Katholieke Universiteit Leuven
Faculteit Wetenschappen
Departement Wiskunde

Joint Modelling of

Longitudinal and Survival Data

Dimitris Rizopoulos

Promotor:	Prof. Geert Verbeke	Proefschrift voorgedragen tot
co-Promotoren:	Prof. Emmanuel Lesaffre	het behalen van het doctoraat
	Prof. Irene Gijbels	de Wetenschappen

Mei 2008

ISBN 978-90-8649-177-3
D/2008/10.705/23

Acknowledgements

My doctoral studies have been an enjoyable journey that matured me as a person and as a scientist. I would like to say a few words of thanks to those who contributed to my achievements in one or another way.

First, I would like to thank my promotor Prof. Dr. Geert Verbeke for his guidance and support during the years. His critical questions and suggestions have greatly helped me understand in depth the area of joint modelling. I would like also to express my gratitude to my co-promotor Prof. Dr. Emmanuel Lesaffre for his insightful comments and ideas that considerably contributed to the improvement of my work, and for creating opportunities for future collaborations. I would also like to express my gratitude to Prof. Dr. Irène Gijbels for agreeing to act as my co-promotor from the Faculty of Science and for her comments on an earlier draft of my thesis. I also wish to particularly thank Prof. Dr. Geert Molenberghs for his interest in my work and his valuable input in various aspects of my research.

I greatly appreciate the time, effort, comments and suggestions of the members of the doctoral examination committee: Prof. Dr. Geert Molenberghs, Prof. Dr. Marc Aerts, Prof. Dr. Lea Vermeire, Prof. Dr. An Carbonez, and Prof. Dr. Mia Hubert that improved the final version of this thesis.

Many thanks also to my colleagues from the Biostatistical Centre, and especially to Steffen for our fruitful and enjoyable discussions over biostatistics and others and for bringing to my attention the renal graft failure data set that motivated me to investigate a number of issues in the joint modelling area, and Kris for the technical support.

Moreover, I gratefully acknowledge the permission of Prof. Dr. Yves Vanrenterghem to use the renal graft failure data in my thesis, and thank him for providing background information on this project.

I would like to thank my parents, my brother, my uncle and my grandmother in

Greece for their continual support throughout my undergraduate and postgraduate years of study and for providing me with the opportunity to be where I am.

Last but not least, I would especially like to thank my fiancée, Roula. Her unconditional love and support in difficult situations have helped and guided me during all these years.

Dimitris Rizopoulos

Leuven, May 2008

This thesis is based on the following papers:

Chapter 2: Rizopoulos, D., Verbeke, G. and Molenberghs, G. (2008). Shared parameter models under random effects misspecification. *Biometrika* **95**, 63 – 74.

Chapter 3: Rizopoulos, D., Verbeke, G., Lesaffre, E. and Vanrenterghem, Y. (2007). A two-part joint model for the analysis of survival and longitudinal binary data with excess zeros. *Biometrics*, doi: 10.1111/j.1541-0420.2007.00894.x.

Chapter 4: Rizopoulos, D., Verbeke, G. and Lesaffre, E. (2007). Fully exponential Laplace approximation for the joint modelling of survival and longitudinal data. Under review.

The author of this thesis is also an author in the following papers:

Rizopoulos, D. and Moustaki, I. (2008). Generalized latent variable model with nonlinear effects. *British Journal of Mathematical and Statistical Psychology*, doi: 10.1348/000711007X213963.

Tsonaka, R., Rizopoulos, D., Verbeke, G. and Lesaffre, E. (2008). Marginalized semiparametric shared parameter models for incomplete longitudinal ordinal responses. Under revision.

Rizopoulos, D., Verbeke, G. and Lesaffre, E. (2008). Sensitivity analysis in pattern mixture models using the extrapolation method. Under revision.

Lesaffre, E., Rizopoulos, D. and Tsonaka, S. (2007). The logistic transform for bounded outcome scores. *Biostatistics* **8**, 72 – 85.

Rizopoulos, D. (2006). ltm: An R package for latent variable modelling and item response theory analyses. *Journal of Statistical Software* **17**(5), 1 – 25.

Tsonaka, S., Rizopoulos, D. and Lesaffre, E. (2006). Power and sample size calculations for discrete bounded outcome scores. *Statistics in Medicine* **25**, 4241 – 4252.

Contents

Index to Notation xi

1 General Introduction 1

1.1 Joint Modelling for Longitudinal & Survival Data 1

1.2 Motivating Case Study 4

1.3 Joint Modelling of Survival and Longitudinal Data in the Literature 9

1.4 Issues in Joint Modelling & Aims of this Thesis 11

1.4.1 Random Effects Misspecification 12

1.4.2 Investigation of the Association Structure 13

1.4.3 Choice of Copula and Sensitivity Analysis 14

1.4.4 High Dimensional Random Effects 14

1.4.5 Estimation of Standard Errors 16

1.4.6 Software 17

References 17

2 Shared Parameter Models under Random Effects Misspecification 21

2.1 Introduction 22

2.2 Shared Parameter Models Framework 23

2.2.1 Model Specification 23

2.2.2 Two Parameterizations 25

2.3 Random Effects Misspecification 27

2.3.1 Preamble 27
2.3.2 Parameter Estimators 27
2.3.3 Standard Errors 29
2.4 A Simulation Study 30
2.4.1 Study Set-up 30
2.4.2 Fitted Models 31
2.4.3 Results 32
2.5 Application 33
2.6 Discussion 37
2.7 Appendix A 39
2.8 Appendix B 42
References 44

3 A Two-Part Joint Model for the Analysis of Survival and Longitudinal Binary Data with excess Zeros 47

3.1 Introduction 48
3.2 The Two-Part Shared Parameter Model Formulation 52
3.2.1 Submodels Specification 52
3.2.2 Goodness-of-Fit, Choice of Copula and Sensitivity Analysis 55
3.3 EM Algorithm 57
3.4 Renal Graft Failure Analysis 60
3.5 Simulation Study 64
3.6 Conclusion 65
3.7 Supplementary Material 66
3.7.1 Simulation Study Design 66
3.7.2 Simulation Analysis Models 68
3.7.3 Simulation Results 68

3.8 Appendix A 69
3.9 Appendix B 74
3.10 Appendix C 76
References 76

4 Fully Exponential Laplace Approximations for the Joint Modelling of Survival and Longitudinal Data 81

4.1 Introduction 82
4.2 Joint Modelling Framework 86
4.2.1 Shared Parameter Model 86
4.2.2 Longitudinal Responses Submodel 87
4.2.3 Survival Submodel 88
4.3 Estimation via an EM Algorithm 90
4.3.1 Score Vector and EM algorithm 90
4.3.2 Laplace Approximations for Posterior Moments 92
4.3.3 Standard Errors 95
4.4 Asymptotic Behaviour of the Laplace Estimators 96
4.5 Simulation Study 97
4.6 Application 98
4.7 Discussion 101
4.8 Supplementary Material 103
4.8.1 Simulation Study Set-up 103
4.8.2 Simulation Study Results 104
4.9 Appendix A 107
4.10 Appendix B 108
4.11 Appendix C 110
References 111

5 General Conclusion **115**

5.1 Discussion 115

5.1.1 Choice of Parameterization & Questions of Interest 115

5.1.2 Random Effects Misspecification 116

5.1.3 Choice of Numerical Integration Method 117

5.1.4 Importance of Underlying Assumptions 119

5.2 Topics for Future Research 121

5.2.1 More than One Marker 121

5.2.2 Sensitivity Analysis 122

5.2.3 Residuals Analysis 125

References 126

A Software **129**

A.1 The JM Package 129

A.2 JM Package Extensions 131

References 132

Notation

In this section we give brief explanation of the abbreviations and notations used in this thesis.

AIC	Akaike's Information Criterion
BIC	Bayesian Information Criterion
EB	Empirical Bayes
GFR	Glomerular Filtration Rate
HIV	Human Immune Virus
MAR	Missing At Random
NMAR	Not Missing At Random
y_i^o	the observed part of the longitudinal response vector of the ith subject
y_i^m	the missing part of the longitudinal response vector of the ith subject
$\top$	matrix transpose
$\text{diag}(\cdot)$	diagonal matrix
$\text{tr}(\cdot)$	the trace of a matrix
$\det(\cdot)$	the determinant of a matrix
$\|\cdot\|$	the Euclidean norm
$\Pr(\cdot)$	probability
$p(\cdot)$	probability density function
$O(n)$	$f : \Re \to \Re$ is $O(n)$ if and only if $\exists n_0$, $\exists M > 0$ such that $\|f(n)\| \leq Mn$, for $n > n_0$
$\xrightarrow{P}$	convergence in probability
$\in$	$a \in \mathcal{A}$ denotes that a belongs to the set $\mathcal{A}$

CHAPTER 1 General Introduction

1.1 Joint Modelling for Longitudinal & Survival Data

In medical studies often two types of outcomes are considered, namely a set of longitudinal response measurements and the time to an event of interest. These two outcomes are usually separately analyzed; however, in three settings a joint modelling approach is required. First, in a survival analysis context, in order to measure the effect of a time-dependent covariate measured with error or in order to exploit longitudinal markers as surrogates for survival. Second, in longitudinal studies in order to adjust derived inferences for possibly outcome-dependent dropout. Finally, in investigating the association structure between the longitudinal and event processes. Joint models for survival and longitudinal data have recently become quite popular in HIV and cancer vaccines studies. In particular, in HIV studies, interest lies in the relation between longitudinal measurements of a biomarker such as the CD4 lymphocyte count or the estimated viral load, and the time to seroconversion or death. The number and timing of these longitudinal measurements usually vary between patients, resulting in highly unbalanced data

sets, in which dropout is quite common. In cancer vaccine (immunotherapy) trials the time-to-event endpoint of interest is often the disease progression or time to death. Vaccinations are given to patients to raise patient's antibody levels against tumor cells. Therefore a successful vaccine activates the patient's immune system against future tumor growth by increasing the antibodies' production and strength. Longitudinal measurements of concentration of antibodies help clinicians to monitor the progress of the immunity level, and thus these measurements are expected to be highly predictive for the time-to-event endpoint.

When interest is in the survival outcome, traditional approaches, including the partial likelihood for the Cox proportional hazards model, encounter several difficulties (Tsiatis, DeGruttola, and Wulfsohn, 1995; Wulfsohn and Tsiatis, 1997) with longitudinal time-dependent covariates. If the entire history of the time-dependent covariate is available, then partial likelihood could be readily employed and there would be no complication in modelling the survival times. However, the longitudinal measurements are typically intermittently collected at some set of times and are not available at all time points, and especially not at the event time. Moreover, the observed values may not be the "true" values since usually the longitudinal responses contain measurement error. A final complication is that the longitudinal measurements are, in fact, the output of a stochastic process that is generated by the individual, and is directly related to the failure mechanism. In particular, in this case we know that the individual has not yet experienced the event, provided that we were able to collect his measurement. Such time-dependent covariates are termed internal covariates (Kalbfleisch and Prentice, 2002, Sect. 6.3.2) and must be specially treated in a survival analysis setting, since in this case the hazard function is not directly related to the survival function (Kalbfleisch and Prentice, 2002, Sect. 6.3). In order to account for these special features of the longitudinal time-dependent covariate, a model for the joint distribution of the longitudinal and survival outcomes is required.

When interest is in the longitudinal outcome, the occurrence of events induces incompleteness since no longitudinal measurements are available at and after the event time. This type of incompleteness is known in the missing data literature as dropout or attrition. When the dropout is nonignorable, i.e., when it corresponds to a Not Missing At Random (NMAR) missing data mechanism as defined in Little and Rubin (2002, Sect. 1.3), then it is required that the longitudinal and dropout processes are jointly modelled in order to obtain valid inferences. One of the modelling frameworks for handing nonignorable dropout is the shared parameter model (Wu and Carroll, 1988; Wu and Bailey, 1989; Follmann and Wu, 1995), which postulates a time-to-dropout submodel for the missingness process and a mixed effects submodel for the longitudinal responses. The specification of shared parameter models allows the probability of dropout at time k to depend on values of the outcome at both past and future time points, through a set of random effects. This kind of models have two appealing features. First, for some medical studies, they provide a conceptually simpler framework for handing nonignorable dropout than other commonly used models, such as the selection and pattern mixture models (for a thorough description of the selection and pattern mixture models see Molenberghs and Kenward, 2007). That is, in some applications it is reasonable to assume that subjects which show steep increases in their longitudinal profiles may be more (or less) likely to dropout. Second, such models can easily handle the case of intermittent or nonmonotone missingness (Tsonaka, Verbeke, and Lesaffre, 2008; Tsonaka et al., 2007), which is a frequent type of missing data especially in longitudinal studies with few planned measurements.

When interest is in the dependence structure between the two processes, common measures of association (e.g., nonparametric measures based on ranks) cannot be readily applied. This is due to the multivariate nature and incompleteness in the longitudinal outcome, and censoring in the event outcome. In such situations joint likelihood models that assume underlying latent variable constructs can be

used to investigate the association structure between the two processes (Henderson, Diggle, and Dobson, 2000). In particular, conditionally on these latent variables the two processes are regarded as independent (conditional independence assumption). This formulation facilitates the investigation of dependence, since association is measured at the low-dimensional latent variables space.

1.2 Motivating Case Study

Chronic kidney disease, also known as chronic renal disease, is a progressive loss of renal function over a period of months or years through five stages. Each stage is a progression through an abnormally low and progressively worse glomerular filtration rate. Patients with chronic kidney disease experience complications such as high blood pressure, anemia, weak bones, poor nutritional health and nerve damage. Furthermore, when kidney disease progresses, this may eventually lead to renal failure, which requires dialysis or a kidney transplantation to maintain life. Many studies have been conducted to investigate which factors play a role in the progression of chronic kidney diseases. The motivating example that is used throughout this thesis concerns a study on patients that underwent, between 1/21/1983 and 8/16/2000, a primary renal transplantation with a graft from a deceased or living donor in the University Hospital Gasthuisberg of the Catholic University of Leuven (Belgium). The clinical interest lies in the long term performance of the new graft, and especially in graft survival for at least a period of ten years. During the follow-up period, patients were periodically tested for the condition and performance of their graft. Here we consider three markers that are known to be related to graft functioning and graft survival. These are the Glomerular Filtration Rate (GFR), the proteinuria, and the blood haematocrit level. More information regarding these markers is given in the following.

Creatinine is a waste product of creatine, an important substance of muscles

and meat. Creatinine is almost entirely cleared from the blood by the kidney by filtration through the glomerular filter. The serum creatinine level is therefore used as an indicator of the Glomerular Filtration Rate. However, the creatinine generation rate is heavily influenced by dietary protein intake, muscle mass and muscle activity, and thus methods to calculate the GFR from the serum creatinine level that correct for this variation are used. In this study, the Cockcroft-Gault formula (Cockcroft and Gault, 1976) is applied to transform the creatinine level into a (calculated) GFR. The Cockcroft-Gault formula is an inverse function of the creatinine level, adjusted for gender, body weight and age; these three covariates are used as surrogates of dietary protein intake, muscle mass, and muscle activity.

Proteins are large organic compounds made of amino acids, which are essential parts of the human body and participate in every process within cells. Most proteins in the blood are too big to pass through the kidneys' filters into the urine unless the kidneys are damaged. Thus, the condition in which the urine contains an abnormal amount of protein is called proteinuria, and is an indication of renal graft malfunctioning. In the current study, the presence of more than 1 gr of protein in a 24 hours urine collection has been recorded as a positive finding of proteinuria.

The haematocrit is defined as the percentage of the volume of blood that is occupied by red blood cells. Red blood cells are crucial for the transportation of oxygen and a decrease of their number results in anemia. One of the functions of the kidneys is to produce a sufficient amount of the hormone erythropoietin to ensure that the blood haematocrit level remains within normal limits. However, damaged kidneys produce smaller amounts of erythropoietin, resulting thereby in low haematocrit levels. Therefore, the blood haematocrit level has been also considered as an indicator for the functioning of a renal graft in the present study.

From preliminary descriptive analyses there is evidence that both the average and the subject-specific longitudinal trajectories are of relevance for predicting a

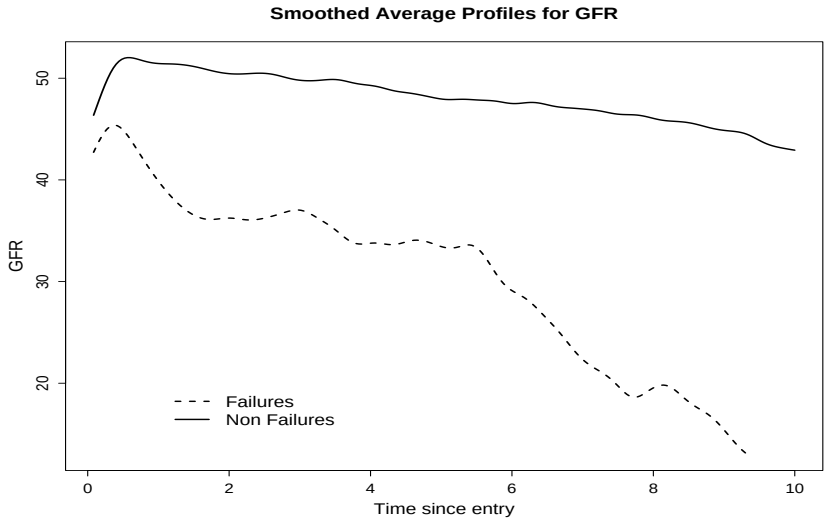

Figure 1.1: *Smooth average longitudinal profiles for GFR, obtained using a Nadaraya-Watson kernel regression estimate.*

graft failure. Smooth versions of the average longitudinal profiles are depicted in Figures 1.1, 1.2 and 1.3, and subject-specific profiles for eight randomly selected patients from the sample at hand are illustrated in Figures 1.4, 1.5 and 1.6.

In particular, we observe that patients who experienced a graft failure, present relatively different time evolutions than patients who did not exhibit the event. Hence, the resulting patient-specific longitudinal profiles contain a wealth of information for the clinician who wants to monitor the progression of the risk status of a patient. In this thesis we are interested in investigating the association structure between the time to graft failure and the longitudinal outcomes, and in determining whether the longitudinal measurements aid in the prediction of events. To tackle these questions a joint modelling approach of the longitudinal markers and the time to graft failure is required.

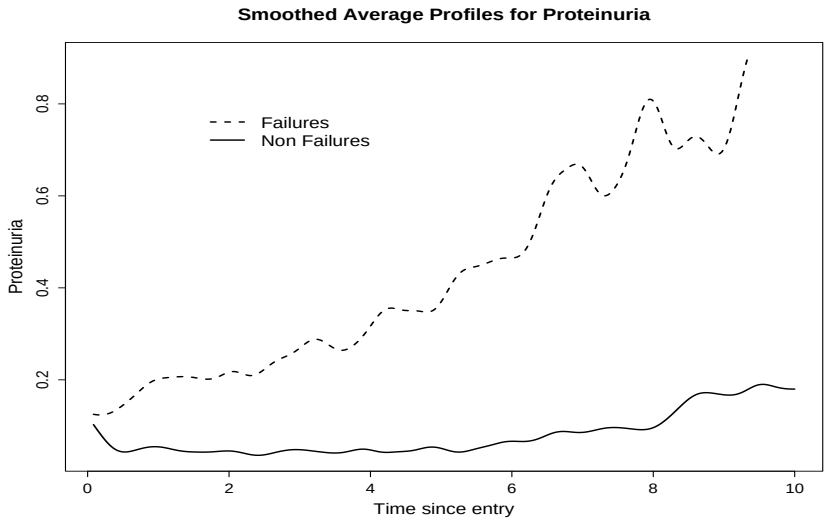

Figure 1.2: *Smooth average longitudinal profiles for Proteinuria, obtained using a Nadaraya-Watson kernel regression estimate.*

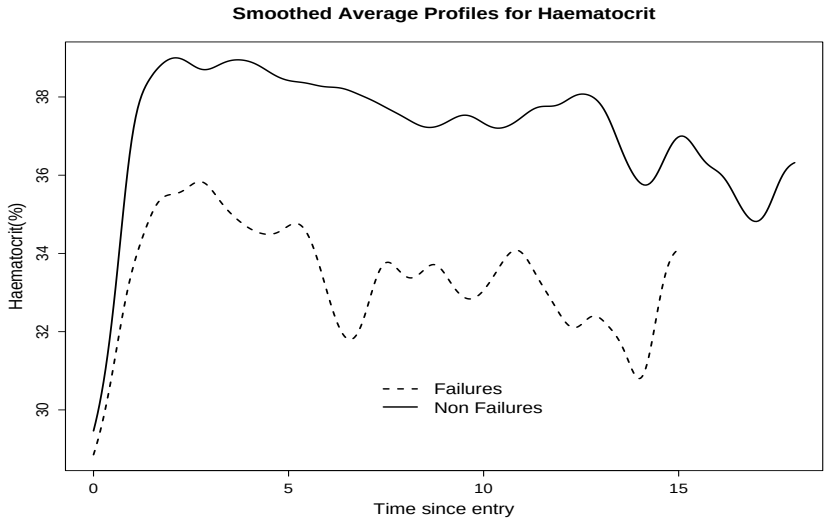

Figure 1.3: *Smooth average longitudinal profiles for Haematocrit, obtained using a Nadaraya-Watson kernel regression estimate.*

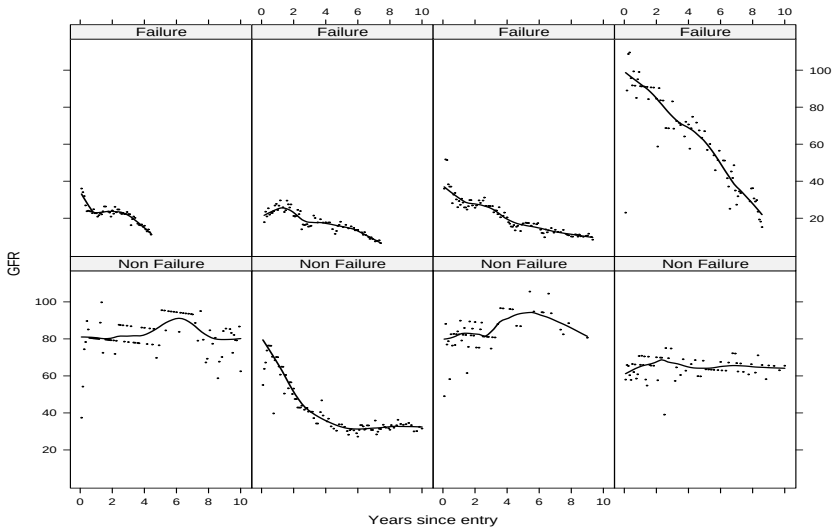

Figure 1.4: *GFR sample subject-specific longitudinal trajectories for eight randomly selected patients. The top-row panels depict patients who experience graft failure and the bottom-row panels censored patients. The superimposed lines represent fitted curves using the loess smoother.*

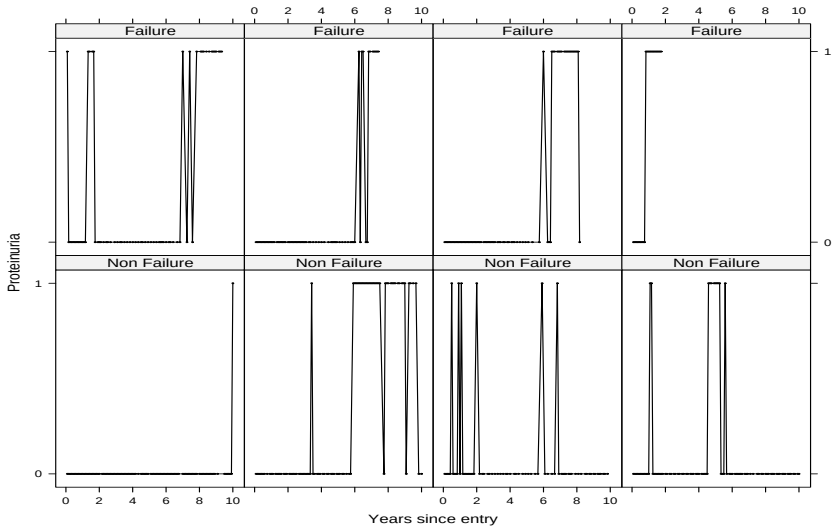

Figure 1.5: *Proteinuria sample subject-specific longitudinal trajectories for eight randomly selected patients. The top-row panels depict patients who experience graft failure and the bottom-row panels censored patients.*

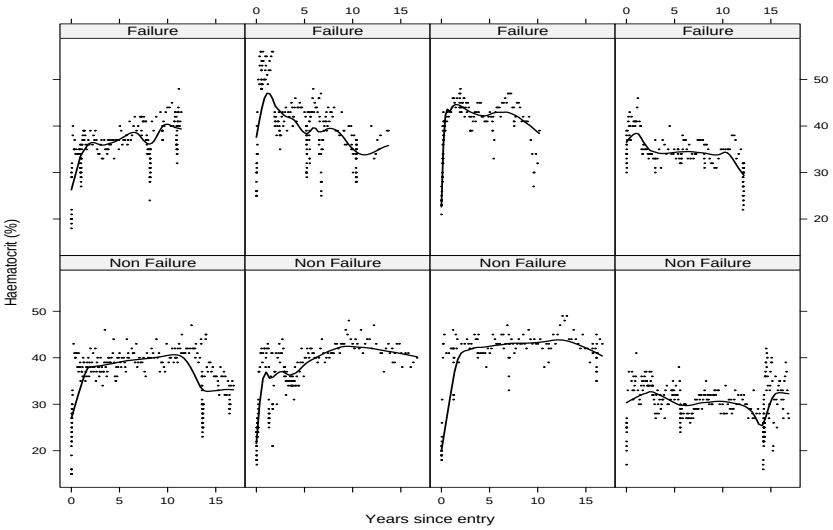

Figure 1.6: *Haematocrit sample subject-specific longitudinal trajectories for eight randomly selected patients. The top-row panels depict patients who experience graft failure and the bottom-row panels censored patients. The superimposed lines represent fitted curves using the loess smoother.*

1.3 Joint Modelling of Survival and Longitudinal Data in the Literature

Shared parameter models offer an appealing framework for the joint modelling of survival and longitudinal processes. In particular, they assume that a latent process, expressed by a set of time-invariant random effects, induces the dependence between the two explicitly observed processes. Excellent overviews of the joint modelling literature are given by Hogan and Laird (1997) and more recently by Tsiatis and Davidian (2004) and Yu et al. (2004). Here we review some of the most important contributions.

Wu and Carroll (1988) first formulated a shared parameter model in order to correct for informative censoring in the longitudinal process by including the subject-specific slopes as a covariate in a probit model for the dropout (i.e., censo-

ring) process. Follmann and Wu (1995) developed a conditional approximation to the full shared parameter model, in order to avoid the requirement for numerical integration. Their model is defined as the integral over the random effects, of the product of two terms, namely, the conditional distribution of the longitudinal responses given the random effects, and the conditional distribution of the random effects given the covariates. Shared parameter models for longitudinal binary data with nonignorable dropout have been considered by Pulkstenis, Ten Have, and Landis (1998) and Ten Have et al. (1998) in the case of dropout, whereas Faucett and Thomas (1996) presented a model for intermittently observed time-dependent binary responses.

Schluchter (1992), Pawitan and Self (1993) and DeGruttola and Tu (1995) considered a joint model in which the time-to-event is modelled parametrically, which facilitates straightforward likelihood inference. Tsiatis et al. (1995) proposed a two-stage approach in which, based on an approximation to the hazard function for the event times, the usual partial likelihood for the Cox model can be used. In this approach the observed covariate history is estimated using empirical Bayes methodology, which requires fitting as many mixed effects models as there are event times in the data set. Dafni and Tsiatis (1998) investigated the performance of this approach via simulation and found that this approximate method yields estimators that reduce but do not completely eliminate the bias.

Wulfsohn and Tsiatis (1997) considered a full likelihood approach for a joint model based on a linear mixed model for the longitudinal process and a proportional hazards model, with infinite-dimensional baseline hazard for the event process. Wang and Taylor (2001), instead of shared random effects, posited a shared integrated Ornstein-Uhlenbeck process. Henderson et al. (2000) proposed a more general joint model by postulating two correlated stationary Gaussian processes, including both random effects and serial correlation, for the longitudinal measurements and survival times, respectively. As noted by Wang and Taylor (2001) and

Henderson et al. (2000), serial correlation processes allow the trend to vary with time and induce a within-subject autocorrelation structure that may be thought of as arising from evolving biological fluctuations in the process about a smooth trend. Tsiatis and Davidian (2004, Sect. 2.2) provide an interesting contradiction between the two approaches (i.e., random effects versus serial correlation).

Tsiatis and Davidian (2001) and Song, Davidian, and Tsiatis (2002) focused on minimizing the impact that erroneous distributional assumptions for the random effects could have in the derived inferences. The former proposed a conditional score approach and developed a set of unbiased estimating equations, the latter considered the model of Wulfsohn and Tsiatis (1997) but relaxed the assumption of normality of the random effects to one requiring only that these random effects have a distribution with a smooth density.

Finally, Tseng, Hsieh, and Wang (2005) have recently developed a joint modelling framework in which for the survival process an accelerated failure time model is postulated, with the corresponding baseline hazard function expressed as a step function.

1.4 Issues in Joint Modelling & Aims of this Thesis

In this thesis we deal with a number of issues in the joint modelling area. In particular, we investigate the effect of misspecifying the random effects distribution in parameter estimators and standard errors, we propose an alternative parameterization for joint models using copulas for the random effects, we discuss sensitivity analysis issues, we use a new type of Laplace approximation to handle multidimensional random effects vectors in joint models, and we consider a flexible model for the event outcome that results in a straightforward estimation of the standard errors. More details regarding these contributions can be found in the following sections.

1.4.1 Random Effects Misspecification

The key assumption of the joint modelling framework is the existence of a set of random effects that underlies both observed processes and induces dependence. These random effects are usually assumed to be normally distributed, even though this choice is not made on the grounds of computational simplicity. Some authors have questioned the Gaussian assumption, in the sense that the resulting inferences can be sensitive to assumptions not easily verifiable from the available data; see for example the discussion in Scharfstein, Rotnitzky, and Robins (1999). To this end, some approaches have been proposed that either relax the distributional assumptions (Song et al., 2002) or make no parametric assumption at all (Tsiatis and Davidian, 2001) for the random effects distribution. However, the main empirical result from these approaches is that the parameter estimates are rather robust to random effects misspecification.

In Chapter 2 we formally investigate this robustness phenomenon. In particular, we will show that the score vector corresponding to the observed data likelihood can be written as the expected value of the score vector conditional on the random effects with respect to the posterior distribution of the random effects given the observed data. Using arguments similar to asymptotic Bayesian theory (Cox and Hinkley, 1974) we show that, as the number of repeated measurements per individuals grows, the longitudinal measurement model is the dominating part in the posterior distribution of the random effects. Thus, misspecification of the random effects distribution does not affect consistency, as more information per subject becomes available. However, the effect of misspecifying the random effects distribution is more prominent in the estimation of standard errors. To account for this a sandwich-type estimator is proposed.

1.4.2 Investigation of the Association Structure

The typical formulation of joint models assumes that the fixed and random effects parts of the longitudinal mixed model are considered as the time-dependent covariate in the survival. This implies that the longitudinal and event processes share a common set of random effects. An implicit feature of this formulation is that it assumes perfect linear correlation between the latent structures of the two processes, since the same random effects are shared. This could be regarded as a rather restrictive assumption that may not be desirable, especially in settings in which the association structure between the measurement and event processes is of interest.

In Chapters 2 and 3 we consider a more flexible parameterization that considers two separate sets of random effects for the two processes. This formulation is in the spirit of the approach proposed by Henderson et al. (2000) who postulate a bivariate Gaussian latent process that drives the two observed outcomes. In this case, the association between the explicitly observed processes is measured through the association of the two random effects. Furthermore, in order to extend the common normality assumption for the random effects, we postulate a joint distribution for the two random effects components using copulas. Copulas (Nelsen, 1999) are multivariate cumulative distribution functions with uniform marginals, and due to Sklar's theorem (Nelsen, 1999) they provide a convenient approach to construct joint distributions by linking univariate marginals to their full multivariate distribution. The advantage of the copula formulation is that it allows for separate modelling of the association structure and the marginals, facilitating thus exploration of the dependence structure.

1.4.3 Choice of Copula and Sensitivity Analysis

The copula formulation of the random effects joint distribution provides a valuable alternative to the common normality assumption. However, a valid question in this setting is which copula best describes the association structure between the longitudinal and survival data. Unfortunately, the random effects are in fact latent variables, and thus checking the appropriateness of the assumed random effects model (i.e., both copula function and marginal distributions) can be challenging. Alternatively, the performance of the assumed random effects model can be implicitly investigated by checking the fit of the joint model to the observed data, and information criteria, such as the AIC or BIC, can be used to select the best fitting copula.

However, we would like to note that the use of measures, based on the observed data, for identifying the best fitting copula should be done with caution. The reason for this lies in the close relationship between the joint modelling of survival and longitudinal measurements and the missing data framework. In particular, as we already mentioned in Section 1.1, joint models correspond to a NMAR missing data mechanism (Little and Rubin, 2002). As it is known in the missing data literature, in NMAR settings the observed data do not contain enough information to distinguish between certain models, since a lot of information is implicitly provided through modelling assumptions. This feature motivated us in Chapters 2 and 3 to performed a sensitivity analysis in order to investigate the effect of the choice of the copula function in the size of the association between the two processes.

1.4.4 High Dimensional Random Effects

A further issue in the joint modelling framework is the requirement for numerical integration over the random effects. In particular, even though the use of

random effects facilitates the use of flexible submodels for the involved processes, calculation of the joint density of the longitudinal and event outcomes involves intractable integrals. This feature makes the estimation of joint models rather computationally demanding. Common numerical integration techniques such as Gaussian quadrature and Monte Carlo have been successfully applied in the joint modelling framework (Song et al., 2002; Henderson et al., 2000). However, in such methods the computational burden increases exponentially with the dimensionality of the integration, which renders the consideration of complex random effects structures (e.g., modelling nonlinear subject-specific trajectories with splines or high-order polynomials) rather prohibitive. A practical alternative in such settings is the Laplace method for integrals (De Bruijn, 1981). The main computational complexity of the Laplace approximation is the requirement of locating the mode of the integrand with respect to the random effects. Furthermore, even though the Laplace approximation is appealing in high dimensional settings, the order of the error of this approximation is $O(n_i^{-1})$, with n_i denoting the number of repeated measurements for the ith subject. This implies that in order for the Laplace method to work satisfactorily, many repeated measurements per subject are required.

In Chapter 4 we have considered a new type of Laplace approximation for the joint modelling of survival and longitudinal data that is of order $O(n_i^{-2})$. The proposed approximation requires locating the same modes as in the $O(n_i^{-1})$ approximation and the computation of a correction term based on these modes. This yields a better approximation to the integral, while keeping the computational complexity at the same order as in the standard Laplace method. The main idea is to apply a fully exponential Laplace approximation to the score vector of the shared parameter model, which, as mentioned above, is expressed as the expected value of the score vector conditional on the random effects, with respect to the posterior distribution of the random effects given the observed data. The main

strength of this approach is that it effectively copes with high-dimensional random effects structures without increasing substantially the computational burden.

1.4.5 Estimation of Standard Errors

In Chapters 2 and 3 we have considered parametric survival models for the event process. However, in the joint modelling context, an unspecified baseline hazard function is typically assumed for the event process in order to protect derived inferences against misspecification. In this case the estimation of standard errors is based on the profile likelihood that is defined conditionally on the nonparametric maximum likelihood estimate (NPMLE) for the baseline hazard. However, recently Hsieh, Tseng, and Wang (2006) noted that the use of the profile likelihood approach leads to underestimation of standard errors. The reason is that in joint models the NPMLE for the baseline hazard cannot be obtained explicitly under the random effect structure, and thus the profile likelihood remains a function of the baseline hazard. Hsieh et al. (2006) have therefore recommended the use of the Bootstrap method, which renders joint models even more computationally demanding.

In order to avoid this problem, in Chapter 4 we propose a flexible but parametric model for the cumulative baseline hazard function by expanding it into B-splines basis functions. Under this formulation, the computation of standard errors derives from standard maximum likelihood theory. In particular, the model is formulated for the logarithm of the cumulative hazard function for the ith subject, and is related to the time-varying Cox model with covariates affecting the log cumulative hazard ratio instead of the log hazard ratio. Our motivation for postulating a flexible model for the logarithm of the cumulative baseline hazard function is that it is typically gently curved or nearly linear as a function of the logarithm of time, and is usually very smooth. Furthermore, the B-splines based formulation allows for a straightforward reparameterization of the spline coefficients in order

to satisfy the constraint that the logarithm of the cumulative baseline hazard is a nondecreasing function of time.

1.4.6 Software

Unfortunately, procedures or functions for fitting joint models for longitudinal and survival data are not currently available on standard software. Thus, such kind of models are not usually considered for 'every-day' analyses and more naive methods (Tsiatis et al., 1995) are used instead. In order to fill this gap, the set of R functions that have been used throughout this thesis are available via the package **JM**. This package can fit a variety of joint models and contains extra functions that produce useful output for such models; a brief description can be found in Appendix A.

References

Cockcroft, D. and Gault, M. (1976). Prediction of creatinine clearance from serum creatinine. *Nephron*, **16**, 31–41.

Cox, D. and Hinkley, D. (1974). *Theoretical Statistics*. Chapman & Hall, London.

Dafni, U. and Tsiatis, A. (1998). Evaluating surrogate markers of clinical outcome when measured with error. *Biometrics*, **54**, 1445–1462.

De Bruijn, N. (1981). *Asymptotic Methods in Analysis*. Dover, New York.

DeGruttola, V. and Tu, X.-M. (1995). Modelling progression of CD-4 lymphocyte count and its relationship to survival time. *Biometrics*, **50**, 1003–1014.

Faucett, C. and Thomas, D. (1996). Simultaneously modeling censored survival data and repeatedly measured covariates: a Gibbs sampling approach. *Statistics in Medicine*, **15**, 1663–1685.

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

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

Hogan, J. and Laird, N. (1997). Model-based approaches to analysing incomplete longitudinal and failure time data. *Statistics in Medicine*, **16**, 259–272.

Hsieh, F., Tseng, Y.-K. and Wang, J.-L. (2006). Joint modeling of survival and longitudinal data: Likelihood approach revisited. *Biometrics*, **62**, 1037–1043.

Kalbfleisch, J. and Prentice, R. (2002). *The Statistical Analysis of Failure Time Data*. Wiley, New York, 2nd edition.

Little, R. and Rubin, D. (2002). *Statistical Analysis with Missing Data*. Wiley, New York, 2nd edition.

Molenberghs, G. and Kenward, M. (2007). *Missing Data in Clinical Studies*. Wiley, New York.

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

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

Pulkstenis, E., Ten Have, T. and Landis, R. (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.

Scharfstein, D., Rotnitzky, A. and Robins, J. (1999). Adjusting for nonignorable dropout using semiparametric non-response models (with discussion). *Journal of the American Statistical Association*, **94**, 1096–1120.

Schluchter, M. (1992). Methods for the analysis of informatively censored longitudinal data. *Statistics in Medicine*, **11**, 1861–1870.

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

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

Tseng, Y.-K., Hsieh, F. and Wang, J.-L. (2005). Joint modelling of accelerated failure time and longitudinal data. *Biometrika*, **92**, 587–603.

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

Tsiatis, A. and Davidian, M. (2004). Joint modeling of longitudinal and time-to-event data: an overview. *Statistica Sinica*, **14**, 809–834.

Tsiatis, A., DeGruttola, V. and Wulfsohn, M. (1995). Modeling the relationship of survival to longitudinal data measured with error: Application to survival and CD4 counts in patients with AIDS. *Journal of the American Statistical Association*, **90**, 27–37.

Tsonaka, R., Rizopoulos, D., Verbeke, G. and Lesaffre, E. (2007). Marginalized semiparametric shared parameter models for incomplete ordinal responses. Submitted.

Tsonaka, R., Verbeke, G. and Lesaffre, E. (2008). A semiparametric shared parameter model to handle nonmonotone nonignorable missingness. *Biometrics*, accepted.

Wang, Y. and Taylor, J. (2001). Jointly modeling longitudinal and event time data with application to acquired immunodeficiency syndrome. *Journal of the American Statistical Association*, **96**, 895–905.

Wu, M. and Bailey, K. (1989). Estimation and comparison of changes in the presence of informative right censoring: conditional linear model. *Biometrics*, **45**, 939–955.

Wu, M. and Carroll, R. (1988). Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. *Biometrics*, **44**, 175–188.

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

Yu, M., Law, N., Taylor, J. and Sandler, H. (2004). Joint longitudinal-survival-cure models and their application to prostate cancer. *Statistica Sinica*, **14**, 835–832.

CHAPTER

2

Shared Parameter Models under Random Effects Misspecification

Abstract

A common objective in longitudinal studies is the investigation of the association structure between a longitudinal response process and the time to an event of interest. An attractive paradigm for the joint modelling of longitudinal and survival processes is the shared parameter framework where a set of random effects is assumed to induce their interdependence. In this work, we propose an alternative parameterization for shared parameter models and investigate the effect of misspecifying the random effects distribution in the parameter estimates and their standard errors.

Keywords: Copula function; Dropout; Joint modelling; Sandwich variance estimator.

2.1 Introduction

In follow-up studies, it is common that each subject provides both a sequence of longitudinal response measurements as well as the time to an event of interest. In such studies, the main scientific interest may focus on three distinct aspects, namely, the longitudinal process in which the event occurrence causes nonignorable dropout, the survival process in which the longitudinal measurements are considered as a time-dependent covariate measured with error and the association structure between the two processes. Typical examples in this setting include HIV studies, in which longitudinal measurements of CD4 cell counts or the estimated viral load are predictive for the time to onset of clinical AIDS or death, as well as kidney disease studies where longitudinal glomerular filtration rate measurements are predictive for the time to kidney failure.

Shared parameter models (Wu and Carroll, 1988; Wulfsohn and Tsiatis, 1997; Tsiatis and Davidian, 2004) offer an appealing framework for the joint modelling of survival and longitudinal processes. In particular, in these models it is assumed that a latent process, expressed by a set of time-invariant random effects, induces the dependence between the two explicitly observed processes. These random effects are usually assumed to be normally distributed, even though this choice is not made on the grounds of computational simplicity. Some authors have questioned the Gaussian assumption, in the sense that the resulting inferences can be sensitive to assumptions not easily verifiable from the available data; see for example the discussion of Scharfstein et al. (1999). To this end, some approaches have been proposed that either relax the distributional assumptions (Song et al., 2002) or make no parametric assumption at all (Tsiatis and Davidian, 2001) about the random effects distribution. However, the main empirical result from these approaches is that the parameter estimates are rather robust to random effects misspecification. Huang, Stefanski, and Davidian (2006) have explored similar behaviour in

structural measurement error models.

In this paper, we formally investigate the effect of misspecifying the random effects distribution in shared parameter models. In particular, we show that, as the number of repeated longitudinal measurements per individual grows, the effect of random effects misspecification vanishes for certain parameter estimators. Two types of parameterization for the random effects component are considered, namely a common set and different sets of random effects for the two processes. For the second type, a copula representation of the random effects distribution is proposed, allowing for different types of dependence structure between the underlying processes, enabling thus sensitivity analysis regarding the association structure.

2.2 Shared Parameter Models Framework

2.2.1 Model Specification

Let T_i^* denote the true event time for the ith subject and consider a random sample of n subjects ($i = 1, \ldots, n$). Letting $\mathcal{J}_i$ denote the underlying potential censoring for subject i, one observes $T_i = \min(T_i^*, \mathcal{J}_i)$ and $\delta_i = I(T_i^* \leq \mathcal{J}_i)$, where $I(\cdot)$ is the indicator function. Moreover, let $y_i(t_{ij})$ denote the longitudinal measurement for subject i taken at time t_{ij}, $j = 1, \ldots, n_i$. Clearly, $y_i(t_{ij})$ is observed whenever $t_{ij} \leq T_i$, and generally $y_i(T_i)$ is not available. Let $\mathcal{Y}_i^\top = \{y_i(t_{ij}), j = 1, \ldots, n_i\}$ denote the vector of observed longitudinal responses for the ith subject. Finally, let b_i represent time-independent random effects that underly both the longitudinal measurement and survival processes. Under this setting, the shared parameter model is defined as follows:

$$p(\mathcal{Y}_i, T_i, \delta_i; \theta) = \int p(\mathcal{Y}_i \mid b_i; \theta_y)\, p(T_i, \delta_i \mid b_i; \theta_t)\, p(b_i; \theta_b)\, db_i, \tag{2.1}$$

where $\theta^\top = (\theta_y^\top, \theta_t^\top, \theta_b^\top)$ is the vector containing the parameters of each one of the submodels, with $p(\cdot)$ denoting the appropriate probability density functions. Here $p(T_i, \delta_i \mid b_i; \theta_t) = p_{T_i^*}(T_i \mid b_i; \theta_t)^{\delta_i} S_{T_i^*}(T_i \mid b_i; \theta_t)^{1-\delta_i}$; that is, it equals either the density for the true event times or the survival function for censored observations. Moreover, we assume that, conditionally on b_i, the longitudinal measurements $\mathcal{Y}_i$ are independent, that is

$$p(\mathcal{Y}_i \mid b_i; \theta_y) = \prod_j p\{y_i(t_{ij}) \mid b_i; \theta_y\}. \tag{2.2}$$

An implicit assumption in factorization (2.1) is that both the censoring and the visiting processes are noninformative, i.e., independent of b_i, and can be ignored in the modelling procedure. Although such an assumption might be questionable in certain situations, we adhere to it here and revisit it in §2.6.

Shared parameter models are built under the so-called conditional independence assumption, where the survival and longitudinal processes are assumed independent given the random effects b_i, implying that all association is induced by the random effects. It is customary to assume b_i to follow a normal distribution, even though this does not usually lead to a tractable form for the integral in (2.1), and hence numerical integration remains a requirement to evaluate the associated likelihood. According to (2.1), distributional assumptions about the random effects allegedly play an important role in the model's factorization since the b_i's link the two processes of interest. However, empirical results (Wang and Taylor, 2001; Song et al., 2002; Tsiatis and Davidian, 2004) show that misspecification of the random effects distribution does not have a great impact on the parameter estimates, except for extreme cases such as discrete distributions. We investigate this phenomenon in more detail in §2.3.

2.2.2 Two Parameterizations

The typical shared parameter models assume that the longitudinal and event processes share a common set of random effects. In particular, the conditional submodels for $\mathcal{Y}_i$ and T_i have the form

$$\begin{aligned} \mathcal{Y}_i \mid b_i \sim \mathcal{N}(\eta_{yi}, \sigma_y^2 I_{n_i}), \qquad & \log T_i \mid b_i \sim \mathcal{P} \quad \text{with} \quad E(\log T_i \mid b_i) = \eta_{ti}, \\ \eta_{yi} = X_{yi}\beta + Z_{yi}b_i, \qquad & \eta_{ti} = x_{ti}^{\top}\gamma + (Z_{yi}b_i)^{\top}\alpha, \end{aligned} \tag{2.3}$$

where I_{n_i} denotes the n_i-dimensional identity matrix, $\mathcal{P}$ denotes a parametric distribution, X_{yi} and Z_{yi} are $n_i \times q_{xy}$ and $n_i \times q_z$ known fixed and random effects design matrices, respectively, β is a vector of unknown fixed effects parameters, σ_y^2 is the error variance, x_{ti} is a $q_{xt} \times 1$ vector of covariates for the event process with an associated coefficient vector γ, and α denotes a vector of association parameters linking the survival process with the random effects structure of the measurement process. If $\alpha = 0$, then the two processes are unrelated, implying that joint modelling is not required under the posited model.

An implicit feature of parameterization (2.3) is that it assumes perfect linear correlation between the latent structures of the two processes, since the same random effects are shared. This could be regarded as a rather restrictive assumption that may not be desirable, especially in settings in which the association structure between the measurement and event processes is of interest. Therefore, we propose a more flexible parameterization that considers two separate sets of random effects for the two processes, linking them using a copula function. Copulas (Nelsen, 1999) are multivariate cumulative distribution functions with uniform marginals, and they provide a natural approach to construct joint distributions and explore dependence. The consideration of two separate random effects is in the spirit of the approach proposed by Henderson et al. (2000) who postulate a bivariate Gaussian latent process shared by the two processes. In particular, we

assume that

$$\eta_{yi} = X_{yi}\beta + Z_{yi}b_{yi}, \quad \eta_{ti} = x_{ti}^{\top}\gamma + b_{ti}, \tag{2.4}$$

$$p(b_{yi}, b_{ti}) = c\{F_y(b_{yi}), F_t(b_{ti}); \alpha\}\; p(b_{yi})\; p(b_{ti}), \tag{2.5}$$

where b_{yi} are random effects for the measurement process and b_{ti} is a frailty term for the survival process. The frailty term is assumed to represent an unobserved covariate explaining heterogeneity (Keiding, Andersen, and Klein, 1997). For the joint density of $\{b_{yi}, b_{ti}\}$ given in (2.5) we assume a copula representation, where $c(\cdot)$ denotes the density of a copula function $C(\cdot)$, and $F_y(\cdot)$ and $F_t(\cdot)$ are the marginal cumulative distributions functions for b_{yi} and b_{ti}, respectively. In the case of multivariate b_{yi}, we assume that the copula behind $F_y(\cdot)$ is directly compatible with $C(\cdot)$ (Nelsen, 1999, pp. 85–6). It is important to note that, under (2.4), the association parameter is a parameter of the random effects model and specifically of the copula function, in contrast to (2.3), where α is a parameter of the event process model. The main advantage of parameterization (2.4) is the flexibility in considering different dependence structures between the two processes by using different copula functions while keeping all other aspects of the model fixed. For instance, under the usual normality assumption for b_i, parameterization (2.3) is a special case of (2.4) with $C(\cdot)$ being the Gaussian copula with a restricted correlation matrix assuming $\text{corr}(b_{yi}, b_{ti}) = \pm 1$ depending on the sign of α under (2.3), and Gaussian marginals $F_y(\cdot)$ and $F_t(\cdot)$. In this example, $b_{ti} = \alpha b_{yi}$; that is, α^2 is merely a rescaling factor for the variance of b_{yi}.

However, even though the latter parameterization offers increased flexibility for the association structure between the two processes, we should note that shared parameter models, in general, imply a restrictive representation of the marginal joint distribution of $\{\mathcal{Y}_i, T_i\}$. To see this, consider the following simple but instructive example. Assume no censoring and moreover that all processes involved, namely

$\{\mathcal{Y}_i \mid b_{yi}\}$, $\{\log T_i \mid b_{ti}\}$ and $\{b_{yi}, b_{ti}\}$ follow normal distributions. Then the covariance for the marginal distribution of $\{\mathcal{Y}_i, \log T_i\}$ is of the form $V = \tilde{Z} D \tilde{Z}^\top + \Sigma$, where $\tilde{Z} = \text{diag}(Z_y, 1)$, D is the covariance matrix for the joint distribution of $\{b_{yi}, b_{ti}\}$, and Σ is the residual covariance matrix for the joint distribution of $\{\mathcal{Y}_i, \log T_i \mid b_{yi}, b_{ti}\} = \{\mathcal{Y}_i \mid b_{yi}\}\{\log T_i \mid b_{ti}\}$. Clearly, V is of a specific form assuming positive correlation and not a general variance-covariance matrix.

2.3 Random Effects Misspecification

2.3.1 Preamble

In this section, we investigate the effect of misspecifying the random effects distribution in parameter estimators and standard errors under the shared parameter models framework. Unless explicitly stated, we will denote by b_i the set of random effects under both parameterizations (2.3) and (2.4); in the latter case $b_i^\top = (b_{yi}^\top, b_{ti})$. We assume that the true random effects probability density function is $p(b_i)$, whereas the fitted one is $f(b_i; \theta_b)$, where both $p(b_i)$ and $f(b_i; \theta_b)$ are absolutely continuous. Moreover, we assume that there is no $\theta_b \in \Theta_b$ such that $f(b_i; \theta_b) = p(b_i)$, where Θ_b is the parameter space of θ_b. Finally, the conditional models for the longitudinal measurement and event processes, $p(\mathcal{Y}_i \mid b_i; \theta_y)$ and $p(T_i, \delta_i \mid b_i; \theta_t)$, respectively, are assumed correctly specified.

2.3.2 Parameter Estimators

We will distinguish between two sets of parameters, namely $\theta_{yt}^\top = (\theta_y^\top, \theta_t^\top)$ and θ_b.

Theorem 1 *For fixed sample size n and as the number n_i of repeated measurements per individual in the longitudinal process $\mathcal{Y}_i$ increases, the maximum likelihood estimator $\tilde{\theta}_{yt}$ under $f(b_i; \theta_b)$ converges to the maximum likelihood estimator $\hat{\theta}_{yt}$ under the correct model $p(b_i)$.*

The proof and the formal conditions under which Theorem 1 holds can be found in Appendix A. The key to the argument is that, as n_i grows, the longitudinal measurement model $p(\mathcal{Y}_i \mid b_i; \theta_y)$ becomes the dominating part in the posterior distribution of the random effects $p(b_i \mid \mathcal{Y}_i, T_i, \delta_i; \theta)$, implying that the choice between $p(b_i)$ or $f(b_i; \theta_b)$ is of minimal importance. However, the above theorem does not hold for θ_b, and in this case the effect of misspecification will be more prominent. According to White (1982), the maximum likelihood estimator $\tilde{\theta}_b$ will converge in probability to the value θ_b^o that minimizes the Kullback-Leibler distance $\mathcal{D}(p : f; \theta_b) = \int\int p(\mathcal{Y}, T, \delta) \ \log\{p(\mathcal{Y}, T, \delta)/f(\mathcal{Y}, T, \delta; \theta_b)\} \ d\mathcal{Y}dT$.

Two remarks based on the above theorem are worth making. First, in many clinical examples the main interest lies in the degree of the association between the longitudinal measurements and the survival process. As noted in §2.2, in the standard parameterization (2.3) that assumes perfect correlation, the association parameter α is, in fact, a parameter of the survival model or a parameter of the longitudinal model, if (2.3) were written as $\eta_{yi} = X_{yi}\beta + Z_{yi}b_i^*$, $\eta_{ti} = x_{ti}^\top\gamma + b_i$, with $b_i^* = \alpha b_i$. Thus, under Theorem 1, α will be minimally affected by misspecification of the random effects distribution, which explains the empirical results reported by other authors (Wang and Taylor, 2001; Song et al., 2002; Tsiatis and Davidian, 2004). However, under parameterization (2.4), α is a parameter of the copula function, which is a part of the random effects model. Thus, even for large n_i, we may observe some sensitivity in the estimation of α under different choices for $C(\cdot)$. Secondly, a straightforward extension of Theorem 1 shows that θ_y will be unbiasedly estimated, even if the event process model is misspecified. This has a direct impact in the missing data context where shared parameter models are also used to correct for nonignorable dropout (Follmann and Wu, 1995). In particular, if the informative censoring mechanism producing the missing data in the longitudinal process is described by a shared parameter model, then the effect of misspecifying both the survival and random effects models will be minimal as

the number of repeated longitudinal measurements per individual increases.

2.3.3 Standard Errors

As we argued in the previous section, misspecification does not affect consistency. However, the effect of misspecifying the random effects distribution will be more prominent in the estimation of standard errors of $\tilde{\theta}_{yt}$. This becomes more transparent if we examine the form, under model (2.1), of the negative of the inverse Hessian matrix, which we would have used as a consistent estimator of the asymptotic inverse Fisher Information matrix if misspecification had been ignored. In particular, with the notation introduced in Appendix A and for $k, k' = y, t, b$, we let $H_{kk'} = n^{-1} \sum_i \partial L_i^f(\tilde{\theta}_k)/\partial \theta_{k'}$ denote the corresponding block of the Hessian matrix H, where

$$\frac{\partial L_i^f(\tilde{\theta}_k)}{\partial \theta_{k'}} = \begin{cases} E_f[\partial h(\cdot;\tilde{\theta}_k)/\partial \theta_k] + E_f[h(\cdot;\tilde{\theta}_k)\{h(\cdot;\tilde{\theta}_k) - L_i^f(\tilde{\theta}_k)\}^\top], & k' = k \\ E_f[h(\cdot;\tilde{\theta}_k)\{h(\cdot;\tilde{\theta}_{k'}) - L_i^f(\tilde{\theta}_{k'})\}^\top], & k' \neq k, \end{cases} \tag{2.6}$$

with E_f denoting the expectation with respect to the posterior distribution $f(b_i \mid \mathcal{Y}_i, T_i, \delta_i; \theta)$. If we assume that H^{-1} exists, the asymptotic variance matrix of $\tilde{\theta}_{yt}$ under standard likelihood methods has the form $\hat{\text{var}}(\tilde{\theta}_{yt}) = -(H_{yt} - H_{yt,b}H_{bb}^{-1}H_{yt,b}^\top)^{-1}$, where $H_{yt,b}^\top = \{H_{yb}^\top, H_{tb}^\top\}$. The second part in the parenthesis is clearly affected by misspecification. To see this we focus on the H_{yb} block of $H_{yt,b}$, with the results for H_{tb} and H_{bb} following similarly. For H_{yb}, (2.6) can be rewritten as

$$H_{yb} = \frac{1}{n}\sum_{i=1}^{n} E_f \left[\left\{ \sum_{j=1}^{n_i} \frac{\partial}{\partial \theta_y} \log p(y_i(t_{ij})|b_i;\tilde{\theta}_y) \right\} \left\{ \frac{\partial}{\partial \theta_b} \log f(b_i;\tilde{\theta}_b) \right\}^\top \right] - \left\{ L_i^f(\tilde{\theta}_y) \right\} \left\{ L_i^f(\tilde{\theta}_b) \right\}^\top .$$

If we let n_i grow, then $E_f \to E_p$; that is, in the corresponding expectations the true posterior is used. However, note that both parts of H_{yb} still depend on the

misspecified random effects model, since $L_i^p(\theta_b) = \int\{\partial \log f(b_i;\theta_b)/\partial\theta_b\}\ p(b_i \mid \mathcal{Y}_i, T_i;\theta)\ db_i$, which will result in some bias in the standard error estimators.

By standard maximum likelihood theory under misspecification (White, 1982), the asymptotic covariance matrix for $\tilde{\theta}$ is $\text{var}(\tilde{\theta}) = K^{-1}DK^{-1}$, where $K = E\{-H(\theta^*_{yt}, \theta^o_b)\}$, $D = E\{L^f(\theta^*_{yt}, \theta^o_b)L^f(\theta^*_{yt}, \theta^o_b)^\top\}$, and the expectations are taken with respect to the true distribution $p(\mathcal{Y}, T, \delta; \theta^*)$. Using the sandwich variance estimator as a consistent estimator for this covariance matrix, we obtain

$$\tilde{\text{var}}(\tilde{\theta}_{yt}) = \mathcal{A}X\mathcal{A} + 2\mathcal{B}Z^\top\mathcal{A} + \mathcal{B}W\mathcal{B}^\top, \tag{2.7}$$

where $\mathcal{A} = \hat{\text{var}}(\tilde{\theta}_{yt}) = -(H_{yt} - H_{yt,b}H_{bb}^{-1}H_{yt,b}^\top)^{-1}$, $\mathcal{B} = -H_{yt}^{-1}H_{yt,b}(H_b - H_{yt,b}^\top H_{yt}^{-1}$ $H_{yt,b})^{-1}$, $\mathcal{X} = n^{-1}\sum_i L_i^f(\tilde{\theta}_{yt})L_i^f(\tilde{\theta}_{yt})^\top$, $\mathcal{Z} = n^{-1}\sum_i L_i^f(\tilde{\theta}_{yt})L_i^f(\tilde{\theta}_b)^\top$ and $\mathcal{W} = n^{-1}\sum_i L_i^f(\tilde{\theta}_b)\ L_i^f(\tilde{\theta}_b)^\top$. Note that, as $n \to \infty$ and if the correct model $p(b_i)$ had been used, then $\|\tilde{\text{var}}(\tilde{\theta}_{yt}) - \hat{\text{var}}(\tilde{\theta}_{yt})\| \to 0$. Straightforward algebra then implies that $\text{diag}(\mathcal{A}X + 2\mathcal{B}Z^\top + \mathcal{B}W\mathcal{B}^\top A^{-1}) - 1$ quantifies the extra variance for $\tilde{\theta}_{yt}$ owing to misspecification.

2.4 A Simulation Study

2.4.1 Study Set-up

A small simulation study was performed to corroborate empirically the arguments unfolded in §2.3. Since the case of one random effect has been extensively studied in the literature, for example by Song et al. (2002), here we investigate the effect of misspecifying the random effects distribution in parameter estimates and standard errors for case in (2.5) of two random effects. The study considers a two-group comparison with $n = 200$. In particular, for the linear mixed effects model in (2.3), we assume measurement error variance $\sigma_y^2 = 0{\cdot}5^2$, and a linear predictor given by $\eta_{yi} = (\beta_0 + b_{yi}) + \beta_1\mathcal{T}_i + \beta_2 t_{ij} + \beta_3 t_{ij}^2 + \beta_4\mathcal{T}_i t_{ij} + \beta_5\mathcal{T}_i t_{ij}^2$,

where b_{yi} denotes a random-intercepts term, $\mathcal{T}_i$ is the binary treatment indicator and $(\beta_0, \ldots, \beta_5)^\top = (1, 0, 1{\cdot}5, 2{\cdot}5, -0{\cdot}5, -1)$. For the survival model in (2.3), we assume that $\mathcal{P}$ follows the extreme value distribution with scale parameter $\sigma_t = 0{\cdot}5$, and the linear predictor given by $\eta_{ti} = (\gamma_0 + b_{ti}) + \gamma_1 \mathcal{T}_i$, where b_{ti} is a frailty term and $(\gamma_0, \gamma_1)^\top = (2, 1{\cdot}5)$. The censoring mechanism follows an exponential distribution with mean 20, resulting in about 50% censoring, and the visiting times t_{ij} are random. For the random effects model in (2.4) the following three scenarios are considered: (i) a bimodal mixture distribution $0{\cdot}45 \times \mathcal{N}\{(-2, -2{\cdot}1)^\top, \Sigma\} + 0{\cdot}55 \times \mathcal{N}\{(1{\cdot}636, 1{\cdot}718)^\top, \Sigma\}$, with $\Sigma = \text{vech}(1{\cdot}5^2, 1^2, 0{\cdot}5)$, where, in $\text{vech}(s_1^2, s_2^2, \rho)$, s_1^2 and s_2^2 denote the two variances and ρ the correlation of the covariance matrix Σ; (ii) a unimodal skewed mixture distribution $0{\cdot}7 \times \mathcal{N}\{(1{\cdot}3, 0{\cdot}9)^\top, \Sigma\} + 0{\cdot}3 \times \mathcal{N}\{(-3{\cdot}033, 2{\cdot}1)^\top, \Sigma\}$, with $\Sigma = \text{vech}(1{\cdot}6^2, 1{\cdot}7^2, 0{\cdot}7)$; and (iii) a normal distribution $\mathcal{N}(0, \Sigma)$, with $\Sigma = \text{vech}(2{\cdot}5^2, 2{\cdot}2^2, 0{\cdot}82)$. The parameter values have been chosen such that the variances σ_{by}^2 and σ_{bt}^2, and the degree of association in terms of Kendall's τ for the random effects, are of the same magnitude for all three scenarios. For n_i, two cases are considered, namely the large-n_i case, in which $\max_i(n_i) = 15$ with 10 measurements per subject on average, and the small-n_i case, in which $\max_i(n_i) = 4$ with 2·5 measurements per subject on average. Finally, for each scenario and for each choice of n_i, 100 data sets are simulated.

2.4.2 Fitted Models

For each simulated data set four joint models are fitted. In particular, for the longitudinal $p(\mathcal{Y}_i \mid b_{yi}; \theta_y)$ and the survival $p(T_i, \delta_i \mid b_{ti}; \theta_t)$ processes the correct models are assumed, whereas for the random effects model four copulas are considered, namely the Frank, Gumbel, Gaussian and Student's-t_4 copulas, with normal marginals. Thus, under scenarios (i) and (ii) all fitted models are misspecified, whereas for scenario (iii) the normal copula random effects model corresponds to the true

joint model from which we simulated. Furthermore, the quality of the model-based standard errors $\hat{\text{var}}(\tilde{\theta})$, i.e., assuming the random effects distribution had been correctly specified, and the sandwich-estimator standard errors $\tilde{\text{var}}(\tilde{\theta})$ is compared to the empirical standard errors given by $\left\{\sum_{m=1}^{M}(\hat{\theta}_m - \overline{\theta})^2/(M-1)\right\}^{1/2}$, where $\hat{\theta}_m$ denotes the maximum likelihood estimates in the mth simulated dataset, $\overline{\theta} = \sum_{m=1}^{M} \hat{\theta}_m/M$ and $M = 100$. The models are fitted using an EM algorithm in which the random effects are treated as missing values; more details can be found in Appendix B. All computations have been performed in R (R Development Core Team, 2007).

2.4.3 Results

Tables 2.1, 2.2 and 2.3 present the results, under scenarios (i), (ii) and (iii), respectively. For all scenarios we observe that the parameter estimators for the longitudinal and survival models are rather robust to random effects misspecification. In contrast, the estimators for the random effects model, and especially the estimator of the association parameter, show greater sensitivity regarding the choice of the copula $C(\cdot)$. Furthermore, the small-n_i case yielded relatively more sensitive results for the estimators, in accordance with Theorem 1. An interesting feature is that the Gaussian copula performed rather well under misspecification. This feature can be explained by the concept of local dependence introduced by Holland and Wang (1987). The local dependence function equals $\partial^2 \log p(b_{yi}, b_{ti})/\partial b_{yi}\partial b_{ti}$ and is used to quantify dependence when both the degree and the direction of the dependence are different in different regions of the plane (Jones, 1996). A numerical comparison between the values of the local dependence function of the true random effects densities under scenarios (i) and (ii), and the corresponding values of the assumed copulas, reveals that the Gaussian copula is on average closer to the true densities than the other copulas. Finally, regarding the estimation of standard errors, we observe that the average of the sandwich estimators is closer

to the empirical standard errors than the corresponding model-based ones, with the exception of scenario (iii) under the Gaussian copula, where the model-based standard errors, as expected, show good behaviour.

2.5 Application

In this section, we present the analysis of data coming from a longitudinal study on patients who received a kidney transplant. The main scientific focus lies in the time a patient can maintain the new graft. In this case, a good marker for the kidneys' performance is the level of serum creatinine in blood. However, since the observed levels of this marker are directly influenced by a person's muscle activity, the glomerular filtration rate, GFR, is typically used, which is an inverse function of serum creatinine correcting also for sex, weight and age.

During the 10 year follow-up period GFR measurements are regularly taken and our aim here is to explore the association structure between longitudinal GFR measurements and the time to graft failure. Out of the 432 patients, 91 (21·1%) experienced the event; moreover, patients made on average 72 visits, with a standard deviation 22·4 visits, resulting in a total of 31062 records. Based on descriptive measures and plots we adopted the following models for the two processes. For the longitudinal process a linear random-intercepts model is assumed with fixed-effects quadratic time trends for the first 6 months, followed by linear time trends for the remaining follow-up period. For the survival process we include the age, weight and sex as main effects, and a frailty term related to the random-intercept term of the measurement model.

To investigate the influence of parametric assumptions on the size of the association between the two processes, we performed a sensitivity analysis under different copula functions and assuming normal marginals for the joint distribution of the random effects, and different survival distributions. In particular, we con-

Table 2.1: *Simulation study. Parameter estimates and standard errors under the bimodal scenario (i). The top part contains the results for the large-n_i case, with $\max_i(n_i) = 15$, whereas the bottom part contains the results for the small-n_i case, with $\max_i(n_i) = 4$. The 'Std. Err.' columns contain the empirical/average sandwich/average model-based standard errors, respectively.*

		Frank		Gumbel		Gaussian		Student's-t_4	
	True	Mean	Std. Err.	Mean	Std. Err.	Mean	Std. Err.	Mean	Std. Err.
β_0	1	0·979	0·141/0·111/0·081	0·966	0·146/0·122/0·089	0·982	0·145/0·136/0·098	0·980	0·146/0·133/0·093
β_1	0	0·038	0·215/0·191/0·174	0·038	0·206/0·183/0·107	0·029	0·201/0·194/0·155	0·024	0·216/0·191/0·166
β_2	1·5	1·502	0·022/0·023/0·014	1·503	0·022/0·023/0·018	1·502	0·022/0·023/0·013	1·502	0·022/0·024/0·014
β_3	2·5	2·499	0·002/0·002/0·001	2·500	0·002/0·002/0·001	2·500	0·002/0·002/0·001	2·500	0·002/0·002/0·001
β_4	-0·5	-0·502	0·031/0·030/0·020	-0·503	0·031/0·030/0·022	-0·503	0·031/0·030/0·019	-0·502	0·031/0·031/0·018
β_5	-1	-1·000	0·002/0·002/0·001	-0·999	0·002/0·002/0·001	-1·001	0·002/0·002/0·001	-1·000	0·002/0·002/0·001
σ_y	0·5	0·498	0·013/0·010/0·008	0·497	0·015/0·017/0·009	0·499	0·013/0·016/0·009	0·498	0·014/0·013/0·009
γ_0	2	1·978	0·210/0·198/0·188	1·992	0·209/0·199/0·172	2·005	0·201/0·191/0·172	1·982	0·201/0·203/0·176
γ_1	1·5	1·496	0·272/0·264/0·167	1·510	0·260/0·242/0·188	1·492	0·269/0·272/0·196	1·492	0·290/0·273/0·198
σ_t	0·5	0·418	0·199/0·188/0·173	0·610	0·308/0·313/0·266	0·429	0·218/0·190/0·150	0·669	0·210/0·207/0·183
τ	0·62	0·619	0·075/0·065/0·036	0·724	0·117/0·108/0·098	0·634	0·067/0·059/0·042	0·693	0·098/0·100/0·068
σ_{by}	2·35	2·305	0·104/0·108/0·025	2·319	0·140/0·129/0·099	2·349	0·105/0·110/0·078	2·372	0·116/0·117/0·073
σ_{bt}	2·15	2·172	0·143/0·145/0·136	2·141	0·199/0·172/0·145	2·205	0·164/0·146/0·101	2·121	0·197/0·192/0·151
β_0	1	0·952	0·195/0·136/0·134	0·905	0·259/0·130/0·131	0·964	0·210/0·192/0·104	0·941	0·214/0.183/0.120
β_1	0	0·083	0·290/0·186/0·170	0·021	0·339/0·257/0·190	0·060	0·309/0·287/0·193	0·053	0·308/0.259/0.189
β_2	1·5	1·505	0·029/0·029/0·006	1·508	0·029/0·029/0·011	1·506	0·029/0·029/0·009	1·505	0·029/0.029/0.006
β_3	2·5	2·500	0·002/0·003/0·001	2·500	0·003/0·003/0·001	2·500	0·002/0·001/0·003	2·500	0·003/0.002/0.001
β_4	-0·5	-0·505	0·035/0·038/0·008	-0·507	0·035/0·038/0·012	-0·506	0·035/0·038/0·012	-0·506	0·035/0.038/0.008
β_5	-1	-1·000	0·003/0·003/0·001	-0·999	0·003/0·003/0·001	-1·000	0·003/0·003/0·003	-1·000	0·003/0.003/0.001
σ_y	0·5	0·497	0·023/0·015/0·225	0·498	0·023/0·023/0·015	0·497	0·022/0·019/0·015	0·497	0·022/0.015/0.008
γ_0	2	2·032	0·215/0·139/0·061	1·989	0·277/0·156/0·087	2·059	0·251/0·185/0·097	2·036	0·256/0.196/0.090
γ_1	1·5	1·511	0·321/0·201/0·080	1·479	0·378/0·219/0·091	1·495	0·348/0·219/0·091	1·473	0·354/0.344/0.272
σ_t	0·5	0·327	0·128/0·095/0·033	0·580	0·358/0·124/0·092	0·411	0·203/0·174/0·122	0·615	0·285/0.254/0.174
τ	0·62	0·663	0·045/0·039/0·013	0·758	0·131/0·107/0·067	0·660	0·045/0·056/0·012	0·727	0·106/0.095/0.046
σ_{by}	2·35	2·294	0·107/0·114/0·037	2·266	0·215/0·196/0·116	2·338	0·108/0·120/0·033	2·408	0·222/0.201/0.171
σ_{bt}	2·15	2·298	0·154/0·162/0·025	2·191	0·239/0·210/0·173	2·279	0·179/0·171/0·055	2·256	0·295/0.265/0.191

Table 2.2: *Simulation study. Parameter estimates and standard errors under the skewed scenario (ii). The top part contains the results for the large-n_i case, with $\max_i(n_i) = 15$, whereas the bottom part contains the results for the small-n_i case, with $\max_i(n_i) = 4$. The 'Std. Err.' columns contain the empirical/average sandwich/average model-based standard errors, respectively.*

		Frank		Gumbel		Gaussian		Student's-t_4	
	True	Mean	Std. Err.	Mean	Std. Err.	Mean	Std. Err.	Mean	Std. Err.
β_0	1	0·992	0·169/0·151/0·142	0·993	0·171/0·166/0·156	0·995	0·173/0·161/0·155	0·995	0·177/0·167/0·118
β_1	0	-0·014	0·256/0·210/0·199	-0·009	0·260/0·231/0·220	-0·024	0·269/0·272/0·251	-0·022	0·269/0·256/0·193
β_2	1·5	1·501	0·022/0·021/0·017	1·502	0·022/0·022/0·011	1·501	0·022/0·021/0·016	1·502	0·022/0·022/0·011
β_3	2·5	2·499	0·002/0·002/0·001	2·500	0·002/0·002/0·001	2·500	0·002/0·002/0·001	2·500	0·002/0·002/0·001
β_4	-0·5	-0·499	0·029/0·028/0·019	-0·500	0·029/0·028/0·015	-0·500	0·029/0·028/0·015	-0·501	0·030/0·028/0·013
β_5	-1	-1·000	0·002/0·002/0·001	-1·000	0·002/0·002/0·001	-1·000	0·002/0·001/0·001	-1·000	0·002/0·002/0·001
σ_y	0·5	0·504	0·013/0·010/0·007	0·506	0·021/0·018/0·010	0·506	0·021/0·019/0·010	0·506	0·021/0·018/0·012
γ_0	2	1·947	0·226/0·235/0·195	1·980	0·222/0·207/0·174	2·028	0·238/0·219/0·178	2·004	0·222/0·209/0·188
γ_1	1·5	1·502	0·326/0·308/0·268	1·518	0·330/0·336/0·273	1·506	0·352/0·326/0·296	1·501	0·374/0·354/0·308
σ_t	0·5	0·582	0·260/0·269/0·203	0·574	0·282/0·257/0·199	0·471	0·226/0·203/0·177	0·589	0·244/0·223/0·197
τ	0·63	0·654	0·103/0·101/0·085	0·684	0·093/0·100/0·077	0·626	0·060/0·055/0·033	0·661	0·082/0·065/0·051
σ_{by}	2·56	2·549	0·109/0·107/0·072	2·523	0·121/0·131/0·067	2·527	0·108/0·118/0·083	2·530	0·107/0·117/0·088
σ_{bt}	2·19	2·105	0·218/0·221/0·179	2·110	0·209/0·196/0·149	2·159	0·205/0·193/0·127	2·092	0·222/0·202/0·176
β_0	1	0·981	0·272/0·213/0·190	1·027	0·249/0·189/0·154	1·022	0·251/0·199/0·178	1·004	0·260/0·213/0·143
β_1	0	0·006	0·314/0·295/0·269	-0·006	0·327/0·295/0·191	0·003	0·329/0·310/0·299	0·018	0·337/0·298/0·234
β_2	1·5	1·502	0·028/0·029/0·006	1·502	0·028/0·029/0·007	1·502	0·028/0·029/0·006	1·502	0·028/0·029/0·007
β_3	2·5	2·500	0·003/0·003/0·001	2·500	0·003/0·003/0·001	2·500	0·003/0·003/0·001	2·500	0·003/0·003/0·001
β_4	-0·5	-0·504	0·036/0·038/0·008	-0·504	0·036/0·038/0·008	-0·504	0·036/0·038/0·007	-0·504	0·036/0·038/0·008
β_5	-1	-1·000	0·003/0·003/0·001	-0·999	0·003/0·003/0·001	-1·000	0·003/0·003/0·001	-1·000	0·003/0·003/0·001
σ_y	0·5	0·496	0·021/0·015/0·009	0·496	0·021/0·015/0·008	0·496	0·021/0·015/0·009	0·496	0·021/0·017/0·009
γ_0	2	1·937	0·233/0·191/0·112	1·996	0·204/0·196/0·101	2·022	0·228/0·184/0·088	1·983	0·225/0·197/0·123
γ_1	1·5	1·539	0·320/0·299/0·111	1·511	0·317/0·297/0·133	1·544	0·332/0·310/0·185	1·557	0·352/0·294/0·193
σ_t	0·5	0·552	0·290/0·244/0·096	0·535	0·271/0·178/0·088	0·431	0·203/0·163/0·082	0·628	0·246/0·191/0·091
τ	0·63	0·687	0·105/0·089/0·021	0·709	0·081/0·068/0·014	0·649	0·048/0·035/0·012	0·672	0·081/0·063/0·013
σ_{by}	2·56	2·570	0·138/0·117/0·057	2·543	0·134/0·131/0·040	2·539	0·123/0·130/0·034	2·545	0·124/0·130/0·041
σ_{bt}	2·19	2·180	0·213/0·179/0·085	2·159	0·190/0·160/0·024	2·203	0·177/0·173/0·076	2·098	0·202/0·196/0·088

Table 2.3: *Simulation study. Parameter estimates and standard errors under the normal scenario (iii). The top part contains the results for the large-n_i case, with $\max_i(n_i) = 15$, whereas the bottom part contains the results for the small-n_i case, with $\max_i(n_i) = 4$. The 'Std. Err.' columns contain the empirical/average sandwich/average model-based standard errors, respectively.*

		Frank		Gumbel		Gaussian		Student's-t_4	
	True	Mean	Std. Err.	Mean	Std. Err.	Mean	Std. Err.	Mean	Std. Err.
β_0	1	1·056	0·215/0·199/0·118	1·055	0·216/0·194/0·110	1·004	0·212/0·190/0·223	1·066	0·214/0·193/0·163
β_1	0	0·077	0·220/0·113/0·089	0·073	0·225/0·175/0·102	0·022	0·215/0·222/0·210	0·062	0·230/0·198/0·120
β_2	1·5	1·498	0·019/0·024/0·014	1·498	0·019/0·024/0·017	1·497	0·018/0·026/0·019	1·498	0·018/0·021/0·014
β_3	2·5	2·500	0·001/0·002/0·001	2·500	0·001/0·002/0·001	2·500	0·002/0·002/0·002	2·500	0·002/0·002/0·001
β_4	-0·5	-0·496	0·024/0·031/0·015	-0·496	0·024/0·028/0·013	-0·496	0·024/0·032/0·025	-0·497	0·024/0·030/0·015
β_5	-1	-1·000	0·002/0·002/0·002	-1·000	0·002/0·002/0·002	-1·000	0·002/0·003/0·002	-1·000	0·002/0·002/0·001
σ_y	0·5	0·501	0·013/0·011/0·007	0·501	0·014/0·017/0·008	0·501	0·014/0·019/0·011	0·501	0·013/0·012/0·009
γ_0	2	2·000	0·246/0·226/0·193	2·028	0·271/0·236/0·189	2·034	0·264/0·279/0·257	2·026	0·275/0·202/0·172
γ_1	1·5	1·528	0·438/0·374/0·253	1·542	0·464/0·423/0·396	1·506	0·448/0·460/0·438	1·538	0·497/0·461/0·395
σ_t	0·5	0·682	0·301/0·282/0·220	0·655	0·340/0·335/0·271	0·527	0·255/0·234/0·228	0·641	0·259/0·233/0·196
τ	0·61	0·712	0·124/0·096/0·054	0·730	0·130/0·116/0·096	0·610	0·036/0·040/0·033	0·676	0·114/0·127/0·085
σ_{by}	2·5	2·519	0·135/0·123/0·094	2·486	0·131/0·137/0·109	2·487	0·127/0·132/0·126	2·487	0·128/0·121/0·099
σ_{bt}	2·2	2·049	0·184/0·158/0·112	2·103	0·159/0·146/0·117	2·174	0·160/0·215/0·148	2·098	0·191/0·198/0·138
β_0	1	0·966	0·344/0·331/0·293	0·997	0·332/0·303/0·343	1·043	0·323/0·353/0·344	0·994	0·323/0·335/0·216
β_1	0	0·028	0·321/0·280/0·149	-0·017	0·293/0·254/0·190	0·098	0·311/0·373/0·354	0·098	0·303/0·285/0·196
β_2	1·5	1·504	0·029/0·030/0·005	1·505	0·029/0·030/0·016	1·505	0·029/0·030/0·026	1·505	0·029/0·030/0·007
β_3	2·5	2·500	0·003/0·003/0·001	2·500	0·003/0·003/0·002	2·500	0·003/0·003/0·001	2·500	0·003/0·003/0·001
β_4	-0·5	-0·506	0·037/0·038/0·007	-0·507	0·037/0·038/0·011	-0·508	0·037/0·045/0·038	-0·508	0·037/0·038/0·008
β_5	-1	-0·999	0·003/0·003/0·001	-0·999	0·003/0·003/0·001	-0·999	0·003/0·003/0·001	-0·999	0·003/0·003/0·001
σ_y	0·5	0·495	0·019/0·016/0·009	0·495	0·019/0·016/0·010	0·495	0·019/0·023/0·017	0·495	0·019/0·015/0·010
γ_0	2	1·976	0·250/0·215/0·165	2·007	0·249/0·216/0·136	2·024	0·252/0·291/0·235	1·994	0·242/0·222/0·164
γ_1	1·5	1·580	0·294/0·236/0·142	1·534	0·290/0·247/0·197	1·572	0·300/0·283/0·293	1·551	0·289/0·253/0·195
σ_t	0·5	0·494	0·263/0·236/0·123	0·502	0·280/0·258/0·213	0·424	0·197/0·250/0·207	0·559	0·245/0·246/0·183
τ	0·61	0·644	0·094/0·071/0·027	0·666	0·091/0·075/0·048	0·604	0·064/0·082/0·061	0·642	0·074/0·058/0·019
σ_{by}	2·5	2·515	0·147/0·126/0·091	2·478	0·158/0·146/0·055	2·482	0·136/0·167/0·159	2·492	0·135/0·127/0·092
σ_{bt}	2·2	2·205	0·223/0·189/0·109	2·178	0·199/0·171/0·124	2·156	0·189/0·198/0·179	2·149	0·207/0·199/0·121

sidered the Frank, Gumbel, Gaussian, and Student's-t_4 copulas, and the Weibull, log-normal and log-logistic distributions as survival distributions. The estimates of Kendall's τ for each scenario are presented in Table 2.4. For the entire analysis we observed results similar to those in §2.4. In particular, the main effects for both the linear mixed and survival models were minimally affected by different assumptions regarding the random effects, whereas the degree of the association between the two processes was influenced to a much larger extent by the choice of the copula function. The results suggest a moderate positive association between the underlying latent processes, ranging from 0·56 to 0·86. However, note that this is far from the perfect correlation that the common parameterization (2.3) assumes.

2.6 Discussion

As the number n_i of repeated longitudinal measurements per individual increases, the effect of misspecifying the random effects distribution became minimal for certain estimators. However, estimation of the standard errors under the misspecified model will generally be affected, and thus the use of the sandwich estimator is recommended. How large n_i has to be depends on the type of information for b_i that is included in $\mathcal{Y}_i$. In particular, we expect that for linear mixed models, smaller values of n_i will suffice, as opposed to generalized or nonlinear mixed models, because in the former case $\log p(\mathcal{Y}_i \mid b_i; \theta_y)$ will be quadratic in b_i, which implies that convergence of $p(b_i \mid \mathcal{Y}_i, T_i, \delta_i; \theta)$ to a normal distribution will be faster. In addition, note that Theorem 1 requires all subjects to have a relatively large number of repeated measurements. Thus, if some groups of subjects have very few measurements, perhaps because of dropout, choosing the correct random effects distribution will be important. Moreover, our results are based on the assumption that both $p(b_i)$ and $f(b_i; \theta_b)$ are continuous densities, excluding the case in which

Table 2.4: *Kidney transplant data. Estimated Kendall's τ with sandwich/model-based standard errors in parenthesis for the association between time to graft failure longitudinal measurements of glomerular filtration rate under different copulas and survival models.*

Model	Frank	Gumbel	Gaussian	Student's-t_4
Weibull	0·569 (0·091/0·062)	0·803 (0·051/0·021)	0·855 (0·022/0·011)	0·657 (0·068/0·030)
log-normal	0·564 (0·103/0·064)	0·802 (0·062/0·022)	0·629 (0·042/0·019)	0·747 (0·066/0·026)
log-logistic	0·566 (0·088/0·066)	0·802 (0·048/0·022)	0·747 (0·075/0·040)	0·591 (0·071/0·031)

the true random effects distribution is discrete, with few support points. In that setting we would expect the robustness of $\tilde{\theta}_{yt}$ to be seriously affected.

Moreover, the formulation of the shared parameter model presented in §2.2 assumed a noninformative visiting process, which enabled an easier likelihood construction. However, in cases where such an assumption is erroneous, ignoring the visiting process may influence results considerably since each subject will have n_i measurement occasions, leading to a multivariate model. Thus, the posterior distribution of the random effects will then depend heavily on both the longitudinal and visit process models.

Finally, we have assumed that the parameter space of the survival model is of finite dimension. This excludes the commonly used semiparametric framework in which the baseline hazard is left unspecified. Extensions of the results presented here for this case are under consideration.

2.7 Appendix A

Let $p(\mathcal{Y}, T, \delta; \theta)$ and $f(\mathcal{Y}, T, \delta; \theta)$ denote the marginal densities under the correctly specified $p(b)$ and the misspecified $f(b; \theta_b)$ random effects distributions, respectively. First, we work under parameterization (2.3) with a common random effect b_i for the two processes. We make the following assumptions: (i) both $p(\mathcal{Y}, T, \delta; \theta)$ and $f(\mathcal{Y}, T, \delta; \theta)$ are well-defined densities under the usual regularity conditions (Cox and Hinkley, 1974, p. 281); (ii) for fixed n we define the log likelihood functions $\ell_n^p(\theta) = n^{-1} \sum_{i=1}^n \log p(\mathcal{Y}_i, T_i, \delta_i; \theta)$, $\ell_n^f(\theta) = n^{-1} \sum_{i=1}^n \log f(\mathcal{Y}_i, T_i, \delta_i; \theta)$, and in addition, we assume that $\ell_n^p(\theta)$ and $\ell_n^f(\theta)$ have unique maxima at $\hat{\theta}_{yt} \in \Theta_{yt}$ and $\tilde{\theta}_{yt} \in \Theta_{yt}$, respectively, with θ_y and θ_t having disjoint parameter spaces, i.e., $\Theta_{yt} = \Theta_y \times \Theta_t$; (iii) for the score vectors $L_n^p(\theta) = \partial \ell_n^p(\theta)/\partial\theta$, and $L_n^f(\theta) = \partial \ell_n^f(\theta)/\partial\theta$ we assume that the required conditions hold which allow differentiation to be taken inside the integral sign; (iv) finally, we assume that both $\log p(b_i)$ and

$\log f(b_i;\theta_b)$ are bounded and smooth functions of b_i around the neighbourhood of the mode $\hat{b}_i$ of $\log p(\mathcal{Y}_i \mid b_i;\theta_y) = \sum_{j=1}^{n_i} \log p\{y_i(t_{ij}) \mid b_i;\theta_y\}$.

Next we note that, under assumption (iii), the score vector takes the form

$$\begin{aligned} L_n^p(\theta) &= \sum_{i=1}^{n} \frac{\partial}{\partial\theta} \log \int p(\mathcal{Y}_i \mid b_i;\theta_y)\; p(T_i,\delta_i \mid b_i;\theta_t)\; p(b_i)\; db_i \\ &= \sum_{i=1}^{n} \int h(\cdot;\theta)\; p(b_i \mid \mathcal{Y}_i, T_i, \delta_i;\theta)\; db_i, \end{aligned} \tag{2.8}$$

where $h(\cdot;\theta)$ denotes the corresponding score vector of each of the submodels; for instance, for the measurement process, $L_n^p(\theta_y)$ requires $h(\cdot;\theta) = \partial \log p(\mathcal{Y}_i \mid b_i;\theta_y)/\partial\theta_y$. Analogously, the misspecified score vector takes the form

$$L_n^f(\theta) = \sum_{i=1}^{n} \int h(\cdot;\theta)\; f(b_i \mid \mathcal{Y}_i, T_i, \delta_i;\theta)\; db_i. \tag{2.9}$$

Equation (2.9) differs from (2.8) in that $f(b_i \mid \mathcal{Y}_i, T_i, \delta_i;\theta)$ is the posterior under $f(b_i;\theta_b)$, but also that for $L_n^f(\theta_b)$, $h(\cdot;\theta) = \partial \log f(b_i;\theta_b)/\partial\theta_b$. For fixed n, the score vectors are functions of the number of repeated measurements n_i. Henceforth we assume that, for all i, $n_i \to \infty$. Then, under assumptions (i) and (iv) both posterior distributions $p_{n_i}(b_i \mid \mathcal{Y}_i, T_i, \delta_i;\theta)$ and $f_{n_i}(b_i \mid \mathcal{Y}_i, T_i, \delta_i;\theta)$ will be concentrated in the neighbourhood of the mode $\hat{b}_i$ of the correctly specified longitudinal model $\log p(\mathcal{Y}_i \mid b_i;\theta_y) = \sum_{j=1}^{n_i} \log p\{y_i(t_{ij}) \mid b_i;\theta_y\}$ (Cox and Hinkley, 1974, pp. 399–400), which implies that, as $n_i \to \infty$, $|f_{n_i}(b_i \mid \mathcal{Y}_i, T_i, \delta_i;\theta) - p_{n_i}(b_i \mid \mathcal{Y}_i, T_i, \delta_i;\theta)| \to 0$. Based on this result, we have that for every $\epsilon > 0$ there exists an integer m such that for all $n_i \geq m$ and for all $\theta_{yt} \in \Theta_{yt}$ we obtain

$$
\begin{aligned}
&\|L^f_{n_i}(\theta_{yt}) - L^p_{n_i}(\theta_{yt})\| \\
&= \|\sum_i \int h(\cdot;\theta_{yt}) \{f_{n_i}(b_i \mid \mathcal{Y}_i, T_i, \delta_i; \theta) - p_{n_i}(b_i \mid \mathcal{Y}_i, T_i, \delta_i; \theta)\}\ db_i\| \\
&\leq \sum_i \int \|h(\cdot;\theta_{yt}) \{f_{n_i}(b_i \mid \mathcal{Y}_i, T_i, \delta_i; \theta) - p_{n_i}(b_i \mid \mathcal{Y}_i, T_i, \delta_i; \theta)\}\| \ db_i \\
&\leq \sum_i \int \|h(\cdot;\theta_{yt})\| \cdot \mid f_{n_i}(b_i \mid \mathcal{Y}_i, T_i, \delta_i; \theta) - p_{n_i}(b_i \mid \mathcal{Y}_i, T_i, \delta_i; \theta) \mid \ db_i < \epsilon,
\end{aligned}
$$

where $\|\cdot\|$ denotes the Euclidean vector norm. The last statement combined with the application of the mean value theorem to either $L^f_{n_i}(\theta_{yt})$ or $L^p_{n_i}(\theta_{yt})$ implies that $\|\tilde{\theta}_{yt} - \hat{\theta}_{yt}\| \to 0$.

Under the two random effects parameterization (2.4), the arguments raised above can be adapted accordingly to show that $\tilde{\theta}_y$ will converge to $\hat{\theta}_y$. However, for $\tilde{\theta}_t$ we have that

$$
\begin{aligned}
L^f_n(\theta_t) &= \sum_{i=1}^n \iint \frac{\partial}{\partial \theta_t} \log p(T_i, \delta_i \mid b_{ti}; \theta_t)\ f(b_{yi}, b_{ti} \mid \mathcal{Y}_i, T_i, \delta_i; \theta)\ db_{yi} db_{ti} \\
&= \sum_{i=1}^n \int \frac{\partial}{\partial \theta_t} \log p(T_i, \delta_i \mid b_{ti}; \theta_t)\ f(b_{ti} \mid \mathcal{Y}_i, T_i, \delta_i; \theta)\ db_{ti},
\end{aligned}
$$

where $f(b_{ti} \mid \mathcal{Y}_i, T_i, \delta_i; \theta) \propto f(\mathcal{Y}_i \mid b_{ti}; \theta_y, \theta_b) p(T_i, \delta_i \mid b_{ti}; \theta_t) f(b_{ti}; \theta_b)$, with $f(\mathcal{Y}_i \mid b_{ti}; \theta_y, \theta_b) = \int p(\mathcal{Y}_i \mid b_{yi}; \theta_y) f(b_{yi} \mid b_{ti}; \theta_b) db_{yi}$. Heuristically, as long as $f(b_{yi} \mid b_{ti}; \theta_b) \neq f(b_{yi}; \theta_b)$, then as n_i increases the $p(\mathcal{Y}_i \mid b_{yi}; \theta_y)$ part of $f(\mathcal{Y}_i \mid b_{ti}; \theta_y, \theta_b)$ provides increasing information regarding b_{ti}, and this information becomes greater as the association between b_{yi} and b_{ti} gets stronger. Formally, note that, for $\tilde{\theta}_t$ to converge to $\hat{\theta}_t$ as $n_i \to \infty$, we require that $|f_{n_i}(b_{ti} \mid \mathcal{Y}_i, T_i, \delta_i; \theta) - p_{n_i}(b_{ti} \mid \mathcal{Y}_i, T_i, \delta_i; \theta)| \to 0$. This would be the case so long as $f_{n_i}(b_{ti} \mid \mathcal{Y}_i, T_i, \delta_i; \theta)$ is concentrated in the neighbourhood of the mode $\hat{b}_{ti}$ of $f_{n_i}(\mathcal{Y}_i \mid b_{ti}; \theta_y, \theta_b)$. To ensure this, we also adopt the regularity conditions of Heyde and Johnstone (1979), under

which both $f_{n_i}(b_{ti} \mid \mathcal{Y}_i, T_i, \delta_i; \theta)$ and $p_{n_i}(b_{ti} \mid \mathcal{Y}_i, T_i, \delta_i; \theta)$ will converge to the same normal distribution with mean $\hat{b}_{ti}$, even though conditional independence does not hold in this case, i.e., $f(\mathcal{Y}_i \mid b_{ti}; \theta_y, \theta_b) \neq \prod_{j=1}^{n_i} f\{y_i(t_{ij}) \mid b_{ti}; \theta_y, \theta_b\}$.

Finally, to make the above results probabilistic in nature and to ensure that $\tilde{\theta}_{yt}$ converges in probability to the true parameter vector θ^*_{yt}, we assume that $n \to \infty$ and that $\hat{\theta}_{yt}$ is a consistent estimator of θ^*_{yt}.

2.8 Appendix B

The maximum likelihood estimates for the parameter vector θ are obtained using an EM algorithm, where b_{yi} and b_{ti} are treated as missing data. We assume the following submodels for the processes involved in the specification of the shared parameter model:

$$\mathcal{Y}_i = X_{yi}\beta + Z_{yi}b_{yi} + \varepsilon_{yi}, \qquad \log T_i = x_{ti}^{\top}\gamma + b_{ti} + \sigma_t^{-1}\varepsilon_{ti},$$

where $\varepsilon_{yi} \sim \mathcal{N}_{n_i}(0, V_i = \sigma_y^2 Q_i)$ with Q_i being a correlation matrix with an associated parameter vector κ, $\varepsilon_{ti} \sim \mathcal{P}$ where $\mathcal{P}$ denotes an appropriate distribution function with corresponding survival function S and density function p, and σ_t is a scale parameter (Kalbfleisch and Prentice, 2002, Ch. 3). Finally, the joint density of $\{b_{yi}, b_{ti}\}$ follows (2.5), with copulas belonging to either the Archimedean or elliptical classes and Gaussian marginals.

For the E-step we let $\ddot{A}$ denote $E\{A(b_{yi}, b_{ti}) \mid \mathcal{Y}_i, T_i, \delta_i; \theta\}$, where the required integrals are approximated using a Gauss-Hermite quadrature rule. For the parameters with no closed-form solution, we let $\ell(\cdot)$ denote the score vector of the complete-data log likelihood. The expected value $\ddot{\ell}(\cdot)$ of $\ell(\cdot)$, with respect to $p(b_{yi}, b_{ti} \mid \mathcal{Y}_i, T_i, \delta_i; \theta)$, is used to maximize numerically the expected value of the complete-data log likelihood, based on a quasi-Newton algorithm. In particular,

the following expressions define the M-step.

For the longitudinal measurement model we have

$$\begin{aligned}
\beta &= \left(\sum_{i=1}^{n} X_{yi}^{\top} V_i^{-1} X_{yi}\right)^{-1} \left\{\sum_{i=1}^{n} X_{yi}^{\top} V_i^{-1} (y_i - Z_{yi}\ddot{b}_{yi})\right\}, \\
\sigma_y^2 &= \frac{1}{N}\sum_{i=1}^{n} \mu_{yi}^{\top} Q_i^{-1}(\mu_{yi} - 2Z_{yi}\ddot{b}_{yi}) + \mathrm{tr}(Z_{yi}^{\top} Q_i^{-1} Z_{yi} \ddot{vb}_{yi}) + \ddot{b}_{yi}^{\top} Z_{yi}^{\top} Q_i^{-1} Z_{yi} \ddot{b}_{yi}, \\
\ddot{\ell}(\kappa) &= \frac{1}{2}\sum_{i=1}^{n} \mathrm{tr}(-Q_i^{-1} W_i) + \mu_{yi}^{\top} K_i(\mu_{yi} - 2Z_{yi}\ddot{b}_{yi}) + \mathrm{tr}(M_i \ddot{vb}_{yi}) + \ddot{b}_{yi}^{\top} M_i \ddot{b}_{yi},
\end{aligned}$$

where $N = \sum_{i=1}^{n} n_i$, $\mu_{yi} = y_i - X_{yi}\beta$, $\ddot{b}_{yi} = E(b_{yi} \mid \mathcal{Y}_i, T_i, \delta_i; \theta)$, $\ddot{vb}_{yi} = \mathrm{var}(b_{yi} \mid \mathcal{Y}_i, T_i, \delta_i; \theta)$, $W_i = \partial Q_i / \partial \kappa$, $K_i = Q_i^{-1} W_i Q_i^{-1}$, $M_i = Z_{yi}^{\top} K_i Z_{yi}$.

For the event process model we have

$$\ell(\gamma) = \sigma_t^{-1} \sum_{i=1}^{n} x_{ti} a_i \quad \text{and} \quad \ell(\sigma_t) = \sigma_t^{-1} \sum_{i=1}^{n} \omega_i a_i - \delta_i,$$

where $a_i = -\delta_i\{\partial \log p(\omega_i)/\partial \omega_i\} - (1-\delta_i)\{\partial \log S(\omega_i)/\partial \omega_i\}$ and $\omega_i = (\log T_i - x_{ti}^{\top}\gamma - b_{ti})/\sigma_t$.

For the random effects model we distinguish between the following cases. In the first case, the Gaussian copula combined with normal marginals results in a multivariate normal distribution with known derivatives for the variance components. In the second case, the Student's-t copula involves the inverse distribution function of the Student's-t distribution and thus numerical derivatives are used. Finally, for Archimedean copulas, $\ell(\alpha)$ is derived for each particular copula separately, whereas, for the parameters θ_{by} and θ_{bt} of the marginal models for b_{yi} and b_{ti}, the following general formula is used:

$$\begin{aligned}
\ell(\theta_{by}) = \sum_{i=1}^{n} & \frac{g^{(3)}\{C(u_i, v_i)\}}{g^{(2)}\{C(u_i, v_i)\}} - 3\frac{g^{(2)}\{C(u_i, v_i)\}}{g^{(1)}\{C(u_i, v_i)\}} \; c_u(v_i) + \frac{g^{(2)}(u_i)}{g^{(1)}(u_i)} \; \frac{\partial u}{\partial \theta_{by}} + \\
& \frac{\partial \log p(b_{yi}; \theta_{by})}{\partial \theta_{by}},
\end{aligned}$$

where $g(\cdot)$ is the generator function of the archimedean copula with $g^{(l)}(\cdot)$ denoting its lth derivative, $c_u(v) = \partial C(u,v)/\partial u$, u and v are the distribution functions of the marginal Gaussian distributions for b_{yi} and b_{ti}, respectively, and $\ell(\theta_{bt})$ is derived analogously.

References

Cox, D. and Hinkley, D. (1974). *Theoretical Statistics*. Chapman & Hall, London.

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

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

Heyde, C. and Johnstone, I. (1979). On asymptotic posterior normality for stochastic processes. *Journal of the Royal Statistical Society, Series B*, **41**, 184–189.

Holland, P. and Wang, Y.-J. (1987). Dependence function for continuous bivariate densities. *Comm. Statist., Theory Methods*, **16**, 863–876.

Huang, X., Stefanski, L. and Davidian, M. (2006). Latent-model robustness in structural measurement error models. *Biometrika*, **93**, 53–64.

Jones, M. (1996). The local dependence function. *Biometrika*, **83**, 899–904.

Kalbfleisch, J. and Prentice, R. (2002). *The Statistical Analysis of Failure Time Data*. Wiley, New York, 2nd edition.

Keiding, N., Andersen, P. and Klein, J. (1997). The role of frailty models and accelerated failure time models in describing heterogeneity due to omitted covariates. *Statistics in Medicine*, **16**, 215–224.

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

R Development Core Team (2007). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL `http://www.R-project.org`. ISBN 3-900051-07-0.

Scharfstein, D., Rotnitzky, A. and Robins, J. (1999). Adjusting for nonignorable dropout using semiparametric non-response models (with discussion). *Journal of the American Statistical Association*, **94**, 1096–1120.

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

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

Tsiatis, A. and Davidian, M. (2004). Joint modeling of longitudinal and time-to-event data: an overview. *Statistica Sinica*, **14**, 809–834.

Wang, Y. and Taylor, J. (2001). Jointly modeling longitudinal and event time data with application to acquired immunodeficiency syndrome. *Journal of the American Statistical Association*, **96**, 895–905.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, **50**, 1–26.

Wu, M. and Carroll, R. (1988). Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. *Biometrics*, **44**, 175–188.

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

CHAPTER

3

A Two-Part Joint Model for the Analysis of Survival and Longitudinal Binary Data with excess Zeros

Abstract

Many longitudinal studies generate both the time to some event of interest and repeated measures data. This paper is motivated by a study on patients with a renal allograft, in which interest lies in the association between longitudinal proteinuria (a dichotomous variable) measurements and the time to renal graft failure. An interesting feature of the sample at hand is that nearly half of the patients were never tested positive for proteinuria ($\geq$ 1 gr/day) during follow-up, which introduces a degenerate part in the random effects density for the longitudinal process. In this paper we propose a two-part shared parameter model framework that effectively takes this feature into account, and we investigate sensitivity to the various dependence structures used to describe the association between the longitudinal measurements of proteinuria and the time to renal graft failure.

Keywords: Copulas; Joint modelling; Sensitivity analysis; Shared parameter model.

3.1 Introduction

Chronic kidney diseases affect one in nine US adults, and may lead to complications such as high blood pressure, anemia, weak bones, poor nutritional health and nerve damage. Furthermore, when kidney diseases progress, this may eventually lead to renal failure, which requires dialysis or a kidney transplantation to maintain life. Many studies have been conducted to investigate which factors may play a role in the progression of chronic kidney diseases.

Our research has been motivated by a study on patients that underwent, between 1/21/1983 and 8/16/2000, a primary renal transplantation with a graft from a deceased or living donor in the University Hospital Gasthuisberg from the Catholic University of Leuven (Belgium). We consider the 432 patients for whom the new graft has survived for at least one year. The clinical interest lies in the long term performance of the new graft, and especially in the graft survival for a ten year period. Out of the 432 patients considered, 91 (21.1%) experienced a graft failure. The corresponding Kaplan-Meier estimate for the time to graft failure is depicted in the top-left panel of Figure 3.1. The estimated graft survival function shows that the renal graft survival rate at ten years equals 0.79 (95% CI: 0.75, 0.83). During the ten year follow-up period, the patients were periodically tested for the performance of the graft. One of the outcomes measuring this performance is the presence of proteinuria. Proteinuria is the condition in which the urine contains an abnormal amount of protein, which is an indication of renal graft malfunctioning. For the current analysis proteinuria was defined as the presence of 1 gr of protein in a 24 hours urine collection. An interesting feature of the sample at

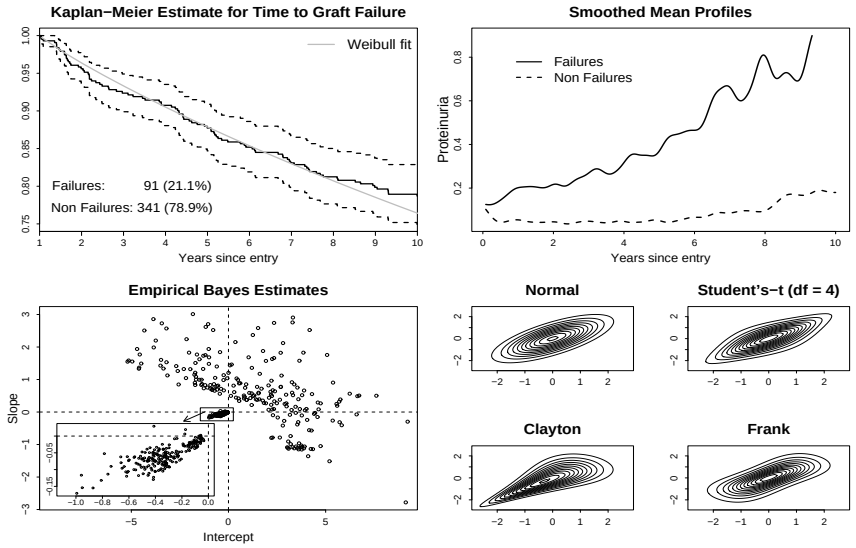

Figure 3.1: *Top left panel: Kaplan-Meier estimate (with associated 95% CI) for time to graft failure, with superimposed Weibull fit. Top right panel: sample smooth average profiles (obtained using a Nadaraya-Watson kernel regression estimate) for the probability of proteinuria versus years since entry, for patients with at least one finding of proteinuria during follow-up. Bottom left panel: empirical Bayes estimates under an ignorable random slopes logistic regression for proteinuria, including all patients. The rectangle around zero contains the patients with no proteinuria history and it is magnified in the third quadrant. Bottom right panel: contour plots of the Normal, Student's-t (df* $= 4$*), Clayton, and Frank copula for standard normal marginals and Kendall's* $\tau = 0.5$.

Table 3.1: *Contingency table for findings of proteinuria versus renal graft failure.*

Proteinuria	Failure	No Failure	Total
at least once	72 (32.4%)	150 (67.6%)	222
never	19 (9%)	191 (91%)	210

hand is that for nearly half of the patients, proteinuria of more than 1 gr/day has never been observed. Table 3.1 presents the frequencies of at least one positive finding of proteinuria during follow-up versus failure status. We observe that the use of at least one finding of proteinuria as a prognostic factor for graft failure would result in a very high negative predictive value, since 91% (95% CI: 87.1%, 94.8%) of the patients with no proteinuria history did not experience a graft failure. On the contrary, the positive predictive value is low 32.4% (95% CI: 26.3%, 38.6%) implying that at least one finding of proteinuria is not indicative of graft failure. However, the sample smooth average profiles (obtained using a Nadaraya-Watson kernel regression estimate) for the patients with at least one positive diagnosis of proteinuria, presented in the top-right panel of Figure 3.1, show a steep increase for failures. This feature suggests that exploration of the longitudinal evolution of proteinuria could be more insightful for the time to graft failure. Thus, our aim here is to investigate the association structure between these two processes.

The setting described above connects to the framework of joint modelling of longitudinal and time to event data (see Tsiatis and Davidian, 2004, for a review). The majority of the research in this area has focused on continuous longitudinal responses motivated by HIV and cancer studies. Joint models for cases where the longitudinal measured outcome is binary have been considered for instance by Faucett, Schenker, and Elashoff (1998) and Larsen (2004), and have also been applied in the missing data context (Pulkstenis et al., 1998; Albert, 2000). Joint models are constructed under the conditional independence assumption, which posits that the event process and the longitudinal responses are independent conditionally on a latent process expressed by a set of random effects. These random

effects are typically assumed to be normally distributed, but relaxations of the normality assumption have been proposed, for instance by Song et al. (2002). However, note that a normal or another smooth random effects density might be unrealistic for our data, since half of the patients never showed proteinuria during follow-up. This feature, in fact, induces a bimodality in the random effects density, which is also evident in the plot of the empirical Bayes (EB) estimates, obtained by the ignorable (i.e., ignoring the survival process) mixed-effects logistic regression, presented in the bottom-left panel of Figure 3.1. This model includes as fixed-effects linear time trends with some additional baseline covariates that will be introduced in Section 3.4, while intercepts and slopes are used in the random effects component. In particular, we observe that the random effects estimates for the patients with no proteinuria are concentrated around zero, with very small dispersion compared to the estimates for the other subjects. To overcome this problem, we propose a two-part shared parameter model which assumes that the distribution of the longitudinal process is a two-component mixture with a degenerate component for patients with no proteinuria history and a mixed-effects logistic regression component for the remaining patients. This formulation allows to investigate separately the effect of, first, the longitudinal evolution of proteinuria and, second, the history of proteinuria, to the time to graft failure. In addition, inference for the whole population can easily be made by mixing the probability distribution for the two parts. Such mixture models have been proposed in various contexts in the statistical literature. Zero-inflated Poisson and negative binomial count models are presented in Ridout, Hinde, and Demetrio (2001), whereas two-part models for longitudinal data have been proposed by Olsen and Schafer (2001) and Kowalski et al. (2003). Furthermore, joint modelling with cure-rate survival models is reviewed in Yu et al. (2004).

A final issue that we tackle in this work is the sensitivity of inference to parametric assumptions for the association structure between the survival and lon-

gitudinal processes. Sensitivity might be expected from experience related to the joint modelling context (i.e., missing data framework). In particular, proteinuria measurements are not available at the observed graft failure times, and can only be identified using modelling assumptions. Thus, investigation of robustness of inference to these assumptions is required. Here we follow a copula parameterization for the joint distribution of the underlying random effects, which allows investigation of dependence, and we perform a sensitivity analysis by considering different copula functions.

The remaining of the paper is organised as follows: Section 3.2 presents the two-part shared parameter model, discusses its features and refers to goodness-of-fit, choice of copula, and sensitivity analysis issues. Section 3.3 presents an EM algorithm for obtaining the maximum likelihood estimates under the proposed model. Finally, Section 3.4 presents the analysis of the renal graft failure data, Section 3.5 discusses some simulation results, and Section 3.6 concludes the paper.

3.2 The Two-Part Shared Parameter Model Formulation

3.2.1 Submodels Specification

Joint models typically consist of three submodels, namely the longitudinal, the survival, and the random effects models. In our formulation however, we introduce a fourth component that accounts for the patients with no proteinuria history. In particular, let T_i be the observed failure time for the ith patient ($i = 1, \ldots, n$), which is the minimum of the true failure time T_i^* and the censoring time $\mathcal{J}_i$. Set δ_i to be the failure indicator that equals one for true events and zero otherwise, i.e., $\delta_i = I(T_i^* \leq \mathcal{J}_i)$, where $I(\cdot)$ is the indicator function. Let y_i denote the $n_i \times 1$ vector of binary indicators for proteinuria, and let d_i be an indicator variable that equals one if the ith patient showed clinically important proteinuria at least once during follow-up and zero otherwise, i.e., $d_i = I(y_{ij} = 1;\ \text{for some } j = 1, \ldots, n_i)$. The

two-part shared parameter model, omitting covariates in the notation, is defined as

$$p(T_i, \delta_i, y_i; \theta) = \sum_{d_i} p(d_i; \theta) p(T_i, \delta_i, y_i \mid d_i; \theta) \tag{3.1}$$

with

$$\begin{aligned} & p(T_i, \delta_i, y_i \mid d_i; \theta) \\ &= \int\int \breve{p}(T_i, \delta_i \mid b_{ti}, d_i; \theta_t)\, p(y_i \mid b_{yi}, d_i; \theta_y) \\ &\qquad\qquad p(b_{yi}, b_{ti} \mid d_i; \theta_b) db_{yi} db_{ti}, \end{aligned}$$

where $\theta^\top = (\theta_d^\top, \theta_t^\top, \theta_y^\top, \theta_b^\top)$ is the vector of the parameters in each one of the submodels and let also $A^\top$ denote the transpose of A. Further, let $p(\cdot)$ denote the appropriate probability density functions for the longitudinal and random effects parts, whereas for the event process we set $\breve{p}(T_i, \delta_i \mid b_{ti}, d_i; \theta_t) = p(T_i \mid b_{ti}, d_i; \theta_t)^{\delta_i} S(T_i \mid b_{ti}, d_i; \theta_t)^{1-\delta_i}$, i.e., equal to either the density for the true event times or the survival function for censored observations. Factorization (3.1) resembles the pattern mixture models factorization used in the missing data context (Little and Rubin, 2002) that posits an inherent heterogeneity, which deterministically groups individuals according to their proteinuria history. The model for d_i is a simple logistic regression, which will be described in Section 3.4.

For the survival process we assume a mixed-effects accelerated failure time model defined as

$$\log T_i = w_i^\top \gamma + d_i \gamma_d + b_{ti} + \sigma_t \varepsilon_i, \qquad \varepsilon_i \sim \mathcal{P}, \tag{3.2}$$

where $\theta_t^\top = (\gamma^\top, \gamma_d, \sigma_t)$, and w_i is a $q_w \times 1$ vector of baseline covariates. Parameter γ_d measures the effect of proteinuria history in the logarithm of time to graft failure, which, according to Table 3.1, is expected to be highly significant. The random effect b_{ti} represents a frailty term that captures unobserved heterogeneity

induced, for instance, by omitted covariates (Keiding et al., 1997). The errors ε_i are assumed to follow the distribution function $\mathcal{P}$, with corresponding survival function S and density function p, and σ_t denotes a scale parameter (Kalbfleisch and Prentice, 2002, Ch. 3). In this work we consider parametric models for $\mathcal{P}$; non-parametric alternatives in the joint modelling framework have been proposed by Tseng et al. (2005).

The model for the longitudinal process conditionally on d_i contains a degenerate part in order to account for the fact that $y_{ij} = 0$, for all j when $d_i = 0$. For the patients with proteinuria history, we model the longitudinal evolution of proteinuria findings using a mixed-effects logistic regression. In particular, we assume that

$$\begin{cases} \Pr(y_{ij} = 0, \forall j) = 1, & \text{if } d_i = 0 \\ \Pr(y_{ij} = 1 \mid b_{yi}) = \pi_{ij} = 1/\left[1 + \exp\{-(x_{ij}^\top \beta + z_{ij}^\top b_{yi})\}\right], & \text{if } d_i = 1, \end{cases} \tag{3.3}$$

where $\theta_y = \beta$ is the vector of regression coefficients, y_{ij} equals one if the ith patient had a proteinuria finding at the jth time, and zero otherwise, b_{yi} are subject-specific random effects dictating patient's longitudinal trajectories, and X_i and Z_i are $n_i \times q_x$ and $n_i \times q_z$ design matrices for the fixed- and random effects, respectively.

The common parameterization used in joint models postulates that $b_{ti} = \alpha^\top b_{yi}$, where α denotes an association parameter. That is, the longitudinal and survival processes share, in fact, the same random-effect b_{yi}, with α^2 being a rescaling factor for the variance of b_{yi}. However, this parameterization assumes perfect correlation between the underlying random effects, which may be unrealistic in many applications. Therefore we relax this assumption and estimate the correlation between the random effects of the two processes. This parameterization is similar to the joint model of Henderson et al. (2000) who considered two correlated Gaussian processes to induce dependence. In particular, for the patients with

proteinuria history we use a copula representation for the joint distribution of b_{yi} and b_{ti}. Copulas (Nelsen, 1999) are multivariate distribution functions with uniform marginals that can be used to construct multivariate densities and investigate dependence. Under (3.1) the random effects density then takes the form

$$p(b_{yi}, b_{ti} \mid d_i; \theta_b) = \begin{cases} p(b_{ti}; \omega_t), & \text{if } d_i = 0 \\ \mathcal{C}\ p(b_{yi}; \omega_y) p(b_{ti}; \omega_t), & \text{if } d_i = 1, \end{cases} \tag{3.4}$$

where $\mathcal{C} = c\{H_y(b_{yi}; \omega_y), H_t(b_{ti}; \omega_t); \alpha\}$, $c(\cdot)$ is the density of the copula $C(\cdot)$, $H_y(\cdot)$ and $p(b_{yi})$ are the marginal cumulative distribution function and the probability density function for b_{yi}, respectively, and $H_t(\cdot)$ and $p(b_{ti})$ are defined analogously for b_{ti}. The parameter vector for the random effects density is $\theta_b^\top = (\alpha, \omega_y^\top, \omega_t^\top)$, where α is the association parameter of the copula, and ω_y and ω_t are the parameter vectors for the two marginals. The advantage of the copula parameterization is that it allows for separate modelling of the association structure and the marginals, thus facilitating exploration of dependence. In particular, the $c\{H_y(b_{yi}; \omega_y), H_t(b_{ti}; \omega_t); \alpha\}$ part of (3.4) is the function that specifies the association type between the two marginals $H_y(\cdot)$ and $H_t(\cdot)$.

3.2.2 Goodness-of-Fit, Choice of Copula and Sensitivity Analysis

The use of correlated random effects between the two processes enables the specification of flexible joint models. However, since the random effects are in fact latent variables, checking the appropriateness of the assumed random effects model (i.e., copula function and marginal distributions) can be challenging. This implies that proposed methods for goodness-of-fit in copula models (e.g., Genest and Rivest, 1993) can be difficult to apply since the empirical cumulative distribution function cannot be readily computed for the pair (b_{yi}, b_{ti}). Moreover, the use of the Empirical Bayes estimates to uncover either the shape of the copula

distribution $C\{H_y(b_{yi}), H_t(b_{ti})\}$ or the shape of the marginal $H_y(b_{yi})$ and $H_t(b_{ti})$, can be misleading due to shrinkage (Fitzmaurice, Laird, and Ware, 2004; Verbeke and Molenberghs, 2000). Alternatively, the performance of the assumed random effects model can be implicitly investigated by checking the fit of the joint model to the observed data. In particular, a plot of the fitted marginal survival function versus the Kaplan-Meier estimate, and a plot of the fitted average longitudinal profiles versus the smoothed (as in the top-right panel of Figure 3.1) sample average profiles, could show potential model miss-fit. Moreover, information criteria, such as the AIC, could be also employed to select the best fitting copula.

However, we would like to note that the use of measures, based on the observed data, for identifying the best fitting copula should be done with caution. The reason for this lies in the close relationship between the joint modelling of survival and longitudinal measurements and the missing data framework. To see this more clearly, let y_i^o and y_i^m denote the set of observed and missing longitudinal measurements for the ith individual, before and at the observed event time T_i, respectively. Then the conditional distribution of the missingness process (i.e., the event process) given the complete vector of longitudinal measurements (y_i^o, y_i^m), that is used to characterize the missing data mechanism, has the form

$$\begin{aligned} p(T_i \mid y_i^o, y_i^m) &= \frac{\int p(T_i \mid b_i)\, p(y_i^o, y_i^m \mid b_i)\, p(b_i)\, db_i}{\int p(y_i^o, y_i^m \mid b_i)\, p(b_i)\, db_i} \\ &= \int p(T_i \mid b_i)\, p(b_i \mid y_i^o, y_i^m)\, db_i, \end{aligned} \tag{3.5}$$

where $b_i^\top = (b_{yi}^\top, b_{ti})$. According to Little and Rubin (2002), since this distribution depends on y_i^m, joint models imply a Not Missing At Random (NMAR) missing data mechanism. As it is known in the missing data literature, in NMAR settings the observed data do not contain enough information to distinguish between certain models, since a lot of information is implicitly provided through modelling assumptions. In such cases it is advisable to perform a sensitivity analysis un-

der different model formulations rather than rely on goodness-of-fit measures and criteria that depend on the observed data only (see e.g., discussion of Diggle and Kenward, 1994; Copas and Li, 1997; Little and Rubin, 2002; Jansen et al., 2006). In our proposed model, the posterior distribution of the random effects in (3.5) is analogous to $p(b_{yi}, b_{ti} \mid y_i^o, y_i^m) \propto p(y_i^o, y_i^m \mid b_{yi}) p(b_{yi}, b_{ti})$, which according to (3.4) implies that the copula is the key part that describes the association between the missingness and longitudinal processes. Varying the choice of the copula function leads to different shapes of association structure. This is illustrated in the bottom-right panel of Figure 3.1, which depicts the contours of four copulas assuming standard normal marginals. In order to obtain comparable contour plots, we have chosen the copula parameter α such that the association between the two standard normal marginals equals 0.5 in terms of Kendall's τ. However, we observe that the copula function can significantly alter the shape of the association, even though all the other components (i.e., marginals and global association measure) of the bivariate densities remain the same. Thus, in the analysis of the proteinuria data presented in Section 3.4, in addition to the proposed methods for goodness-of-fit and choosing copula described above, we have also performed a sensitivity analysis in order to investigate the effect of the choice of the copula function in the size of the association between the two processes.

3.3 EM Algorithm

In this section we focus on the estimation of $\theta^* = (\theta_t^\top, \theta_y^\top, \theta_b^\top)^\top$, since estimates for θ_d are easily obtained by fitting separately the logistic regression for $\Pr(d_i = 1; \theta_d)$. The maximum likelihood estimates for the model parameters θ^* are obtained using an EM algorithm, in which b_{yi} and b_{ti} are treated as missing data.

For the E-step, denote $E\{A(b_{yi}, b_{ti}) \mid y_i, T_i, \delta_i; \theta\}$ as $\tilde{A}$, i.e., the expected value of any function $A(\cdot)$ of b_{yi} and b_{ti} with respect to $p(b_{yi}, b_{ti} \mid y_i, T_i, \delta_i; \theta)$. These

expectations are approximated using a Gauss-Hermite quadrature rule; more details can be found in Appendix A. For the M-step, unfortunately the complete data log-likelihood for the two-part shared parameter model does not have closed form solutions with respect to θ. Thus, the expected value of the complete data log-likelihood is numerically maximized using a quasi-Newton algorithm. This procedure requires the expected score vector of the complete data log-likelihood, given d_i, which we denote by $\tilde{\ell}(\cdot)$. The expressions of $\tilde{\ell}(\cdot)$ for β, γ, γ_d, σ_t have the form

$$\begin{aligned}
\tilde{\ell}(\beta) &= \sum_{i=1}^{n} X_i^\top (y_i - \tilde{\pi}_i) \\
\tilde{\ell}\{(\gamma^\top, \gamma_d)\} &= \sigma_t^{-1} \sum_{i=1}^{n} \tilde{a}_i \ddot{w}_i \\
\tilde{\ell}(\sigma_t) &= \sigma_t^{-1} \sum_{i=1}^{n} \widetilde{\zeta_i a_i} - \delta_i,
\end{aligned}$$

where $\tilde{\pi}_i = \int p(b_{yi} \mid y_i, T_i, \delta_i)/[1 + \exp\{-(X_i\beta + Z_i b_{yi})\}]db_{yi}$, $\ddot{w}_i^\top = (w_i^\top, d_i)$, $a_i = -\delta_i\{\partial \log p(\zeta_i)/\partial\zeta_i\} - (1-\delta_i)\{\partial \log S(\zeta_i)/\partial\zeta_i\}$, and $\zeta_i = (\log T_i - w_i^\top\gamma - d_i\gamma_d - b_{ti})/\sigma_t$.

To define the expression of $\tilde{\ell}(\cdot)$ for the parameters $\theta_b^\top = (\alpha, \omega_y^\top, \omega_t^\top)$ of the random effects model, we assume normal marginals with mean zero, and we distinguish the following cases. First, we consider the elliptical copulas class and specifically the normal and Student's-t copulas. The normal copula combined with normal marginals results in a multivariate normal distribution with derivatives for the variance components given by

$$\begin{aligned}
\tilde{\ell}(\theta_b) = \frac{1}{2} \sum_{i=1}^{n} & \operatorname{tr}(-\Sigma^{-1}K) \\
+ \quad & \operatorname{tr}(\Sigma^{-1}K\Sigma^{-1}\widetilde{vb_i}) + \tilde{b}_i^\top \Sigma^{-1} K \Sigma^{-1} \tilde{b}_i,
\end{aligned}$$

where $b_i^\top = (b_{yi}^\top, b_{ti})$, Σ is the covariance matrix of $p(b_{yi}, b_{ti})$ parameterized through θ_b, $K = \partial\Sigma/\partial\theta_b$, $\tilde{b}_i = \int b_i p(b_i \mid y_i, T_i, \delta_i) db_i$, and $\widetilde{vb}_i = \int [b_i - \tilde{b}_i]^2 p(b_i \mid y_i, T_i, \delta_i) db_i$. The Student's-$t$ copula involves the inverse cumulative distribution function of the Student's-t distribution and thus $\tilde{\ell}(\cdot)$ is approximated numerically using a central difference approximation. Second, for archimedean copulas, $\tilde{\ell}(\alpha)$ is derived for each particular copula separately, whereas for the parameters ω_y and ω_t of the marginal models we use the result (Nelsen, 1999, Ch. 4) that the density of the copula function has the form

$$c(u,v) = -\frac{g^{(2)}\{C(u,v)\} g^{(1)}(u) g^{(1)}(v)}{[g^{(1)}\{C(u,v)\}]^3},$$

which leads to the following general formulae

$$\begin{aligned} \tilde{\ell}(\omega_y) &= \tilde{\ell}_1(\omega_y) + \tilde{\ell}_2(\omega_y) \\ \ell_1(\omega_y) &= \sum_{i=1}^{n} \left\{ \mathcal{G}\, c_u(v_i) + \frac{g^{(2)}(u_i)}{g^{(1)}(u_i)} \right\} \frac{\partial u}{\partial \omega_y} \\ \tilde{\ell}_2(\omega_y) &= \frac{1}{2} \sum_{i=1}^{n} \operatorname{tr}(-D^{-1}Q) + \operatorname{tr}(D^{-1}QD^{-1}\widetilde{vb}_{yi}) \\ &\qquad + \tilde{b}_{yi}^\top D^{-1} Q D^{-1} \tilde{b}_{yi}, \end{aligned} \tag{3.6}$$

with

$$\mathcal{G} = \frac{g^{(3)}\{C(u_i,v_i)\}}{g^{(2)}\{C(u_i,v_i)\}} - 3\frac{g^{(2)}\{C(u_i,v_i)\}}{g^{(1)}\{C(u_i,v_i)\}},$$

where $g(\cdot)$ is the generator function of the archimedean copula with $g^{(l)}(\cdot)$ denoting its lth derivative, $c_u(v) = \partial C(u,v)/\partial u$ is the conditional distribution function for V given $U = u$, $U = H_y(b_{yi}; \omega_y)$ and $V = H_t(b_{ti}; \omega_t)$, D is the covariance matrix of the normal marginal for b_{yi}, $Q = \partial D/\partial\omega_y$, $\tilde{b}_{yi} = \int b_{yi} p(b_{yi} \mid y_i, T_i, \delta_i) db_{yi}$, $\widetilde{vb}_{yi} = \int [b_{yi} - \tilde{b}_{yi}]^2 p(b_{yi} \mid y_i, T_i, \delta_i) db_{yi}$, and $\tilde{\ell}(\omega_t)$ is derived analogously. The form of $\partial u/\partial\omega_y$, for the univariate and the bivariate case, is presented in Appendix B. Finally, based on the above expression both $\tilde{\ell}_1(\omega_y)$, using $\ell_1(\omega_y)$ from (3.6), and

$\tilde{\ell}_1(\omega_t)$ are numerically approximated using the procedure described in Appendix A.

3.4 Renal Graft Failure Analysis

We continue with the analysis of the renal graft failure study which was introduced in Section 3.1. In total, the patients made on average 62.8 visits (standard deviation 21.9 visits), resulting in 27147 records. The specification of the components of the two-part shared parameter model (3.1) is as follows. First, for the history of proteinuria a logistic regression is used. Second, for the survival process a Weibull model is assumed, which seems to provide a relatively reasonable fit to the survival function, according to the top-left panel of Figure 3.1. For completeness the derivatives for the M-step under the Weibull model are presented in Appendix C. Third, for the longitudinal process and based on the ignorable analysis (i.e., ignoring the event process), a mixed-effects logistic regression is adopted, with random-intercepts and -slopes. The covariate effects that are considered in all the above submodels are gender, weight, tobacco group (no-smoker, smoker, ex-smoker), age (older than 55), and long dialysis (if dialysis before transplant). Furthermore, for the longitudinal model the interaction between time (i.e., years since entry) and gender is considered as well. Finally, for the random effects model and in order to investigate the influence of parametric assumptions on the size of the association between the two processes, we performed a sensitivity analysis using the normal, Student's-t (df = 4), Clayton, and Frank copula functions assuming normal marginals. All models were fitted using the EM algorithm described in Section 3.3, and all computations have been performed in `R` (R Development Core Team, 2007). Due to the large sample size of this application nine quadrature points are used in the Gauss-Hermite rule; however, we expect that the procedure described in the Appendix A provides parameter estimates and standard errors of

Table 3.2: *Parameter estimates with standard errors in parenthesis, under the normal, Student's-t (df = 4), Clayton, and Frank copulas, for the logistic regression for proteinuria history, the survival and longitudinal processes, and for the random effects model.*

		Normal	Student's-t	Clayton	Frank
Proteinuria History	Intercept	0.08 (0.21)	0.08 (0.21)	0.08 (0.21)	0.08 (0.21)
	Gender (female)	−0.36 (0.23)	−0.36 (0.23)	−0.36 (0.23)	−0.36 (0.23)
	Weight	−0.02 (0.01)	−0.02 (0.01)	−0.02 (0.01)	−0.02 (0.01)
	Tob. Group (smoker)	−0.55 (0.49)	−0.55 (0.49)	−0.55 (0.49)	−0.55 (0.49)
	Tob. Group (ex-smoker)	−0.10 (0.22)	−0.10 (0.22)	−0.10 (0.22)	−0.10 (0.22)
	Age	1.26 (0.29)	1.26 (0.29)	1.26 (0.29)	1.26 (0.29)
	Dialyses	−0.21 (0.20)	−0.21 (0.20)	−0.21 (0.20)	−0.21 (0.20)
Survival Processes	Intercept	2.53 (0.17)	2.34 (0.17)	3.51 (0.19)	1.73 (0.18)
	No Prtn History	1.52 (0.21)	1.44 (0.22)	0.71 (0.22)	2.47 (0.34)
	Gender (female)	0.53 (0.19)	0.54 (0.19)	0.52 (0.22)	0.49 (0.20)
	Weight	−0.01 (0.01)	−0.01 (0.01)	−0.01 (0.01)	−0.01 (0.01)
	Tob. Group (smoker)	−0.45 (0.30)	−0.48 (0.31)	−0.57 (0.37)	−0.37 (0.33)
	Tob. Group (ex-smoker)	0.36 (0.19)	0.44 (0.18)	0.44 (0.21)	0.44 (0.19)
	Age	−0.19 (0.27)	−0.29 (0.26)	−0.09 (0.30)	−0.17 (0.27)
	Dialyses	0.05 (0.16)	0.02 (0.16)	−0.04 (0.19)	−0.01 (0.17)
	scale	0.86 (0.07)	0.87 (0.07)	0.97 (0.08)	0.92 (0.08)
Longitudinal Process	Intercept	−3.53 (0.24)	−3.96 (0.18)	−2.32 (0.25)	−4.20 (0.23)
	Year Snce Entry	0.36 (0.04)	0.31 (0.03)	0.45 (0.03)	0.30 (0.04)
	Gender (female)	0.76 (0.35)	0.57 (0.21)	1.01 (0.25)	0.73 (0.27)
	Weight	0.04 (0.01)	0.06 (0.01)	0.01 (0.01)	0.06 (0.01)
	Tob. Group (smoker)	1.07 (0.21)	1.37 (0.23)	1.34 (0.20)	0.77 (0.35)
	Tob. Group (ex-smoker)	−0.03 (0.38)	−0.47 (0.11)	0.03 (0.13)	−0.33 (0.13)
	Age	0.76 (0.19)	1.19 (0.15)	0.59 (0.18)	1.22 (0.17)
	Dialyses	−0.83 (0.21)	−0.46 (0.11)	−0.94 (0.10)	−0.53 (0.12)
	Year Snce Entry:Age	−0.38 (0.05)	−0.31 (0.03)	−0.39 (0.03)	−0.30 (0.05)
Random-Effects	Long. Intercept	2.45 (0.24)	2.25 (0.14)	2.75 (1.22)	4.46 (0.28)
	Long. Slope	0.73 (0.02)	0.67 (0.04)	0.82 (0.34)	1.35 (0.10)
	Long. correlatn	−0.74 (0.02)	−0.70 (0.03)	−0.88 (0.11)	−0.86 (0.02)
	Surv. frailty	0.54 (0.05)	0.56 (0.05)	0.50 (0.07)	0.61 (0.03)
	Kendall's-τ	−0.24 (0.09)	−0.25 (0.10)	−0.18 (0.04)	−0.54 (0.07)

good quality.

The parameter estimates and standard errors under the scenarios considered are presented in Table 3.2. As can been seen, the choice of the copula function has a direct impact on certain parameter estimates. For instance, the smoker effect is lower for the Frank copula compared to the Student's-t copula. Moreover, the association between the survival and longitudinal processes varies from -0.18 ($s.e. = 0.04$) to -0.54 ($s.e. = 0.07$), which is different from the common perfect correlation assumption discussed in Section 3.2.1. As expected, the estimated association is negative suggesting that the lower the probability of proteinuria findings, the longer the graft survives. In addition, for all copulas we observe a significant effect of proteinuria history, indicating that patients with no proteinuria findings during follow-up maintain their graft longer. The effects of the copula function are also apparent in the plots of the EB estimates for the random effects of the longitudinal process, the marginal survival function for the event process, and the marginal average longitudinal evolutions for the probability of proteinuria,

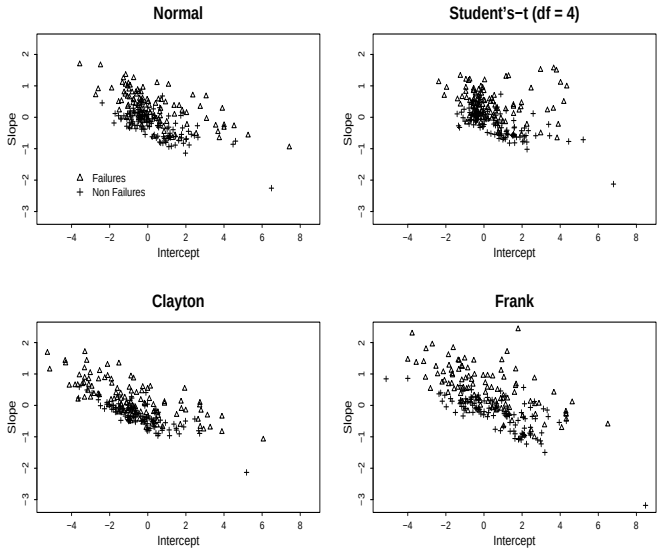

Figure 3.2: *Empirical Bayes estimates for the random effects in the longitudinal processes under the normal, Student's-t (df = 4), Clayton, and Frank copulas, for the patients with proteinuria history.*

presented in Figures 3.2, 3.3 and 3.4. The EB estimates are defined as the posterior modes, i.e.,

$$\begin{aligned}
&\arg\max_{b_{yi},b_{ti}} p(b_{yi}, b_{ti} \mid y_i, T_i, \delta_i, d_i; \hat{\theta}) \\
&= \arg\max_{b_{yi},b_{ti}} \{\log \breve{p}(T_i, \delta_i \mid b_{ti}, d_i; \hat{\theta}_t) + \log p(y_i \mid b_{yi}, d_i; \hat{\theta}_y) \\
&+ \log p(b_{yi}, b_{ti} \mid d_i; \hat{\theta}_b)\},
\end{aligned}$$

whereas the marginal survival function is computed by

$$\hat{S}(T_i) = \sum_d p(d_i; \hat{\theta}_d) \int S(T_i \mid b_{ti}, d_i; \hat{\theta}_t)\; p(b_{ti} \mid d_i; \hat{\theta}_b)\; db_{ti}$$

Figure 3.2 shows that the EB estimates are generally higher for failures than for non failures. This indicates that patients who experience graft failure either start with

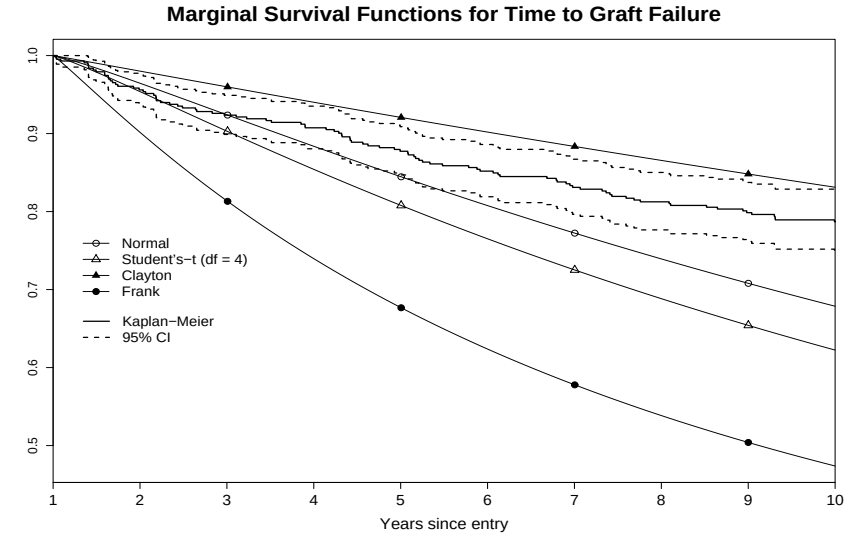

Figure 3.3: *Fitted marginal survival functions under the normal, Student's-t (df = 4), Clayton, and Frank copulas, with superimposed Kaplan-Meier estimate and associated 95% CI.*

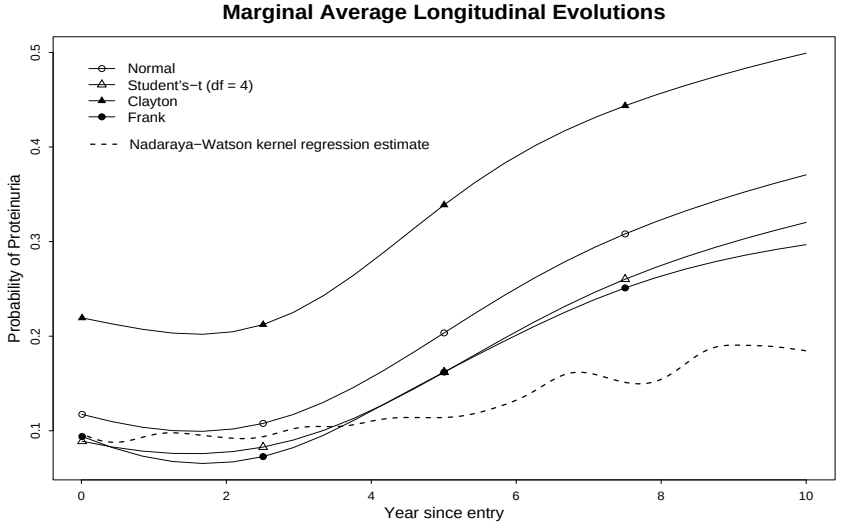

Figure 3.4: *Fitted marginal average longitudinal evolutions under the normal, Student's-t (df = 4), Clayton, and Frank copulas, with superimposed the Nadaraya-Watson kernel regression estimate for the sample profiles.*

low probability of showing clinically important proteinuria and quickly develop it or they start with relative high probability of showing proteinuria and maintain it. The marginal survival functions and the marginal average evolutions under each copula have been marginalized over the covariate values as well. Both Figures 3.3 and 3.4 suggest that the fitted models do not capture perfectly the observed data. The AIC values (smaller is better) for the four copulas are 2555.504, 2201.598, 3471.406 and 4841.606 for the normal, Student's-t, Clayton and Frank copula, respectively, which identify the Student's-t as the best of the four.

However, we would like to note that Figures 3.3 and 3.4 do not necessarily imply that the model does not fit the data. This is due to the fact that a comparison of the fitted model with the observed data is only valid under Missing Completely At Random missing data mechanisms, which is certainly not the case for our application since the association between the two processes is significant for all copulas (i.e., Table 3.2, Kendall's-τ estimates). For instance, in Figure 3.4 the model successfully acknowledges that if the patients with graft failure have not failed, the average evolution would yield higher values than observed ones for the last years. In conclusion, the variability we observe in the overall results under the different copulas could be regarded as variability due to modelling assumptions, which is a clear indication that distributional assumptions for the random effects may prove difficult to verify.

3.5 Simulation Study

A simulation study has been performed to empirically investigate the finite sample performance of the proposed model, and in addition to explore the effect of copula misspecification. In particular, we considered four simulation scenarios corresponding to the normal, Student's-t (df = 4), Clayton and Frank copulas, and two sample size settings, namely a large one with $n = 200$ and a small one with

$n = 50$. Under each scenario and sample size setting 500 data-sets were simulated, and each data-set was fitted under four two-part joint models. For the degenerate component, the event process, and the longitudinal measurement process the correct model specifications have been assumed. In order to investigate random effects misspecification, the normal, the Student't-t (df $= 4$), the Clayton and the Frank copulas are fitted for each data-set. The study's set-up is presented in detail in Section 3.7. The results, presented in Tables 3.3 to 3.10 and discussed in Section 3.7.3, showed an overall good performance of the proposed model but also some sensitivity issues that can be attributed to the arguments raised in Section 3.2.2. Moreover, the use of the AIC for choosing the best fitting copula revealed that even though in the majority of times the true random effects model was selected, the number of times another copula was selected was not negligible.

3.6 Conclusion

We have proposed a new shared parameter model for the joint modelling of longitudinal binary measurements and time to event data, and demonstrated its use through a real data example. The main strength of this framework is that it effectively handles the existence of excess zeros patterns in the binary responses by assuming a degenerate part in the longitudinal response model. In addition, it was shown in the application that the shared parameter models with binary responses are not robust with respect to the assumptions for the random effects distribution, and thus a sensitivity analysis should be performed. A potential drawback of the proposed model is that the logistic regression part in the two-part longitudinal process defined in (3.3), does not impose the constraint that $\Pr(y_{ij} = 0, \forall j) = 0$. We expect that this feature could lead to some bias, especially for small n_i, but this is not the case for our application.

Several extensions of the proposed model can be considered. First, the para-

metric distributional assumptions for the survival process can be relaxed either by considering a Cox-type proportional hazards model or by extending the approach of Tseng et al. (2005), in order to account for a longitudinal binary covariate with excess zeros and by postulating two separate random effects components for the two processes. Second, for ordinal longitudinal measurements, the degenerate component formulation can be extended to handle several excess levels as well, by positing a multinomial model for d_i. Third, other types of longitudinal responses (e.g., semicontinuous random variables with point masses at one or more locations) can be easily handled under the proposed framework by simply changing the appropriate parts in the EM algorithm. Finally, in our sensitivity analysis we have concentrated on the effect of the copula part of the random effects distribution since this is the part that describes the association between the two processes. However, in a larger scale sensitivity analysis it would be useful to examine the effect of the assumptions for the marginal distributions for b_{yi} and b_{ti} as well.

3.7 Supplementary Material

3.7.1 Simulation Study Design

A simulation study has been performed to empirically investigate the performance of the maximum likelihood estimates under the proposed two-part joint model in finite samples as well as the effect of copula misspecification. In particular, a two group comparison is considered with two choices for the sample size, a moderate one with $n = 200$ and a small one with $n = 50$. The submodels specification for the two-part joint model is as follows. First, the degenerate group indicator d_i is simulated according to the logistic model

$$\text{logit}\left\{\Pr(d_i = 1; \theta_d)\right\} = \theta_{d0} + \theta_{d1}\mathcal{T}_i,$$

where $\mathcal{T}_i$ denotes the treatment indicator, and $(\theta_{d0}, \theta_{d1}) = (1, 0.8)$. Second, for the survival process a Weibull model with a frailty term is assumed

$$\log T_i = \gamma_0 + \gamma_1 \mathcal{T}_i + \gamma_d d_i + b_{ti} + \sigma_t \varepsilon_i,$$

where $(\gamma_0, \gamma_1, \gamma_d) = (0.5, 1.5, -0.5)$, $\sigma_t = 0.5$, and ε_i follows an extreme value distribution. The censoring mechanism follows an exponential distribution with mean 11, which results in 30% censoring on average. Third, the non-degenerate part of the longitudinal process has the form

$$\text{logit}\left\{\Pr(y_{ij} = 1 \mid b_{yi}, d_i = 1)\right\} = \beta_0 + \beta_1 t_{ij} + \beta_2 \mathcal{T}_i t_{ij} + b_{y0i} + t_{ij} b_{y1i},$$

where t_{ij} denotes the time points at which the y_{ij} measurements are taken, $\mathcal{T}_i t_{ij}$ is the interaction term between $\mathcal{T}_i$ and t_{ij}, and $(\beta_0, \beta_1, \beta_2) = (0.1, 1.5, -1)$. The maximum number of repeated measurements per individual is 20, with $t_{ij} = \text{seq}(0, 4, 20)$, where $\text{seq}(a, b, c)$ denotes a regular sequence from a to b of length c (e.g., $\text{seq}(0, 2, 5) = 0, 0.5, 1, 1.5, 2$). Taking into account the censoring and the degenerate individuals (i.e., $\{i \in (1, 2, \ldots) : d_i = 0\}$), the average number of measurements per individual is 9.1 with standard deviation 7.6 measurements. Finally, for the random effects model the following scenarios are considered: (i) the normal copula, with correlation $\alpha = \sin(0.5\pi/2) = 0.707$; (ii) the Student's-$t$ copula, with 4 degrees of freedom and correlation $\alpha = \sin(0.5\pi/2) = 0.707$; (iii) the Clayton copula, with parameter $\alpha = 2$; (iv) the Frank copula, with parameter $\alpha = 5.736$. The value of α for each copula has been chosen such that the association between b_{yi} and b_{ti} equals 0.5 in terms of Kendall's-τ. Moreover, both marginal distributions $H_y(b_{yi}; \omega_y)$ and $H_t(b_{ti}; \omega_t)$ are taken to be normal with zero mean, and with covariance matrix parameters $\omega_y = \{\sigma^2_{by0} = \text{var}(b_{y0i}) = 4; \sigma^2_{by1} = \text{var}(b_{y1i}) = 1; \text{cor}_{by} = \text{cor}(b_{y0i}, b_{y1i}) = 0.4\}$ and variance $\omega_t^2 = 4$, respectively. For each scenario and each sample size setting 500 data-sets have been simulated.

3.7.2 Simulation Analysis Models

Each simulated data-set has been analyzed under four two-part joint models. In particular, for the degenerate component, the event process and the longitudinal measurement process the correct model specifications, as described in the previous section, are assumed. However, for the random effects the normal, the Student't-t (df = 4), the Clayton and the Frank copulas are fitted for each data-set. Thus, under each scenario (i) to (iv), the true and three misspecified random effects models are fitted. The MLEs for the joint models' parameters are obtained using the EM algorithm described in Section 3.3, using 13 quadrature points for the Gauss-Hermite rule.

3.7.3 Simulation Results

For the data-sets considered in the simulation study the average computer time for fitting each model was 84.27 min. (standard deviation 63.58 min.) for $n = 200$, and 41.01 min. (standard deviation 43.87 min.) for $n = 50$. Computations have been performed in `R` version 2.4.1, on an AMD Opteron Cluster, consisting of 164 dual Opteron250 servers running Linux (kernel version 2.6.15.7), with 2GB RAM, several nodes with 16GB RAM, and 4 to 8 CPUs with frequencies varying from 1.8 to 2.6 GHz. Thus, we feel that our proposed model requires reasonable computing time considering its complexity.

The bias and root mean square error for each parameter of the two-part joint model, presented in Section 3.7.1, are given in Tables 3.3 to 3.6 for sample size $n = 200$, and Tables 3.7 to 3.10 for sample size $n = 50$. For all scenarios we observe that the true model has a relatively good performance compared to the joint models with misspecified copula function. Moreover, the performance of the two elliptical copulas (i.e., normal and Student's-t) seems to be better than the one of the Archimedean ones (i.e., Clayton and Frank). For example, in Table 3.5 we

observe that the Clayton copula, even though is the correct one, shows greater bias for γ_d, β_1 and β_2 than the elliptical copulas; however, for the variance components it performs better than the other copulas. In addition, some sensitivity of the results regarding the copula choice is apparent in the parameter estimates of all three submodels. For instance, in Table 3.3 the estimated association parameter τ has less bias and root mean square error for the misspecified Student's-t copula rather than for the normal copula, which is the true one. Another example can be found in Table 3.7 in which the Clayton and Frank copulas provide better estimates for γ_0 than the true normal copula. For sample size $n = 50$, we observed the same behaviour for the two-part joint models but, as expected, with larger values for the root mean square error, and in some cases with also more bias. Finally, the sensitivity regarding the choice of copula is also evident in the number of times each copula has been selected as the best fitting copula according to the AIC, for each scenario. Even though the AIC the majority of times selects the true model, the number of times it fails to do so is not negligible. These observations reinforce our statement, presented in Section 3.2.2, that sensitivity analysis is necessary in the joint modelling framework in order to investigate the impact of the modelling assumptions.

3.8 Appendix A

The integrals involved in the specification of the E-step do not have a closed form solution and thus are approximated using the Gauss-Hermite quadrature rule. In particular,

$$\begin{aligned} E\left\{A(b_{yi}, b_{ti}) \mid y_i, T_i, \delta_i\right\} &= \int\int A(b_{yi}, b_{ti})\; p(b_{yi}, b_{ti} \mid y_i, T_i, \delta_i)\; db_{yi} db_{ti} \\ &\approx 2^{q/2} \sum_{t_1 \cdots t_q} h_t A(t\sqrt{2})\; p(t\sqrt{2} \mid y_i, T_i, \delta_i) \exp(\|t\|^2), \end{aligned}$$

Table 3.3: *Simulation results based on 500 data-sets under the normal scenario (i) with sample size $n = 200$. The bias and root mean square error for each parameter are presented. The line 'AIC' denotes the number of times each copula has been selected as the best fitting copula according to the AIC.*

		Normal		Student's-t		Clayton		Frank	
	True	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
θ_{d0}	1.0	0.0052	0.2309	0.0052	0.2309	0.0052	0.2309	0.0052	0.2309
θ_{d1}	0.8	0.0444	0.3789	0.0444	0.3789	0.0444	0.3789	0.0444	0.3789
γ_0	0.5	0.0522	0.3660	0.0607	0.3735	−0.0025	0.3078	0.0346	0.3425
γ_1	1.5	−0.0191	0.3178	−0.0227	0.3178	0.0192	0.3268	−0.0152	0.3392
γ_d	−0.5	−0.0043	0.4084	0.0059	0.4088	0.5516	0.6870	0.5068	0.6778
σ_t	0.5	−0.0195	0.1850	−0.0070	0.1944	−0.1712	0.1976	−0.1227	0.2075
β_0	0.1	0.0129	0.2129	0.0151	0.2125	−0.0966	0.2209	−0.0872	0.2148
β_1	1.5	0.1079	0.4617	0.1358	0.4788	0.1500	0.4843	0.1820	0.5417
β_2	−1.0	−0.0901	0.4875	−0.1093	0.4979	−0.1669	0.4954	−0.1822	0.5677
σ_{by0}	2.0	−0.0553	0.1930	−0.0283	0.1844	−0.0317	0.1815	−0.2248	0.3040
σ_{by1}	1.0	0.0305	0.1909	0.0447	0.1833	0.0160	0.1267	−0.0471	0.1671
cor_{by}	0.4	0.0168	0.2045	0.0539	0.1965	0.2226	0.2923	0.3004	0.3465
ω_t	2.0	−0.0147	0.1496	−0.0197	0.1569	−0.0445	0.1316	0.0790	0.1577
τ	0.5	0.0117	0.0654	0.0066	0.0560	0.0158	0.0829	0.0948	0.1149
AIC		236 (47.2%)		87 (17.4%)		103 (20.6%)		74 (14.8%)	

Table 3.4: *Simulation results based on 500 data-sets under the Student's-t scenario (ii) with sample size $n = 200$. The bias and root mean square error for each parameter are presented. The line 'AIC' denotes the number of times each copula has been selected as the best fitting copula according to the AIC.*

		Normal		Student's-t		Clayton		Frank	
	True	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
θ_{d0}	1.0	0.0079	0.2318	0.0079	0.2318	0.0079	0.2318	0.0079	0.2318
θ_{d1}	0.8	0.0361	0.3821	0.0361	0.3821	0.0361	0.3821	0.0361	0.3821
γ_0	0.5	0.0540	0.3847	0.0452	0.3779	0.0071	0.3437	0.0413	0.3545
γ_1	1.5	−0.0109	0.3112	−0.0111	0.3154	0.0398	0.3332	0.0016	0.3256
γ_d	−0.5	−0.0119	0.4097	−0.0052	0.4080	0.5787	0.7398	0.5208	0.6890
σ_t	0.5	−0.0096	0.1787	−0.0149	0.1777	−0.1511	0.2035	−0.1224	0.2034
β_0	0.1	0.0201	0.2222	0.0171	0.2198	−0.0914	0.2301	−0.0732	0.2236
β_1	1.5	0.0406	0.4726	0.0803	0.4724	0.1016	0.4780	0.1473	0.5403
β_2	−1.0	−0.0185	0.5068	−0.0595	0.5033	−0.1153	0.4975	−0.1391	0.5686
σ_{by0}	2.0	−0.0654	0.2142	−0.0411	0.1989	−0.0511	0.1918	−0.2479	0.3326
σ_{by1}	1.0	0.0052	0.1717	0.0198	0.1607	0.0182	0.1224	−0.0671	0.1682
cor_{by}	0.4	−0.0343	0.2233	−0.0037	0.2030	0.1625	0.2640	0.2611	0.3259
ω_t	2.0	−0.0167	0.1531	−0.0133	0.1536	−0.0564	0.1376	0.0912	0.1625
τ	0.5	−0.0109	0.0352	0.0012	0.0609	0.0031	0.0898	0.0831	0.1093
AIC		128 (25.6%)		223 (44.6%)		84 (16.8%)		65 (13.0%)	

Table 3.5: *Simulation results based on 500 data-sets under the Clayton scenario (iii) with sample size $n = 200$. The bias and root mean square error for each parameter are presented. The line 'AIC' denotes the number of times each copula has been selected as the best fitting copula according to the AIC.*

		Normal		Student's-t		Clayton		Frank	
	True	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
θ_{d0}	1.0	0.0063	0.2340	0.0063	0.2340	0.0063	0.2340	0.0063	0.2340
θ_{d1}	0.8	0.0364	0.3815	0.0364	0.3815	0.0364	0.3815	0.0364	0.3815
γ_0	0.5	−0.9074	0.9867	−0.8978	0.9766	−0.9308	1.0074	−0.8897	0.9733
γ_1	1.5	−0.0256	0.3079	−0.0177	0.3059	0.0024	0.3142	−0.0247	0.3271
γ_d	−0.5	−0.0149	0.4165	0.0055	0.4164	0.9261	1.0355	0.6666	0.8279
σ_t	0.5	0.0858	0.2000	0.1374	0.2298	0.0138	0.1862	0.0239	0.1941
β_0	0.1	0.0469	0.2131	0.0538	0.2149	0.0159	0.2197	0.0020	0.2056
β_1	1.5	0.0512	0.6562	0.1196	0.6958	0.2036	0.6437	0.2624	0.7458
β_2	−1.0	0.0250	0.6671	−0.0366	0.6916	−0.1358	0.6388	−0.1465	0.7092
σ_{by0}	2.0	−0.1458	0.2693	−0.1221	0.2452	−0.0119	0.1824	−0.2570	0.3388
σ_{by1}	1.0	−0.0132	0.2044	−0.0247	0.1864	−0.0128	0.1113	−0.0970	0.1696
cor_{by}	0.4	0.0342	0.2560	0.0481	0.2475	−0.0104	0.2593	0.1790	0.3224
ω_t	2.0	0.0164	0.1444	−0.0054	0.1419	−0.0284	0.1273	0.1439	0.2082
τ	0.5	−0.0363	0.0802	−0.0185	0.0738	−0.0094	0.0755	0.0491	0.0865
AIC		54 (10.8%)		31 (6.2%)		368 (73.6%)		47 (9.4%)	

Table 3.6: *Simulation results based on 500 data-sets under the Frank scenario (iv) with sample size $n = 200$. The bias and root mean square error for each parameter are presented. The line 'AIC' denotes the number of times each copula has been selected as the best fitting copula according to the AIC.*

		Normal		Student's-t		Clayton		Frank	
	True	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
θ_{d0}	1.0	0.0110	0.2303	0.0110	0.2303	0.0110	0.2303	0.0110	0.2303
θ_{d1}	0.8	0.0297	0.3842	0.0297	0.3842	0.0297	0.3842	0.0297	0.3842
γ_0	0.5	−0.7247	0.8020	−0.7156	0.7910	−0.7484	0.8306	−0.6969	0.7295
γ_1	1.5	−0.0061	0.2793	−0.0159	0.2790	0.0006	0.2931	−0.0028	0.2830
γ_d	−0.5	−0.0080	0.3714	−0.0038	0.3663	0.6785	0.8148	0.5733	0.6033
σ_t	0.5	0.2216	0.3021	0.2387	0.3125	0.1080	0.2734	0.0249	0.2032
β_0	0.1	−0.0127	0.2109	−0.0188	0.2139	−0.0193	0.2294	0.0054	0.2099
β_1	1.5	0.1235	0.6934	0.1786	0.7251	0.1683	0.6808	0.1372	0.7012
β_2	−1.0	−0.0784	0.7199	−0.1338	0.7476	−0.1667	0.6976	−0.0364	0.6205
σ_{by0}	2.0	−0.0746	0.2363	−0.0469	0.2161	0.0966	0.2197	−0.0604	0.2129
σ_{by1}	1.0	0.1819	0.3337	0.1332	0.2677	0.1007	0.2276	0.0153	0.1604
cor_{by}	0.4	0.1590	0.2890	0.1646	0.2784	0.0349	0.3089	0.0135	0.2626
ω_t	2.0	−0.3134	0.3540	−0.3200	0.3612	−0.2951	0.3424	−0.0679	0.1887
τ	0.5	−0.0921	0.1204	−0.0813	0.1109	−0.0910	0.1446	0.0011	0.0654
AIC		59 (11.8%)		58 (11.6%)		76 (15.2%)		307 (61.4%)	

Table 3.7: *Simulation results based on 500 data-sets under the normal scenario (i) with sample size $n = 50$. The bias and root mean square error for each parameter are presented. The line 'AIC' denotes the number of times each copula has been selected as the best fitting copula according to the AIC.*

		Normal		Student's-t		Clayton		Frank	
	True	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
θ_{d0}	1.0	0.0569	0.5193	0.0569	0.5193	0.0569	0.5193	0.0569	0.5193
θ_{d1}	0.8	0.5450	3.1888	0.5450	3.1888	0.5450	3.1888	0.5450	3.1888
γ_0	0.5	0.1478	0.7275	0.1718	0.7530	0.0923	0.4826	0.0945	0.5770
γ_1	1.5	−0.0299	0.5141	−0.0443	0.5317	0.0618	0.5096	0.0386	0.5061
γ_d	−0.5	−0.0451	0.7833	−0.0383	0.7822	0.3322	0.6532	0.3175	0.7076
σ_t	0.5	−0.1299	0.2849	−0.1157	0.2958	−0.2599	0.2799	−0.2487	0.2740
β_0	0.1	0.0627	0.4166	0.0641	0.4248	−0.0890	0.4461	−0.1083	0.4168
β_1	1.5	0.4324	1.3088	0.4719	1.3201	0.4441	1.1714	0.4895	1.3950
β_2	−1.0	−0.3439	1.3625	−0.3727	1.3704	−0.4032	1.1774	−0.4558	1.4572
σ_{by0}	2.0	−0.0970	0.3786	−0.0516	0.3566	−0.0498	0.3245	−0.1843	0.4111
σ_{by1}	1.0	0.0049	0.2126	0.0889	0.3360	0.0132	0.2004	−0.0054	0.2726
cor_{by}	0.4	−0.0154	0.3930	−0.0382	0.3756	0.1278	0.3916	0.2110	0.3852
ω_t	2.0	−0.0222	0.2972	−0.0350	0.3178	−0.0821	0.2403	0.0348	0.2331
τ	0.5	−0.0008	0.1228	−0.0137	0.1165	−0.0331	0.1649	0.0626	0.1376
AIC		159 (31.8%)		89 (17.8%)		143 (28.6%)		109 (21.8%)	

Table 3.8: *Simulation results based on 500 data-sets under the Student's-t scenario (ii) with sample size $n = 50$. The bias and root mean square error for each parameter are presented. The line 'AIC' denotes the number of times each copula has been selected as the best fitting copula according to the AIC.*

		Normal		Student's-t		Clayton		Frank	
	True	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
θ_{d0}	1.0	0.0561	0.5186	0.0561	0.5186	0.0561	0.5186	0.0561	0.5186
θ_{d1}	0.8	0.4898	2.9970	0.4898	2.9970	0.4898	2.9970	0.4898	2.9970
γ_0	0.5	0.1174	0.6226	0.1085	0.6409	0.0624	0.4779	0.0683	0.5070
γ_1	1.5	0.0195	0.5584	0.0106	0.5583	0.0373	0.4598	0.0456	0.5133
γ_d	−0.5	−0.0263	0.6964	−0.0137	0.6897	0.3703	0.6730	0.3729	0.6805
σ_t	0.5	−0.1503	0.2844	−0.1292	0.2903	−0.2620	0.2844	−0.2354	0.2873
β_0	0.1	0.0856	0.4101	0.0793	0.4139	−0.0735	0.4963	−0.0891	0.4241
β_1	1.5	0.3159	1.2349	0.3659	1.2150	0.4589	1.2562	0.4759	1.4233
β_2	−1.0	−0.2008	1.2479	−0.2622	1.2009	−0.3999	1.2391	−0.4309	1.3975
σ_{by0}	2.0	−0.0787	0.3575	−0.0517	0.3494	−0.0677	0.3128	−0.2005	0.4125
σ_{by1}	1.0	0.1112	0.3709	0.0601	0.2856	0.0079	0.2269	−0.0362	0.2335
cor_{by}	0.4	−0.0570	0.4077	−0.0594	0.3779	0.1004	0.4284	0.1780	0.3910
ω_t	2.0	−0.0107	0.2918	−0.0317	0.2961	−0.0807	0.2471	0.0568	0.2430
τ	0.5	−0.0165	0.1314	−0.0230	0.1207	−0.0464	0.1848	0.0531	0.1395
AIC		120 (24.0%)		115 (23.0%)		141 (28.2%)		124 (24.8%)	

Table 3.9: *Simulation results based on 500 data-sets under the Clayton scenario (iii) with sample size $n = 50$. The bias and root mean square error for each parameter are presented. The line 'AIC' denotes the number of times each copula has been selected as the best fitting copula according to the AIC.*

		Normal		Student's-t		Clayton		Frank	
	True	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
θ_{d0}	1.0	0.0619	0.5552	0.0619	0.5552	0.0619	0.5552	0.0619	0.5552
θ_{d1}	0.8	0.5262	3.2102	0.5262	3.2102	0.5262	3.2102	0.5262	3.2102
γ_0	0.5	−0.8270	1.1638	−0.8221	1.1497	−0.7270	1.0631	−0.6868	1.0321
γ_1	1.5	0.0006	0.6456	0.0151	0.6448	0.0361	0.5801	−0.0255	0.5954
γ_d	−0.5	−0.0126	0.8695	0.0059	0.8495	0.8588	1.1989	0.6774	1.0846
σ_t	0.5	0.0150	0.2911	0.0686	0.3130	−0.1864	0.2761	−0.1455	0.2825
β_0	0.1	0.0898	0.4642	0.0955	0.4695	0.0140	0.4694	0.0068	0.4289
β_1	1.5	0.8313	3.5437	0.6959	2.9828	0.8175	2.7667	0.9618	3.0770
β_2	−1.0	−0.6773	3.5421	−0.5415	2.9658	−0.7066	2.7117	−0.7944	3.0267
σ_{by0}	2.0	−0.1557	0.3971	−0.1241	0.3584	−0.0055	0.3103	−0.2347	0.4302
σ_{by1}	1.0	0.1622	0.4481	0.0361	0.2930	−0.0163	0.2262	−0.0675	0.2225
cor_{by}	0.4	−0.0303	0.4604	−0.0865	0.4280	−0.2586	0.5070	−0.0943	0.4599
ω_t	2.0	−0.0193	0.2997	−0.0454	0.3046	−0.0328	0.2178	0.1734	0.3119
τ	0.5	−0.0307	0.1397	−0.0430	0.1309	−0.0173	0.1633	0.0660	0.1338
AIC		59 (11.8%)		39 (7.8%)		293 (58.6%)		109 (21.8%)	

Table 3.10: *Simulation results based on 500 data-sets under the Frank scenario (iv) with sample size $n = 50$. The bias and root mean square error for each parameter are presented. The line 'AIC' denotes the number of times each copula has been selected as the best fitting copula according to the AIC.*

		Normal		Student's-t		Clayton		Frank	
	True	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
θ_{d0}	1.0	0.0554	0.5080	0.0554	0.5080	0.0554	0.5080	0.0554	0.5080
θ_{d1}	0.8	0.5424	3.1396	0.5424	3.1396	0.5424	3.1396	0.5424	3.1396
γ_0	0.5	−0.6909	1.0083	−0.6779	0.9904	−0.4908	0.8235	−0.5567	0.8750
γ_1	1.5	−0.0125	0.5684	−0.0157	0.5770	−0.0046	0.4988	−0.0009	0.5045
γ_d	−0.5	0.0031	0.7896	0.0021	0.7840	0.5861	0.9640	0.6450	0.9698
σ_t	0.5	0.1157	0.3347	0.1539	0.3590	−0.1596	0.2813	−0.1434	0.2788
β_0	0.1	0.0256	0.4435	0.0178	0.4645	0.0129	0.5623	-0.0020	0.4464
β_1	1.5	0.5263	2.4957	0.5694	2.5227	0.6159	2.5402	0.5721	2.3889
β_2	-1.0	-0.4654	2.5157	-0.5100	2.5411	-0.6006	2.5242	-0.5129	2.3896
σ_{by0}	2.0	−0.0946	0.3773	−0.0490	0.3455	0.1055	0.3502	−0.0692	0.3565
σ_{by1}	1.0	0.2997	0.5695	0.1555	0.3952	0.0681	0.3026	0.0351	0.2795
cor_{by}	0.4	0.0628	0.4454	0.0183	0.4132	−0.2068	0.4819	−0.1503	0.4266
ω_t	2.0	−0.3318	0.4395	−0.3433	0.4462	−0.1968	0.3047	−0.0242	0.2555
τ	0.5	−0.0930	0.1687	−0.0978	0.1672	−0.1037	0.2057	0.0240	0.1141
AIC		55 (11.0%)		53 (10.6%)		169 (33.8%)		223 (44.6%)	

where q denotes the integral dimension, $\sum_{t_1 \cdots t_q}$ is used as shorthand for $\sum_{t_1} \cdots \sum_{t_q}$, $t^\top = (t_1, \ldots, t_q)$ are the abscissas with corresponding weights h_t, and $\|\cdot\|^2$ denotes the square of the Euclidean distance.

A known problem of the Gaussian-Hermite rule (Pinheiro and Bates, 1995), is that it assumes that the main mass of the integrand is around zero, which might not be the case for certain individuals. The adaptive Gauss-Hermite rule solves this problem by centering and rescaling the integrand in each iteration, increasing however dramatically the computational burden. In order to avoid both the poor approximation of the simple Gauss-Hermite rule and the computational complexity of the adaptive rule, we use the EB estimates and their standard error from the ignorable models, to center and scale the integrand. Even though this procedure is not a fully adaptive rule, we expect that the ignorable EB estimates provide a good approximation to the patients' standing in the random effects dimension, resulting in an acceptable integral approximation with a moderate number of quadrature points.

3.9 Appendix B

Here we present the form of $\partial u/\partial \omega_y = \partial H_y(b_{yi}; \omega_y)/\partial \omega_y$, used in the M-step of the EM algorithm, where $H_y(b_{yi}; \omega_y)$ denotes the normal cumulative distribution function (cdf) with zero mean and variance components parameterized through ω_y. We present two cases; univariate and bivariate random effects. First, in the univariate case, with b_{yi} representing a random intercepts term, we get

$$\frac{\partial}{\partial \omega_y} H_y(b_{yi}; \omega_y) = -\frac{b_{yi}}{\omega_y} p(b_{yi}; \omega_y),$$

where $p(b_{yi}; \omega_y)$ denotes the normal probability density function with zero mean and standard deviation ω_y. Second, in the bivariate case, where $b_{yi} = (b_{y1i}, b_{y2i})$, we use the parameterization of the bivariate normal cdf considered in Drezner and

Wesolowsky (1989):

$$
\begin{aligned}
& H_y(b_{y1i}, b_{y2i}; \omega_{y1}, \omega_{y2}, \rho) \\
&= \frac{(\omega_{y1}\omega_{y2})^{-1}}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{b_{y1i}} \int_{-\infty}^{b_{y2i}} \exp \quad -\frac{h_1^2/\omega_{y1}^2 + h_2^2/\omega_{y2}^2 - 2\rho h_1 h_2/\omega_{y1}\omega_{y2}}{2(1-\rho^2)} \quad dh_1 dh_2 \\
&= H_y(b_{y1i}; \omega_{y1}) H_y(b_{y2i}; \omega_{y2}) + \\
&\qquad \frac{1}{2\pi} \int_0^{\rho} \frac{\exp\{-(b_{y1i}^2/\omega_{y1}^2 + b_{y2i}^2/\omega_{y2}^2 - 2r b_{y1i} b_{y2i}/\omega_{y1}\omega_{y2})/2(1-r^2)\}}{\sqrt{1-r^2}} dr,
\end{aligned}
$$

which leads to the following expressions for the partial derivatives with respect to the correlation ρ, and the standard deviations ω_{y1}, and ω_{y2}

$$
\frac{\partial H_y(b_{y1i}, b_{y2i}; \omega_{y1}, \omega_{y2}, \rho)}{\partial \rho} = \frac{\exp\{-(b_{y1i}^2/\omega_{y1}^2 + b_{y2i}^2/\omega_{y2}^2 - 2\rho b_{y1i} b_{y2i}/\omega_{y1}\omega_{y2})/2(1-\rho^2)\}}{2\pi\sqrt{1-\rho^2}},
$$

$$
\begin{aligned}
\frac{\partial H_y(b_{y1i}, b_{y2i}; \omega_{y1}, \omega_{y2}, \rho)}{\partial \omega_{y1}} &= -\frac{b_{y1i}}{\omega_{y1}} p(b_{y1i}; \omega_{y1}) H_y(b_{y2i}; \omega_{y2}) \\
+ \quad & \int_0^{\rho} \frac{B(r) \exp\{-(b_{y1i}^2/\omega_{y1}^2 + b_{y2i}^2/\omega_{y2}^2 - 2r b_{y1i} b_{y2i}/\omega_{y1}\omega_{y2})/2(1-r^2)\}}{2\pi\sqrt{1-r^2}} dr,
\end{aligned}
$$

where

$$
B(r) = \frac{2 b_{y1i}}{\omega_{y1}^2\sqrt{1-r^2}} \quad \frac{b_{y1i}}{\omega_{y1}} - \frac{r b_{y2i}}{\omega_{y2}} \quad ,
$$

and $\partial H_y(b_{y1i}, b_{y2i}; \omega_{y1}, \omega_{y2}, \rho)/\partial \omega_{y2}$ is derived analogously. The integral over r can be easily approximated using an adaptive Gauss-Kronrod rule (Piessens et al., 1983).

3.10 Appendix C

The form of $\tilde{\ell}\{(\gamma^\top, \gamma_d)\}$ and $\tilde{\ell}(\sigma_t)$ under the Weibull model is

$$\begin{aligned} \tilde{\ell}\{(\gamma^\top, \gamma_d)\} &= \sigma_t^{-1} \sum_{i=1}^{n} \{\exp(\tilde{\zeta}_i) - \delta_i\} \ddot{w}_i \\ \tilde{\ell}(\sigma_t) &= \sigma_t^{-1} \sum_{i=1}^{n} \tilde{A}_i - (1 + \tilde{\zeta})\delta_i, \end{aligned}$$

where $\tilde{\zeta}_i = (\log T_i - w_i^\top \gamma - d_i \gamma_d - \tilde{b}_{ti})/\sigma_t$, with $\tilde{b}_{ti} = \int b_{ti}\ p(b_{yi} \mid y_i, T_i, \delta_i)\ db_{ti}$, and $A_i = \zeta_i \exp(\zeta_i)$.

References

Albert, P. (2000). A transitional model for longitudinal binary data subject to nonignorable missing data. *Biometrics*, **56**, 602–608.

Copas, J. and Li, H. (1997). Inference for non-random samples (with discussion). *Journal of the Royal Statistical Society, Series B*, **59**, 55–95.

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

Drezner, Z. and Wesolowsky, G. (1989). On the computation of the bivariate normal integral. *Journal of Statistical Computation and Simulation*, **35**, 277–279.

Faucett, C., Schenker, N. and Elashoff, R. (1998). Analysis of censored survival data with intermittently observed time-dependent binary covariates. *Journal of the American Statistical Association*, **93**, 427–437.

Fitzmaurice, G., Laird, N. and Ware, J. (2004). *Applied Longitudinal Analysis*. Wiley, Hoboken, New Jersey.

Genest, C. and Rivest, L.-P. (1993). Statistical inference procedures for bivariate Archimedean copulas. *Journal of the American Statistical Association*, **88**, 1034–1043.

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

Jansen, I., Hens, N., Molenberghs, G., Aerts, M., Verbeke, G. and Kenward, M. (2006). The nature of sensitivity in monotone missing not at random models. *Computational Statistics & Data Analysis*, **50**, 830–858.

Kalbfleisch, J. and Prentice, R. (2002). *The Statistical Analysis of Failure Time Data*. Wiley, New York, 2nd edition.

Keiding, N., Andersen, P. and Klein, J. (1997). The role of frailty models and accelerated failure time models in describing heterogeneity due to omitted covariates. *Statistics in Medicine*, **16**, 215–224.

Kowalski, K., McFadyen, L., Hutmacher, M., Frame, B. and Miller, R. (2003). A two-part mixture model for longitudinal adverse event severity data. *Journal of Pharmacokinetics and Pharmacodynamics*, **30**, 315–335.

Larsen, K. (2004). Joint analysis of time-to-event and multiple binary indicators of latent classes. *Biometrics*, **60**, 85–92.

Little, R. and Rubin, D. (2002). *Statistical Analysis with Missing Data*. Wiley, New York, 2nd edition.

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

Olsen, M. and Schafer, J. (2001). A two-part random-effects models for semicontinuous longitudinal data. *Journal of the American Statistical Association*, **96**, 730–745.

Piessens, R., deDoncker-Kapenga, E., Uberhuber, C. and Kahaner, D. (1983). *Quadpack: a Subroutine Package for Automatic Integration.* Springer, New York.

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

Pulkstenis, E., Ten Have, T. and Landis, R. (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.

R Development Core Team (2007). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria. URL `http://www.R-project.org`. ISBN 3-900051-07-0.

Ridout, M., Hinde, J. and Demetrio, C. (2001). A score test for testing a zero-inflated poisson regression model against zero-inflated negative binomial alternatives. *Biometrics*, **57**, 219–223.

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

Tseng, Y.-K., Hsieh, F. and Wang, J.-L. (2005). Joint modelling of accelerated failure time and longitudinal data. *Biometrika*, **92**, 587–603.

Tsiatis, A. and Davidian, M. (2004). Joint modeling of longitudinal and time-to-event data: an overview. *Statistica Sinica*, **14**, 809–834.

Verbeke, G. and Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. Springer-Verlag, New York.

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

Yu, M., Law, N., Taylor, J. and Sandler, H. (2004). Joint longitudinal-survival-cure models and their application to prostate cancer. *Statistica Sinica*, **14**, 835–832.

CHAPTER

4

Fully Exponential Laplace Approximations for the Joint Modelling of Survival and Longitudinal Data

Abstract

A common objective in longitudinal studies is the joint modelling of a longitudinal response with a time-to-event outcome. Random effects are typically used in the joint modelling framework to explain the interrelationships between these two processes. However, estimation in the presence of random effects involves intractable integrals requiring therefore numerical integration. In this paper we propose a new computational approach for fitting such models based on the Laplace method for integrals that makes the consideration of high dimensional random effects structures feasible. Contrary to the common Laplace approximation, our method requires much less repeated measurements per individual in order to produce reliable results.

Keywords: B-splines; EM algorithm; Dropout; Longitudinal models; Shared parameter model; Survival models.

4.1 Introduction

Longitudinal studies often produce two types of outcomes, namely a set of longitudinal response measurements and the time to an event of interest. Joint modelling of these two processes is required mainly in three settings. First, in a survival analysis context in order to measure the effect of a time-dependent covariate measured with error. Secondly, in longitudinal studies in order to adjust derived inferences for possibly outcome-dependent dropout, and finally, in investigating the association structure between the longitudinal and event time processes. A well known example is found in AIDS research in which a biomarker such as CD4 lymphocyte count is measured intermittently and its relationship with time to seroconversion or death is of interest (Tsiatis and Davidian, 2004).

Shared parameter models (Tsiatis and Davidian, 2004; Follmann and Wu, 1995) provide an appealing framework for the joint modelling of survival and longitudinal data. In particular, in these models it is assumed that a latent process, expressed by a set of time-invariant random effects, induces the dependence between the two explicitly observed processes. Even though the use of random effects facilitates the use of flexible submodels for the involved processes, the need for numerical integration makes the estimation of shared parameter models rather computationally demanding. Common numerical integration techniques such as Gaussian quadrature and Monte Carlo have been successfully applied in the joint modelling framework (Song et al., 2002; Henderson et al., 2000). However, in such methods the computational burden increases exponentially with the dimensionality of the integration, which renders the consideration of complex random effects structures (e.g., modelling nonlinear subject-specific trajectories with splines or high-order

polynomials) rather prohibitive. A practical alternative in such settings is the Laplace method for integrals (De Bruijn, 1981). The main computational complexity of the Laplace approximation is the requirement of locating the mode of the integrand with respect to the random effects. Thus, when the dimensionality of the integration is increased, the computational burden of the Laplace method is smaller compared to the quadrature or Monte Carlo integration approaches. Even though the Laplace approximation is appealing in high dimensional settings, the order of the approximation error is $O(n_i^{-1})$, with n_i denoting the number of repeated measurements for the ith subject. This implies that in order for the Laplace method to work satisfactorily, many repeated measurements per subject are required.

In this paper we consider a new type of Laplace approximation for the joint modelling of survival and longitudinal data that is of order $O(n_i^{-2})$. The proposed approximation requires locating the same modes as in the original $O(n_i^{-1})$ approximation and the computation of a correction term based on these modes. Thereby a better approximation is achieved with computational complexity of the same order as in the common Laplace method. The main idea is to apply the Laplace approximation to the score vector of the shared parameter model, which is expressed as the expected value of the score vector conditional on the random effects, with respect to the posterior distribution of the random effects given the observed data. Our method is based on the fully exponential Laplace approximation, proposed by Tierney, Kass, and Kadane (1989) for approximating posterior moments of nonpositive functions in a Bayesian analysis context. A similar approach has been also considered by Steel (1996) for the estimation of generalized mixed models. An additional contribution of our work is the formulation of the cumulative baseline hazard function for the survival outcome. In the joint modelling context, an unspecified baseline hazard function is typically assumed for the event process to protect the derived inferences against misspecification. However, recently Hsieh

et al. (2006) noted that the use of the profile likelihood approach, which is required in this case, leads to underestimation of standard errors. Thus, here we postulate a flexible but parametric model for the cumulative baseline hazard function by expanding it into B-splines basis functions. Under this formulation, the computation of standard errors derives from standard maximum likelihood theory. The required monotonicity constraint is imposed by reparameterizing the basis coefficients similarly to the reparameterization of the threshold coefficients in the proportional odds model. Finally, we also investigate the asymptotic properties of the proposed Laplace based maximum likelihood estimators when both the sample size and number of repeated measurements per individual grow to infinity.

This research is motivated by a study on 407[1] patients that underwent, between 1/21/1983 and 8/16/2000, a primary renal transplantation with a graft from a deceased or living donor in the University Hospital Gasthuisberg of the Catholic University of Leuven (Belgium). The clinical interest lies in the long term performance of the new graft, and especially in the graft survival for over a ten year period. During the follow-up period patients were periodically tested for the performance of the graft. In the present analysis interest lies in the use of longitudinal haematocrit measurements as a prognostic factor for time to graft failure. An interesting feature of the data is that the longitudinal subject-specific profiles are highly nonlinear, as it is also illustrated in Figure 4.1 for eight randomly selected patients. For a preliminary ignorable analysis (i.e., ignoring the survival process) of the data, a linear mixed model is posited, in which subject-specific time evolutions are modelled nonlinearly using natural cubic splines with seven degrees of freedom. Table 4.1 shows a comparison of this model with simpler linear mixed models assuming linear and quadratic subject-specific time evolutions, respectively. Both AIC and BIC values support the use of natural cubic splines

[1]The difference with the number of patients in Chapters 2 and 3 is due to the fact that haematocrit measurements were not available for all 432 patients.

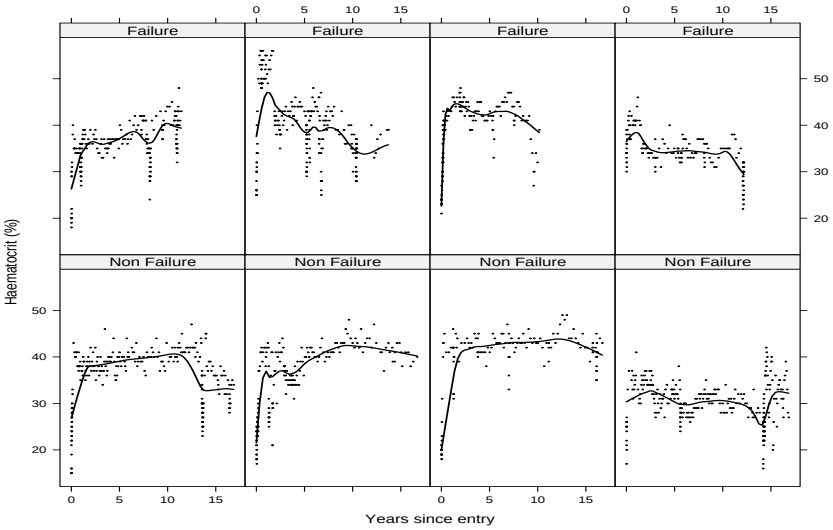

Figure 4.1: *Haematocrit subject-specific longitudinal trajectories for eight randomly selected patients. The top-row panels depict patients who experience graft failure and the bottom-row panels censored patients. The superimposed lines represent fitted curves using the loess smoother.*

Table 4.1: *Ignorable analysis of the renal graft failure data. Three linear mixed models are fitted assuming linear, quadratic and nonlinear (using natural cubic splines with 7 degrees of freedom) subject-specific evolutions in time. '#' denotes the number of parameters in the model, 'AIC' and 'BIC' the Akaike and Bayesian Information Criteria values, respectively, and 'log-Lik' the log-likelihood value. AIC and BIC are defined such that smaller values are better.*

Model	#	AIC	BIC	log-Lik
Linear	7	369089	369152	−184537.5
Quadratic	11	354553	354652	−177265.5
Natural splines	46	310231	310644	−155069.5

for modelling the time evolutions. This suggests that in the joint modelling analysis of the data we should allow for such nonlinearities. In order to fit the joint model under this high-dimensional random effects structure the proposed Laplace approximation is used.

The remaining of the paper is organised as follows: Section 4.2 presents the shared parameter model framework and the specification of the submodels for the involved processes. Section 4.3 presents the Laplace based EM algorithm for obtaining the maximum likelihood estimates under the proposed model and Section 4.4 discusses the asymptotic behaviour of the Laplace estimators. Finally, Section 4.5 discusses some simulation results and Section 4.6 presents the analysis of the renal graft failure data.

4.2 Joint Modelling Framework

4.2.1 Shared Parameter Model

Let T_i^* denote the true failure time for the ith subject ($i = 1, \ldots, n$) and $T_i = \min(T_i^*, \mathcal{J}_i)$ the observed failure time, where $\mathcal{J}_i$ is the censoring time. Define the event indicator as $\delta_i = I(T_i^* \leq \mathcal{J}_i)$, where $I(\cdot)$ is the indicator function. Let $y_{ij} = \{y_i(t_{ij}), j = 1, \ldots, n_i\}$ denote the longitudinal response measurements for the ith subject taken at time points t_{ij}. Finally, let b_i represent time-independent random effects that underly both the longitudinal measurement and survival processes. The joint likelihood contribution of the ith subject for the two outcomes, omitting covariates in the notation, is defined as

$$p(T_i, \delta_i, y_i; \theta) = \int \left\{ p(T_i \mid b_i; \theta)^{\delta_i} \mathcal{S}(T_i \mid b_i; \theta)^{1-\delta_i} \right\} p(y_i \mid b_i; \theta) p(b_i; \theta) \, db_i, \qquad (4.1)$$

where θ is the parameter vector, y_i is the vector of longitudinal responses of the ith subject, and $p(\cdot)$ denotes appropriate probability density functions. For the

event process, $\mathcal{S}(\cdot)$ denotes the survival function conditional on the random effects. Furthermore, we make the assumption that both the censoring and the visiting processes are noninformative, i.e., independent of b_i, T_i and y_i, and can thus be ignored from the modelling procedure (Tsiatis and Davidian, 2004).

4.2.2 Longitudinal Responses Submodel

For the longitudinal process we postulate a linear mixed effects model formulated as

$$y_i(t_{ij}) = W_i(t_{ij}) + \varepsilon_i(t_{ij}), \;\; \varepsilon_i(t_{ij}) \sim \mathcal{N}(0, \sigma^2), \tag{4.2}$$

where $W_i(t_{ij}) = x_i^\top(t_{ij})\beta + z_i^\top(t_{ij})b_i$, X_i and Z_i are the $n_i \times q_x$ and $n_i \times q_z$ design matrices (with corresponding row vectors $x_i^\top(t_{ij})$ and $z_i^\top(t_{ij})$) for the fixed and random effects, β and b_i, respectively. In order to model flexibly the subject-specific profiles, we assume that both $X_i(t)$ and $Z_i(t)$ contain a potentially high-dimensional vector of functions $f(t)$ of time t, expressed in terms of high-order polynomials or splines. The term $\varepsilon_i(t_{ij})$ denotes measurement error independent of b_i. In addition, we assume that the correlation between the repeated measurements in the longitudinal process is captured by the random effects, i.e., $\text{cov}\{\varepsilon_i(t), \varepsilon_i(t')\} = 0$, for $t \neq t'$, and thus the conditional independence assumption $p(y_i \mid b_i; \theta) = \prod_j p\{y_i(t_{ij}) \mid b_i; \theta\}$, holds for all i. Extensions to more complex error structures including serial correlation terms are straightforward but are not considered here. Finally, regarding the random effects, Song et al. (2002) have explored the need for more flexible assumptions for their distribution in the joint modelling framework; however, they discovered that parameter estimates and standard errors are rather robust to misspecifications, a feature that has been also theoretically corroborated by Rizopoulos, Verbeke, and Molenberghs (2008) and Hsieh et al. (2006). We therefore assume b_i to follow a multivariate normal distribution with mean zero and variance-covariance matrix D, without requiring

further investigation of this assumption.

4.2.3 Survival Submodel

The submodel for the event process accommodates the effects of the longitudinal time-dependent covariate and additional baseline covariates. In particular, the model is formulated for the logarithm of the cumulative hazard function for the ith subject as

$$\log H_i(t \mid b_i) = \log H_0(t) + \alpha W_i(t) + \gamma^\top x_{ti}, \tag{4.3}$$

where x_{ti} denotes the $q_{xt} \times 1$ vector of baseline covariates, α and γ are regression coefficients, and $H_0(\cdot)$ is the cumulative baseline hazard. This model is related to the time-varying Cox model with covariates affecting the log cumulative hazard ratio instead of the log hazard ratio. In order to allow for flexibility in the specification of the survival model we expand $\log H_0(t)$ into B-spline basis functions given by

$$\log H_0(t) = s(\log t; \omega) = \omega_0 + \sum_{k=1}^{m} \omega_k B_k(\log t, q), \tag{4.4}$$

where $\omega^\top = (\omega_0, \omega_1, \ldots, \omega_m)$, q denotes the degree of the B-splines and $m = \ddot{m} + q - 1$, with $\ddot{m}$ denoting the number of interior knots. Our motivation for postulating a flexible model for the logarithm of the cumulative baseline hazard function is that $\log H_0$ is typically gently curved or nearly linear as a function of $\log t$, and is usually very smooth. Thus, few knots will be required to capture its shape. To complete the definition of the event model we need to assure that $\log H_0(t)$ is a nondecreasing function of t or equivalently of $\log t$. Monotonicity can be achieved in the following way. The derivative $\partial s(\log t; \omega)/\partial \log t$ can be written as (Dierckx, 1995):

$$\begin{aligned}\partial s(\log t;\omega)/\partial \log t &= \sum_k \omega_k \big\{\partial B_k(\log t, q)/\partial \log t\big\} \\ &= q\sum_k \frac{\omega_{k+1}-\omega_k}{\lambda_{k+q+1}-\lambda_k} B_k(\log t, q-1), \qquad (4.5)\end{aligned}$$

where $\{\lambda_k\}$ denotes a vector of knot positions with nondecreasing values. Then, since by definition $B_k(\log t, q-1) \geq 0$ (Dierckx, 1995), monotonicity is guaranteed if the sequence of coefficients ω_k is nondecreasing, i.e., $\omega_{k+1} \geq \omega_k$, for all k. To avoid the imposition of inequality constraints, we make the following reparameterization: $\omega_1 = \omega_1^*$, $\omega_k = \omega_{k-1} + \exp(\omega_k^*)$, for $k > 1$, where ω_k^*'s are unconstrained. In the remainder of the paper and for notational simplicity ω_k^* will be denoted as ω_k. Similar models have been considered by Rosenberg (1995) who proposed to estimate the hazard function using B-splines, and by Royston and Parmar (2002) who modelled $\log H_0(t)$ using natural cubic splines but without imposing the required monotonicity constraint.

According to our experience, the placement of the boundary and internal knots does not appear critical for a good fit of the model. In particular, we suggest that the boundary knots are placed at the extreme log survival times, and the internal knots at quantiles of the uncensored log survival times. Four or five internal knots usually provide a reasonably good fit. The rationale for placing the internal knots according to quantiles of the uncensored survival times is to allow the data to be most closely modelled in the region of greatest density.

4.3 Estimation via an EM Algorithm

4.3.1 Score Vector and EM algorithm

The maximum likelihood estimates in the joint modelling framework are typically obtained using standard maximization algorithms such as the EM or the Newton-Raphson (Lange, 2004). For the former, b_i's are treated as missing data. The key component for applying either of these two algorithms in joint models is the score vector of the observed data log-likelihood function $\ell(\theta) = \sum_i \log p(T_i, \delta_i, y_i; \theta)$. According to (4.1) and the submodels specification presented in Section 4.2, $\ell(\theta)$ is easily found to be proportional to (i.e., constant terms are excluded)

$$\begin{aligned}\ell(\theta) \quad \propto \quad & \sum_i \log \int \left\{ \frac{1}{T_i} \frac{\partial s(\log T_i; \omega)}{\partial \log T_i} \exp\big[\eta_i(b_i) - \exp\{\eta_i(b_i)\}\big] \right\}^{\delta_i} \\ & \Big\{ \exp\big[-\exp\{\eta_i(b_i)\}\big] \Big\}^{1-\delta_i} \times (\sigma^2)^{-n_i/2} \exp\big\{ - \parallel y_i - X_i\beta - Z_i b_i \parallel^2 / 2\sigma^2 \big\} \\ & \times \det(D)^{-1/2} \exp\big(-b_i^\top D^{-1} b_i / 2\big)\, db_i, \end{aligned} \tag{4.6}$$

where $\eta_i(b_i) = s(\log T_i; \omega) + \alpha W_i(T_i) + \gamma^\top x_{ti}$, $\partial s(\log t; \omega)/\partial \log t$ is given in (4.5), and $\| \cdot \|$ denotes the Euclidean vector norm. It can be easily shown that the score vector under (4.6) is written as

$$\begin{aligned} S(\theta) \quad &= \quad \frac{\partial \ell(\theta)}{\partial \theta^\top} = \sum_i \frac{\partial}{\partial \theta^\top} \log \int p(T_i, \delta_i \mid b_i; \theta) p(y_i \mid b_i; \theta) p(b_i; \theta)\, db_i \\ &= \quad \sum_i \int h(\theta, b_i) p(b_i \mid T_i, \delta_i, y_i; \theta)\, db_i, \end{aligned} \tag{4.7}$$

where $p(T_i, \delta_i \mid b_i; \theta) = p(T_i \mid b_i; \theta)^{\delta_i} \mathcal{S}(T_i \mid b_i; \theta)^{1-\delta_i}$, and $h(\cdot)$ denotes the complete data score vector given by $h(\theta, b_i) = \partial \log p(T_i, \delta_i, y_i, b_i)/\partial \theta^\top = \partial\{\log p(T_i, \delta_i \mid b_i; \theta) + \log p(y_i \mid b_i; \theta) + \log p(b_i; \theta)\}/\partial \theta^\top$. Note 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. This implies that (4.7) can

play a double role. In particular, if the score equations corresponding to (4.7) are solved with respect to θ with $p(b_i \mid T_i, \delta_i, y_i; \theta)$ fixed at the θ value of the previous iteration, then this corresponds to an EM algorithm, whereas if the score equations are solved with respect to θ considering $p(b_i \mid T_i, \delta_i, y_i; \theta)$ also as a function of θ, then this corresponds to a maximization of the observed data log-likelihood $\ell(\theta)$. As we will discuss further, this appealing feature allows for an easy interchange between maximization algorithms and in a straightforward calculation of standard errors.

Here the EM algorithm is used; the use of a Newton-type algorithm is briefly presented in Section 4.7. The measurement error variance in the longitudinal measurement model and the covariance matrix of the random effects are updated in the M-step according to the closed-form expressions

$$\begin{aligned}
\hat{\sigma}^2 &= N^{-1} \sum_i \int \left\{ (y_i - X_i\beta - Z_i b_i)^\top (y_i - X_i\beta - Z_i b_i) \right\} p(b_i \mid T_i, \delta_i, y_i; \theta)\, db_i \\
&= N^{-1} \sum_i (y_i - X_i\beta)^\top (y_i - X_i\beta - 2Z_i\tilde{b}_i) + \mathrm{tr}(Z_i^\top Z_i \widetilde{vb}_i) + \tilde{b}_i^\top Z_i^\top Z_i \tilde{b}_i, \\
\hat{D} &= n^{-1} \sum_i \widetilde{vb}_i + \tilde{b}_i \tilde{b}_i^\top,
\end{aligned}$$

where $N = \sum_i n_i$, $\tilde{b}_i = E(b_i \mid T_i, \delta_i, y_i; \theta)$, $\widetilde{vb}_i = \mathrm{var}(b_i \mid T_i, \delta_i, y_i; \theta)$. For the parameters $\theta_t^\top = (\omega^\top, \gamma^\top, \alpha)$ of the survival process closed-form solutions of the score equations do not exist and thus the M-step is implemented via an one-step Newton-Raphson update, given by

$$\hat{\beta}^{cr+1} = \hat{\beta}^{cr} - \left\{ \partial S(\hat{\beta}^{cr}) / \partial \beta \right\}^{-1} S(\hat{\beta}^{cr}) \tag{4.8}$$

$$\hat{\theta}_t^{cr+1} = \hat{\theta}_t^{cr} - \left\{ \partial S(\hat{\theta}_t^{cr}) / \partial \theta_t \right\}^{-1} S(\hat{\theta}_t^{cr}), \tag{4.9}$$

where $\hat{\beta}$ and $\hat{\theta}_t^{cr}$ denotes the values of β and θ_t at the current iteration, respectively, and $\partial S(\hat{\beta}^{cr})/\partial\beta$ and $\partial S(\hat{\theta}_t^{cr})/\partial\theta_t$ denotes the Hessian matrices evaluated at $\hat{\beta}^{cr}$

and $\hat{\theta}_t^{cr}$, respectively. The score vectors for β and θ_t have the form

$$S(\beta) = \sum_i X_i^\top \{y_i - X_i\beta - Z_i\tilde{b}_i\}/\sigma^2 + \int \Big(\alpha x_i(T_i)[\delta_i - \exp\{\eta_i(b_i)\}]\Big) p(b_i \mid T_i, \delta_i, y_i; \theta)\, db_i, \quad (4.10)$$

$$S(\theta_t) = \sum_i \int \left[\delta_i \frac{\partial}{\partial \theta_t^\top} \log \frac{\partial s(\log T_i; \omega)}{\partial \log T_i} + [\delta_i - \exp\{\eta_i(b_i)\}] \frac{\partial \eta_i(b_i)}{\partial \theta_t^\top}\right] p(b_i \mid T_i, \delta_i, y_i; \theta)\, db_i. \quad (4.11)$$

The Hessian matrices $\partial S(\beta)/\partial\beta$ and $\partial S(\theta_t)/\partial\theta_t$ are computed using a central difference approximation (Press et al., 2007).

4.3.2 Laplace Approximations for Posterior Moments

The calculation of the M-step updates, presented in Section 4.3.1, requires the specification of the E-step expectations. Under the score vector (4.7) these expectations are of the form

$$E\{A(b_i) \mid T_i, \delta_i, y_i; \theta\} = \int A(b_i) p(b_i \mid T_i, \delta_i, y_i; \theta)\, db_i \quad (4.12)$$

$$= \frac{\int A(b_i) p(T_i, \delta_i \mid b_i; \theta) p(y_i \mid b_i; \theta) p(b_i)\, db_i}{\int p(T_i, \delta_i \mid b_i; \theta) p(y_i \mid b_i; \theta) p(b_i)\, db_i}, \quad (4.13)$$

where $A(\cdot)$ denotes a function of the random effects, and θ in the posterior distribution $p(b_i \mid T_i, \delta_i, y_i; \theta)$ is fixed to its value from the previous iteration. For the M-step updates presented in Section 4.3.1, $A(\cdot)$ takes the following forms: $A(b_i) = b_i$, $A(b_i) = \exp\{\eta_i(b_i)\}$, and $A(b_i) = \exp\{\eta_i(b_i)\} z_i^\top(t) b_i$. Furthermore, for the closed-form updates of $\hat{\sigma}^2$ and $\hat{D}$ we also need $\widetilde{vb}_i = \text{var}\{A(b_i) \mid T_i, \delta_i, y_i; \theta\}$, with $A(b_i) = b_i$.

Unfortunately, the integrals in (4.12) and (4.13) do not have closed-form solutions and therefore numerical integration methods are usually employed for their

calculation. The Laplace method for integrals provides an appealing approximation in such settings, especially when the dimensionality of the integration is high. The Laplace approximation is typically applied in (4.12) by expanding the logarithm of the integrand using a second-order Taylor series around the mode with respect to b_i. The error of this approximation is of order $O(n_i^{-1})$. Here we apply a fully exponential Laplace approximation in the numerator and denominator of (4.13), as it has been proposed by Tierney et al. (1989) for approximating posterior moments in the Bayesian analysis context. Let $E\{A(b_i)\} = E\{A(b_i) \mid T_i, \delta_i, y_i; \theta\}$; Tierney et al.'s approximation is applied in positive functions and thus the cumulant generating function $\log E\big[\exp\{c^\top A(b_i)\}\big]$ is used, where $A(b_i)$ is assumed to have a nonzero derivative at $\hat{b}_i = \hat{b}_i^{(0)}$, with $\hat{b}_i^{(c)} = \arg\max_b\{\log p(T_i, \delta_i, y_i, b) + c^\top A(b)\}$. Then the required expectations are given as $E\{A(b_i)\} = \partial \log E\big[\exp\{c^\top A(b_i)\}\big]/\partial c^\top \mid_{c=0}$. The approximation is performed in two steps. First, the modes $\hat{b}_i$ are obtained for each subject using the Newton-Raphson scheme

$$\hat{b}_i^{it+1} = \hat{b}_i^{it} - \Sigma_i^{-1} L(\hat{b}_i^{it}), \tag{4.14}$$

where it denotes the iteration,

$$\begin{aligned} L(b_i) &= -\frac{\partial\big\{\log p(T_i, \delta_i \mid b_i) + \log p(y_i \mid b_i) + \log p(b_i)\big\}}{\partial b_i^\top} \\ &= -\alpha\big[\delta_i - \exp\{\eta_i(b_i)\}\big] z_i(t) - Z_i^\top (y_i - X_i\beta - Z_i b_i)/\sigma^2 + D^{-1} b_i, \end{aligned}$$

and $\Sigma_i = \Sigma_i^{(c)} \mid_{(c,b)=(0,\hat{b}_i)}$, with

$$\begin{aligned} \Sigma_i^{(c)} &= -\frac{\partial^2\big\{\log p(T_i, \delta_i \mid b_i) + \log p(y_i \mid b_i) + \log p(b_i) + c^\top A(b_i)\big\}}{\partial b_i^\top \partial b_i} \\ &= \alpha^2 \exp\{\eta_i(b_i)\} z_i(t) z_i^\top(t) + Z_i^\top Z_i/\sigma^2 + D^{-1} - \\ &\qquad \partial^2 c^\top A(b_i)/\partial b_i^\top \partial b_i. \end{aligned} \tag{4.15}$$

The Newton-Raphson algorithm in (4.14) converges very fast, since in this case the objective function is $g(b_i) = \log p(T_i, \delta_i \mid b_i) + \log p(y_i \mid b_i) + \log p(b_i)$, for which, especially as n_i increases, the leading term is $\log p(y_i \mid b_i) = \sum_j \log p\{y_i(t_{ij}) \mid b_i\}$, which is quadratic in b_i. In the second step, a fully exponential Laplace approximation is applied to $\log E\big[\exp\{c^\top A(b_i)\}\big]$; by differentiating the result and evaluating at $c = 0$ we obtain the following formulas for approximating $E\{A(b_i)\}$ and $\mathrm{var}\{A(b_i)\}$, respectively:

$$\begin{aligned} E\{A(b_i)\} &= A(\hat{b}_i) + \frac{\partial \log\big\{\det\big(\Sigma_i^{(c)}\big)\big\}^{-1/2}}{\partial c^\top} \mid_{(c,b)=(0,\hat{b}_i)} + O(n_i^{-2}) \\ &= A(\hat{b}_i) - \frac{1}{2}\mathrm{tr}(\mathcal{V}) + O(n_i^{-2}), \end{aligned} \tag{4.16}$$

$$\begin{aligned} \mathrm{var}\{A(b_i)\} &= A'^\top(\hat{b}_i)\Sigma_i^{-1}A'(\hat{b}_i) + \frac{\partial^2 \log\big\{\det\big(\Sigma_i^{(c)}\big)\big\}^{-1/2}}{\partial c^\top \partial c} \mid_{(c,b)=(0,\hat{b}_i)} + O(n_i^{-3}) \\ &= A'^\top(\hat{b}_i)\Sigma_i^{-1}A'(\hat{b}_i) - \\ &\quad \frac{1}{2}\mathrm{tr}\Big\{-\mathcal{V}\mathcal{V}^\top + \Sigma_i^{-1}\frac{\partial^2}{\partial c^\top \partial c}\Sigma_i^{(c)} \mid_{(c,b)=(0,\hat{b}_i)}\Big\} + O(n_i^{-3}), \end{aligned} \tag{4.17}$$

where $A'(b_i) = \partial A(b_i)/\partial b_i^\top$, and $\mathcal{V} = \Sigma_i^{-1}\{\partial \Sigma_i^{(c)}/\partial c^\top\} \mid_{(c,b)=(0,\hat{b}_i)}$. Thus, given the $\hat{b}_i$ values, the approximations require evaluation of $A(\hat{b}_i)$ and some extra correction terms the calculation of which is presented in Appendix A.

The appealing feature of the fully exponential Laplace approximation is that equations (4.16) and (4.17) are computed at $\hat{b}_i$, implying that only one optimization with respect to b_i is needed in each iteration. In order to appreciate the importance of this feature we compare this approximation with the common Laplace approximation, when an EM algorithm or a direct maximization of the log-likelihood is used to derive the maximum likelihood estimates. In particular, applying the common Laplace in the EM algorithm described above, requires an optimization with respect to b_i for *each* different $A(\cdot)$ function, increasing thus the computational burden considerably. On the other hand, in a direct maximi-

zation of the observed data log-likelihood only the integral in (4.6) needs to be approximated that requires one optimization. Note that in the common Laplace approximation of $\ell(\theta)$ the same $\hat{b}_i$'s are required as in the fully exponential Laplace approximation described above. Thus, in this case, the computational burden of the $O(n_i^{-2})$ and $O(n_i^{-1})$ approximations is of the same magnitude, since the calculation of the correction terms is not computationally demanding. However, the approximation error in the fully exponential Laplace method in (4.16) and (4.17) is of order $O(n_i^{-2})$ and $O(n_i^{-3})$, respectively, implying that the derived parameter estimators will enjoy better asymptotic properties, as we also discuss in Section 4.4, compared to the $O(n_i^{-1})$ approximation.

4.3.3 Standard Errors

One of the features of our model is that a flexible but parametric cumulative baseline hazard has been considered using B-splines. An advantage of this approach is that the calculation of standard errors is much simpler and follows from standard theory. This contradicts joint models in which the baseline hazard is left unspecified. In particular in this case, the use of the profile likelihood leads to underestimation of standard errors (Hsieh et al., 2006), and therefore the Bootstrap method is recommended, which renders joint models even more computationally demanding. Furthermore, even though we have used the EM algorithm for obtaining the maximum likelihood estimates (MLE), the calculation of standard errors is straightforward and does not require, for instance Louis (1982) formula. This is because, as we noted in Section 4.3.1, the complete data score vector used in the EM is also the observed data score vector that can be used to derive the standard errors. In particular, let $\hat{\theta}$ denote the MLEs, then from standard likelihood theory we have that $\text{var}(\hat{\theta}) = \left[E\left\{-\partial S(\theta)/\partial\theta\right\}\right]^{-1}$, where the expectation is taken with respect to the true joint density. Here we use the negative of the observed Hessian matrix $\mathcal{H}(\theta) = n^{-1}\sum_i \partial S_i(\theta)/\partial\theta$, evaluated at $\hat{\theta}$, as a consistent estimator of the

Fisher information matrix. Under (4.7) it can be easily shown that the elements of $\mathcal{H}$ are written in the following form

$$\frac{\partial S_i(\hat{\theta}_u)}{\partial \theta_{u'}} = E\Big\{\partial h(\hat{\theta}_u, b_i)/\partial \theta_{u'}\Big\} + E\Big[h(\hat{\theta}_u, b_i)\Big\{h(\hat{\theta}_{u'}, b_i) - S_i(\hat{\theta}_{u'})\Big\}^\top\Big],$$

where $u, u' = t, y, b$, $\theta^\top = (\theta_t^\top, \theta_y^\top, \theta_b^\top)$ is the vector of the parameters in the survival, longitudinal and random effect submodels, respectively, and the expectation is taken with respect to the posterior distribution of the random effects. We observe that the Hessian matrix is also written as the expectation of terms with nonzero derivative at $\hat{b}_i$, with respect to $p(b_i \mid T_i, \delta_i, y_i; \theta)$. This implies that $\mathcal{H}(\hat{\theta})$ can be sufficiently approximated using a numerical differentiation algorithm (e.g., a forward or central difference approximation) in $S(\hat{\theta})$. The elements of the score vector are presented in Section 4.7. Finally, it is often the case in joint modelling that the baseline hazard function is not of primary interest, which implies that ω in (4.4) should be treated as a vector of nuisance parameters. In this setting let $\theta_{-\omega}$ denote the vector of all other parameters excluding ω, then the covariance matrix of $\hat{\theta}_{-\omega}$ can be consistently estimated by $\hat{\text{var}}(\hat{\theta}_{-\omega}) = -\big\{\mathcal{H}(\hat{\theta}_{-\omega}) - \mathcal{H}(\hat{\theta}_{-\omega}, \hat{\omega})\mathcal{H}^{-1}(\hat{\omega})\mathcal{H}^\top(\hat{\theta}_{-\omega}, \hat{\omega})\big\}^{-1}$, where $\mathcal{H}(\theta_{-\omega}, \omega) = n^{-1}\sum_i \partial S_i(\theta_{-\omega})/\partial\omega$.

4.4 Asymptotic Behaviour of the Laplace Estimators

In this section we investigate the asymptotic properties of the Laplace maximum likelihood estimators $\hat{\theta}$ presented in Section 4.3.1. Our arguments are similar to those of Vonesh (1996) who discussed the asymptotic properties of the $O(n_i^{-1})$ Laplace estimators for nonlinear mixed models. In particular, let θ_0 denote the true parameter value, then under suitable regularity conditions it is shown in

Appendix C that

$$\hat{\theta} - \theta_0 = O_p[\max\{n^{-1/2}, \min(n_i)^{-2}\}]. \tag{4.18}$$

Thus, $\hat{\theta}$ will be consistent as long as both n and $\min(n_i)$ grow to infinity. Intuitively, the $n^{-1/2}$ term comes from standard asymptotic theory, while the $\min(n_i)^{-2}$ term comes from the Laplace approximation. Moreover, based on the derivation of equation (4.18) presented in Appendix C, we can deduce that if $\min(n_i) = O(n^{\rho})$ for $\rho > 1/2$, then the fully exponential Laplace maximum likelihood estimators will be asymptotically equivalent to the true maximum likelihood estimators that solve $S(\theta) = 0$. Finally, under the same regularity conditions, it can be shown in a similar manner that the standard Laplace method produces estimators with approximation error of order $O_p[\max\{n^{-1/2}, \min(n_i)^{-1}\}]$. This clearly demonstrates the superiority of approximations (4.16) and (4.17), since the common Laplace requires $\min(n_i) = O(n^{\rho})$ for $\rho > 1$ in order the MLEs not to be affected by approximation error.

4.5 Simulation Study

A simulation study has been performed to empirically investigate the finite sample performance of the Laplace based maximum likelihood estimators. In particular, we considered two simulation scenarios corresponding to high- and low-dimensional random effect structures, and three $\bar{n}_i$ settings, namely $\bar{n}_i = 5$, $\bar{n}_i = 12$ and $\bar{n}_i = 25$, where $\bar{n}_i$ denotes the average number of repeated measurements per subject. In the low-dimensional scenario random intercepts and random slopes have been considered, whereas in the high-dimensional scenario a fourth degree polynomial was posited. Under each scenario and $\bar{n}_i$ setting, 500 data sets were simulated with $n = 200$ subjects. Parameter estimates in each simulated data set were obtained using the EM algorithm described in Section 4.3.1. In the low-dimensional scenario we also fitted the model using the Gauss-Hermite quadrature

rule to approximate the E-step expectations and compare it to the Laplace based estimates. A detailed description of the simulation set-up and all tables with the simulation results can be found in Section 4.8. The results showed an overall good performance of the proposed Laplace approximation, even in the small $\bar{n}_i$ setting. In addition regarding computing time, we observed that in the low-dimensional scenario the Gauss-Hermite integration rule is faster that the Laplace method. However, in our simulations we have considered the simple Gauss-Hermite, which is expected to work satisfactorily when the integrand is proportional to a normal density with covariance matrix of a magnitude $2^{-1}I$, with I denoting here the identity matrix. When this is not the case, an adaptive Gauss-Hermite rule would be required making thus the computational burden much heavier.

4.6 Application

We continue with the analysis of the renal graft failure study which was introduced in Section 4.1. Out of the 407 patients considered, 126 experienced a graft failure resulting in about 70% censoring. The patients made on average 143.8 visits (standard deviation 57.1 visits), resulting in 58531 records. The minimum number of repeated measurement is 14, which according to (4.18) implies that the derived Laplace MLEs have the same accuracy as the true MLEs, since $\max\{407^{-1/2}, 14^{-2}\} = 407^{-1/2}$; however, note that this is not the case under the common Laplace approximation in which the error is of order $\max\{407^{-1/2}, 14^{-1}\} = 14^{-1}$. The submodels specification of the joint model (4.1) is as follows. In the survival model four internal knots (placed at the 20%, 40%, 60% and 80% quantiles of the uncensored log survival times) are used to capture the shape of the log cumulative baseline hazard. In the longitudinal process, and based on the ignorable analysis discussed in Section 4.1, a natural cubic spline with seven degrees of freedom is postulated to model the subject-specific trajectories. This formulation results in

Table 4.2: *Renal graft failure data. Parameter estimates with standard errors in parenthesis. For the longitudinal process 'NCS 1' - 'NCS 7' denote the parameters of the natural cubic spline. For the event process 'Assoct.' denotes the association parameter α. For the random effects submodel 'D_{11}' - 'D_{88}' denote the diagonal elements of the covariance matrix D.*

Survival Process		Longitudinal Process		Random Effects	
Param.	Value (Std. Err.)	Param.	Value (Std. Err.)	Param.	Value (Std. Err.)
Intercept	24.49 (0.085)	Gender	0.43 (0.200)	D_{11}	22.46 (1.629)
NCS 1	14.78 (0.405)	Weight	0.02 (0.009)	D_{22}	67.12 (4.962)
NCS 2	13.83 (0.008)	Assoct.	−0.17 (0.027)	D_{33}	61.22 (4.725)
NCS 3	14.22 (0.327)			D_{44}	66.05 (5.099)
NCS 4	13.16 (0.140)			D_{55}	59.08 (4.723)
NCS 5	8.13 (0.417)			D_{66}	76.34 (6.303)
NCS 6	22.19 (0.843)			D_{77}	120.94 (9.782)
NCS 7	11.22 (1.262)			D_{88}	289.47 (25.114)
Gender	0.73 (0.204)				
Weight	0.05 (0.016)				
σ	3.17 (0.010)				

an eight-dimensional random effects vector with respect to which the joint density (4.1) of the observed process is defined. Finally, the gender and weight covariate effects are considered in both submodels.

The model is fitted using the EM algorithm described in Section 4.3.1, and all computations have been performed in R. Initial values were obtained by fitting the ignorable linear mixed model for the longitudinal process, and model (4.3) for the event process, in which the calculation of $W_i(t)$ was based on the empirical Bayes estimates of the ignorable linear mixed model. As an illustration of the relative computational performance of our model we would like to mention that under the eight-dimensional random effects distribution the computing time for the E-step of one EM iteration using the Gauss-Hermite rule with only three quadrature points is greater than 25 min. (i.e., we stopped the program after 25 min.), whereas the proposed Laplace approximation takes only 12 sec. The parameter estimates and associated standard errors are presented in Table 4.2. The estimated association parameter α is negative and highly significant indicating that the higher the average haematocrit values, the longer the graft survives. Furthermore, the estimates of the natural cubic spline coefficients suggest a strong nonlinear time effect for the

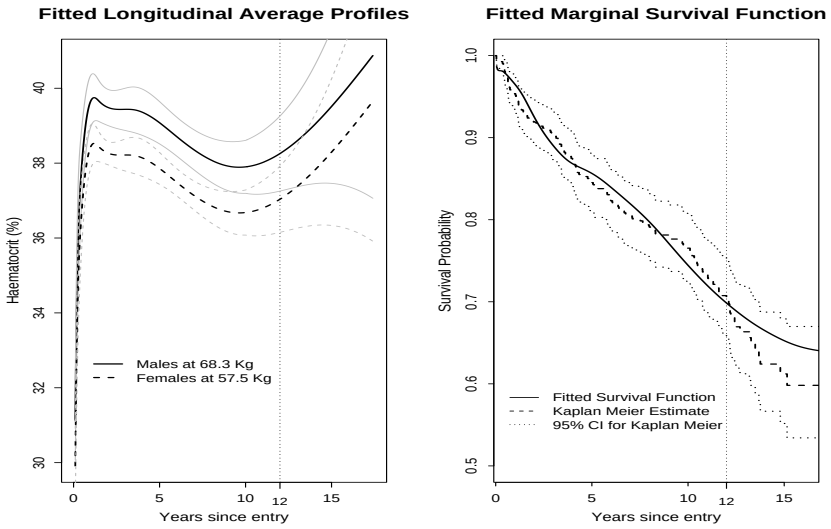

Figure 4.2: *Renal graft failure data. Fitted longitudinal average profiles (with associated 95% pointwise CI) and marginal survival function. The average longitudinal profiles have been computed for males and females at the median weight. In the plot of the marginal survival function the Kaplan-Meier estimate (with associated 95% CI) for the time to graft failure is also superimposed.*

average longitudinal evolutions that is illustrated in Figure 4.2. We observe that the model captures the initial and expected increase in haematocrit levels after the transplant as well as the decline in the following years due to the kidney problems. The unexpected increase in the average haematocrit levels after the twelfth year since entry is attributed to the small amount of available information. In particular, only 4.5% of the total number of measurements were taken after twelve years, which results in very wide confidence intervals for the average haematocrit levels in this area. Figure 4.2 depicts as well the fitted marginal survival function for the time to graft failure that is computed according to the following expression

$$\mathcal{S}(t) = \int \mathcal{S}(t \mid b_i; \hat{\theta}) p(b_i; \hat{\theta})\, db_i \approx n^{-1} \sum_i \mathcal{S}(t \mid \hat{b}_i; \hat{\theta}).$$

We observe some discrepancy between the fitted survival function after the twelfth

year since entry and the Kaplan-Meier estimate. This discrepancy is explained first by the fact that the Kaplan-Meier estimate does not account for the informative censoring, and second by the small number of events in this area (i.e., the last internal knot is placed at 10.5 years), and thus it does not pose a major concern.

4.7 Discussion

We have proposed a new computational approach for the joint modelling of longitudinal measurements and time-to-event data, and demonstrated its use through a real data example. The main strength of this framework is that it effectively copes with high-dimensional random effects structures without increasing substantially the computational burden. Furthermore, the B-splines formulation of the log cumulative baseline hazard allows for great flexibility in the specification of the survival submodel, whereas it does not pose complications in the calculation of standard errors. Our main motivation to flexibly model the log cumulative baseline hazard using B-splines was the easy in incorporating the required monotonicity constraint; however, other smoothing methods (e.g., kernels or local regression) can be considered as well. Another straightforward extension of the proposed framework is to handle categorical longitudinal responses by postulating a generalized mixed model for the longitudinal process. This would require to adapt the M-step update for β and approximate the expectation of different $A(\cdot)$ functions depending on the type of categorical responses (e.g., binomial, poisson, etc.).

Furthermore, in this paper we have used the EM algorithm to obtain the maximum likelihood estimates. A direct maximization of the observed data log-likelihood can be also considered using a quasi-Newton algorithm, such as the BFGS (Lange, 2004). The BFGS algorithm requires the specification of the observed data score vector corresponding to $\ell(\theta)$. However, as noted in Section 4.3.1,

$S(\theta)$ used to derive the M-step updates is also the observed data score vector. Thus, similarly to the M-step updates, we can easily derive the components of the score vector $S(\theta)$ corresponding to σ^2 and θ_b (i.e., the vector parameterizing the covariance matrix D), which take the form:

$$\begin{aligned} S(\sigma^2) &= -\frac{N}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_i (y_i - X_i\beta)^\top \{y_i - X_i\beta - 2Z_i\tilde{b}_i(\theta)\} + \\ &\qquad \operatorname{tr}\big\{Z_i^\top Z_i \widetilde{vb}_i(\theta)\big\} + \tilde{b}_i^\top(\theta) Z_i^\top Z_i \tilde{b}_i(\theta), \\ S(\theta_b) &= -\frac{n}{2}\operatorname{tr}(D^{-1}\partial D/\partial\theta_b) + \frac{1}{2}\sum_i \operatorname{tr}\big\{D^{-1}(\partial D/\partial\theta_b)D^{-1}\widetilde{vb}_i(\theta)\big\} + \\ &\qquad \tilde{b}_i^\top(\theta) D^{-1}\frac{\partial D}{\partial\theta_b} D^{-1}\tilde{b}_i(\theta), \end{aligned}$$

while for the score vectors for β and the parameters of the event process are given in (4.10) and (4.11), respectively. Furthermore, in the above equations we have used the notation $\tilde{b}_i(\theta)$ and $\widetilde{vb}_i(\theta)$ to stress that these are also treated as functions of θ.

Finally, an issue with the Laplace approximation (4.16) is that it cannot be applied for functions $A(b_i)$ for which $A'(\hat{b}_i) = 0$. This excludes the possibility of applying this approximation to compute the log-likelihood, which could be required for instance to perform a likelihood ratio test. In this case $\ell(\theta)$ can be easily approximated with the common Laplace approximation (i.e., using the values of $\hat{b}_i$ and Σ_i that have been already computed) by

$$\begin{aligned} \ell(\theta) &= \frac{n\kappa}{2}\log 2\pi + \Big[\sum_i \log p(T_i, \delta_i \mid \hat{b}_i) + \log p(y_i \mid \hat{b}_i) + \log p(\hat{b}_i) \\ &\qquad -\frac{1}{2}\log\{\det(\Sigma_i)\}\Big] + O\big\{n \min(n_i)^{-1}\big\}, \end{aligned}$$

where κ denotes the dimensionality of the random effects vector.

4.8 Supplementary Material

4.8.1 Simulation Study Set-up

A simulation study has been performed to empirically investigate the finite sample performance of the proposed model. In all cases, for each of $n = 200$ subjects, the event times were simulated according to the model

$$\log H_i(t \mid b_i) = \log H_0(t) + \alpha W_i(t),$$

where $\alpha = -1$ and $H_0(t) = 10^{-5}t$. The censoring mechanism follows an exponential distribution with mean 6.5, which results in about 30% censoring on average. For the longitudinal process we consider two simulation scenarios corresponding to high- and low-dimensional random effects structures, and three $\bar{n}_i$ settings, namely $\bar{n}_i = 5$, $\bar{n}_i = 12$ and $\bar{n}_i = 25$, where $\bar{n}_i$ denotes the average number of repeated measurements per subject. In particular, for the low-dimensional case we take

$$W_i(t_{ij}) = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i})t_{ij},$$

where $(\beta_0, \beta_1) = (2, -6.4)$, and $t_{ij} = \text{seq}(0, 5, 15)$[2] for $\bar{n}_i = 5$, $t_{ij} = \text{seq}(0, 5, 30)$ for $\bar{n}_i = 12$ and $t_{ij} = \text{seq}(0, 5, 70)$ for $\bar{n}_i = 25$. The measurement error variance is taken $\sigma^2 = 0.6$, and the elements of the covariance D for the random effects are $\text{var}(b_{0i}) = 5$, $\text{var}(b_{1i}) = 1.8$ and $\text{cov}(b_{0i}, b_{1i}) = -0.7$. For the high-dimensional scenario, we simulate censoring times from an exponential distribution with mean 2.5 (in order to have again 30% censoring on average), and for the longitudinal process we take

$$W_i(t_{ij}) = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i})t_{ij} + (\beta_2 + b_{2i})t_{ij}^2 + (\beta_3 + b_{3i})t_{ij}^3,$$

[2] where $\text{seq}(a, b, c)$ denotes a regular sequence from a to b of length c, e.g., $\text{seq}(0, 2, 5) = 0, 0.5, 1, 1.5, 2$.

where $(\beta_0, \beta_1, \beta_2, \beta_3) = (2, -6.4, -4.5, -2.8)$, and $t_{ij} = \text{seq}(0, 3, 20)$ for $\bar{n}_i = 5$, $t_{ij} = \text{seq}(0, 3, 45)$ for $\bar{n}_i = 12$ and $t_{ij} = \text{seq}(0, 3, 90)$ for $\bar{n}_i = 25$. The measurement error variance is taken $\sigma^2 = 0.6$, and the elements of the covariance D for the random effects are

$$D = \begin{bmatrix} 5 & -0.3969 & -0.3182 & -0.2372 \\ -0.3969 & 1.4 & 0.1122 & 0.0837 \\ -0.3182 & 0.1122 & 0.9 & 0.0671 \\ -0.2372 & 0.0837 & 0.0671 & 0.5 \end{bmatrix}.$$

Finally, under each scenario and $\bar{n}_i$ setting, 500 data sets were simulated.

In both scenarios we fitted the EM algorithm described in Section 3 of the paper, using the the proposed Laplace approximation. In the low-dimensional scenario we also fitted the model using the Gauss-Hermite quadrature rule (with 35 quadrature points for each dimension) to approximate the E-step expectations, and compare it to the Laplace based estimates.

4.8.2 Simulation Study Results

Descriptives for the computing time (in minutes) under the different scenarios and sample size settings are presented in Table 4.3. Computations have been performed in R version 2.5.0, on an AMD Opteron Cluster, consisting of 164 dual Opteron250 servers running Linux (kernel version 2.6.15.7), with 2GB RAM, several nodes with 16GB RAM, and 4 to 8 CPUs with clock speeds varying from 1.8 to 2.6 GHz. We observe that in the low-dimensional scenario the Gauss-Hermite integration rule is faster than the Laplace method. However, in our simulations we have considered the simple Gauss-Hermite rule, which is expected to work satisfactorily when the integrand is proportional to a normal density with covariance matrix of a magnitude $2^{-1}I$, with I denoting the identity matrix. When this is not the case, an adaptive Gauss-Hermite rule would be required making thus the computational

burden much heavier. The average computing time required in the Laplace method for the second scenario is more than the corresponding computing time for the first one. We would like to note that the reason for this is that in the second scenario we have nine more parameters to estimate, which inevitably requires more iterations for the maximization algorithm. However, the average computing time required per iteration for the computation of the modes $\hat{b}_i$ is in the two scenarios of the same magnitude. Finally, an interesting feature for the Laplace method is that computing times are decreasing with increasing $\bar{n}_i$. After careful investigation, we discovered that this feature is explained as follows. As also noted above, the computing time for the Laplace approximation, performed in the E-step, does not substantially increase with $\bar{n}_i$. However, for small $\bar{n}_i$ more EM iterations are required to locate the MLEs, since less information is available in the data.

Tables 4.4 and 4.5 present the bias and root mean square error (RMSE) for the association parameter, the fixed-effects parameters of the longitudinal model, and the measurement error standard deviation. For all scenarios the proposed model showed a good performance in terms of bias and RMSE. In addition and as expected, the RMSE for the Laplace based estimates decreases with increasing $\bar{n}_i$. Furthermore, in the low-dimensional scenario we observe that the Laplace based estimates performs slightly better than the estimates based on the Gauss-Hermite method. As we noted above, this could be attributed to the fact that an adaptive Gauss-Hermite rule might be required since in this case. In the high-dimensional scenario and for the quadratic β_2 and cubic β_2 fixed effect parameters, we observed relatively larger bias and RMSE compared to the linear effect β_1, especially for $\bar{n}_i = 5$. This is attributed to the fact that when few repeated measurements per individual are available, there is considerably less information in the data for β_2 and β_2 compared to the other $\bar{n}_i$ settings.

Table 4.3: *Descriptives for the computing time (in minutes) under the two scenarios and the three $\bar{n}_i$ settings. The top and middle parts contain the results for the Gauss-Hermite rule (with 35 quadrature points) and the Laplace approximation, respectively, under the low-dimensional scenario; the bottom part contains the results for the Laplace approximation under the high-dimensional scenario. '1st Qu.' and '3rd Qu.' denote the first and third quartiles, respectively.*

	$\bar{n}_i = 5$	$\bar{n}_i = 12$	$\bar{n}_i = 25$
1st Qu.	12.317	12.756	14.075
Median	12.898	13.586	15.231
Mean	13.270	13.921	15.811
3rd Qu.	14.528	15.360	18.146
1st Qu.	16.230	6.408	5.855
Median	47.030	21.670	19.726
Mean	48.100	37.140	36.060
3rd Qu.	71.510	60.870	53.150
1st Qu.	34.460	19.250	19.450
Median	54.200	37.390	26.140
Mean	59.600	51.820	51.860
3rd Qu.	78.320	75.320	73.600

Table 4.4: *Simulation results for the low-dimensional scenario based on 500 data sets with sample size $n = 200$. The bias and root mean square error (RMSE) for parameters of the survival and longitudinal models are presented. The top part contains the results for the Gauss-Hermite rule with 35 quadrature points, whereas the bottom part contains the results for the Laplace approximation.*

	True	$\bar{n}_i = 5$		$\bar{n}_i = 12$		$\bar{n}_i = 25$	
	Value	Bias	RMSE	Bias	RMSE	Bias	RMSE
α	-1.0	-0.1936	1.4585	-0.0854	0.1156	-0.1526	1.2505
β_0	2.0	-0.0813	0.1880	-0.0712	0.1751	-0.0311	0.2020
β_1	-6.4	0.1756	0.2413	0.0949	0.1711	0.0348	0.1703
σ	0.8	0.0142	0.1393	0.0022	0.0656	0.0137	0.1333
α	-1.0	-0.0861	0.1394	-0.0512	0.0984	-0.0249	0.0694
β_0	2.0	-0.0767	0.1571	-0.0449	0.1413	-0.0269	0.1192
β_1	-6.4	0.1108	0.1633	0.0618	0.1209	0.0225	0.0959
σ	0.8	0.0026	0.0217	0.0001	0.0137	0.0001	0.0083

Table 4.5: *Simulation results for the high-dimensional scenario using the Laplace approximation based on 500 data sets with sample size $n = 200$. The bias and root mean square error (RMSE) for parameters of the survival and longitudinal models are presented.*

	True	$\bar{n}_i = 5$		$\bar{n}_i = 12$		$\bar{n}_i = 25$	
	Value	Bias	RMSE	Bias	RMSE	Bias	RMSE
α	-1.0	0.0961	0.2391	-0.0793	0.1759	0.0438	0.0932
β_0	2.0	-0.0515	0.1198	-0.0659	0.1371	-0.0597	0.1247
β_1	-6.4	0.9052	1.0919	0.3299	0.8953	0.1678	0.6093
β_2	-4.5	1.3369	1.7968	0.7107	1.1712	0.0930	0.8035
β_3	-2.8	1.1307	1.4540	0.5192	1.2381	0.1562	0.7636
σ	0.8	0.0104	0.0263	0.0021	0.0128	0.0001	0.0086

4.9 Appendix A

For the calculation of the correction terms we first require the following results

$$\begin{aligned}
\frac{\partial \Sigma_i^{(c)}}{\partial c^\top} \mid_{(c,b)=(0,\hat{b}_i)} &= \frac{\partial}{\partial \hat{b}_i^{(c)}} \Sigma_i^{(c)} \frac{\partial \hat{b}_i^{(c)}}{\partial c^\top} \mid_{(c,b)=(0,\hat{b}_i)} = \frac{\partial \Sigma_i}{\partial b^\top} \frac{\partial \hat{b}_i^{(c)}}{\partial c^\top} \mid_{c=0}, \\
\frac{\partial^2 \Sigma_i^{(c)}}{\partial c^\top \partial c} \mid_{(c,b)=(0,\hat{b}_i)} &= \frac{\partial^2 \Sigma_i}{\partial b^\top \partial b} \left\{ \frac{\partial \hat{b}_i^{(c)}}{\partial c} \frac{\partial \hat{b}_i^{(c)}}{\partial c^\top} \right\}_{c=0} + \frac{\partial \Sigma_i}{\partial b^\top} \frac{\partial^2 \hat{b}_i^{(c)}}{\partial c^\top \partial c} \mid_{c=0},
\end{aligned}$$

where, in Appendix B, we show that $\partial \hat{b}_i^{(c)} / \partial c^\top \mid_{c=0} = \Sigma_i^{-1} A'(\hat{b}_i)$, and $\partial^2 \hat{b}_i^{(c)} / \partial c^\top \partial c \mid_{c=0}$ $= \Sigma_i^{-1} \big[A''(\hat{b}_i)\, \Sigma_i^{-1} A'^\top(\hat{b}_i) + A''(\hat{b}_i) \Sigma_i^{-1} A'(\hat{b}_i) - (\partial \Sigma_i / \partial b_i) \{ \Sigma_i^{-1} A'^\top(\hat{b}_i) \} \{ \Sigma_i^{-1} A'(\hat{b}_i) \} \big]$, with $A''(b_i) = \partial^2 A(b_i) / \partial b_i^\top \partial b_i$.

For the calculation of the terms $\{\partial \Sigma_i^{(c)} / \partial c^\top\}$ and $\{\partial^2 \Sigma_i^{(c)} / \partial c^\top \partial c\}$ at $(c, b) = (0, \hat{b}_i)$, we observe from (4.15) that only $\exp\{\eta_i(b)\}$ and $\partial^2 c^\top A(b) / \partial b^\top \partial b$ are fu-

nctions of b, for which we obtain

$$\begin{aligned}
\left[\frac{\partial}{\partial c^\top}\Big\{\frac{\partial^2 c^\top A\{\hat{b}_i^{(c)}\}}{\partial b^\top \partial b}\Big\}\right]_{c=0} &= \frac{\partial^2 A(\hat{b}_i)}{\partial b^\top \partial b} + \left[c^\top\Big\{\frac{\partial}{\partial c^\top}\frac{\partial^2 A\{\hat{b}_i^{(c)}\}}{\partial b^\top \partial b}\Big\}\right]_{c=0} \\
&= \frac{\partial^2 A(\hat{b}_i)}{\partial b^\top \partial b}, \\
\left[\alpha^2 z_i(t) z_i^\top(t)\frac{\partial}{\partial c^\top}\exp\{\eta_i(\hat{b}_i^{(c)})\}\right]_{c=0} &= \alpha^3 z_i(t) z_i^\top(t)\exp\{\eta_i(\hat{b}_i)\}\mathcal{R}, \\
\left[\alpha^2 z_i(t) z_i^\top(t)\frac{\partial^2}{\partial c^\top \partial c}\exp\{\eta_i(\hat{b}_i^{(c)})\}\right]_{c=0} &= \alpha^3 z_i(t) z_i^\top(t)\exp\{\eta_i(\hat{b}_i)\}\Big[\alpha\mathcal{R}\mathcal{R}^\top + \\
&\qquad z_i^\top(t)\frac{\partial^2 \hat{b}_i^{(c)}}{\partial c^\top \partial c}\,|_{c=0}\Big],
\end{aligned}$$

where $\mathcal{R} = z_i^\top(t)\Sigma_i^{-1}A'(\hat{b}_i)$. We have not presented the calculation of the term

$$\left[\frac{\partial^2}{\partial c^\top \partial c}\Big\{\frac{\partial^2 c^\top A\{\hat{b}_i^{(c)}\}}{\partial b^\top \partial b}\Big\}\right]_{c=0},$$

since this involves the product of $A'''(b_i) = \partial A''(b_i)/\partial b_i$ with other terms and needs to be calculated only for $A(b_i) = b_i$, which implies that it drops. Finally, in order to apply (4.16) and (4.17) in the M-step updates presented in Section 4.3.1, $A'(b_i)$ and $A''(b_i)$ are required for $A(b_i) = \exp\{\eta_i(b_i)\}$, and $A(b_i) = \exp\{\eta_i(b_i)\}z_i^\top(t)b_i$; for completeness these are also presented in Appendix B.

4.10 Appendix B

Here we present the calculation of the terms $\partial\hat{b}_i^{(c)}/\partial c^\top$ and $\partial^2\hat{b}_i^{(c)}/\partial c^\top \partial c$ evaluated at $c = 0$. In particular, let $\mathcal{K}(b_i) = \log p(T_i, \delta_i \mid b_i) + \log p(y_i \mid b_i) + \log p(b_i)$, then $\hat{b}_i^{(c)}$ is chosen such that

$$\frac{\partial}{\partial b_i}\big\{\mathcal{K}(b_i) + A^\top(b_i)\, c\big\}_{b_i=\hat{b}_i^{(c)}} = 0,$$

from which we obtain

$$\mathcal{K}'(\hat{b}_i^{(c)}) + A'^{\top}(\hat{b}_i^{(c)})\, c = 0 \Rightarrow \frac{\partial}{\partial c^{\top}}\left\{\mathcal{K}'(\hat{b}_i^{(c)}) + A'^{\top}(\hat{b}_i^{(c)})\, c\right\} = 0 \Rightarrow$$

$$\mathcal{K}''(\hat{b}_i^{(c)})\frac{\partial \hat{b}_i^{(c)}}{\partial c^{\top}} + A'(\hat{b}_i^{(c)}) + c^{\top} A''(\hat{b}_i^{(c)})\frac{\partial \hat{b}_i^{(c)}}{\partial c^{\top}} = 0 \overset{c=0}{\Longrightarrow} \tag{4.19}$$

$$\frac{\partial \hat{b}_i^{(c)}}{\partial c^{\top}} \mid_{c=0} = \left\{-\mathcal{K}''(\hat{b}_i)\right\}^{-1} A'(\hat{b}_i) = \Sigma_i^{-1} A'(\hat{b}_i).$$

where $\mathcal{K}'(b_i) = \partial \mathcal{K}(b_i)/\partial b_i^{\top}$, and $\mathcal{K}''(b_i) = \partial^2 \mathcal{K}(b_i)/\partial b_i^{\top}\partial b_i$. Similarly, from (4.19) we derive

$$\frac{\partial}{\partial c}\left\{\mathcal{K}''(\hat{b}_i^{(c)})\frac{\partial \hat{b}_i^{(c)}}{\partial c^{\top}} + A'(\hat{b}_i^{(c)}) + c^{\top} A''(\hat{b}_i^{(c)})\frac{\partial \hat{b}_i^{(c)}}{\partial c^{\top}}\right\} = 0 \Rightarrow$$

$$\mathcal{K}'''(\hat{b}_i^{(c)})\frac{\partial \hat{b}_i^{(c)}}{\partial c}\frac{\partial \hat{b}_i^{(c)}}{\partial c^{\top}} + \mathcal{K}''(\hat{b}_i^{(c)})\frac{\partial^2 \hat{b}_i^{(c)}}{\partial c^{\top}\partial c} + A''(\hat{b}_i^{(c)})\frac{\partial \hat{b}_i^{(c)}}{\partial c} + A''(\hat{b}_i^{(c)})\frac{\partial \hat{b}_i^{(c)}}{\partial c^{\top}} + c^{\top}\frac{\partial}{\partial c}\left\{A''(\hat{b}_i^{(c)})\frac{\partial \hat{b}_i^{(c)}}{\partial c^{\top}}\right\} \overset{c=0}{\Longrightarrow}$$

$$\frac{\partial^2 \hat{b}_i^{(c)}}{\partial c^{\top}\partial c} \mid_{c=0} = \Sigma_i^{-1}\Big[\mathcal{K}'''(\hat{b}_i)\Sigma_i^{-1}A'^{\top}(\hat{b}_i)\Sigma_i^{-1}A'(\hat{b}_i) + A''(\hat{b}_i)\Sigma_i^{-1}A'^{\top}(\hat{b}_i) + A''(\hat{b}_i)\Sigma_i^{-1}A'(\hat{b}_i)\Big],$$

where $\mathcal{K}'''(b_i) = \partial \mathcal{K}''(b_i)/\partial b_i = -\partial \Sigma_i/\partial b_i$. Note that $\partial^2 \hat{b}_i^{(c)}/\partial c^{\top}\partial c \mid_{c=0}$ is required for the calculation of the term $\partial^2 \Sigma_i^{(c)}/\partial c^{\top}\partial c$ in equation (15), which is only computed for $A(b_i) = b_i$. Thus, we obtain the following simplification

$$\frac{\partial^2 \hat{b}_i^{(c)}}{\partial c^{\top}\partial c} \mid_{c=0} = \Sigma_i^{-1}\big[\mathcal{K}'''(\hat{b}_i)\Sigma_i^{-1}\mathbf{1}^{\top}\Sigma_i^{-1}\mathbf{1}\big],$$

where $\mathbf{1}$ denotes the summing vector of conforming dimensions. Finally, the first and second order partial derivatives for $A(b_i) = \exp\{\eta_i(b_i)\}$ and $A(b_i) = \exp\{\eta_i(b_i)\}z_i^\top(t)b_i$ considered in Section 4.3.2 are given by: $\partial \exp\{\eta_i(b_i)\}/\partial b_i^\top = \alpha \exp\{\eta_i(b_i)\}z_i(t)$, $\partial^2 \exp\{\eta_i(b_i)\}/\partial b_i^\top \partial b_i = \alpha^2 \exp\{\eta_i(b_i)\}z_i(t)z_i^\top(t)$, $\partial\big[\exp\{\eta_i(b_i)\}z_i^\top(t)b_i\big]/\partial b_i^\top = \exp\{\eta_i(b_i)\}\big\{\alpha z_i^\top(t)b_i+1\big\}z_i(t)$, and $\partial^2\big[\exp\{\eta_i(b_i)\}z_i^\top(t)b_i\big]/\partial b_i^\top \partial b_i = \alpha \exp\{\eta_i(b_i)\}z_i(t)z_i^\top(t)\big\{\alpha z_i^\top(t)b_i + 2\big\}$.

4.11 Appendix C

We work under the following assumptions: (i) $p(T_i, \delta_i, y_i; \theta)$ is a well-defined density under the usual regularity conditions (Cox and Hinkley, 1974, p. 281); (ii) the true parameter value θ_0 is an interior point of the parameter space, and the Laplace estimator $\hat{\theta}$ is an interior point in a neighbourhood containing θ_0; (iii) $\hat{b}_i = \arg\max_b\{\log p(T_i, \delta_i, y_i, b)\}$ exists for all i. Let $\hat{S}(\cdot)$ denote the approximated score vector according to (4.16), then we obtain

$$\begin{aligned} S(\hat{\theta}) &= \sum_i \big\{\hat{S}_i(\hat{\theta}) + O(n_i^{-2})\big\} \Rightarrow \\ n^{-1}S(\hat{\theta}) &= n^{-1}\hat{S}(\hat{\theta}) + O\big\{\min(n_i)^{-2}\big\}, \end{aligned} \tag{4.20}$$

where $\hat{S}(\hat{\theta}) = 0$. Under the regularity conditions in (i) we can apply a Taylor series expansion in $S(\theta)$ around θ_0

$$S(\hat{\theta}) = S(\theta_0) + \mathcal{H}(\theta^*)(\hat{\theta} - \theta_0), \tag{4.21}$$

where θ^* lies on the segment joining θ_0 and $\hat{\theta}$. From (4.21) and (4.20) we obtain

$$\begin{aligned} n^{1/2}(\hat{\theta} - \theta_0) &= -\Big[n^{-1}\sum_i \mathcal{H}_i(\theta^*)\Big]^{-1}\Big\{n^{-1/2}\sum_i S_i(\theta_0) - S_i(\hat{\theta})\Big\} \\ &= -\Big[n^{-1}\sum_i \mathcal{H}_i(\theta^*)\Big]^{-1}\Big[n^{-1/2}S(\theta_0) + O\{n^{1/2}\min(n_i)^{-2}\}\Big]. \end{aligned} \tag{4.22}$$

In addition, under assumptions (i) and (ii) we have that as $n \to \infty$, $n^{-1}\mathcal{H}(\theta^*) \xrightarrow{P} E\{\mathcal{H}(\theta_0)\}$, where the expectation is taken with respect to $p(T_i, \delta_i, y_i; \theta_0)$, and $\mathcal{H}(\theta) = \sum_i \mathcal{H}_i(\theta)$. By further assuming that $E\{\mathcal{H}(\theta_0)\}$ is nonsingular we obtain

$$\big\{n^{-1}\mathcal{H}(\theta^*)\big\}^{-1} \xrightarrow{P} \big[E\{\mathcal{H}(\theta_0)\}\big]^{-1}. \tag{4.23}$$

Therefore, combining (4.22) and (4.23) we obtain the desired result

$$\begin{aligned} n^{1/2}(\hat{\theta} - \theta_0) &= -\big[E\{\mathcal{H}(\theta_0)\}\big]^{-1}\Big[n^{-1/2}S(\theta_0) + O_p\big\{n^{1/2}\min(n_i)^{-2}\big\}\Big] \Rightarrow \\ (\hat{\theta} - \theta_0) &= -\big[E\{\mathcal{H}(\theta_0)\}\big]^{-1}\Big[n^{-1}S(\theta_0) + O_p\big\{\min(n_i)^{-2}\big\}\Big] \\ &= O_p[\max\{n^{-1/2}, \min(n_i)^{-2}\}], \end{aligned}$$

where in the last step we use the fact that under the regularity conditions (i), $n^{-1}S(\theta_0) = O_p(n^{-1/2})$, and $E\{\mathcal{H}(\theta_0)\} = O_p(1)$.

References

Cox, D. and Hinkley, D. (1974). *Theoretical Statistics.* Chapman & Hall, London.

De Bruijn, N. (1981). *Asymptotic Methods in Analysis.* Dover, New York.

Dierckx, P. (1995). *Curve and Surface Fitting with Splines.* Oxford University Press, New York.

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

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

Hsieh, F., Tseng, Y.-K. and Wang, J.-L. (2006). Joint modeling of survival and longitudinal data: Likelihood approach revisited. *Biometrics*, **62**, 1037–1043.

Lange, K. (2004). *Optimization*. Springer-Verlag, New York.

Louis, T. (1982). Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society, Series B*, **44**, 226–233.

Press, W., Teukolsky, S., Vetterling, W. and Flannery, B. (2007). *Numerical Recipes: The Art of Scientific Computing*. Cambridge University Press, New York, 3rd edition.

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

Rosenberg, P. (1995). Hazard function estimation using B-splines. *Biometrics*, **51**, 874–887.

Royston, P. and Parmar, M. (2002). Flexible parametric propostional-hazards and proportional-odds models for censored survival data with application to prognostic modelling and estimation of treatment effects. *Statistics in Medicine*, **21**, 2175–2197.

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

Steel, B. (1996). A modified EM algorithm for estimation in generalized mixed models. *Biometrics*, **52**, 1295–1310.

Tierney, L., Kass, R. and Kadane, J. (1989). Fully exponential Laplace approximations to expectations and variances of nonpositive functions. *Journal of the American Statistical Association*, **84**, 710–716.

Tsiatis, A. and Davidian, M. (2004). Joint modeling of longitudinal and time-to-event data: an overview. *Statistica Sinica*, **14**, 809–834.

Vonesh, E. (1996). A note on the use of Laplace's approximation for nonlinear mixed-effects models. *Biometrika*, **83**, 447–452.

CHAPTER 5 General Conclusion

5.1 Discussion

5.1.1 Choice of Parameterization & Questions of Interest

The typical parameterization in the joint modelling framework assumes that the linear predictor of the longitudinal mixed model conditional on the random effects is included as a time-dependent covariate in the linear predictor of the model for the event process. This implies that a single set of random effects is shared by the two processes. In Chapters 2 and 3 we have considered a more flexible parameterization using two sets of random effects, one for each process. The association in this case is modelled through the joint distribution function of the two random effects.

The choice between these two parameterizations depends on the question of interest, and in particular, whether the focus of inference is on measuring the effect of the longitudinal time-dependent covariate in the survival outcome, or on investigating the association structure between the two processes. In particular,

the standard parameterization is given by

$$\eta_{ti} = x_{ti}^{\top}\gamma + \alpha(x_{yi}^{\top}\beta + z_{yi}^{\top}b_i), \tag{5.1}$$

where i denotes the subject, x_{ti} and x_{yi} are the covariate vectors for the fixed effects in the longitudinal and dropout processes, respectively, and z_{yi} is the design matrix for the random effects b_i in the longitudinal process. Under (5.3), the association parameter *directly* quantifies the effect of the longitudinal process in the event process, and thus this parameterization is preferable when this effect is the focus of inference. However, when interest is in the association structure between the two processes, then the two random effects parameterization offers greater flexibility. To justify this, note that the one set random effects parameterization is a special case of the two random effects parameterization when the two random effects are perfectly correlated. Furthermore, the copula formulation of the random effects joint distribution facilitates the investigation of the dependence structure and in particular, checking whether different copulas may influence the estimated size of the association, as we observed for the GFR and Proteinuria markers in Chapters 2 and 3, respectively.

5.1.2 Random Effects Misspecification

In Chapter 2 we have shown that misspecification of the random effects distribution does not affect consistency as the number of repeated measurements per individual grows. This argument is based on two facts. First, the score vector corresponding to the observed data log-likelihood can be written as the expected value of the score vector corresponding to the complete data log-likelihood with respect to the posterior distribution of the random effects given the observed data. Second, as n_i grows, the model for the longitudinal process, which was assumed correctly specified, becomes the leading term in the posterior distribution of the random

effects, which implies that misspecification of the prior distribution for the random effects does not influence the results very much.

In practice, however longitudinal measurements are taken at a finite set of time points. Furthermore, even though the theorem presented in Chapter 2 applies for all well defined densities, the actual number of repeated measurements required in order for the above result to hold will depend on the type of the posited longitudinal model. In particular, a small n_i will be sufficient as long as $\log p(y_i \mid b_i)$ can be well approximated by a quadratic function of b_i, near its mode $\hat{b}_i$. For instance, in the exponential family we expect that normal longitudinal responses will require smaller n_i compared to categorical responses, and among the members of the family for categorical responses, Poisson data will need smaller n_i than binary outcomes.

In order to protect the resulting inferences against misspecification of the random effects distribution, an adequate specification of the longitudinal model is of major importance. Thus, in addition to a good data exploration for the choice of the mean and association structures, relatively flexible models should be used, which are able to capture the characteristics of the data at hand. For instance, for continuous baseline covariates splines could be employed to capture potential nonlinear effects, and meaningful interactions with categorical covariates should be included.

5.1.3 Choice of Numerical Integration Method

An important issue in the shared parameter models is the need for numerical integration, which unfortunately makes fitting such models a computationally demanding task. In Chapter 4 we have proposed a new computational approach based on the fully exponential Laplace approximation, which makes fitting joint models feasible, even with high-dimensional random effect structures. We should mention however that we do not suggest that the use of the Gauss-Hermite rule should be

abandoned. In fact, there are cases, in which the Gauss-Hermite rule should be preferred over the Laplace approximation. In particular, for low-dimensional random effects structures, such as random intercepts and random slopes, we have seen in the simulation results in Chapter 4 that the Gauss-Hermite rule produced reliable results and it was faster than the Laplace approximation. Thus, as a guideline regarding the choice of numerical integration method, we suggest the following. First, if high-dimensional random effects structures are to be considered, then it is known that the Gauss-Hermite rule will be very time consuming, and thus the enhanced Laplace approximation of Chapter 4 should be preferred. Second, for low dimensional random effects the choice between integration methods should be based on the empirical Bayes estimates of the random effects, and their estimated variance covariance matrix, under the ignorable (i.e., ignoring the event process) linear mixed effects model. In particular, the empirical Bayes estimates are given by (Verbeke and Molenberghs, 2000)

$$\hat{b}_i = DZ_i^\top W_i(y_i - X_i\beta),$$

and their covariance matrix is of the form

$$\text{var}(\hat{b}_i) = DZ_i^\top \left\{ W_i - W_i X_i \Big(\sum_i X_i^\top W_i X_i \Big)^{-1} X_i^\top W_i \right\} Z_i D$$

where, y_i is the longitudinal response vector for the ith subject, X_i and Z_i are the design matrices, for the fixed effects β and random effects b_i, respectively, D is the covariance matrix of the random effects, W_i equals V_i^{-1}, and $V_i = Z_i D Z_i^\top + \sigma^2 I$ (I denotes the identity matrix). Even though these estimates ignore the influence of the survival outcome, they provide a good indication about the behaviour of the integrand. This argument is based on the fact that, as we have also seen in Chapter 2, as n_i increases the longitudinal model becomes the leading term of the posterior

distribution of the random effect, which is proportional to this integrand. If for many of the individuals in the sample $\hat{b}_i$ is far more zero, and more importantly the elements of $\text{var}(\hat{b}_i)$ are not of magnitude $2^{-1}I$, then the Laplace method should be used, otherwise the faster Gauss-Hermite could be employed. This is due to the fact that the Gauss-Hermite weight function is proportional to the $\mathcal{N}(0, 2^{-1}I)$ density, which implies that the Gauss-Hermite rule will only work satisfactorily when $\text{var}(\hat{b}_i) \approx 2^{-1}I$.

5.1.4 Importance of Underlying Assumptions

The formulation of the joint models that have been considered in this thesis was based on three assumptions. In particular, we have assumed that the visiting and censoring processes are noninformative, and that given the random effects the longitudinal and survival processes are independent (conditional independence). By visiting process we refer to the stochastic mechanism that generates the time points at which the longitudinal measurements are collected (Lipsitz et al., 2002). In this section we comment on the importance and nature of these assumptions, and on the difficulties in checking them.

The assumptions for noninformative visiting and censoring processes are similar in spirit to the Missing At Random (MAR) assumption in the missing data framework (Little and Rubin, 2002). In particular, in order to ignore these two processes from the definition of the shared parameter model, we have assumed that the probabilities of visiting and censoring at time point k depend only on observed history but not on the event times and future longitudinal measurements themselves. As observed history we define all available information for the longitudinal process prior to time point k. Practically speaking, these assumptions imply the belief that decisions on whether a subject withdraws from the study or appears at the clinic for a longitudinal measurement depend on the observed past history (longitudinal measurements and baseline covariates), but there is no ad-

ditional dependence on underlying, latent subject characteristics associated with prognosis. Evaluating the plausibility of the non-informativeness for the visiting and censoring processes usually requires external information from subject-matter experts, since the observed data do not contain enough information to suggest otherwise. For instance, if the censoring is informative, then it may depend on the unobserved true event times and/or future longitudinal measurements.

As we have explained in Chapter 3, there is a direct connection between the joint modelling of longitudinal and survival data, and the missing data framework. That is, the requirement for joint modelling corresponds to NMAR dropout mechanism. The conditional independence assumption in fact directly relates to the nature of the missing data mechanism. In particular, the dropout mechanism under the shared parameter model is written as

$$\begin{aligned} p(T_i \mid y_i^o, y_i^m) &= \int p(T_i \mid b_i, y_i^o, y_i^m) p(b_i \mid y_i^o, y_i^m)\, db_i \\ &= \int p(T_i \mid b_i) p(b_i \mid y_i^o, y_i^m)\, db_i, \end{aligned}$$

where y_i^m and y_i^o denote the missing and observed components of the longitudinal response vector for the ith subject, respectively. If conditional independence does not hold then $p(T_i \mid b_i, y_i^o, y_i^m) \neq p(T_i \mid b_i)$, implying that the probability of event (i.e., dropout) may still depend on unobserved responses y_i^m. The last statement elucidates the difficulty to check this assumption in practice. That is, as we have noted in Chapter 3, inferences regarding the missing data mechanism should be made with caution, since information is implicitly provided through modelling assumptions. The only pragmatic approach to investigate the impact of violating the conditional independence is to perform a sensitivity analysis. Further comments and possibilities for sensitivity analysis are discussed in Section 5.2.2.

5.2 Topics for Future Research

5.2.1 More than One Marker

In this thesis we have considered joint modelling of a single longitudinal marker with a time-to-event outcome. However, in our case study, it would be also interesting to account for the effects of GFR, proteinuria and haematocrit simultaneously in the survival model (Fieuws et al., 2007).

The formulation of a joint model with more than one longitudinal outcome is conceptually straightforward. However, practical difficulties with such extensions arise for two reasons. First, each longitudinal response will require a different set of possibly high-dimensional random effects. For instance, according to the models for the GFR, proteinuria and haematocrit that we have considered throughout the thesis, the calculation of the full joint model would require a 13-dimensional numerical integration. A possible solution to this problem is the Laplace approximation that we have discussed in Chapter 4, and which could be adapted to handle the full joint model with more than one longitudinal marker. The second problem is that the dimensionality of the parameter vector will inevitably increase. Maximization of the observed data log-likelihood in so many dimensions is a numerically and computationally challenging task. Under a full conditional independence assumption (i.e., given the random effects, the event process is independent from the longitudinal markers, and the longitudinal markers are independent of each other) the EM algorithm has the advantage that the majority (except for the survival model) of the M-step updates are in closed-form. However, the EM algorithm has a linear convergence rate (Little and Rubin, 2002), which implies that a large number of iterations will usually be required before convergence. Alternatively, algorithms with a superlinear convergence rate, such as the BFGS (Lange, 2004), could be used instead. Even though quasi-Newton algorithms will converge faster than the EM, the requirement of inverting in each iteration an approximation to

the Hessian matrix is also computationally intensive, since in the full joint model this will be a high-dimensional matrix. The best approach would be a combination of the two algorithms, namely a prespecified number of EM iteration could be used as a refinement of the starting values before initiating the main optimization routine based on a quasi-Newton algorithm.

Finally, in the related framework of longitudinal data analysis, Fieuws and Verbeke (2006) have recently proposed a pairwise approach for the joint modelling of multivariate longitudinal profiles. Investigating the possibility of extending such an approach to the joint modelling of multivariate longitudinal and time-to-event data is worth considering.

5.2.2 Sensitivity Analysis

The focus of inference in Chapters 2 and 3, was the association structure between the longitudinal and event outcomes. As we illustrated in detail in Chapter 3, joint models assume a Not Missing At Random dropout mechanism. This motivated us to perform a sensitivity analysis around the choice of the copula function, which models the dependence between the two processes. However, a larger scale sensitivity analysis would be also worth considering. In particular, since the assumptions for the missing data mechanism are impossible to be tested from the data at hand (see e.g., Molenberghs and Kenward, 2007, Sect. 1.3; Copas and Li, 1997), it would be of interest to investigate the sensitivity of inferences based on the shared parameter models when several aspects of the longitudinal, survival and random effects submodels are altered. In the following we mention some possibilities for sensitivity analysis that are worth further investigation:

- The specification of the random effects component of the joint model requires to make appropriate choices for the random effects distribution and the random effects design matrix Z. Based on the argument raised in Chapter 2, we

expect that an inappropriate assumption for the random effects distribution will not affect inference for large enough n_i. However, we would expect that inferences could be relatively more sensitive to the choice of Z. There are mainly two reasons for this argument. First, the design matrix of the random effect is used in the specification of the conditional density $p(y_i \mid b_i)$, which as noted in Chapter 2, is the dominating term of the posterior distribution of the random effects, and thus greatly influences inferences. Second, in the NMAR context we require an appropriate model for the conditional density $p(y^m \mid y^o, T)$. This implies that not only the mean but also the correlation structure of the longitudinal joint model $p(y^m, y^o)$, which is also influenced by the choice of Z, needs to be correctly specified.

- In relation to the point above, it is worth considering whether including an extra serial correlation term in the error component of the longitudinal mixed model, and further the type of the serial correlation model (e.g., exponential, Gaussian, etc.), could affect inferences. For instance, Verbeke and Molenberghs (2000), in the selection models framework, observed that after having included a serial correlation structure, including an extra measurement error term, altered considerably the resulting p-value for testing MAR. We should also mention that models with an elaborate random effects structure (e.g., high-order polynomials or splines in Z) and no serial correlation terms may produce fits in the observed data similar to the ones from models with a simple random effects structure (e.g., only random intercepts) and a serial correlation term. However, the fit in the missing longitudinal responses y^m could be substantially different, since the dropout process depends usually only on the random effects and not the correlation structure terms (an exception of this is the model of Henderson et al., 2000).

- Several options for the time-to-dropout model may also be considered in a

sensitivity analysis. For instance, inferences under a proportional hazards model can be compared with the corresponding ones from an accelerated failure time model. Furthermore, different kinds of parameterization may be used that associate the longitudinal and the dropout processes. The standard parameterization is given in (5.3); some further possibilities are:

$$\eta_{ti} = x_{ti}^\top \gamma + \alpha z_{yi}^\top b_i, \tag{5.2}$$

$$\eta_{ti} = x_{ti}^\top \gamma + z_{yi}^\top \{A(\alpha)\, b_i\}, \tag{5.3}$$

$$\eta_{ti} = x_{ti}^\top \gamma + b_{ti}, \tag{5.4}$$

where i denotes the subject, and b_{ti} denotes a frailty term. The association parameter α plays a different role in each of the equations (5.1) to (5.4). In particular, in parameterization (5.1) both the fixed and random effects parts of the longitudinal model affect the time-to-dropout, whereas in parameterizations (5.2) and (5.3) only the random effects part does so. In addition, in parameterization (5.3), $A(\alpha)$ denotes a lower triangular matrix, parameterized by the vector of association parameters α, which acts as a rescaling factor for the covariance matrix of the random effects. Finally, parameterization (5.4) is in fact the parameterization that we have used in Chapters 2 and 3, in which association is measured via the joint distribution $\{b_i, b_{ti}\}$.

- Shared parameter models postulate that the probability of dropout at any time point depends on values of the outcome at both past and future time points, through the random effects. Such models are plausible when subjects which show steep increases in their profiles may be more (or less) likely to dropout. However, in some applications, the probability of dropout may directly depend on the actual unobserved longitudinal responses. Furthermore, in some case-studies there is no subject-matter knowledge to justify one missing data mechanism over an other. Thus, a broader sensitivity analysis in

which different NMAR models (i.e., selection and pattern mixture models) are fitted could be utilized.

5.2.3 Residuals Analysis

A basic approach of goodness-of-fit testing is to display or summarize the observed data, and compare this to what might have been expected from the fitted model. If there are systematic discrepancies between the data summaries and their reference distribution under the assumed model, this implies a misfit of the model to the data. A traditional approach to check model assumptions is the inspection of residual plots. Properties and features of residuals, when longitudinal and survival outcomes are separately modelled, have been extensively studied in the literature. For instance, different types of residuals for linear mixed models are discussed in Nobre and Singer (2007) and Verbeke and Molenberghs (2000), whereas residuals for parametric and semiparametric survival models are presented in Therneau and Grambsch (2000).

However, in the case of joint modelling, the reference distribution for the above mentioned types of residuals is not directly evident. Complications arise due to the nonignorable dropout in the longitudinal process caused by the occurrence of events. That is the observed data, upon which the residuals are calculated, are not a random sample of the target population. This implies for instance that a potential systematic behaviour of the residuals versus the fitted values is not necessarily indicative of a model misfit. Thus, conclusions from common residual plots in the joint model framework should be extracted with caution. A promising approach to handle this issue is the use of multiple imputation for posterior predictive checks, as it has been presented by Gelman et al. (2005). The idea is to form random versions of the complete data set by imputing y_i^m using random draws from the predictive distribution $p(y_i^m \mid y_i^o, T_i, \delta_i; \theta)$. These versions of the complete data set can be used to calculate residuals, which could then be

compared to the null hypothesis (e.g., zero mean and independence). Adapting this approach to the case of joint modelling is currently under investigation.

References

Copas, J. and Li, H. (1997). Inference for non-random samples (with discussion). *Journal of the Royal Statistical Society, Series B*, **59**, 55–95.

Fieuws, S. and Verbeke, G. (2006). Pairwise fitting of mixed models for the joint modelling of multivariate longitudinal profiles. *Biometrics*, **62**, 424–431.

Fieuws, S., Verbeke, G., Maes, B. and Vanrenterghem, Y. (2007). Predicting renal graft failure using multivariate longitudinal profiles. *Biostatistics*, doi: 10.1093/biostatistics/kxm041.

Gelman, A., Van Mechelen, I., Verbeke, G., Heitjan, D. and Meulders, M. (2005). Multiple imputation for model checking: completed-data plots with missing and latent data. *Biometrics*, **61**, 74–85.

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

Lange, K. (2004). *Optimization*. Springer-Verlag, New York.

Lipsitz, S., Fitzmaurice, G., Ibrahim, J., Gelber, R. and Lipshultz, S. (2002). Parameter estimation in longitudinal studies with outcome-dependent follow-up. *Biometrics*, **58**, 621–630.

Little, R. and Rubin, D. (2002). *Statistical Analysis with Missing Data*. Wiley, New York, 2nd edition.

Molenberghs, G. and Kenward, M. (2007). *Missing Data in Clinical Studies*. Wiley, New York.

Nobre, J. and Singer, J. (2007). Residuals analysis for linear mixed models. *Biometrical Journal*, **49**, 863–875.

Therneau, T. and Grambsch, P. (2000). *Modeling Survival Data: Extending the Cox Model*. Springer-Verlag, New York.

Verbeke, G. and Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. Springer-Verlag, New York.

APPENDIX A Software

A.1 The JM Package

Even though joint modelling of longitudinal and time-to-event data has received considerable interest in the statistical literature in the recent years, there is no publicly available software to fit these kind of models. Guo and Carlin (2004) have provided WinBUGS and SAS `NLMIXED` code to fit joint models under an exponential distribution for the event times. However, this code requires first, non user-friendly modifications in order to fit other types of survival models (e.g., the model considered in Chapter 4), and second, considerable effort in order user-required quantities to be computed. Thus, there is an evident need for flexible and user-friendly joint modelling software.

The R package **JM** (available from the Comprehensive R Archive Network at `http://cran.r-project.org/`) has been developed to fill this gap. The package has a single model-fitting function named `jointModel()`, which accepts as main arguments a linear mixed effects object fit returned by function `lme()` from package **nlme**, and a survival object fit returned by either function `coxph()` or

function `survreg()` from package **survival**. In addition, the `method` argument of `jointModel()` specifies the type of the survival submodel to be fitted and type of numerical integration method; available options are:

`method = "ph-GH"`: the time-dependent version of a proportional hazards model with unspecified baseline hazard function. The Gauss-Hermite integration rule is used to approximate the required integrals. (This option corresponds to the joint model proposed by Wulfsohn and Tsiatis, 1997)

`method = "weibull-GH"`: the Weibull model under the accelerated failure time formulation. The Gauss-Hermite integration rule is used to approximate the required integrals.

`method = "ch-GH"`: the log cumulative hazard formulation presented in Chapter 4; the log cumulative baseline hazard is approximated using B-splines. The Gauss-Hermite integration rule is used to approximate the required integrals.

`method = "ch-Laplace"`: the log cumulative hazard formulation presented in Chapter 4; the log cumulative baseline hazard is approximated using B-splines. The Laplace approximation method developed in Chapter 4 is used to approximate the required integrals. (This option corresponds to the joint model proposed in Chapter 4)

A sample syntax is as follows:

```
# linear mixed effects model fit
fitLME <- lme(y ~ drug * time, random = ~ time | id, data)

# Cox proportional hazards model fit
fitSURV <- coxph(Surv(ftime, d) ~ drug, data = data, x = TRUE)

# joint model fit
jointModel(fitLME, fitSURV, timeVar = "time", method = "weibull-GH")
```

Several supporting functions are available that extract or calculate several quantities based on the fitted joint model, such as model summary and statistical significance for the estimated coefficients, empirical Bayes estimates (and their standard error), fitted and residuals values, etc. In particular, the function `jointModel()` return objects of class `jointModel`, for which the following methods are available: `print()`, `coef()`, `fixef()`, `ranef()`, `fitted()`, `residuals()`, `summary()`, `plot()`, `vcov()`, and `logLik()`. A detailed description of these functions is available at the on-line help files.

A.2 JM Package Extensions

The **JM** package is still under active development and a number of extensions is already in the planning and design stages. These extensions can be summarized in the following:

- The package currently fits joint models under the standard parameterization (see e.g., (5.3)). In a future release parameterizations (5.1) to (5.4) will be added. Apart from the enhanced flexibility in the joint model definition, this feature will also facilitate the investigation of sensitivity of inferences under different formulations of the random effects structure.

- A `predict()` method will be included that calculates the expected future lifetime. This is defined as the expected value of the event time given that this subject has not yet exhibited the event by the time his last longitudinal measurement was collected, and can be used to provide predictions of subject-specific event times for future individuals in the study.

- Implementation of the posterior predictive residuals as discussed in Section 5.2.3. In particular, an appropriate Monte Carlo algorithm (e.g., Metropolis-Hastings) needs to be considered in order to sample multiple imputations y_i^m

from the nonstandard density $p(y_i^m \mid y_i^o, T_i)$.

- Extension of the Laplace approximation method for the Weibull and time-dependent proportional hazards models, in order to handle high-dimensional random effects structures.

References

Guo, X. and Carlin, B. (2004). Separate and joint modeling of longitudinal and event time data using standard computer packages. *The American Statistician*, **58**, 16–24.

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