747	1.766	0.034	52.216	< 0.001
Time	-0.031	-0.022	0.003	-6.620	< 0.001
Group	0.069	-0.014	0.044	-0.311	0.756
Time * Group	0.003	-0.010	0.005	-2.121	0.034
NJM, $\widehat{\alpha} = -0.297$ and $\omega = 0.6$					
Intercept	6.951	7.062	0.173	40.863	< 0.001
Time	-0.159	-0.179	0.022	-8.259	< 0.001
Group	0.482	0.299	0.269	1.108	0.268
Time * Group	0.021	0.004	0.124	1.151	0.901

Table 5.3: Longitudinal parameter estimates for the Poisson joint model (PJM) and the linear mixed model (NJM).

Both results suggest that dropout process is ignorable at least approximately, while the parameter estimates are on a different scale, see Table 5.3.

5.5 MMT Data

In this Section, we consider a dataset of $n = 136$ heroin users enrolled in a methadone maintenance treatment (MMT) program at a clinic in western Sidney in 1986; they have been observed once a week for 26 weeks, see Alfó and Aitkin (2000). At the end of the study, 51 events are oberved, resulting in 62.5% censoring. The response is the recorded test, which can be positive, $y_i(t) = 1$, or negative, $y_{it} = 0$, to morphine, the biological marker of heroin use. This data were previously analysed by Chan *et al.* (1998), with the aim at investigating the relationship between varying daily dose of methadone, duration of treatment and heroin use as detected by urine testing. They noted that beyond the first six months of treatment non-random drop-outs begin to appear, with patients who continued regular heroin use being more likely to leave the programme. This results have been confirmed by Alfó and Aitkin (2000) through the use of a first order autoregressive model with random effects.

Our approach is different, and consider the *whole* history of the response, $m_i(t)$, instead of only considering the response at time $t - 1$. In the following, we will apply the generalized linear mixed model to the MMT data and make a comparison with the estimates of the autoregressive model discussed in Alfó and Aitkin (2000).

In the Joint Model approach, we consider a random coefficient associated to the dose variable. Hence, the individual heterogeneity in the effect of the methadone dose is also

assumed to be shared by the process that generates the time-to-event. The following GLMJM holds:

$$\begin{cases} logit(m_i(t)) = (\beta_0 + b_{i0}) + (\beta_1 + b_{i1}) \log(dose_i(t)) + \beta_2 time_i(t) \\ h_i(t \mid M_i(t), \mathbf{W}_i) = \xi t_i^{\xi-1} \exp\{\gamma_1 dose_i + \alpha m_i(t)\}, \end{cases} \tag{5.8}$$

We compare the GLMJM parameter estimates to the first order autoregressive model in Alfó and Aitkin (2000), which does not account for non-ignorability of the drop-ou process. The results are shown in tables 5.4, together with longitudinal parameter estimates for the MAR model .

	$\widehat{\beta}_{\text{MAR}}$	$\widehat{\beta}_{\text{auto}}$	$\widehat{\beta}_{\text{JM}}$	p-value
Intercept	-0.111	-1.498	-0.1633	0.842
	(0.401)	*(0.335)*	*(0.818)*	
log(Time)	-0.512*	**-0.271***	-0.136	0.460
	(0.071)	*(0.069)*	*(0.184)*	
Dose	-0.018*	-0.028	-0.027	0.694
	(0.005)	*(0.023)*	*(0.069)*	
$y_i(t-1)$		**1.431***		
		(0.139)		

Table 5.4: MMT data: longitudinal parameter estimates for the Bernoulli joint model and the corresponding autoregressive and MAR model (standard errors in brackets. $\widehat{\alpha} = \mathbf{5.602}^*$. The symbol $*$ stand for significant coefficients.

The autoregressive and the joint models both lead to intuitive results. While the dose effect is not significant and close to zero, the time effect is significant for the autoregressive model and not significant for the joint model. The role of this effect, together with the one corresponding to the response measured at the previous time point $(t-1)$, is coherent with the non-ignorability parameter that is estimated as significantly different from zero. This in fact suggests that in the autoregressive model, where the drop-out is not considered, the time each patient spent in the follow-up is a relevant explanation variable, while when a non-ignorability parameter is considered, the time effect is included in the time-to-event process. Therefore, the time may be not significant in itself, but rather due to the non-ignorability of the drop-out process.

5.6 Discussion

In this Chapter, a Generalized Linear Mixed Model has been proposed to deal with discrete longitudinal outcomes. While Rizopoulos and Ghosh (2011) consider a multivariate response which can be either discrete or continuous, our approach is focused on the formulation of a model when interest lies in the joint modeling of one discrete

longitudinal outcome and a time-to-event. The main findings concern the behaviour of the GLMJM, that have been seen to be more reliable than the MAR model when the ignorability parameter is different from zero, while presenting a similar empirical distribution when $\alpha = 0$. Applications to benchmark datasets have suggested that the GLMJM lead to intuitive results as far as the assessment of the ignorability of the drop-out process is concerned. Further it highlighted the possible fields of application of the proposed model.

Nevertheless, some computational limitations have been encountered, mainly due to the fact that the expected value of the random variable is always positive and, in some cases, this could lead to an overly high risk of the event. Potential developments refer to consider the assumption that the *linear predictor* of the response influences the hazard function.

Chapter 6

Concluding remarks

In this thesis our aim was at describing a general approach to deal with longitudinal data in presence of drop-out, which represents a common statistical issue. The basic concept underlying the missing data models is that *drop-out* represents an information in itself, and that inferences that do not take this information into account results in less efficient (by not considering the process underlying the drop-out) and, sometimes, biased estimates. This may happen when the drop-out mechanism is *non-ignorable*, which means that it has an influence on (and thus modify) overall model inferences. Several model frameworks have been proposed to take into account the dependence between the longitudinal and the drop-out mechanisms. We focused on *shared parameter models*, introduced by Wu and Carrol (1988) and further developed by Follmann and Wu (1995), Wulfsohn and Tsiatis (1997), Henderson *et al.* (2000) and Rizopoulos *et al.* (2008b), just to mention a few. The key idea is that unobservable individual specific sources of heterogeneity, described by a set of a random coefficient, is shared by the longitudinal and the drop-out processes; here and conditionally on the random coefficients the two processes are independent. Thus, the ML estimation conditional on a set of values for the random effects can be based on standard (univariate) methods. This model structure presents some relevant advantages but also some disadvantages. First, non-ignorability is assessed by a single parameter, and this leads to a straightforward interpretation; on the other hand, the corresponding parameter estimates could be affected by the non-ignorability assumption. Second, sharing the same random coefficients results in a parsimonious model, but the assumption of a perfect correlation between the random effects in the two processes may be not always realistic. Third, in the joint model parametrization the effect on the hazard of the whole past history of the response is assumed to be summarized by the current expected value, thus the effect of the response at a given past time $t - j$ is not directly captured. Finally, while plenty of theory and software is available to deal with shared parameter models with continuous Gaussian longitudinal responses, not so much has been done for discrete outcomes.

In this manuscript, we have focused mainly on the first and the last issues, discussed in Chapter 4 and 5, respectively. The main results are reported above.

Chapter 4 dealt with the issue of defining an Index of Local Sensitivity to Non-Ignorability when a shared parameter model is considered. The index was based on the ISNI proposed by Troxel *et al.* (2004) and was proposed in the literature only for selection models; however, the simple dependence structure which is implied by using the SPMs and the availability of relevant computational routines for fitting such models pushed to the need for defining reliable sensitivity tools also in this context. The topic was covered by extending the ISNI to SPMs, by defining alternative relative versions of the index and by studying the corresponding empirical behaviour through simulation and real studies.

In Chapter 5, a Generalized Linear Mixed Model was proposed to deal with discrete longitudinal outcomes. While Rizopoulos and Ghosh (2011) considered multivariate responses which can be either discrete or continuous, our approach was focused on the formulation of a model where the main interest lies in the joint modeling of one discrete longitudinal outcome and a time-to-event. The main findings concerned the behaviour of the GLMJM, that was seen to be more precise than the corresponding MAR model in simulated samples when the ignorability parameter is assumed to be different from zero, while presented a similar empirical distribution when $\alpha = 0$. Applications to benchmark datasets suggested that the GLMJM leads to intuitive results as far as the assessment of the ignorability of the drop-out process was concerned and highlighted possible fields of application of the proposed model.
Nevertheless, some computational limitations were encountered, mainly due to the fact that the expected values of the response variable are always positive and, in some cases, this could lead to an overly high risk of the drop-out event. Potential developments are to consider that the *linear predictor* for the longitudinal response influences the hazard function, to consider more general failure time processes (e.g. piecewise constant baseline hazard models), and to take into account the past (observed) response history in the hazard specification.

Appendix A

Calculations for Chapter 4

A.1 Calculating the ISNI for the Shared Parameter Models

The score vectors with respect to the longitudinal fixed effects and the association parameter for the joint model assume the form

$$\begin{aligned} s^{\mathrm{JM}}(\beta) &= \sum_{i=1}^{n} \delta_i \alpha X_i^{\mathsf{T}}(T_i) - \int_{b_i} \int_0^{T_i} h_0(s_i) \alpha X_i^{\mathsf{T}}(s_i) \exp\left\{\gamma^{\mathsf{T}} W_i + \alpha m_i(s_i)\right\} ds_i \\ &\qquad + X_i^{\mathsf{T}} \sigma^{-2} \left(Y_i - m_i(t_i)\right) p(b_i \mid T_i, \delta_i, y_i; \theta) db_i, \end{aligned}$$

and

$$\begin{aligned} s^{\mathrm{JM}}(\alpha) &= \sum_{i=1}^{n} \delta_i m_i(t_i) - \int_{b_i} \int_0^{T_i} h_0(s_i) m_i(s_i) \exp\left\{\gamma^{\mathsf{T}} W_i + \alpha m_i(s_i)\right\} ds_i \\ &\qquad p(b_i \mid T_i, \delta_i, y_i; \theta) db_i. \end{aligned}$$

For the shared parameter model the score vectors with respect to β and α can be written as

$$s^{\mathrm{SPM}}(\beta) = \sum_{i=1}^{n} \int_{b_i} X_i^{\mathsf{T}} \sigma^{-2} \left(Y_i - m_i(t_i)\right) p(b_i \mid T_i, \delta_i, y_i; \theta) db_i,$$

and

$$s^{\mathrm{SPM}}(\alpha) = \sum_{i=1}^{n} \int_{b_i} \left(\delta_i b_i - \int_0^{T_i} h_0(s_i) b_i \exp\left\{\gamma^{\mathsf{T}} W_i + \alpha^{\mathsf{T}} b_i\right\} \right) p(b_i \mid T_i, \delta_i, y_i; \theta) db_i,$$

respectively. The posterior distribution of the random effects is given by

$$p(b_i \mid T_i, \delta_i, Y_i; \theta) = \frac{p(T_i, \delta_i \mid b_i; \theta) p(y_i \mid b_i; \theta) p(b_i)}{p(T_i, \delta_i, y_i)}.$$

We derive the ISNI for the joint model as follows:

$$\begin{aligned}\text{ISNI}_{\text{JM}}(\beta) &= -\left(\sigma^{-2}X_i^{\mathsf{T}}X_i - \sum_{i=1}^{n}\int_0^{T_i} h_0(s_i)\alpha^2 X_i(s_i)^{\mathsf{T}}X_i(s_i)\mu_{1,b_i}(s_i)ds_i + I_1\right)^{-1} \\ &\quad \times \sum_{i=1}^{n}\delta_i X_i(t_i) - \int_0^{T_i}\left[h_0(s_i)X_i(s_i)\mu_{1,b_i}(s_i)\right]\left[1+\alpha m_i(s_i)\right]ds_i + I_2,\end{aligned}$$

where $\mu_{1,b_i}(t_i) = \mathbb{E}_{b_i|T_i,\delta_i,y_i}\left[\exp\left\{\gamma^{\mathsf{T}}W_i + \alpha m_i(t_i)\right\}\right]$,

$$I_1 = \int_{b_i} q(\beta,b_i)\left\{q(\beta,b_i) - s_i^{\text{JM}}(\beta)\right\}^{\mathsf{T}} p(b_i \mid T_i,\delta_i,y_i),$$

and

$$\begin{aligned}q(\beta,b_i) &= \alpha\delta_i X_i - \int_0^{T_i} h_0(s_i)\alpha X_i(s_i)\exp\left\{\gamma^{\mathsf{T}}W_i + \alpha m_i(s_i)\right\}ds_i \\ &\quad + X_i^{\mathsf{T}}\sigma^{-2}(Y_i - m_i(t_i)),\end{aligned}$$

while $I_2 = \int_{b_i} q(\beta,b_i)\left\{q(\alpha,b_i) - s_i^{\text{JM}}(\alpha)\right\}^{\mathsf{T}} p(b_i \mid T_i,\delta_i,y_i)$ and

$$q(\alpha,b_i) = \delta_i m_i(t_i) - \int_0^{T_i} h_0(s_i)m_i(s_i)\exp\left\{\gamma^{\mathsf{T}}W_i + \alpha m_i(s_i)\right\}ds_i.$$

The ISNI for the shared parameter model can be written as

$$\begin{aligned}\text{ISNI}_{\text{SPM}}(\beta) &= -\left(\sum_{i=1}^{n}\sigma^{-2}X_i^{\mathsf{T}}X_i + I_3\right)^{-1} \\ &\quad \times \sum_{i=1}^{n}\sigma^{-2}X_i^{\mathsf{T}}\left[\mu_{2,b_i} - \mu_{3,b_i}\int_0^{T_i} h(s_i)ds_i\right],\end{aligned}$$

where $\mu_{2,b_i} = \mathbb{E}_{b_i|T_i,\delta_i,y_i}\left[b_i(y_i - m_i(t_i))\right]$, $\mu_{3,b_i} = \mathbb{E}_{b_i|T_i,\delta_i,y_i}\left[b_i\exp\left\{\gamma^{\mathsf{T}}W_i\alpha^{\mathsf{T}}b_i\right\}(y_i - m_i(t_i))\right]$,

$$I_3 = \int_{b_i} q^*(\beta,b_i)\left\{q^*(\beta,b_i) - s_i^{\text{SPM}}(\beta)\right\}^{\mathsf{T}} p(b_i \mid T_i,\delta_i,y_i),$$

and

$$q^*(\beta,b_i) = \sigma^{-2}X_i^{\mathsf{T}}(y_i - m_i(t_i)).$$

In the latter formulation, the ISNI represents a matrix with columns equal to the number of random effects inserted in the survival model. For instance, if we consider a model with p fixed effects $\beta_1,\ldots,\beta_p$ and two random effects b_{1i} and b_{2i}, the association parameter is a vector $\alpha = (\alpha_1,\alpha_2)$ and the ISNI is a matrix whose $\{i,k\}$th element is given by

$$\text{ISNI}_{\text{JM}}(\beta) = \left\{-\left(\frac{\partial^2 L}{\partial\beta_i\partial\beta_j}\right)^{-1}\frac{\partial^2 L}{\partial\beta_i\partial\alpha_k}\right\}$$

where $i, j = 1, \ldots, p$ and $k = 1, 2$.
Finally, it is relevant to note that the ISNI for the shared parameter model depends on α only through the first derivative of the posterior distribution of the random effects. This implies that the random effects covariance matrix may have an effect on the index computation.

Bibliography

M. Alfó and M. Aitkin (2000). Random coefficient model for binary longitudinal responses with attrition. *Statistics and Computing*, **10**, 279–287.

M. Alfó, L. Nieddu, and C. Vitiello (2010). Finite mixture models for longitudinal responses with attrition: a pattern mixture specification. In: *ERCIM*.

E. B. Andersen (1973). *Conditional inference and models for measuring*. Mentalhygiejnisk forlag, Copenhagen.

D. A. Anderson and K. Aitkin (1985). Variance component models with binary response: interviewer variability. *Journal of the Royal Statistical Society, B*, **47**, 203–210.

D. Calvetti, G. H. Golub, W. B. Gragg, and L. Reichel (2000). Computation of gauss-kronrod quadrature rules. *Math. Comput.*, **69**, 1035–1052.

B. P. Carlin and T. A. Louis (2009). *Bayesian Methods for Data Analysis*. Boca Raton, FL: Chapman and Hall - CRC Press.

J. S. K. Chan, A. Y. C. Kuk, J. Bell, and C. Mc Gilchrist (1998). The analysis of methadone clinic data using marginal and conditional logistic models with mixture or random effects. *Australian and NewZealand Journal of Statistics*, **40**, 1–10.

D. R. Cox (1972). Regression models and life tables (with discussion). *Journal of the Royal Statistical Society, Series B*, **34**, 187–220.

A. Creemers, N. Hens, M. Aerts, G. Molenberghs, G. Verbeke, and M. Kenward (2010). A sensitivity analysis for shared-parameter models for incomplete longitudinal data. *Biometrical Journal*, **52**, 111–125.

M. Davidian, A. A. Tsiatis, and S. Leon (2005). Semiparametric estimation of treatment effect in a pretest-posttest study with missing data. *Statistical Science*, **20**, 261–301.

P. J. Davis and P. Rabinowitz (1967). *Numerical Integration*. Waltham, Mass.

A.P. Dempster, N.M. Laird, and D.B. Rubin (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society, Series B*, **39**, 1–38.

A.P. Dempster and D.B. Rubin (1987). *Incomplete Data in Sample Surveys, Vol. II: Theory and Annotated Bibliography*. New York: Academic Press.

D. Diggle, P and. Farewell and R. Henderson (2007). Analysis of longitudinal data with dropout: objectives, assumptions and a proposal. *Applied Statistics*, **56**, 499–550.

P. J. Diggle (1990). *Time series: a Biostatistical Introduction*. Oxford University Press.

P. J. Diggle (1994). Informative dropout in longitudinal data analysis. *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, **43**, 49–93.

P. J. Diggle, P. J. Heagerty, K. Liang, and S. L. Zeger (1994). *Analysis of Longitudinal Data*. Oxford Statistical Science.

P. J. Diggle and M. G. Kenward (1994). Informative dropout in longitudinal data analysis. *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, **43**, 49–93.

C. Fawcett and D. Thomas (1996). Simultaneously modelling censored survival data and repeatedly measured covariates. *Statistics in Medicine*, **15**, 1663–168.

G. M. Fitzmaurice, N. M. Laird, and J. H. Ware (2004). *Applied Longitudinal Analysis*. Wiley Series in Probability and Statistics.

D. Follmann and M. Wu (1995). An approximate generalized linear model with random effects for informative missing data. *Biometrics*, **51**, 151–168.

A. Goldman, B. Carlin, L. Crane, C. Launer, J. Korvick, L. Deyton, and D. Abrams (1996). Response of CD4+ and clinical consequences to treatment using ddI or ddC in patients with advanced HIV infection. *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology*, **11**, 161–169.

R. Henderson, P. Diggle, and A. Dobson (2000). Joint modelling of longitudinal measurements and event time data. *Biostatistics*, **1**, 465–480.

J. W. Hogan and P. J. Laird (1997). Mixture models for the joint distribution of repeated measures and event times. *Statistics in Medicine*, **16**, 259–272.

J. W. Hogan, J. Roy, and C. Korkontyelou (2004). Tutorial in biostatistics - handling drop-out in longitudinal studies. *Statistics in Medicine*, **23**, 1455–1497.

D. G. Horvitz and D. J. Thompson (1952). A generalization of sampling without replacement from a finite population. *Journal of the American Statistics Association*, **47**, 663–685.

Z. Huang, L. A. Stefanski, and M. Davidian (2009). Latent-model robustness in joint models for a primary endpoint and a longitudinal process. *Biometrics*, **65**, 719–727.

F. James (1980). Monte carlo theory and practice. *Reports on Progress in Physics*, **43**, 1065–1190.

E. L. Korn and A. S. Whittermore (1979). Methods for analyzing panel studiesof acute health effects of air pollution. *Biometrics*, **35**, 795–802.

A. S. Kronrod (1964). Doklady akad (in russian). *Nauk SSSR*, **154**, 283–286.

P. J. Kurland, B. F. aand Heagerty (2005). Direct parametrized regression conditioning on being alive: analysis of longitudinal data truncated by deaths. *Biostatistics*, **6**, 241–258.

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

Y. Lee and J. A. Nelder (2009). Likelihood inference for models with unosservables: Another view. *Statistical Science*, **3**, 255–269.

K. Y. Liang and S. L. Zeger (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, **73**, 1–38.

R. J. A. Little (1993). Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association*, **88**, 125–134.

R. J. A. Little (1995). Modelling the dropout mechanism in repeated-measures studies. *Journal of American Statical Association*, **90**, 1112–1121.

R. J. A. Little and D. B. Rubin (2002). *Statistical Analysis with Missing Data*. Wiley Series in Probability and Statistics.

Q. Liu and D. A. Pierce (1994). A note on gauss-hermite quadrature. *Biometrika*, **81**, 624–629.

G. Ma, A. B. Troxel, and F. Heitjan (2005). An index of local sensitivity to nonognorabile drop-out in longitudinal modelling. *Statistics in Medicine*, **24**, 2129–2150.

P. McCullagh and J. A. Nelder (1989). *Generalized linear models*. Chapman and Hall, New York.

G. Molemberghs and M. G. Kenward (2007). *Missing Data in Clinical Studies*. John Wiley and Sons, Ltd.

G. Molemberghs, G. Verbeke, C. G. B. Demetrio, and M. C. Vieira (2011). A family of generalized linear models for repeated measures with normal and conjugate random effects. *Statistical Science*, **25**, 325–347.

G. Molenberghs and G. Verbeke (2005). *Models for Discrete Longitudinal Data*. New York: Springer.

P. Murtaugh, E. Dickinson, G. Van Dam, M. Malincho, P. Grambsch, A. Langworthy, and C. Gips (2002). Primary billiary cirrhosis: prediction of short-term survival based on repeated patient visit. *Hepatology*, **20**, 126–134.

J.C. Naylor and A. F. M. Smith (1982). Applications of a method for the efficient computation of posterior distributions. *Applied Statistics*, **31**, 214–225.

R. Nelsen (1999). *An Introduction to Copulas*. Springer-Verlag, New York.

Y. Pawitan and S. Self (1993). Modeling desease marker process in aids. *Journal of the American Statistical Association*, **83**, 719–726.

R. Prentice (1989). Surrogate endpoints in clinical trials: Definition and operation criteria. *Statistics in Medicine*, **8**, 431–440.

R. L. Prentice (1988). Correlated binary regression with covariates specific to each binary observation. *Biometrics*, **44**, 1033–1048.

Ten Have T. Pulkstenis, E. P and and J. R. Landis (1998). Model for the analysis of binary longitudinal pain data subject to informative dropout through remedication. *Journal of the American Statistical Association*, **93**, 438–450.

Y. Qian and H. Xie (2010). Measuring the impact of nonignorability in panel data with nonmonotone nonresponse. *Journal of Applied Econometrics*.

M. S. Ridout (1994). Reader reaction: testing for random dropouts in repeated measurements data. *Biometrics*, **47**, 1255–1258.

D. Rizopoulos (2010). Jm: An r package for the joint modelling of longitudinal and time-to-event data. *Journal of Statistical Software*, **35**, 9.

D. Rizopoulos and P. Ghosh (2011). A bayesian semiparametric multivariate joint model for multiple longitudinal outcomes and a time-to-event. *Statistics in Medicine*, **30**, 1366–1380.

D. Rizopoulos, G. Verbeke, and E. Lesaffre (2009). Fully exponential laplace approximations for the joint modelling of survival and longitudinal data. *Journal of the Royal Statistical Society, Series B*, **71**, 637–654.

D. Rizopoulos, G. Verbeke, and G. Molenberghs (2008a). Shared parameter models under random effects misspecification. *Biometrika*, **95**, 63–74.

D. Rizopoulos, G. Verbeke, and Y. Vanrenterghem (2008b). A two-part joint model for the analysis of survival and longitudinal binary data with excess zeros. *Biometrics*, **64**, 611–619.

D.B. Rubin (1976). Inference and missing data (with discussion). *Biometrika*, **63**, 581–592.

H. E. Salzer, R. Zucker, and R. Capuano (1952). Table of zeros and weights factors of the first twenty hermite polynomials. *J. Res. Nat. Bur. Standards*, **48**, 111–116.

S. Self and Y. Pawitan (1992). *AIDS Epidemiology: Metodological Issues.*, chap. Modeling a marker of desease progression and onset of desease. Birkhauser, Boston.

J. L. Shafer (1997). *Analysis of Incomplete Multivariate Data*. Chapman and Hall.

X. Song, M. Davidian, and A. A. Tsiatis (2002a). A semiparametric likelihood approach to joint modeling of longitudinal and time-to-event data. *Biometrics*, **58**, 742–753.

X. Song, M. Davidian, and A.A. Tsiatis (2002b). A semiparametric likelihood approach to joint modeling of longitudinal and time-to-event data. *Biometrics*, **58**, 742–753.

T. R. Ten Have, E. R. Kunselman, A. R. Pulkstenis, and J. R. Landis (1998). Mixed effects logistic regression models for longitudinal binary response data with informative drop-out. *Biometrics*, **54**, 367–383.

A. B. Troxel, G. Ma, and F. Heitjan (2004). Shared parameter models for the joint analysis of longitudinal data and event times. *Statistica Sinica*, **14**, 1221–1237.

A. Tsiatis and M. Davidian (2001). A semiparametric estimator for the proportional hazard model with longitudinal covariates measured with error. *Biometrika*, **88**, 447–458.

A. Tsiatis and M. Davidian (2004). An overview of joint modeling of longitudinal and time-to-event data. *Statistica Sinica*, **14**, 793–818.

G. Verbeke, G. Molemberghs, H. Thijs, E. Lesaffre, and M. G. Kenward (2001). Sensitivity analysis for non-random drop-out: a local influence approach. *Biometrics*, **57**, 7–14.

S. Viviani, D. Rizopoulos, and M. Alfó (2011). Joint models and sensitivity: the isni formulation.

J. H. Ware, S. Lipsitz, and F. E. Speizer (1988). Issues in the analysis of repeated categorical outcomes. *Statistics in Medicine*, **7**, 95–107.

S. P. Willemsen, M. De Ridder, A. Hokken-Koelega, and E. Lesaffre (2011). Modelling height for children born small for gestational age treated with growth hormone. *Statistical Methods in Medical Research*, to appear.

W. H. Wong (1986). Theory of partial likelihood. *Annals of Statistics*, **14**, 88–123.

M. Wu and R. Carrol (1988). Estimation and comparison of changes in presence of informative right censoring by modelling the censoring process. *Biometrics*, **45**, 939–955.

M. Wulfsohn and A. Tsiatis (1997). A joint model for survival and longitudinal data measured with error. *Biometrics*, **53**, 330–339.

H. Xie (2008). A local sensitivity analysis approach to longitudinal non-gaussian data with non-ignorable dropout. *Statistics in Medicine*, **27**, 3155–3177.

H. Xie (2009). Bayesian inference from incomplete longitudinal data: A simple method to quantify sensitivity to nonignorable dropout. *Statistics in Medicine*, **28**, 2725–2747.

S. L. Zeger and K. Y. Liang (1996). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, **42**, 121–130.

Acknowledgements

Ringrazio in primo luogo Marco Alfó, il cui supporto sia professionale che umano, la pazienza e le lunghe discussioni sono stati indispensabili per questa tesi. Per lo stesso motivo ringrazio Dimitris Rizopoulos. I Capitoli 4 e 5 sono stati svolti in collaborazione con loro.
Un ringraziamento all'accoglienza del Dipartimento di Biostatistica di Rotterdam, in particolare a Sten e Magda, che hanno contribuito a rendere il lavoro e la vita in Olanda piú piacevole, e al Dipartimento di Statistica di Roma.
Ringrazio i miei colleghi di Dottorato e gli Assegnisti per le discussioni stimolanti, le risate, i caffé e le serate.
Tutto questo non sarebbe stato possibile senza il coordinatore Fulvio De Santis, una guida diretta ed indiretta in questo percorso.
Un ringraziamento speciale alla mia famiglia, su cui so di poter contare sempre, e agli amici di sempre, Ponga, Gian, Giulia, Alessandra, Alice e Valentina.