Extensions in Joint Modeling of Survival and Longitudinal Outcomes

Magdalena Murawska

Rotterdam, 2013

Extensions in Joint Modeling of Survival and Longitudinal Outcomes

Uitbreidingen van gemengde modellen voor overlevings- en longitudinale data

Proefschrift

ter verkrijging van de graad van doctor aan de
Erasmus University Rotterdam
op gezag van de rector magnificus
rector magnificus
Prof. dr. H.A.P. Pols
en volgens besluit van het College voor Promoties
De openbare verdediging zatplatsvinden op
donderdag 24 april 2014 om 13.30 uur
door

Magdalena Murawska
geboren te Nowy Dwor Mazowiecki, Polen

Promotiecommissie:

Promotor:

Prof.dr. Lesaffre

Overige leden:

Prof.dr. G. Molenberhgs

Prof.dr. J.M.M. Takkenberg

Prof.dr.ing. P.H. Eilers

Copromotor:

Dr. Rizopoulos

To my parents and my fiance

Acknowledgments

I would like to acknowledge the efforts and input of my supervisor, Dr. Dimitris Rizopoulos and Prof. Emmanuel Lesaffre and all my colleagues from the Erasmus Medical Center who were of great help during my research. I would like to thank also Jacqueline M. Smits from Eurotransplant International Foundation in Leiden for her kind support and hospitality during our cooperation.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In many medical studies both longitudinal and event history data are collected for each patient. A well-known and broadly studied example is found in AIDS research, where CD4 cell counts taken at different time points are related to the time-to-death. Often such outcomes are separately analyzed. However, in two particular settings a joint modeling approach is required. First, when focus is on the event times and we wish to study the association between the longitudinal responses and the risk for an event, where we need to take into account the fact that the longitudinal response is also an outcome generated by the same subject. Second, when focus is on the longitudinal outcome and events cause dropout. Under specific circumstances the dropout process needs to be accounted in the analysis in order to obtain valid inferences for the longitudinal outcome.

In the first case, when one is interested in the event time outcome, a standard approach is to use only the last available measurement of the longitudinal biomarker. However, it is evident that by doing so we discard

valuable information. A better alternative is to apply the time-dependent Cox model. This model assumes that the time-dependent covariate process is exogenous. According to Kablfeisch and Prentice [1] the exogenous (also known as external) covariates are the ones for which the path at any future time point t is not affected by the occurrence of an event at time $s \leq t$. This requirement applies in several time-dependent covariates such as environmental factors. But it is clear that biomarkers do not fulfill this requirement. Evidently, biomarkers are the output of the stochastic process generated by the patient himself and as such its value at any time t is influenced by the occurrence if an event at $s \leq t$. To account for these features of biomarkers the framework of joint model has been proposed (Faucett and Thomas [2] and Wulfsohn and Tsiatis [3]). In this approach we estimate the joint distribution of the survival and longitudinal processes. Unlike in the multivariate approach, where we have to interpret the estimates for each longitudinal measurement separately, the joint modeling approach allows to get more insight in the nature of the relation between the two analyzed processes since it assumes some underlying process for the longitudinal measures.

When the focus is on the longitudinal outcome one can distinguish between different types of processes that may cause missingness in that response. Little and Rubin [4] classified the nature of missing data mechanism as Missing Completely At Random (MCAR) when the probability of dropout does not depend on either the observed or unobserved measurements, Missing At Random (MAR), when the probability of dropout depends on the observed data, but not on the unobserved measurements and Missing Not At Random (MNAR), when the probability of missing depends on unobserved and possibly also observed data. For example when a patient does

not come for a visit because she is on vacations the missing data mechanism is probably MCAR. On the other hand, when a doctor advises a patient to leave the study based on a previously observed longitudinal measurement, the mechanism is MAR. Finally, an example of MNAR mechanism is when a patient leaves the study because of her bad condition which is related to her longitudinal profile, including the measurements that would have been observed if she would come for a visit. In the last case, that is under MNAR, the dropout process cannot be ignored, and the joint distribution of the dropout and longitudinal processes needs to be modeled. In this setting three approaches have been proposed, namely pattern mixture, selection and shared parameters models. The first two approaches are mainly applied for discreet time, whereas the last one can handle both discreet and continuous time-to-dropout. An overview of these model classes can be found in Little [5] , Hogan and Laird [6] and Molenberghs and Kenward [7] . In this thesis, motivated by the nature of the data that were collected in a continuous time, we consider only the shared parameter model approach where both submodels for the longitudinal and survival processes share some common parameter. That in fact includes the model postulated by Faucett and Thomas [2] as a special case. This class of models allows to model the dropout mechanism depending on past of future values of the longitudinal outcome using the random effects.

1.1 Literature Review

Several models have been proposed for jointly modeling of longitudinal and survival responses. A nice overview is given by Davidian et al. [8] and Ri-

zopoulos [9]. One of the early motivations for joint modeling has been studies that aimed to evaluate whether the longitudinal marker can act as a surrogate for the survival response. In particular, in early papers joint models have been used in order to model CD4 cell counts together with survival data in HIV/AIDS studies. In order to asses whether the CD4 cell count could be regarded as a surrogate marker for the time to AIDS or death. Tsiatis, DeGruttola and Wulfson [10] studied this surrogacy using the Prentice criteria [11] . They considered a simple linear mixed model for a longitudinal CD4 response taking into account the measurement error and a Cox model for the survival outcome. They proposed a two-stage procedure in which the mixed model was fitted for every event time t separately for all subjects at risk up to time t. In particular, for each time t the empirical Bayes estimate of the biomarker value from the mixed model was used to maximize the partial likelihood of the Cox model. This procedure, motivated by the nonlinearity of the CD4 response, approximated hazard given the observed history of the marker. Dafni and Tsiatis [12] showed via simulation that this two-procedure is still biased but the bias is much reduced compared to naive plug-in methods. Such two-stage approaches have been also considered by Bycott and Taylor [13] .

On the other hand, a growing literature on joint modeling topic was motivated by the non-ignorable missing data problem in longitudinal studies. Wu and Caroll [14] considered a probit model for dropout in order to correct for informative censoring in the longitudinal process comparing unweighted least squares, weighted least squares and a pseudo maximum likelihood approaches. Categorical or ordinal longitudinal responses have received much less attention in the joint modeling framework and the proposed methods

mainly handle nonrandom dropouts for discrete longitudinal data. Molenberghs, Kenward, and Lesaffre considered a multivariate Dale model for longitudinal ordinal data and a logistic regression model for dropout [15]. Kaciroti et al. used a pattern-mixture model to analyze clustered longitudinal ordinal data with non-ignorable missing values [16] . Albert and Follmann [17] extended the model of Wu and Carroll for count data. Viviani et al. [18] considered generalized linear mixed joint models for Poisson and binomial longitudinal responses.

As a further extension Song et al. [19] considered semiparametric joint models relaxing the distributional assumption for the random effects. Tsiatis and Davidian [20] proposed an alternative conditional score approach based on estimating equations that makes no distributional assumption on the underlying random effects. Song et al. [21] extended this approach for the multivariate longitudinal data. Several authors considered the Bayesian approach for fitting joint models. Faucett and Thomas [2] and Xu and Zeger [22] used Markov chain Monte Carlo (MCMC) techniques. Wang and Taylor [23] proposed to account for the association between the longitudinal an the event time outcomes using an integrated Ornstein-Uhlenbeck (IOU) process. Brown and Ibrahim [24] considered a semiparametric Bayesian joint model with no parametric assumption on the random effects. Ibrahim Chen and Sinha [25] proposed a Bayesian generalization of a joint model for the case with multivariate longitudinal data. Bayesian joint models for multivariate longitudinal data have been discussed by several other authors ([26], [27], [28],). Albert and Shih [29] proposed a two-stage procedure using sampling methods to impute missing values.

Joint models with nonlinear mixed-effects submodels have been discussed

by Wu et al. [30] , Hu and Sale [31] and Kaciroti et al. [32] . Finally, Elashoff at al. [33], Li et al. [34] and Williamson et al. [35] extended the joint model for the case of many causes of failures within a competing risks setting using the maximum likelihood approach. Huang et al. [36] handled this problem using Bayesian methods. Some other extensions involved proposing a latent class model ([37], [38], [39]). Within this approach each latent class is characterised by a class-specific marker trajectory and a class-specific risk of the event assuming that a latent class structure entirely captures the correlation between the longitudinal marker trajectory and the risk of the event. Unlike in the shared parameter approach the latent class models do not allow to evaluate specific assumptions regarding the characteristics of the marker trajectory that are the most influential on the event risk.

A great popularity of the joint modeling approach in medical literature was related to the availability of the free software for fitting this kind of models. The R package **JM** developed by Rizopoulos [40] allows to estimate most of the joint models for the normal longitudinal responses and time-to-event under a maximum likelihood approach. Within this package a relative risk survival submodel with different baseline hazards is implemented, as well as Accelerated Failure Time (AFT) model. Competing risks can be also considered. Various options for the survival model and optimization algorithms are provided. As the updated and faster version of the algorithm implemented in **JM** Rizopoulos [41] proposed recently a pseudo-adaptive Gauss-Hermite quadrature rule . Another R package for fitting joint models is **joineR** written by Philipson, Sousa and Diggle [42] . However it allows only for the formulation similar to the one proposed by Wulfson and Tsiatis [3].

1.2 Motivating Data Sets

Our research is mainly motivated by two data sets from transplantation studies. The first data set, analysed in Chapter 2, comes from an international prospective trial on kidney-transplant patients that compared the two main methods of kidney storage prior to the transplantation surgery. In the first arm donors' kidneys were administered to cold storage, whereas in the second arm they were administered to machine perfusion (MP). The advantage of machine perfusion is the possibility of measuring different kidney's parameters reflecting the state of the organ. One of the parameters of interest was renal resistance level (RR), which is an indicator of the kidney condition with respect to its flow.

At the end of the study only 26 graft failures were observed. The renal resistance level (RR) was expected to be an important risk factor for graft failure. It was measured using the perfusion machine at the moment of taking the organ out from a donor ($t = 0$), and thereafter at 10 minutes, 30 minutes, 1 hour, 2 hours, 4 hours and just before transplantation. The time of last measurement was different for different patients and sometimes unknown. Exploratory analysis confirmed the biological expectation that allografts exhibit their highest renal resistance levels just after being extracted from the donor. There was a clear asymptote above zero that was reached after about 5 hours by each patient. This reflected the fact that there is no "perfect flow" through the kidney. Our aim here was to study the association of the renal resistance evolution profile with the risk of graft failure taking into account the baseline characteristics, i.e: the age of the donor, donor's region (3 countries considered) and donor's type (heart-beating or

non-heart-beating).

The second data set also comes from a Eurotransplant registry and consists 2921 heart recipients entering the waiting list in a period of three years. Each recipient was classified to one of the following states: Transplantable (T), Non-Transplantable (NT), Urgent (U) and High Urgent (HU). These states were based on the Eurotransplant urgency-based allocation system [43] and reflected the patients' actual health condition. The evaluation time was different for each recipient and depended on the previous classification in the sense that more severe patients were evaluated more frequently. The first evaluation took place at entry and additional evaluations were performed while the patient remained on the waiting list. Up to the censoring date, 528 patients had died (D) without receiving a transplant, 1565 patients received a transplant (TT) and 239 patients had been removed (R) because of other reasons. The purpose of the study was to predict the future state of the recipient and to estimate the risk of any of the 3 competing events (TT, R, D) based on the history of states on the waiting list with adjusting for baseline covariates. That prediction would allow clinicians for eventual intervention, namely putting the patient with the highest risk of death at the top of the waiting list. This data is analysed using multi-state models techniques in Chapter 3 and using the joint model approach in Chapter 4.

The last data set comes from a study conducted by the Department of Cardio-Thoracic Surgery of the Erasmus Medical Center in the Netherlands. This study includes 285 patients who received a human tissue valve in the aortic position in the hospital in the period from 1987 until 2008 [44]. Aortic allograft implantation are widely used for a variety of aortic valve or aortic root diseases due to their hemodynamic characteristics as a valve sub-

stitute. Using a human tissue in such type of surgery is however related to degeneration and the concomitant that often requires a re-interventions. Due to their complexity these re-operations are unfortunately related to the mortality rate around 4–12%. It is therefore of great interest for clinicians to have prognostic tool for the future prospect of a patient with a human tissue valve in order to plan re-operation and minimize valve-relate morbidity and mortality. In the considered data set 77 (27%) patients received a sub-coronary implantation (SI) and the remaining 208 patients a root replacement (RR). These patients were followed prospectively with the echo examinations scheduled at 6 months and 1 year postoperatively, and biennially thereafter. At each examination the measurements of aortic gradient (mmHg) were taken. By the end of follow-up 59 (20.7%) patients had died, and 73 (25.6%) patients required a re-operation on the allograft. The composite event, re-operation or death, was observed for 125 (43.9%) patients. This data set is analysed in Chapter 5 and our aim here was to provide accurate predictions of re-operation-free survival for future patients taking into account their current longitudinal aortic gradient profile as well as the baseline information, namely age, gender, BMI and the type of operation they underwent.

1.3 Goals of the Thesis

In this thesis we have considered several extensions of joint models that have been proposed in the literature in order to capture the features of the perviously introduced datasets and to answer our scientific questions.

In Chapter 2 we focus on a setting that shares some similarities with the

standard joint modeling framework. In particular, we consider the longitudinal responses that are taken before the actual follow-up for the time-to-event has been initiated. In that setting there is no need for joint modeling since the longitudinal responses did not constitute an endogenous time dependent variable measured at the same period as the time to event. Nevertheless, the problem of measurement error still remains. To handle this problem we propose a two-stage procedure that can be a simpler alternative to joint modeling in similar settings. An additional complexity is the nonlinear character of the longitudinal response that can be handled using this approach compared to joint modeling framework. The procedure can be also generalized for any type of longitudinal responses such as binary or ordinal .

As mentioned in Section 1.1 the joint modeling techniques have been mainly studied only for continuous longitudinal outcome. Categorical longitudinal responses have received less attention in that framework, mainly in the context of handling nonrandom dropouts for discrete longitudinal data. When it comes to competing risks problem the proposed joint models focus mainly on joint analysis of survival and repeated continuous or ordinal responses. We, therefore propose first, in Chapter 3, a simple method to analyze the nominal response in presence of competing risks. In particular, we use the pseudo-values approach introduced by Andersen et al. [45] and apply it for the Aalen-Johansen estimator of the state occupation probabilities of the non-Markov process. This approach allows to study the impact of baseline covariates on the occupation probabilities without modeling the dependence on the history. To address the problem of the competing events we fit a multinomial model for the next state given the previous state observed since the dependence on the previous state was revealed. The proposed

pseudo-values approach is a simpler and straightforward alternative comparing to other non-standard methods available for non-Markov models. In Chapter 4 we analyze the same data using the joint modeling framework and propose a Bayesian model for joint modeling of categorical longitudinal data and time-to-event response taking into account the presence of competing risks.

The majority of prognostic models in the medical literature use only a small fraction of the available biomarker information. Simple approaches discard valuable information not taking into account that the rate of change in the biomarker levels is not only different from patient to patient but also dynamically changes over time for the same patient. Hence, it is medically relevant to investigate whether repeated measurements of a biomarker can provide a better understanding of disease progression. In particular, in transplantation setting, the possibility of updating prediction of the risk of death based on the changing health condition for a specific patient allows to make an intervention by the clinician such as putting her at the top of the waiting list improving the chance of survival of that patient. Motivated by these arguments, in Chapter 4, we present how the joint modeling approach can be used for producing dynamic predictions in the presence of competing risks. In particular, we model the urgency status as a categorical longitudinal response variable, which is assumed to be associated with the competing risks process. We derive the posterior predictive distributions for the longitudinal and event time outcomes. Additionally, we examine how the different parameterizations of the joint model influence the obtained predictions. By different parameterizations we mean different functional relationships between the longitudinal and time-to-event outcomes. The predictions of the

cumulative incidence functions and the categorical longitudinal response are updated as additional measurements of the longitudinal response are available.

Going one step further, in Chapter 5, we compare the joint modeling technique for making dynamic prediction based on the continous longitudinal marker with the older method for producing such predictions, called landmarking. We show how the survival probabilities can be obtained under each method and discuss the differences in the underlying assumptions. In addition, as in the previous chapter, we show how the functional relationship between the two processes may affect predictions. In particular, we consider parametrization in which the risk for an event depends on the rate of increase or decrease of the longitudinal outcome as well as on the whole longitudinal trajectory. To assess the quality of the derived predictions from the two approaches different measures of discrimination and calibration are presented.

Chapter 2

A Two-Stage Joint Model

In this chapter we propose a two-stage approach to for joint modeling nonlinear longitudinal response and time-to-event. At the first stage we summarize the longitudinal information with nonlinear mixed-effects model, and at the second stage we include the Empirical Bayes estimates of the subject-specific parameters as predictors in the Cox model for the time to allograft failure. To take into account that the estimated subject-specific parameters are included in the model, we use a Monte Carlo approach and sample from the posterior distribution of the random effects given the observed data. Our proposal is exemplified on a study of the impact of renal resistance evolution on the graft survival.

This chapter has been published as "A Two-Stage joint model for Nonlinear Longitudinal Response and a Time-to-Event with Application in Transplantation Studies" in *Journal of Probablity and Statistics, 2012* [46].

2.1 Introduction

Many medical studies involve analyzing responses together with event history data collected for each patient. A well-known and broadly studied example can be found in AIDS research, where CD4 cell counts taken at different time points are related to the time to death. These data need to be analyzed using a joint modeling approach in order to properly take into account the association between the longitudinal data and the event times. The requirement for a joint modeling approach is twofold. Namely, when focus is on the longitudinal outcome, events cause nonrandom dropout that needs to be accounted for in order to obtain valid inferences. When focus is on the event times, the longitudinal responses cannot be simply included in a relative risk model because they represent the output of an internal time-dependent covariate [1].

In this paper, we focus on a setting that shares some similarities with the standard joint modeling framework described above, but also has important differences. In particular, we are interested in the association between longitudinal responses taken before the actual follow-up for the time-to-event has been initiated. This setting is frequently encountered in transplantation studies, where patients in the waiting list provide a series of longitudinal outcomes that may be related to events occurring after transplantation. A standard analysis in transplantation studies is to ignore the longitudinal information and use only the last available measurement as a baseline covariate in a model for the allograft survival. It is however evident that such an approach discards valuable information. An alternative straightforward approach is to put all longitudinal measurements as covariates in the survival

model. Nevertheless, there are several disadvantages with this approach. First, it would require spending many additional degrees of freedom, one for each of the longitudinal measurements. Second, patients with at least one missing longitudinal response need to be discarded, resulting in a great loss of efficiency. Finally, we may encounter multicollinearity problems due to the possibly high correlation between the longitudinal measurements at different time points.

Nowadays, when it comes to measuring the association between a longitudinal marker and an event-time outcome, a standard approach is to use the joint model postulated by Faucett and Thomas [2] and Wulfson and Tsiatis [3]. In this setting the longitudinal responses are considered realizations of an endogenous time-dependent covariate (Kabfleish and Prentice [1]), which is measured with error and for which we do not have the complete history of past values available. To account for these features we estimate in the joint modeling framework the joint distribution of the survival and longitudinal processes . Unlike in the multivariate approach, where we have to interpret the estimates for each longitudinal measurement separately, the joint modeling approach allows to get more insight in the nature of the relation between the two analyzed processes since it assumes some underlying process for the longitudinal measures.

However in contrast with the standard joint modeling setting, in our case (i.e., transplantation studies) the longitudinal responses do not constitute an endogenous time dependent variable measured at the same period as the time to event. In particular, since the longitudinal measurements are collected prior to transplantation, occurrence of an event (i.e. graft failure after transplantation) does not cause nonrandom dropout in the longitudi-

nal outcome. Nevertheless, the problem of measurement error still remains. Ignoring the measurement error affects not only the standard errors of the estimates of interest but also it can causes attenuation of the coefficients towards zero [11]. To overcome this we propose a two-stage modeling approach that appropriately summarizes the longitudinal information before the start of follow-up by means of a mixed effects model and then uses this information to model the time to event by including the Empirical Bayes estimates of the subject specific parameters as predictors in the Cox model. To account for the fact that we include the estimates and not the true values of the parameters, we use a Monte Carlo approach and sample from the posterior distribution of the random effects. The proposed method does not require joint maximization neither fitting the random effects model for each event time as in the two-stage approach of Tsiatis, DeGruttola and Wulfsohn [10]. We compare this approach with the "naive" one when the uncertainty about the estimates from the first step is not taken into account, as well as with the full Bayesian approach. Our approach shares similarities with the two-stage approach of Albert and Shih [29]. They considered a model, in which a discrete event time distribution is modeled as a linear function of the random slope of the longitudinal process estimated from the linear mixed model. The bias from informative dropout was reduced by using the conditional distribution of the longitudinal process given the dropout time to construct the complete data set. To account for the measurement error in the mean of the posterior distribution of the random effects, the variance, that incorporates the error in estimating the fixed effects in the longitudinal model, was used. However we use sampling not to impute missing values and correct for non-random dropout but in order to account for

the variability in the predicted longitudinal covariates that are then used in survival model. A method of adjusting for measurement error in covariates, that was used by Albert and Shih, does not apply in our case since it requires the discrete time-to-event and linear model for longitudinal data. The time-to-event could be discretized but still we have a nonlinear model for longitudinal data.

Our research is motivated by data from an international prospective trial on kidney-transplant patients. The study has two arms, where in the first arm donors' kidneys were administered to cold storage, whereas in the second arm they were administered to machine perfusion (MP). The advantage of machine perfusion is the possibility of measuring different kidney's parameters reflecting the state of the organ. One of the parameters of interest is renal resistance level (RR), which has been measured at 10 minutes, 30 minutes, 1 hour, 2 hours, 4 hours and just before transplantation. Our aim here is to study the association of the renal resistance evolution profile with the risk of graft failure. The time of last measurement was different for different patients and often unknown exactly. However based on the medical consult and visual inspection of the individual profiles the last measurement was chosen to be taken at 6 hours for each patient.

The rest of the paper is organized as follows. Section 2.2 provides the general modeling framework with the definition of the two submodels for the longitudinal and survival data, respectively. Section 2.3 describes the estimation methods for the full likelihood and the proposed two-stage approach. In Section 2.4 we apply the two-stage approach to the renal data. Section 2.5 contains the setup and the results for the simulation study. Finally, in Section 2.6 we discuss the proposed methodology.

2.2 Joint Modeling Framework

Let $Y_i(u)$ denote the longitudinal profiles for individual i, $i = 1, 2, \ldots, N$. We assume that $Y_i(u)$ are collected for the ith individual prior to the specified time t_i, $u \in (0, t_i)$. Let $t = 0$ denote the time of the first longitudinal measurement and t_i - the time of the last collected measurement. t_i can be different for different individuals and we denote by m_i the number of longitudinal measurements for subject i collected until time t_i and by u_{ij} the time of jth measurement. Denote by $T_i^* \geq t_i$ the true survival time for individual i. Since the survival time is right censored we observe only $T_i = \min(T_i^*, C_i)$, where $C_i \geq t_i$ is the censoring time with the failure indicator Δ_i, which equals to 1 if the failure is observed and 0 otherwise, i.e. $\Delta_i = I(T_i \leq C_i)$ with $I(\cdot)$ denoting the indicator function. We will assume that censoring is independent of all other survival and covariate information. In addition we assume that the observed longitudinal responses $Y_i(u)$ are measured with error (i.e. biological variation) around the true longitudinal profile $W_i(u)$, i.e.,

$$\begin{aligned} Y_i(u) = W_i(u) + \varepsilon_i(u), \text{with } \varepsilon_i(u) \sim N(0, \sigma^2), \\ \text{and } \operatorname{cov}(\varepsilon_i(u), \varepsilon_i(u')) = 0, u' \neq u. \end{aligned} \tag{2.1}$$

We will consider the longitudinal response that exhibit a nonlinear profiles in time. Therefore, we approximate $W_i(u)$ by means of a nonlinear mixed effects model:

$$W_i(u) = f(u; \boldsymbol{\phi}_i), \text{with } \boldsymbol{\phi}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{\alpha}_i, \tag{2.2}$$

where $f(\cdot)$ is a nonlinear function, parameterized by the vector $\boldsymbol{\phi}_i$. The parameters $\boldsymbol{\phi}_i$ control the shape of the nonlinear function and subscript i denotes that each subject may have its own nonlinear evolution in time in the family $f(\cdot;\boldsymbol{\phi})$. For the subject-specific parameter $\boldsymbol{\phi}_i$ we assume a standard mixed model structure with $\boldsymbol{X}_i$ denoting the fixed effects design matrix with corresponding regression coefficients $\boldsymbol{\beta}$, $\boldsymbol{Z}_i$ the random effects design matrix and $\boldsymbol{\alpha}_i$ the random effects. The random effects $\boldsymbol{\alpha}_i$ are assumed to be independent and normally distributed with mean zero and variance-covariance matrix $\boldsymbol{D}$.

For the event process we postulate the standard relative risk model of the form:

$$\lambda_i(t) = \lambda_0(t)\exp(\boldsymbol{\gamma}^T\boldsymbol{\phi}_i), \tag{2.3}$$

where $\lambda_i(t)$ is the hazard function, $\lambda_0(t)$ is the baseline hazard, which can be modeled parametrically or left completely unspecified. The subject specific parameters $\boldsymbol{\phi}_i$ summarize the longitudinal evolutions of the response for each subject, and therefore coefficients $\boldsymbol{\gamma}$ measure the strength of the association between the different characteristics of the underlying subject-specific nonlinear evolution of the longitudinal profiles and the risk for an event. Within the formulation of the two submodels (2.2) and (2.3) the same random effects now account for both the association between the longitudinal and event outcomes, and the correlation between the repeated measurements in the longitudinal process.

In the particular transplantation setting that will be analyzed in the following study $Y_i(u)$ are the renal resistance level measurements collected

for the ith donor prior to the transplantation time t_i and the same index i is used to denote the allograft transplanted to the ith patient. Time $t = 0$ represents the time that the kidney is removed from the donor and put in cold storage or in a perfusion machine.

2.3 Estimation

2.3.1 Full likelihood framework: Bayesian approach

In the standard joint modeling framework the estimation is typically based on maximum likelihood or Bayesian methods (MCMC). This proceeds under the following set of conditional independence assumptions:

$$p(T_i, \Delta_i, \boldsymbol{Y}_i \mid \boldsymbol{\alpha}_i; \boldsymbol{\theta}) = p(T_i, \Delta_i \mid \boldsymbol{\alpha}_i; \boldsymbol{\theta}_t) p(\boldsymbol{Y}_i \mid \boldsymbol{\alpha}_i; \boldsymbol{\theta}_y)$$
$$p(\boldsymbol{Y}_i \mid \boldsymbol{\alpha}_i; \boldsymbol{\theta}_y) = \prod_{j=1}^{m_i} p(Y_i(u_{ij}) \mid \boldsymbol{\alpha}_i; \boldsymbol{\theta}_y). \tag{2.4}$$

In particular, we assume that given the random effects the longitudinal process is independent from the event times, and moreover, the longitudinal measurements are independent from each other.

Maximum likelihood methods use the joint likelihood and maximize the log-likelihood function $l_i(\boldsymbol{\theta}) = \sum_i \log p(T_i, \Delta_i, \boldsymbol{Y_i}; \boldsymbol{\theta})$. This requires numerical integration and optimization, which makes the fit difficult, especially in high-dimensional random effects settings. Standard options for numerical integration are Gaussian quadrature , Laplace approximation or Monte Carlo sampling ([47], [48]). Maximization of the approximated log-likelihood is based on an EM algorithm ([3], [49], [50], [10], [20]). Several authors proposed a

Bayesian approach (MCMC)([2], [22], [23]). Bayesian estimation, that generalizes a joint model for the case with multivariate longitudinal data, has been discussed by Ibrahim Chen and Sinha [25]. Brown and Ibrahim [24] considered semiparametric model relaxing the distributional assumption for the random effects. In most papers the longitudinal submodel is a linear mixed effects model. Joint models with nonlinear mixed-effects submodels have been less studied in the literature [30]. Nonlinear mixed models are more common in pharmacokinetics and pharmacodynamics, where they are jointly modeled with non-random dropout using NONMEM software. Several authors considered a Bayesian approach with a nonlinear mixed model and informative missingness ([31], [32]).

Here we will proceed under the Bayesian paradigm to estimate the model parameter. Under the conditional independence assumption (2.4) the posterior distribution of the parameters and the latent terms, conditional on the observed data, are derived as:

$$p(\boldsymbol{\theta}, \boldsymbol{\alpha}_i \mid T_i; \Delta_i; \boldsymbol{Y}_i) \propto \prod_{i=1}^{N} \prod_{j=1}^{m_i} \{p(Y_i(u_{ij}) \mid \boldsymbol{\alpha}_i; \boldsymbol{\theta}_y)\}\, p(T_i, \Delta_i \mid \boldsymbol{\alpha}_i; \boldsymbol{\theta}_t)\, p(\boldsymbol{\alpha}_i; \boldsymbol{\theta}_\alpha) p(\boldsymbol{\theta}_y, \boldsymbol{\theta}_t, \boldsymbol{\theta}_\alpha), \tag{2.5}$$

where $\boldsymbol{\theta}^T = (\boldsymbol{\theta}_y^T, \boldsymbol{\theta}_t^T, \boldsymbol{\theta}_\alpha^T)$ is a vector of parameters from the longitudinal and survival models and the vector of the random effects, respectively and $p(\cdot)$ denotes the appropriate probability density function. The likelihood contribution for the ith subject conditionally on the random terms is given by:

$$
\begin{aligned}
p(\boldsymbol{Y_i}, T_i, \Delta_i \mid \boldsymbol{\alpha}_i; \boldsymbol{\theta}) &= p(\boldsymbol{Y_i} \mid \boldsymbol{\alpha}_i; \boldsymbol{\theta}_y) p(T_i, \Delta_i \mid \boldsymbol{\alpha}_i; \boldsymbol{\theta}_t) \\
&= \left[\lambda_0(T_i) \exp\{\boldsymbol{\gamma}^T \boldsymbol{\phi}_i(\boldsymbol{\alpha}_i)\}\right]^{\Delta_i} \exp\left[-\int_0^{T_i} \lambda_0(t) \exp\{\boldsymbol{\gamma}^T \boldsymbol{\phi}_i(\boldsymbol{\alpha}_i)\} dt\right] \\
&\quad \frac{1}{(2\pi\sigma^2)^{m_i/2}} \exp\left[-\sum_{j=1}^{m_i} \frac{\{W_i(u_{ij}, \boldsymbol{\alpha}_i) - Y_i(u_{ij})\}^2}{2\sigma^2}\right].
\end{aligned} \tag{2.6}
$$

The baseline hazard can be assumed of a specific parametric form, e.g. the Weibull hazard. For the priors of the model parameters we make standard assumptions following Ibrahim et al. [25]. In particular, for the regression coefficients $\boldsymbol{\beta}$ of the longitudinal submodel and for the coefficients $\boldsymbol{\gamma}$ of survival submodel we used multivariate normal priors. For variance-covariance matrices we assumed an inverse Wishart distribution and for the variance-covariance parameters we took as a prior an inverse-gamma. For all parameters the vague priors have been chosen.

The implementation of the Cox and piecewise constant hazard models is typically based on the counting process notation introduced by Andersen and Gill [51] and formulated by Clayton [52]. In particular we treat the counting process increments $dN_i(t)$ in the time interval $[t, t + \Delta t]$ as independent Poisson random variables with means $\Lambda_i(t)dt$:

$$
\Lambda_i(t)dt = \omega_i(t) \exp(\boldsymbol{\gamma}^T \boldsymbol{\phi}_i) d\Lambda_0(t), \tag{2.7}
$$

where $\omega_i(t)$ is an observed process taking the value 1 if subject i is observed at time t and 0 otherwise, $d\Lambda_0(t)$ is the increment in the integrated baseline

hazard function occurring during the time interval $[t, t + \Delta t]$. Since the conjugate prior for the Poisson mean is the gamma distribution, we assume the conjugate independent increments prior suggested by Kalbfleisch [53], namely:

$$d\Lambda_0(t) \sim \text{Gamma}(c * d\Lambda_0^*(t), c), \tag{2.8}$$

where $d\Lambda_0^*(t)$ is a prior mean hazard with c being a scaling parameter representing the "strength" of our prior beliefs. The gamma prior was also chosen for the baseline risk parameter of the Weibull distribution in parametric survival model. Alternatively to implement the Cox model in a fully Bayesian approach one may use the "multinomial-Poisson trick" described in the OpenBUGS manual that is equivalent to assuming independent increments in the cumulative hazard function. The increments are treated as failure times and noninformative priors are given for their logarithms. Analogically to the Cox model a piecewise constant hazard model was implemented. We have modeled baseline hazard using a step function with 3 quantiles t_1, t_2 and t_3 as changing points assuring the same number of events in-between. Let t_0 denote the start of the follow up, t_4 the maximum censoring time and $d\Lambda_{0k}(t)$ the increment in the integrated baseline hazard function occurring during the time interval $[t_k, t_{k+1}], k = 0, 1, 2, 3$. Then for different intervals we specify a separate prior hazard mean $d\Lambda_0^*(t)$ and:

$$d\Lambda_{0k}(t) \sim \text{Gamma}(c * d\Lambda_{0k}^*(t), c). \tag{2.9}$$

Similarly as for the Cox model the results were not sensitive with respect to the choice of the hyperparameters as long as the priors were sufficiently

diffuse. The above nonparametric approach can be criticized as having the independent priors for the hazard distribution. However as suggested by Kalbfleisch [53] a mixture of gamma priors can be considered as an alternative. Moreover in a piecewise constant model one can also include change points as unknown parameters in the model as proposed in a Bayesian context by Patra and Dey [54] and applied by Cassellas [55].

In order to assess convergence for the full Bayesian model standard MCMC diagnostic plots were used. The burn-in size was set to 10000 iterations, which was chosen based on the visual inspection of the trace plots, and confirmed by the Raftery and Lewis diagnostics. The same number of iterations were used for constructing the summary statistics. Based on the autocorrelation plots we have chosen every 30th iteration. Therefore in total to obtain 10000 iterations for the final inference 300000 iterations were required after the burn-in part. Additionally we run a second parallel chain and used Gelman and Rubin diagnostic plots to asses the convergence.

2.3.2 Two-stage approach

As mentioned in Section 2.1, the longitudinal measurements in our setting do not constitute an internal time-dependent covariate, since the events took place after the last longitudinal measurement was collected. In particular, since events do not cause nonrandom dropout, the event process does not carry any information for the longitudinal outcome. Mathematically this means that information for the random effects $\boldsymbol{\alpha_i}$ is actually only coming from the longitudinal responses, that is:

$$p(\boldsymbol{\alpha}_i \mid Y_i(u_{ij}); T_i; \Delta_i; \boldsymbol{\theta}_y) = p(\boldsymbol{\alpha}_i \mid Y_i(u_{ij}); \boldsymbol{\theta}_y) \tag{2.10}$$

Thus, we can avoid the computational complexity of the full likelihood framework presented in Section 2.3.1 by employing a two-stage approach. More specifically: At Stage I: we obtain $\hat{\boldsymbol{\theta}}_y$ by maximizing the log-likelihood:

$$l_y(\boldsymbol{\theta}_y) = \sum_{i=1}^{N} \int p(\boldsymbol{Y_i} \mid \boldsymbol{\alpha}_i; \boldsymbol{\theta}_y) p(\boldsymbol{\alpha}_i; \boldsymbol{\theta}_y) d\boldsymbol{\alpha}_i$$

This requires numerical integration and we use a Gaussian quadrature for that purpose. Then we obtain the corresponding empirical Bayes estimates:

$$\hat{\boldsymbol{\alpha}}_i = \arg\max_{\alpha} \left[\log p(Y_i \mid \boldsymbol{\alpha}; \hat{\boldsymbol{\theta}}_y) + \log p(\boldsymbol{\alpha}; \hat{\boldsymbol{\theta}}_y) \right]$$

and compute the predictions:

$$\hat{\boldsymbol{\phi}}_i = \boldsymbol{X}\hat{\boldsymbol{\beta}} + \boldsymbol{Z}_i\hat{\boldsymbol{\alpha}}_i.$$

At Stage II we fit the relative risk model:

$$\lambda_i(t) = \lambda_0(t) \exp\left(\boldsymbol{\gamma}^T \hat{\boldsymbol{\phi}}_i\right).$$

However, a potential problem in the above is that $\hat{\boldsymbol{\phi}}_i$ is not the true subject-specific parameters but rather an estimate with a standard error. If we ignore this measurement error, the regression coefficients $\boldsymbol{\gamma}_i$ will be possibly attenuated. To overcome this problem we propose here a sampling approach to account for the variability in $\hat{\boldsymbol{\phi}}_i$, very close in spirit to the Bayesian ap-

proach of Section 2.3.1. In particular, we use the following sampling scheme:

Step 1: simulate $\boldsymbol{\theta}_y^{(m)} \sim N(\hat{\boldsymbol{\theta}}_y, \widehat{\text{var}}(\hat{\boldsymbol{\theta}}_y))$
Step 2: simulate $\boldsymbol{\alpha}_i^{(m)} \sim \left[\boldsymbol{\alpha}_i \mid \boldsymbol{Y_i}, \boldsymbol{\theta}_y^{(m)}\right]$
Step 3: calculate $\boldsymbol{\phi}_i^{(m)} = \boldsymbol{X}\boldsymbol{\beta}^{(m)} + \boldsymbol{Z}_i\boldsymbol{\alpha}_i^{(m)}$ and fit the relative risk model $\lambda_i(t) = \lambda_0(t)\exp\{\boldsymbol{\gamma}^T\boldsymbol{\phi}_i^{(m)}\}$ from which $\hat{\boldsymbol{\theta}}_{\boldsymbol{t}}^{(\boldsymbol{m})} = \hat{\boldsymbol{\gamma}}^{(m)}$ and $\widehat{\text{var}}(\hat{\boldsymbol{\theta}}_t^{(m)})$ are kept.
Steps 1-3 are repeated $m = 1, \ldots, M$ times.

Step 1 takes into account the variability of the MLEs, and Step 2- the variability of $\boldsymbol{\alpha_i}$. Moreover, because the distribution in Step 2 is not of a standard form, we use a independence Metropolis-Hastings algorithm to sample from it with multivariate t-proposal density centered at an Empirical Bayes estimates $\hat{\boldsymbol{\alpha}}_i$, covariance matrix $\widehat{\text{var}}(\hat{\boldsymbol{\alpha}}_i)$ and df=4. The low number of degrees of freedom was chosen to ensure that the proposal density has heavy tails to provide sufficient coverage of the target density $[\boldsymbol{\alpha}_i \mid \boldsymbol{Y}_i, \boldsymbol{\theta}_y]$. The variance-covariance matrix estimated from the nonlinear mixed model was additionally scaled by some parameter *Scale*. The tuning parameter allows to control the acceptance rate through the range of the proposed distribution. If the range is too narrow, the proposed values will be close to the current ones leading to low rejection rate. On the contrary if the range is too large, the proposed values are far away from the current ones leading to high rejection rate. We chose the acceptance rate to be 0.5 following Carlin [56] that suggests a desirable acceptance rates of Metropolis-Hastings algorithms to be around 1/4 for the dependence (random walk) M-H version and 1/2 for the independent M-H. Roberts et al. [57] recommended to use the acceptance rate close to 1/4 for high dimensions and 1/2 for the mod-

els with dimensions 1 or 2. They discussed this issue in the context of the random walk proposal density. The authors showed that if the target and proposal densities are normal, then the scale of the latter should be tuned so that the acceptance rate is approximately 0.45 in one-dimensional problems and approximately 0.23 as the number of dimensions approaches infinity, with the optimal acceptance rate being around 0.25 in as low as six dimensions. In our case an independence version of Metropolis-Hastings algorithm is applied. The proposal density in the algorithm does not depend on the current point as in the random-walk Metropolis algorithm. Therefore for this sampler to work well, we want to have a proposal density that mimics the target distribution and have the acceptance rate be as high as possible. In order to obtain a desirable acceptance rate one needs to run a sampling algorithm for a number of iterations for a given data set and compute an acceptance rate and then repeat the procedure changing the tuning parameter until the chosen acceptance rate is obtained. Usually a small number of iterations (100-500) is sufficient for the purpose of calibration. More details about the Metropolis-Hastings acceptance-rejection procedure can be found in the supplementary material (section A). A final estimate of $\boldsymbol{\theta}_t$ is obtained using the mean of the estimates from all M iterations:

$$\bar{\hat{\boldsymbol{\theta}}}_t = \sum_{m=1}^{M} \hat{\boldsymbol{\theta}}_t^m \Big/ M. \tag{2.11}$$

To obtain the SE of $\bar{\hat{\boldsymbol{\theta}}}_t$ we use the variance-covariance matrix $\boldsymbol{V}$:

$$\hat{\boldsymbol{V}} = \hat{\boldsymbol{W}} + (M+1)\hat{\boldsymbol{B}} \Big/ M, \tag{2.12}$$

where $\hat{\boldsymbol{W}}$ is the average within-iteration variance and $\hat{\boldsymbol{B}}$ is the between-iteration variance, i.e.,

$$\hat{\boldsymbol{W}} = \sum_{m=1}^{M} \hat{\boldsymbol{U}}^m \Big/ M,$$

and

$$\hat{\boldsymbol{B}} = \frac{1}{M-1} \sum_{m=1}^{M} (\hat{\boldsymbol{\theta}}_t^m - \bar{\hat{\boldsymbol{\theta}}}_t)(\hat{\boldsymbol{\theta}}_t^m - \bar{\hat{\boldsymbol{\theta}}}_t)^T \tag{2.13}$$

$\hat{\boldsymbol{U}}^m$ represents a variance-covariance matrix estimated in iteration m for $\hat{\boldsymbol{\gamma}}^m$.

2.4 Analysis of the RR Data

2.4.1 Models' specification.

We apply the proposed two-stage procedure and a fully Bayesian approach to the transplantation study introduced in Section 2.1. The data was taken from an international prospective trial on 337 kidney pairs, that aimed to compare two different types of storage, namely cold storage and machine perfusion (MP). Here we focus on the second arm. Our main outcome of interest is graft survival time censored after 1 year. At the end of the study only 26 graft failures were observed. The renal resistance level (RR) was expected to be an important risk factor for graft failure. It was measured using the perfusion machine at the moment of taking the organ out from a donor ($t = 0$), and thereafter at 10 minutes, 30 minutes, 1 hour, 2 hours, 4 hours and just before transplantation. As mentioned in the Section 2.1, the time of last measurement was different for different patients and sometimes unknown. However there was a clear asymptote visible from the individual

profiles that was reached after about 5 hours by each patient. Based on that behavior and after the medical consult we chose the last measurement to be taken at 6 hours for each patient. Other variables of interest include the age of the donor, donor's region (3 countries considered) and donor's type (heart-beating or non-heart-beating).

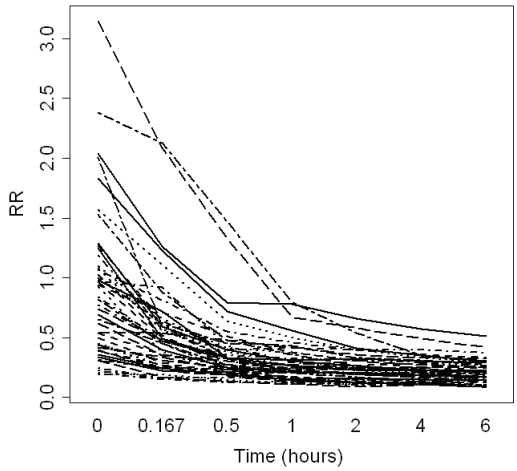


Figure 2.1: *Individual profiles of renal resistance level for 50 sampled donors*

The individual profiles of 50 randomly selected kidney donors are presented in Figure 2.1. This plot confirms the biological expectation that allografts exhibit their highest renal resistance levels just after being extracted from the donor. Thereafter they show a smooth decrease in RR until they reach an asymptote above zero indication that there is no "perfect flow" through the kidney. Furthermore, we observe that the initial RR level, the rate of decrease as well as the final RR level differ from donor to donor. Additional descriptive plots for our data are presented in the supplementary

material (section A).

In the first step of our analysis we aim to describe the evolution of the renal resistance level in time. Motivated by both biological expectation and Figure 2.1 we postulate the following nonlinear function:

$$f(t) = \phi_1 + \phi_2 e^{-\phi_3 t}, \tag{2.14}$$

where ϕ_1 is a lower asymptote, $\phi_1 + \phi_2$ is an initial value for t=0, and ϕ_3 is the rate of decrease from ϕ_2 to ϕ_1 (see also Figure A.2 in Supplementary material).

To accommodate for the shapes of RR evolutions observed in Figure 2.1, we need to constraint ϕ_1, ϕ_2 and ϕ_3 to be positive. Moreover, in order to allow for individual donor effects, we use the following formulation:

$$\begin{aligned} Y_i(t) &= W_i(t) + \varepsilon(t), \text{with} \\ W_i(t) &= f_i(t) = \exp(\phi_{1i}) + \exp(\phi_{2i}) e^{-\exp(\phi_{3i})t}, \end{aligned}$$

where

$$\begin{aligned} \phi_1 &= \beta_{10} + \beta_{11} DonorAge + \beta_{12} DonorType + \beta_{13} DonorReg1 + \beta_{14} DonorReg2 + \alpha_1 \\ \phi_2 &= \beta_{20} + \beta_{21} DonorAge + \beta_{22} DonorType + \beta_{23} DonorReg1 + \beta_{24} DonorReg2 + \alpha_2 \\ \phi_3 &= \beta_{30} + \beta_{31} DonorAge + \beta_{32} DonorType + \beta_{33} DonorReg1 + \beta_{34} DonorReg2 + \alpha_3, \end{aligned}$$

and $\boldsymbol{\alpha_i} \sim N(0, D)$, $\varepsilon(t) \sim N(0, \sigma^2)$ with $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3)$ and $\text{cov}(\alpha_i, \varepsilon(t)) = 0$. In the second step the predicted parameters (ϕ_1, ϕ_2, ϕ_3) summarizing the

RR evolution of the nonlinear mixed model are included in the graft survival model. The final model for graft survival was of the form:

$$\lambda_i(u) = \lambda_0(u) \exp\left(\gamma_1 \hat{\phi}_{1i} + \gamma_2 \hat{\phi}_{2i} + \gamma_3 \hat{\phi}_{3i}\right).$$

To investigate the impact of ignoring that the covariate $\hat{\boldsymbol{\phi}}_i$ is measured with error, we compared the naive approach in which we ignored this measurement error and our proposal that accounts for the uncertainty in $\hat{\boldsymbol{\phi}}_{\boldsymbol{i}}$ via Monte Carlo sampling. We used Metropolis-Hastings algorithm with independent t-proposal and acceptance rate around 50% for the reason given in Section 2.3.2. We simulated $M = 10000$ samples with additional initial step of the scaling parameter calibration in order to achieve the desirable acceptance rate. The final estimates (and SE) of the parameters associated with RR covariates were calculated using the formulas described in the Section 2.3. We compared the results from the two-stage procedure with the estimates obtained from the fully Bayesian joint model fitted for the data using the priors specified in Section 2.3.1.

The analysis was performed using R Statistical Software. Packages *survival* and *nlme* were used for the two submodels fit and a separate code was written by the first author for the sampling part. The fully Bayesian model was fitted using OpenBUGS software with the priors specified in Section 2.3.1. In particular, for the $p \times p$ variance-covariance matrices of multivariate normal priors we used inverse Wishart distribution with p degrees of freedom. For the variance-covariance parameter of the normal longitudinal response we took as a prior an inverse-Gamma(10^{-3},10^{-3}). For the baseline risk parameter of the Weibull distribution in survival submodel a

Gamma(10^{-3},10^{-3}) prior was used. To analyze the data using the fully Bayesian Cox model described in Section 2.3.1 we chose the scaling parameter c in a gamma prior for the independent increments to be equal 0.001 and a prior mean $d\Lambda_0^*(t) = 0.1$. We did not observe any substantial difference for the different values of parameter c as long as c was small enough to keep the prior noninformative. We do not recommend too small values of the scaling parameter c as they can lead to the computation problems. Analogically we have chosen gamma priors for the piecewise constant hazard model. The code for the Bayesian full joint model as well as the R codes for the sampling two-stage procedure are available from the authors on request.

2.4.2 Results

The results for the nonlinear mixed model are presented in Table 2.1, for the two-stage approach and in supplementary material (part A), for the full Bayesian approach with Weibull survival model. The results for the longitudinal part for the full Bayesian approach with Cox and piecewise constant hazard models were similar (not presented). It can be observed, based on the two-stage model results, that only Donor Age had a significant impact on the RR asymptote. Donor Type and Region had a significant impact on the steepness parameter. However we keep all the covariates in the model for the purpose of prediction for the second stage. The mean RR profiles for Heart-Beating and Non-Heart-Beating donors from different regions together with fitted values based on the obtained nonlinear mixed model are given in Supplement A.

Table 2.1: *Parameter estimates, standard errors and 95 % confidence intervals from the non-linear mixed model for RR*

Effect	Parameter	Estimate	SE	(95%CI)
Fixed effects				
ϕ_1				
Constant	β_{10}	2.838	0.094	(2.654; 3.022)
Donor Age	β_{11}	0.005	0.002	(0.001; 0.009)
Donor Type (HB vs NHB)	β_{12}	-0.102	0.068	(-0.235; 0.031)
Donor Region 1 vs 3	β_{13}	-0.078	0.065	(-0.205; 0.049)
Donor Region 2 vs 3	β_{14}	-0.072	0.072	(-0.213; 0.069)
ϕ_2				
Constant	β_{20}	3.510	0.211	(3.096; 3.924)
Donor Age	β_{21}	0.004	0.004	(-0.004; 0.012)
Donor Type (HB vs NHB)	β_{22}	-0.064	0.154	(-0.365; 0.238)
Donor Region 1 vs 3	β_{23}	-0.107	0.147	(-0.395; 0.181)
Donor Region 2 vs 3	β_{24}	0.033	0.163	(-0.286; 0.352)
ϕ_3				
Constant	β_{30}	1.010	0.186	(0.645; 1.375)
Donor Age	β_{31}	0.003	0.003	(-0.003; 0.009)
Donor Type (HB vs NHB)	β_{32}	0.402	0.130	(0.147; 0.657)
Donor Region 1 vs 3	β_{33}	-0.284	0.125	(-0.529; -0.039)
Donor Region 2 vs 3	β_{34}	-0.032	0.138	(-0.302; 0.238)
Random effects				
$se(\alpha_1)$	d_{11}	0.396		
$se(\alpha_2)$	d_{22}	0.955		
$se(\alpha_3)$	d_{33}	0.572		
$cov(\alpha_1, \alpha_2)$	d_{12}	0.257		
$cov(\alpha_1, \alpha_3)$	d_{13}	-0.053		
$cov(\alpha_2, \alpha_3)$	d_{23}	0.023		
$se(\varepsilon_{ij})$	σ	7.507		

In the second step of the analysis we applied at first the naive approach and used the estimates $\hat{\phi_1}, \hat{\phi_2}$ and $\hat{\phi_3}$ from the nonlinear mixed model as fixed covariates in the final Cox models for graft survival. Table 2.2 presents the results for the survival submodel for the all approaches, namely the plug-in method, two-stage approach and the fully Bayesian model. For the fully Bayesian approach the results for the parametric Weibull model together with Cox and piecewise constant hazard models are listed. The results from Table 2.2 indicate that only the RR asymptote could have a significant impact on graft survival.

We observe that the estimates are close or almost identical as in plug-in model. SE of the Cox regression coefficients for the model with sampling are greater than SE from the plug-in model in Table 2.2 (a), especially for the parameter ϕ_3. The increase in SE is somewhat the expected and is caused by the additional variability in covariates captured by the sampling approach. The fully Bayesian model produces similar results to our semi-Bayesian sampling model with somewhat lower SE. We do not observe substantial difference between fully parametric and nonparametric models in a fully Bayesian approach. Since in the analyzed real data the number of events is small the fully Bayesian Cox and piecewise constant hazard Bayesian models were expected to produce similar results. We did not observe any substantial difference for the different values of hyper parameters.

Even though it is hard to compare exactly the computational time for the two approaches, the rough estimation of the total computational time needed to estimate and assess the convergence (2 chains) of the full Bayesian model

Table 2.2: *Parameter estimates, SE and 95 % confidence/credibility intervals from proportional hazards Cox model for graft survival for plug-in method (a), sampled covariates (b) and fully Bayesian approach (c, d, e)*

(a)

Graft Survival - Plug-in

Effect	**Parameter**	**log(HR)**	**SE**	**(95%CI)**
$\exp(\phi_1)$	γ_1	0.052	0.022	(0.009; 0.095)
$\exp(\phi_2)$	γ_2	-0.005	0.005	(-0.015; 0.005)
$\exp(\phi_3)$	γ_3	0.053	0.158	(-0.257; 0.363)

(b)

Graft Survival - Sampling two-stage

Effect	**Parameter**	**log(HR)**	**SE**	**(95%CI)**
$\exp(\phi_1)$	γ_1	0.053	0.024	(0.006; 0.100)
$\exp(\phi_2)$	γ_2	-0.006	0.008	(-0.022; 0.010)
$\exp(\phi_3)$	γ_3	0.055	0.185	(-0.308; 0.418)

(c)

Graft Survival - Fully Bayesian - Weibull

Effect	**Parameter**	**log(HR)**	**SE**	**(95%HPD)**
$\exp(\phi_1)$	γ_1	0.058	0.023	(0.013; 0.103)
$\exp(\phi_2)$	γ_2	-0.005	0.008	(-0.020; 0.011)
$\exp(\phi_3)$	γ_3	0.056	0.180	(-0.299; 0.411)

(d)

Graft Survival - Fully Bayesian - Cox

Effect	**Parameter**	**log(HR)**	**SE**	**(95%HPD)**
$\exp(\phi_1)$	γ_1	0.056	0.023	(0.010; 0.101)
$\exp(\phi_2)$	γ_2	-0.006	0.008	(-0.022; 0.010)
$\exp(\phi_3)$	γ_3	0.055	0.171	(-0.280; 0.390)

(e)

Graft Survival - Fully Bayesian - Piecewise constant hazard

Effect	**Parameter**	**log(HR)**	**SE**	**(95%HPD)**
$\exp(\phi_1)$	γ_1	0.054	0.024	(0.007; 0.102)
$\exp(\phi_2)$	γ_2	-0.005	0.009	(-0.022; 0.012)
$\exp(\phi_3)$	γ_3	0.054	0.179	(-0.297; 0.405)

was about 21.6 hours and depended on the implemented survival model. A similar computational time was needed for the full Bayesian model with the Cox survival model and piecewise constant hazard model with a slightly more time needed for the parametric Weibull model. For the two-stage approach the total computational time was about 10 hours using the Intel(R) Core(TM)2 Duo T9300 2.5 GHz and 3.5 GB RAM.

2.5 Simulations

2.5.1 Design

We have conducted a number of simulations to investigate the performance of our proposed two-stage method. In particular, we compared the plug-in method which uses the Empirical Bayes estimates $\hat{\phi}_i$ from the nonlinear mixed model and ignores the measurement error, the two-stage Monte Carlo sampling approach that accounts for the variability in $\hat{\phi}_i$ and the fully Bayesian approach. In the fully Bayesian approach only the parametric Weibull model was considered.

For the longitudinal part the data were simulated for 500 patients from model (2.15) with $\phi_{1i} = \beta_{10} + \alpha_{1i}$, $\phi_{2i} = \beta_{20} + \alpha_{2i}$ and $\phi_{3i} = \beta_{30} + \alpha_{3i}$, $\boldsymbol{\alpha}_i \sim N(0, \boldsymbol{D})$, $Y \sim N(f(t), \sigma^2)$. The variance-covariance matrix $\boldsymbol{D}$ of the random effects was chosen to be $\boldsymbol{D} = \text{vech}(0.6, 0.01, -0.01, 0.6, 0.01, 0.3)$. We kept 7 measurement points as in the original analyzed data set as well as the nonequal distances between them. The variance of the measurement error σ^2 was chosen to be 0.25, 1 and 25. Survival times were simulated for each patient using the exponential model with the rate parameter equal

$\exp(\lambda)$, where λ:

$$\lambda = \gamma_1 \exp(\phi_1) + \gamma_2 \exp(\phi_2) + \gamma_3 \exp(\phi_3).$$

We considered scenarios with fixed coefficients $\gamma_1 = 0.5, \gamma_2 = 0.5$ and $\gamma_3 = -0.2$. The censoring mechanism was simulated independently using an exponential distribution $Exp(\lambda_C)$. Parameter λ_C was changed in order to control proportion of censored observations. Different scenarios with 40% and 20% of censoring were examined . For each simulated data set we fitted four survival models, namely the gold standard model that uses the true simulated values ϕ_i, the plug-in model, the sampling model and fully Bayesian joint model. Neither nonparametric Cox nor piecewise constant hazard model were considered in a fully Bayesian approach since we have simulated the data from the parametric exponential model and wanted to compare the proposed two-stage approach with the "best" fully Bayesian model. All the prior settings, size of burn-in, number of iterations etc. for the fully Bayesian model were the same as for the real data analysis.

2.5.2 Results

In Table 2.3 we present the average results for 200 simulations of different scenarios are presented. The results from our sampling model were very close to the results obtained for the fully Bayesian model with slightly smaller bias and RMSE for the fully Bayesian approach. That was due to the better estimation of random effects variability in fully Bayesian approach. The plug-in method produced the biggest bias that sometimes with combination with the small variability of the estimates around the biased mean resulted in

larger RMSE than in sampling approach. As the measurement error or the percentage of censored observations increased, the estimates of survival submodel were more biased with larger RMSE for all approaches. The increase in bias was more severe when the measurement error variance was increased rather than when the percentage of to censoring was higher. This bias was however decreased when the number of repeated measures per individual was increased. This has to do with the amount of information that is available in the data for the estimation of $\hat{\phi}_i$. As it is known from the standard mixed models literature [58], the degree of shrinkage in the subject-specific predicted values is proportional to σ and inversely proportional to n_i and σ_α. To compare the relation between variance of the random effects and variance of the measurement error, one can use intra class correlation (ICC) defined as the proportion of the total variability that is explained by the clustering with a given random effect. ICC was similar for the simulated and the real data for the biggest σ and increased in a simulation data as σ decreased.

Since the calculations for the simulation study were highly computationally intensive we have used the cluster with about 20 nodes with AMD Quad-Core Opteron 835X, 4 x 2GHz and 16GB RAM per node. The analysis for the 200 simulated data sets for a single scenario took about 65.5 hours using the Bayesian approach and 31.2 hours using the two-stage approach.

2.6 Discussion

We have proposed a two-stage method that can be used in a joint analysis of longitudinal and time to event data when the longitudinal data are collected

Table 2.3: *Bias and Residual Mean Squared Error (RMSE) for the method with true ϕ_i (GS), Empirical Bayes estimates $\hat{\phi}_i$ (Plug-in), sampled ϕ_i and fully Bayesian approach*

7 time points

% **censoring**	**20**			**40**		
$\sigma = 0.5$	γ_1	γ_2	γ_3	γ_1	γ_2	γ_3
GS	0.00(0.04)	-0.02(0.03)	0.01(0.03)	-0.01(0.04)	0.02(0.04)	-0.02(0.04)
plug-in	-0.05(0.06)	-0.04(0.05)	0.06(0.07)	-0.08(0.09)	-0.04(0.05)	0.12(0.12)
sampling	-0.04(0.05)	0.03(0.08)	0.02(0.07)	-0.05(0.11)	-0.02(0.06)	0.03(0.10)
Bayesian	-0.03(0.04)	-0.02(0.04)	0.01(0.02)	-0.01(0.04)	-0.02(0.04)	0.02(0.07)
$\sigma = 1$						
GS	0.04(0.05)	0.04(0.07)	-0.03(0.07)	-0.05(0.09)	-0.04(0.06)	-0.03(0.05)
plug-in	-0.07(0.08)	-0.08(0.09)	0.07(0.09)	-0.10(0.12)	-0.08(0.09)	0.08(0.11)
sampling	-0.07(0.09)	-0.06(0.10)	-0.02(0.11)	-0.05(0.12)	0.05(0.11)	-0.03(0.12)
Bayesian	0.01(0.03)	0.05(0.06)	-0.03(0.07)	0.05(0.06)	0.04(0.06)	-0.04(0.07)
$\sigma = 5$						
GS	0.04(0.06)	0.05(0.06)	0.04(0.08)	0.05(0.10)	0.01(0.05)	-0.02(0.06)
plug-in	-0.09(0.10)	-0.10(0.11)	0.08(0.11)	-0.20(0.22)	-0.21(0.22)	0.14(0.18)
sampling	0.08(0.13)	0.06(0.12)	-0.05(0.12)	0.07(0.14)	-0.05(0.13)	-0.11(0.18)
Bayesian	0.09(0.10)	0.05(0.09)	-0.09(0.10)	-0.09(0.10)	0.08(0.12)	-0.12(0.18)

14 time points

% **censoring**	**20**			**40**		
$\sigma = 0.5$	γ_1	γ_2	γ_3	γ_1	γ_2	γ_3
GS	-0.03(0.03)	0.00(0.02)	-0.02(0.03)	0.02(0.03)	-0.03(0.04)	0.02(0.04)
plug-in	-0.02(0.03)	-0.03(0.04)	0.05(0.07)	-0.02(0.04)	-0.03(0.04)	0.05(0.06)
sampling	0.03(0.04)	0.02(0.06)	0.02(0.07)	0.02(0.04)	0.04(0.05)	0.02(0.08)
Bayesian	-0.03(0.04)	-0.02(0.04)	-0.02(0.04)	0.02(0.04)	0.03(0.04)	-0.05(0.06)
$\sigma = 1$						
GS	-0.03(0.04)	-0.03(0.04)	-0.01(0.03)	0.00(0.03)	-0.02(0.04)	0.05(0.06)
plug-in	-0.09(0.06)	-0.05(0.06)	0.06(0.07)	-0.02(0.04)	-0.04(0.05)	0.11(0.11)
sampling	0.04(0.08)	0.02(0.08)	-0.02(0.07)	-0.02(0.04)	-0.02(0.08)	0.04(0.09)
Bayesian	-0.03(0.04)	0.04(0.05)	-0.03(0.05)	0.02(0.04)	0.03(0.05)	0.06(0.07)
$\sigma = 5$						
GS	-0.03(0.04)	-0.03(0.04)	0.01(0.04)	-0.01(0.04)	-0.02(0.04)	0.05(0.06)
plug-in	-0.05(0.06)	-0.10(0.11)	0.07(0.09)	-0.10(0.11)	-0.09(0.10)	0.11(0.12)
sampling	0.04(0.09)	0.04(0.11)	-0.05(0.11)	0.07(0.12)	0.05(0.11)	-0.06(0.16)
Bayesian	0.03(0.05)	0.03(0.08)	-0.05(0.10)	0.02(0.04)	0.06(0.10)	-0.09(0.14)

before the start of follow-up for survival and the interest is in estimation of the impact of longitudinal profiles on survival. The modeling strategy is based on specification of two separate submodels for the longitudinal and time to event data. First the longitudinal outcome is modeled using a random effects model. Then the survival outcome is modeled using the Empirical Bayes estimates of the subject specific effects from the first stage. The variability of the estimates from the first stage is properly taken into account using a Monte Carlo approach by sampling from the posterior distribution of the random effects given the data.

As it was demonstrated, ignoring the additional variability of the subject-specific estimates when modeling survival leads to some bias, and in particular, attenuates the regression coefficients towards zero [11]. That was also confirmed by our simulation study. In comparison with the fully Bayesian approach, the proposed partially Bayesian method produced similar results with substantially less number of iterations required. This is due to the fact that sampling was conducted already around the EB estimates and there is no needed for a burn-in part as in a standard fully Bayesian approach. We used 10000 iterations per subject, which was about the size of burn-in needed in the fully Bayesian models. No thinning was used in our approach, based on the visual inspection of the trace plots. Though it is hard compare the fully Bayesian approach and the two-stage approach with respect to the computational time precisely, the rough approximation of the total computational time required for the two-stage approach was about half in comparison with the fully Bayesian approach. The fully Bayesian approach provided similar results with the two-stage approach for the special setting we have considered here. However fitting a fully Bayesian model was a bit of

"overdone" in the sense that by design the longitudinal data could not be affected by the survival. Since in many transplantation studies the longitudinal data are collected before the start of follow-up for survival, therefore using our method in that cases seems to be more appropriate than using a fully Bayesian approach. We recommend the proposed approach not only for the particular transplantation studies but in any setting that shares the similarity of the separated follow-up periods for the two analyzed endpoints. That is for example when the event process does not carry any information for the longitudinal outcome and the condition (2.10) from Section 2.3.2 holds. The simulation results indicate that even if the data come from the real joint setting in which (2.10) may not hold, the proposed two-stage procedure can be comparable to the fully Bayesian approach.

Since the sampling in the proposed method relies strongly on the results of the first part, the accurate estimation of all parameters of nonlinear mixed model is a key feature and should be performed carefully. This can be a problematic when the deviation from normality of the random effects is suspected. However it was shown that even for the non-normal random effects one can still use a standard software such as *nlmixed* in SAS with just a small change in a standard program code. In such cases the probability integral transformation (PIT) proposed by Nelson et al. [59] can be used or the reformulation of the likelihood proposed by Liu and Yu [60] . An alternative is fitting a Bayesian model only to estimate the longitudinal submodel in the first stage, instead of the likelihood methods. Fitting nonlinear mixed models using Bayesian standard software can be however problematic due to the high nonlinearity in random effects that is caused both by the nonlinear function of the longitudinal profiles and by the possible restrictions on

parameters [61].

In comparison with the two-stage approach proposed by Tsiatis, DeGruttola and Wulfsohn [10] our method is less computationally intensive since it does not require fitting as many mixed models as there are event times in the data. An alternative, that is somewhat simpler to implement and does not require any assumption about the distribution on the underlying random effects, is the conditional score approach proposed by Tsiatis and Davidian [20]. However this method is less efficient than the methods based on likelihood. The focus in the discussed approaches is on the association between the longitudinal and event time processes. However in transplantation studies when the two follow-up periods for longitudinal and survival outcomes are often separated the interest is rather in making an inference on the marginal event-time distribution. This is similar to the Bayesian approach proposed by Xu and Zeger [22], that uses the longitudinal data as auxiliary information or surrogate for time-to-event data. Our approach is particulary useful in this setting. Since each of the two submodels is fitted separately, a standard software can be used to implement our method with just a small part of additional programming for Monte Carlo sampling.

Another advantage of the proposed two-stage method is that it can be easily generalized from survival to other types of models as it was applied for the binary Delayed Graft Failure (DGF) indicator (results not shown) in the analysis of the renal data. For that purpose in the second step of the two-stage procedure the survival model was replaced by the logistic regression model for the DGF indicator. The first stage of the proposed approach could be also modified allowing for other types of longitudinal response and other types of mixed models. Therefore instead of using a nonlinear mixed model

a linear mixed model or generalized linear mixed model (GLMMs) can be considered depending on the type and the shape of the longitudinal response. In the presented real data example we have chosen the three parameters that described the evolution of the longitudinal response. However for the particular question of interest one can easily choose the most convenient parametrization for the longitudinal model and use the selected parameters to analyze the non- longitudinal response in the second stage.

Chapter 3

Multi-State Models for Nominal Longitudinal Response

In transplantation studies often several response measurements are collected for patients while they are on the waiting list. In this setting it is often of primary interest to assess whether the available history of a patient can be used for predicting patient survival as well as further performance on the list. In this work we use a multi-state models approach to analyze the performance of patients described by their urgency status that changes in time while waiting for a new organ. We use the pseudo-values approach introduced by Andersen et al. (2003) and apply it for the Aalen-Johansen estimator of the state occupation probabilities since the transition probabilities were found to depend on the history. This approach allows to study the impact of baseline information on the occupation probabilities treating the dependence

on the history as a nuisance. It was found that the previous state, the current state and time from the moment of entering the waiting list had the impact on the future performance of the patient. Depending on those, patients were more likely to come back to the particular status in which they were before, die or get a transplant. To address the problem of those competing events a multinomial approach was used for the next state given the previous state observed.

3.1 Introduction

In transplantation studies the urgency status is measured over time for patients waiting for an organ transplant reflecting their changing disease state. In these setting, it is often of primary interest to investigate whether the available history on the urgency status of a patient can be used for predicting survival as well as the patient's future status on the waiting list. We use a multi-state models approach to address this problem. In particular, we consider data from the Eurotransplant heart transplantation waiting list, where patients are classified as U (Urgent), HU (High Urgent), T (Transplantable) and NT (Not Transplantable) and due to the nature of the last option those categories can be only partially ordered. The absorbing states were death (D), getting a transplant (TT) or removal from other reasons than the previous two (R). A graph depicting all possible transitions between the states is presented in Figure 3.1. The evaluation time was different for each patient and depended on the previous classification meaning that more severe

This chapter has been submitted for publication as "Simple Analysis of Non-Markov Models: A Case Study on Heart Transplant Data" in *Statistical Modelling* [62].

patients were evaluated more frequently.

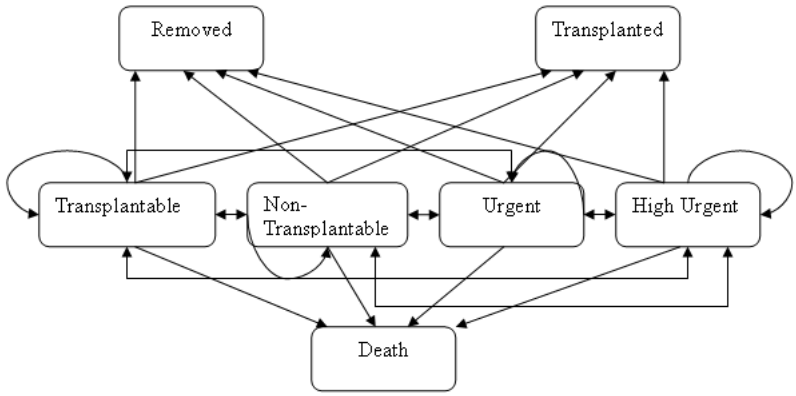


Figure 3.1: *Graph of all possible transitions between the states.*

This type of data is typically analyzed using multi-state models. As standard in this framework, we started our analysis from the assumption that the process is Markov and homogeneous in time (even though by design this assumption was priori questionable). That implied a constant transition (intensity) matrix that does not depend on time or past history of the process (but may depend on time-fixed baseline covariates), and allowed the likelihood methods to be used. The results suggested that the homogeneity assumption does not seem to be satisfied. A standard extension of this framework for relaxing this assumption is to assume a piecewise constant intensity model. However, this still proved not to be sufficient, and in the next step we applied non-parametric methods to estimate transition probabilities using the Aalen-Johansen estimator. In a non-Markov setting it is known that Aalen-Johansen estimators for the transition probabilities are

not consistent and may produce biased results. Alternatively, Datta and Satten [63] showed that for non-Markovian processes, the state occupation probabilities may be still consistently estimated, provided that data are not subject to stage-dependent censoring. In this work we will show how the problem of dependence on the history can be addressed using a multinomial approach. Applied to our data, the multinomial model revealed a dependence on the history, in particular on the state visited before the current one. This motivated us, in the final step of the analysis, to work with occupation probabilities and employ the pseudo-values approach of Andersen et al. [45] to directly model the effect of the baseline covariates on those probabilities, treating the dependence on the history as a nuisance. We should note that the pseudo-values approach was proposed by Andersen et al. in the context of competing risks problem. It uses the pseudo-values from a jackknife statistic constructed for a non-parametric estimator of the transition probabilities. This approach has been applied for cumulative incidence functions in a competing risks model [64,65], to the restricted mean survival time [66] and in simple multistate models [67]. Pseudo-values models were further studied by Graw et al. [68] providing more theoretical justification. However, the consistency and asymptotic normality was proven only in the special case of competing risk models. We extend the univariate pseudo-values approach for the multivariate model on occupation probabilities as suggested by Andersen et al. [45] and present the simulation results for more general non-Markov models.

An additional complexity in the practical use of multi-state models designs is the possible non-ignorability of the observation process and the interval-censoring problem. Regarding the former issue, in our study more

severely ill patients are monitored more close and the next sampling time is chosen on the basis of the current disease state. In particular, given a current state re-evaluation by Eurotransplant audit group is mandatory every fixed number of days and the length of this period depends on the particular current state. For example, for status HU it is 7 days and for state U, 28 days. Therefore, we have the so-called doctor's care sampling scheme. Grüger et al. [69] showed that this scheme is not informative when likelihood methods are used because the likelihood given this examination scheme is proportional to the likelihood obtained when the examination scheme is fixed in advance. Therefore, the parameters of the process can be estimated independently of the parameters of the sampling scheme. For the nonparametric approach the Aalen-Johansen estimator is also valid even if the process is not Markov since the censoring process is not stage-dependent and does not depend on any part of the history prior the current state, neither on the covariates.

The second issue of interval-censoring comes from the fact that the exact transition times are often only observed for the final but not for the intermediate states. In our analysis it was taken into account only in the initial analysis assuming a Markov model, with parametric or eventually piecewise-constant intensities. That was possible due to the fact that for the multi-state model with continuous time the inference problem can be decoupled into several survival problems and assuming the ignorability of the observation process the likelihood for the whole observation of the trajectory can by written [70]. This is not the case for the models with discrete time observations. For nonparametric approaches, in particular, for the pseudo-value approach based on the Aalen-Johansen estimator for the occupation probabilities, we assumed the exact transition times ignoring the interval-

censoring and allowing only for possible right censoring. As an alternative, in the interval-censoring situation, usually the kernel-based methods are proposed in order to obtain nonparametric estimators in multi-state models. Datta and Sundaram [71] suggested smoothed product limit estimators for the occupation probabilities for the current status data when the individuals are not monitored constantly but each individual is observed once at random time point. For non-homogenous Markov models a penalized likelihood approach was also proposed for the simple multi-state models to take into account the interval-censoring problem. Also in the approach of Andersen [45] only the right type of censoring was handled, which was assumed to be stochastically independent of the event times and covariates. The possible consequences of ignoring the interval-censoring have been discussed by Joly et al. [72] for a simple three-state "illness-death" model.

The rest of the paper is organized as follows. In Section 3.2 we describe the methodology for non-Markov models that was applied for the analysis of the real heart transplant data. The obtained results can be found in Section 3.3. In Section 3.4 we present results from a small simulation study. Finally in Section 3.5 we discuss the applied methodology and provide some final conclusions. General framework for multi-state models is provided in Supplementary material B. All the codes and an example of the simulated data set are available at: http://smj.sagepub.com.

3.2 Non-Markov Models

In multi-state models one usually assumes the process is Markov and homogenous in time. Then the transitions probabilities can be derived from

the transition intensities which can be modeled using baseline covariates by means of standard hazard-based models. These may include both multiplicative hazard models and additive hazard models. Provided that there are no loops, i.e., that there is no way to come back to a given state, the explicit expressions for transition intensities are available and yield plug-in methods for the corresponding probabilities, both for Markov and semi-Markov homogenous models. Even though the plug-in methods seem to be simple and straightforward, they often lead to complicated relations between the covariates and transition probabilities. Moreover, they can be used only when the transition intensities are constant or piecewise constant in time. This implies that in regression models for transition intensities only time-fixed covariates are allowed. An attractive alternative involves the direct modeling on transition probabilities using non-parametric Aalen-Johansen (A-J) estimator .

Relaxing the Markov assumption makes the standard parametric approaches inapplicable and restricts the use of nonparametric Aalen-Johansen estimator only for the occupation probabilities. In general, literature on non-Markov multi-state models is scarce and involves mainly nonparametric methods for a direct estimation of the transition probabilities. For a non-Markov illness-death process without recovery Meira-Machado et al. [73] derived nonparametric estimators of the transition probabilities, which they compared with the Aalen-Johansen estimator under different scenarios in a simulation study. The results presented systematic bias of the A-J estimator when the Markov assumption is violated. Nevertheless, the state occupation probabilities can be still estimated using the Aalen-Johansen estimator since the product-integral estimator for those probabilities is consistent without re-

quiring the Markov assumption. The variance of the occupation-probability estimator in a non-Markov setting is not readily available and thus the Bootstrap method is required to estimate it [74]. The above apply for the case of non-Markov process with random censoring. The stage-dependent censoring can be handled via inverse probability of censoring weighting, as proposed by Datta and Satten [75]. In general, as noted by [70], the likelihood or Bayesian approaches seem to be safer than marginal models because of the selection and censoring problems. However, in non-Markov situation, as for our data, the likelihood-based approaches do not apply and one must rely on marginal inference that usually ignores the interval-censoring problem. Some proposed approaches (Cook et al. [76], Sutradhar et al. [77]) allow to handle conditionally Markov progressive model under interval-censoring using random effects approach.

3.2.1 Pseudo-values approach for the non-parametric Aalen-Johansen estimator

In order to assess the effect of covariates on the occupation probabilities we will apply the pseudo-values approach proposed by Andersen et al. [45] to Aalen-Johansen estimator of the occupation probabilities. This approach was originally proposed to model the cumulative incidence function in a competing risk model. To introduce the pseudo-values approach, let $K = \{1, 2, \ldots, N\}$ denote the finite state space for the considered multi-state process X with the time interval $\Gamma = [0, \tau], \tau < \infty$. First, we calculate the so-called pseudo-values for the function of interest $F(X)$, which in our case is the Aalen-Johansen estimator of the occupation probability. More specif-

ically, the pseudo-values are defined as:

$$\hat{\theta}_i(t) = M\hat{F}(X)(t) - (M-1)\hat{F}(X)^{-i}(t), \tag{3.1}$$

where $\hat{F}(X)(t)$ and $\hat{F}(X)^{-i}(t)$ are the estimators of $F(X(t))$ at time t calculated for all individuals and for the subset without individual i, respectively, $i \in 1, \dots M$. The idea of pseudo-values is based on the leave-one-out estimator $\bar{F} = \sum_{i=1}^{M} \hat{F}^{-i}$ [78]. If $\hat{F}$ is unbiased, then $E(\bar{F}) = F$. In particular, it can be shown that $\hat{F}_{Jack} = \hat{F} - b_{Jack}$, where $b_{Jack} = (M-1)(\bar{F} - \hat{F})$, is an unbiased estimate of F up to the second order. That holds, in particular, for the mean of the pseudo-values $\hat{\theta}_i$.

In the pseudo-values approach of Andersen et al. (2003), $\hat{\theta}_i(t)$ are calculated for all individuals i at some arbitrary chosen time points $t_1, \dots, t_k$. Therefore, we obtain k pseudo-values for each subject. Note that due to the fact that $\hat{\theta}_i(t_1), \dots, \hat{\theta}_i(t_k)$ are evaluated for the same subject, they will be correlated. Hence, in the next step, and in order to measure the effects of covariates on $F(X)$, the Generalized Estimating Equations (GEE) approach is utilized with $\hat{\theta}_i$ as a response. Let Z_i denote a vector of covariates of interest. We model the expected value of the pseudo-values, as:

$$E(\theta_i) = g^{-1}(\beta^T Z_i), \tag{3.2}$$

with $g(\cdot)$ denoting a monotonic link function. Estimates of β are based on

the unbiased estimating equations:

$$\sum_i \left\{ \frac{\partial}{\partial \beta} g^{-1}(\beta^T Z_i) \right\}^T V^{-1} \left\{ \hat{\theta}_i - g(\beta^T Z_i) \right\} = \sum_i U_i(\beta) = U(\beta) = 0, \quad (3.3)$$

where V^{-1} is a working covariance matrix. The covariance matrix of $\hat{\beta}$ is obtained using the sandwich estimator [79].
Following Graw et al. [68] here we require that:

$$E(\theta_i \mid Z_i) = g^{-1}(\beta^T Z_i) + O_P(1). \quad (3.4)$$

The above condition of conditional unbiasedness of the pseudo-values given the covariates formulated by Graw et al. [68] relaxes unbiasedness of the pseudo-values, as formulated in Andersen et al. [45] . This relaxation is required in order to be able to use an appropriate theorem of the GEE approach [80] to prove the large sample properties of the GEE solution for β. Condition (3.4) holds trivially when the observations are uncensored, and it was shown by Graw et al. [68] to also hold for the pseudo-values in the right censored situation. The central argument was a second order von Mises expansion of the Aalen-Johansen estimate which leads to an appropriate representation of the jackknife pseudo-values [81]. The interval-censoring was not considered.

We apply the pseudo-values approach to the Aalen-Johansen estimator of the occupation probabilities. We define first the Aalen-Johansen estimator

$\hat{F}_h(t)$ for the probability $p_h(t)$ of occupying state h at time t as:

$$\hat{F}_h(t) = \sum_{k=1}^{N} p_k(0)\hat{p}_{kh}(0,t), \quad h \in K; t \in \Gamma, \tag{3.5}$$

where $p_k(0)$ is the initial distribution of the process at time 0 and $\hat{p}_{kh}(0,t)$ are the estimates of the transition probabilities between states k and h from time 0 to t, that are obtained from the Aalen-Johansen product-integral estimator for the transition probability matrix $P(s,t) = (P_{kh}(s,t))$ for $s = 0$ as:

$$\hat{P}(0,t) = \prod_{(0,t]} (I + d\hat{A}(u)). \tag{3.6}$$

In the above formula $\hat{A}$ denotes the standard Nelson-Aalen estimator for the cumulative transition intensity matrix [79]. Note, as mentioned in the Introduction, that in this nonparametric approach we ignore interval censoring allowing only for potential right censoring when calculating A-J estimator.

$\hat{F}_h(t)$ is calculated for all individuals and we denote by $\hat{F}_h^{-i}(t)$ the A-J estimate for the subset without individual i, $i = 1, \ldots, M$. Then for subject i at time t a pseudo-value is calculated:

$$\hat{\theta}_i(t) = M\hat{F}_h(t) - (M-1)\hat{F}_h^{-i}(t). \tag{3.7}$$

Then, to assess the effects of covariates on the occupation probabilities, we fit a regression model on pseudo-values using the GEE approach according to (3.2) and (3.3), under the assumption (3.4).

Note that in the complete case, i.e. when there is no censoring, $F_h(t)$

could be estimated unbiasely as the proportion of subjects in state h at time t. Then $\theta_i(t)$ is simply $F_h(t)$. With censoring we replace possibly unobserved $F_h(t)$ by the pseudo-value $\hat{\theta}_i(t)$. In the regression model for θ, in which we model $E(\theta_i)$ the $\hat{\theta}_i$ is used as a response. Validity of the assumption (3.4) was analyzed Graw et al. [68] only in the situation of competing risks. However, following their suggestion, this condition can be proved also in a more general model by constructing the pseudo-values using the smoothed mappings of the Nelson-Aalen and the Kaplan-Meier functionals.

We considered the univariate pseudo-values approach, in which we model each occupation probability separately as well as the multivariate pseudo-values approach where all occupation probabilities are modelled together, as suggested by Andersen et al. [67]. Therefore, in the multivariate approach we model the vector $\hat{\theta}_i(t) = (\hat{\theta}_{1i}(t), \ldots, \hat{\theta}_{N_i}(t))$ of the pseudo-values for the occupation probabilities for all N states.

Apart from the formal results of Graw et al. [68] , the pseudo-values approach was mainly evaluated using the simulation methods. The choice of numbers of time points and their location as well as the choice of a link function were shown to have a moderate impact on the results [64].

3.2.2 Multinomial models approach

The regression models based on pseudo-values, presented in the previous section, allow us to study the effect of baseline covariates on the occupation probabilities in a non-Markov model, when the dependence on the history is a nuisance. However, it does not allow to assess the impact of the history on those probabilities. To overcome this limitation we propose here an alternative approach, under which we can study how the observed history

affects the occupation probabilities. This is based on viewing a homogenous Markov model as series of multinomial models for each observation conditionally on the previous observation. These models may be fitted using standard software for multinomial logistic regression. However, when the time is not discrete and the Markovian assumption is violated, we may still fit the multinomial model on the transition level, taking as a response the next state observed and adjusting for the time of transition as well as for baseline covariates.

In particular, for each individual we consider the triples (s_-, s, t) such that the person visited the particular state p at time s before the current state c, i.e : $X(s_-) = p, X(s) = c, X(t) = h,\ h, c, p \in K,\ s_-, s, t \in \Gamma; s_- < s < t$. For these triples we considered the following model:

$$\Pr(X(t) = h \mid X(s) = c; X(s_-) = p) = \frac{\exp(\omega_h)}{1+\Sigma_{k\neq r}\exp(\omega_k)}, h \neq r;$$
$$\omega_h = \gamma_{h0} + \gamma_{h1}t + \gamma_{h2}Z, \tag{3.8}$$

where $X(t)$ is the next state, t-time of transition, Z-baseline covariates. For the reference category r we have:

$$\Pr(X(t) = r \mid X(s) = c; X(s_-) = p) = \frac{1}{1 + \Sigma_{k\neq r}\exp(\omega_k)}. \tag{3.9}$$

In the general setting, current and previous states could be included as covariates in a model (3.8-3.9). In our application, conditioning on c and p in (3.8) and (3.9) is realized by considering the corresponding subset of all transitions for the reasons explained in the next section. We work now on

the transition level, instead of the individual level. Since we pool all individuals together, we need to account for the correlation introduced by using the same patient more than once. For that reason SE for the multinomial models estimates were calculated by means of leave-one-out jackknife estimators estimator [78]. In particular, let π denotes the estimated probability of interest from the multinomial model (3.8-3.9) and π_{-i} denotes the same probability estimated excluding individual i. Then we calculate jackknife SE as simple standard deviation of the $\hat{\pi}_{-i}$ estimates:

$$SE^{Jack}(\hat{\pi}) = \sqrt{\frac{N}{N-1}\sum_{i=1}^{N}(\hat{\pi}_{-i} - \overline{\hat{\pi}}_{-i})^2}, \qquad \overline{\hat{\pi}}_{-i} = \frac{1}{N}\sum_{i=1}^{N}\hat{\pi}_{-i}. \tag{3.10}$$

3.3 Analysis of the Heart Transplant Data

We apply the described methodology to the transplantation study introduced in Section 3.1. The data were taken from an international data base of the Eurotransplant Heart recipient waiting list. We consider 2921 recipients who entered the list in the period from 01.01.2006 to 31.12.2008. Recipients' observation was censored at 31.03.2010. At that time 528 patients have died (18%), 1566 (54%) have received a transplant, 239 (8%) were removed from the list because of other reasons and 588 (20%) were still on the waiting list. At the moment of entering the list some of the baseline information was also collected, namely: country of origin (7 countries), blood group and cardiovascular disease (categorized into Dilated Cardiomyopathy (DCM), Coronary Artery Disease (CAD) and others).

Following a standard approach in the multi-state models framework, we

started our analysis assuming a Markov process and modeled each intensity using the parametric approach. We considered only patients with more than one transition. In addition, due to numerical reasons and in order to be able to fit the model, we also imposed the constraints for the intensities with small number of transitions setting them to zero. To evaluate the fit of the Markov model, we graphically compared the observed and expected prevalence for each state for different models. We did not expect to observe substantial differences when assuming the exact transition times and without that assumption. Plausible reasons for the observed discrepancies include the dependence of transition rates on omitted covariates or on time (non-homogenous Markov process). The other possible reason of the discrepancies is the dependence of transition intensities on the time spent in the current state (semi-Markov) or the history of the process (non-Markov process). We estimated transition probabilities $P_{hr}(0,t)$ using Aalen-Johansen estimator , where h is not an absorbing state and $h \neq r$. $P_{hr}(s,t)$ were relatively close to $P_{hr}(0,t-s)$ indicating that these probabilities seem to depend on the time interval $t-s$. This suggests that the process is homogenous. Nonetheless, as we observed the time-homogenous Markov model had a poor fit. Therefore, the main issue for the data at hand seems to be that the Markov assumption is not satisfied.

Further examination of the history of individual profiles suggested common patterns such as (T, NT, T, NT, ...) or (T, HU, T, HU, T, HU, ...). Since the Aalen-Johnsen estimator cannot be used to investigate the Markov assumption, we decided to apply the multinomial model approach presented in Section 3.2.2 adjusted for blood group, cardiovascular disease and informed consent law (IC) (binary). In our setting, the transition to the same

state is not allowed and additionally for some transitions in the real data we observed very low frequency. Therefore, conditioning on c and p in (3.8) and (3.9) is realized by considering the corresponding subset of all transitions. The death was chosen as the baseline category for most of the models, unless there were none or very few transitions to death from a given current state c, for a given previous state p. Separate models were fitted for given states c and p and jackknife SE were calculated according to formula (3.10). In practice in most of the cases due to the small subgroup sample sizes the adjustment for baseline covariates was not possible.

The results of this analysis are presented in Tables 3.1 and 3.2. None of the effect of baseline covariates, except the effect of blood group for a model with the previous state T and current HU in Table 3.1, was found to be significant. This is probably attributed to the lack of power due to the small size of the subset sample. Therefore, we present only the model adjusted for time and not for baseline characteristics (except the model in Table 3.1). Further results can be found in Supplementary material B. Table 3.1 presents the results from the multinomial model for the previous state Transplantable and the current state HU. Because there were only 3 transitions to state R from HU, they were excluded from the analysis. Table 3.1 presents the results from the multinomial model for the previous state Transplantable and the current state HU. Because there were only 3 transitions to state R from HU, they were excluded from the analysis. Among the patients with the previous state Transplantable and the current state U there was only one case of transition to the state R and it was excluded from the analysis. As can be seen from Table 3.2 a multinomial model grabs the peak in the probability of being in state HU in the early times. This results in the highest

estimate for log(OR)=1.52, which is almost significant (p=0.058). This is due to the fact that most of the transitions to state HU from other states took place in early times of being on the waiting list. For the current state NT there were only 2 transitions to the state HU and one to the state U and they were excluded from the analysis. There were no transitions to TT. As can be concluded, from NT patients were most likely to go back to T. The probability of death was greater than probability of being removed. We analyzed in the same manner other histories. The results for those models are provided in Supplementary material B.

Table 3.1: *Estimates of log(OR) for the effect of current state and time on the probability for the next transition from the current state HU based on the multinomial model with the previous state Transplantable, adjusted for time and baseline covariates. Baseline category is probability of death. Jackknife SE are calculated. The estimates for the time effect are not listed.*

Previous=T

Current=HU

	Intercept	Blood B vs A	Blood AB vs A
P(NT)/P(D)	0.42(1.13)	-0.54(0.58)	-0.9(1.3)
P(HU)/P(D)	-	-	-
P(U)/P(D)	-13.76(6.81)	-1.15(0.62)	-1.44(1.26)
P(T)/P(D)	-0.23(1.01)	-1.28(0.54)	-1.13(1.01)
P(R)/P(D)	-	-	-
P(TT)/P(D)	2.58(0.68)	-0.5(0.48)	0.01(0.76)

As mentioned in the previous section, the A-J estimates for the transition probabilities are consistent only if the Markov assumption is satisfied.

Table 3.2: *Estimates of log(OR) for the effect of the current state and time on the probability for the next transition from the current state U and NT based on the multinomial model with the previous state Transplantable, adjusted for time. Baseline category is probability of death. Jackknife SE are calculated. The estimates for the time effect are not listed.*

Previous=T

	Current=U	**Current=NT**
	Intercept	Intercept
P(NT)/P(D)	-1.25(0.93)	-
P(HU)/P(D)	1.52(0.80)	-
P(U)/P(D)	-	-
P(T)/P(D)	0.52(0.81)	1.52(0.22)
P(R)/P(D)	-	-0.91(0.25)
P(TT)/P(D)	-0.23(0.82)	-

However, for our data, and as shown in the previous analysis using the multinomial approach, the Markov assumption seems to be violated. Therefore, in the final analysis we use the A-J estimates for the state occupation probability, which are consistent regardless of whether the Markov assumption is satisfied. To measure the effect of covariates in these occupation probabilities we employed the pseudo-values approach introduced in Section 3.2.1. To apply this approach we need to first choose the time points, at which the estimators are to be calculated. In the related framework of the competing risks models it has been shown that the choice of the number of grid points can be proven crucial [64]. To be flexible, here we have chosen 10 time points based on the quantiles of the observed time distribution. To take into account the within-individual correlation (for each patient we calculate 10 pseudo-values) we use a GEE model with an unstructured correlation matrix and identity link. This model contained as covariates time, age, blood group, type of the

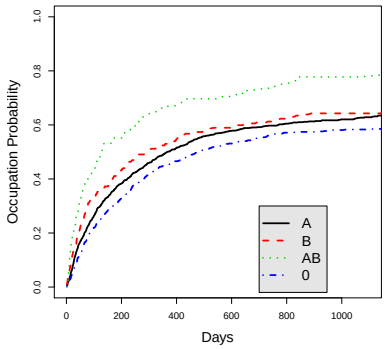
disease (CAD, DCM and other (OD)) and IC indicator. The blood group A and CAD disease were set as the reference levels. The GEE model with time and baseline covariates was fitted for each occupation probability separately in univariate pseudo-values model (3.7) and the multivariate version for all occupation probabilities modeled together. In multivariate GEE model the independence correlation matrix was used because of convergence problems for any other option. The results for the univariate model for state Transplanted are presented in Table 3.3 and illustrated in Figure 3.2. We have analyzed in the same univariate manner the occupation probabilities for the remaining states.

Table 3.3: *Estimates of the effect of baseline covariates on occupation probability for state Transplanted. Results from univariate regression model on pseudo-values.*

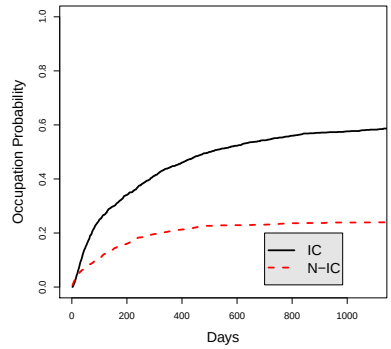
Univariate

	Estimate	**SE**	**p-value**
`Intercept`	-1.090	0.051	< 0.001
`Time (Days)`	-0.723	0.001	< 0.001
`Age`	-0.018	0.001	< 0.001
`Blood B vs A`	0.034	0.027	0.208
`Blood AB vs A`	0.154	0.036	< 0.001
`Blood 0 vs A`	-0.526	0.019	0.006
`DCM vs CAD`	-0.096	0.042	0.021
`Other Disease vs CAD`	-0.046	0.044	0.293
`IConsent (Y vs No)`	-0.182	0.019	< 0.001

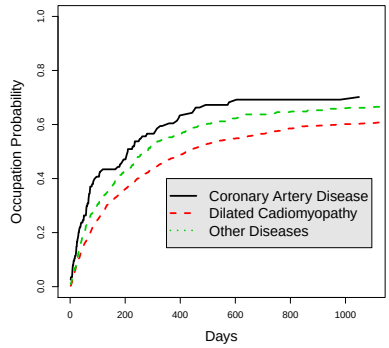
Using univariate pseudo-values approach the probabilities of being in the given state can be compared between the different groups defined in terms of baseline characteristics such us blood group. However, we cannot compare



(a)



(b)



(c)

Figure 3.2: *Aalen-Johansen estimators for occupation probability for state Transplanted for patients with different baseline characteristics.*

the two complementary probabilities of being in state h and r. For example, we can compare if patients with AB blood group have higher probability of getting a transplant than patients with A blood group. Based on the univariate pseudo-values approach we can also compare the probability of death between the patients with A and AB blood group. However, we cannot assess, what is more probable: death or getting a transplant, having the specified blood group. Therefore we also fitted multivariate pseudo-values model. The results regarding the impact of baseline covariates were similar as for the univariate approach and can be found in Supplementary material B. However, based on multivariate model clinician could conclude for example that patient with blood group AB is more likely to get a transplant than die as comparing to patient with blood group A. Similar questions could be also addressed using the multinomial approach.

All computations have been performed in R. In particular, we used function `msm()` from **msm** package for fitting the Markov model. Piecewise-constant intensity models were also fitted using this function. For the Aalen-Johansen estimator calculation the `etm()` function from **etm** package was used. Transitions to the same state are not allowed in the `etm()` function and therefore the times for those transitions were combined. For the multinomial models on the transition-level, the `multinom()` function in the **nnet** package was used with the additional code written for the jackknife standard errors calculation. To apply pseudo-value approach for the occupation probabilities pseudo-values calculation needed to be implemented. Then the `geese()` function from the **geepack** package was used to fit the GEE models on the pseudo-values. All the code used in our analyses is available at: http://smj.sagepub.com.

3.4 Simulation Study

3.4.1 Design

To investigate the performance of the pseudo-values approach for non-Markov models we performed a limited simulation study of 4 different scenarios with 500 samples per each scenario. In all scenarios 4 transient and 3 final states corresponding to the analyzed real data set were considered. For each data set, and for N=1500 subjects we simulated the waiting times T_{hr} for a transition from state h to r from an exponential distribution with rate λ_{hr}. If the simulated waiting time T_{hr} was smaller than the simulated waiting time T_{hs}, then the individual moved from h to r. Each subject was initially in state T and could have at maximum 10 transitions. Individuals being in any of the non-final states after 10th transitions were considered as censored. In Scenario I we simulated data from a homogenous Markov model with no group effect. Individual waiting times were drawn from an exponential distribution with equal rate parameters $\lambda = \exp(-1)$ for all possible transitions so that the average waiting time for any transition was around $\exp(1) = 2.7$. In Scenario II we simulated data from Markov model with a group effect on the intensity of transition to state TT. To simulate the group effect in each data set the individual was randomly chosen to be in one of the 2 groups with probability 0.5. Then for all individuals in group 1 the transition times were drawn from an exponential distribution with equal rate parameters $\lambda = \exp(-1)$ for all possible transitions, whereas in group 2 the waiting times were for the transition times to TT were twice shorter. In Scenario III the non-Markov model was simulated in which the group effect reflected the dependence on the history. The transition times were drawn

from an exponential distribution with equal rate parameters $\lambda = \exp(-1)$ for all possible transitions except that in group 2 the waiting times for a transition to state r were twice shorter if the individual visited state r just prior the current state. Finally, in Scenario IV, we simulated both group effects together, the group effect on the transition intensity to state TT and the group effect related to the history. Therefore, for group 1 the waiting times to state TT were simulated to be twice shorter than in the other group and additionally for the subjects from the same Group 1 the waiting times to the previously visited state were twice shorter than to the other states.

The regression models on univariate and multivariate pseudo-values were fitted for each data set as well as multinomial models. To estimate a true group effect on occupation probabilities within pseudo-values approach framework we fitted GEE models on the true occupation probabilities. Since we estimated twice longer waiting times as a group effect on the history or on the transition to state TT we use $\log(2)$ as the true group effect in a multinomial model on a given intensity. For non-Markov model from Scenario IV we compare the results with and without conditioning on the previous state for the estimated effect on the transition to state TT. As a summary for each scenario bias and root mean squared error (RMSE) were reported. In the regression models on pseudo-values we used $k = 10$ time points in a GEE model placed in the quantiles of the transition times distribution and unstructured variance-covariance matrix. Additionally, for the Scenarios II and IV we have examined the impact of different number of time points in GEE model when increasing sample size. In particular, we have fitted the Markov model from Scenario II and non-Markov model from Scenario IV with k=5 and k=10 time points for the simulated samples of size N=500

and N=1000.

3.4.2 Results

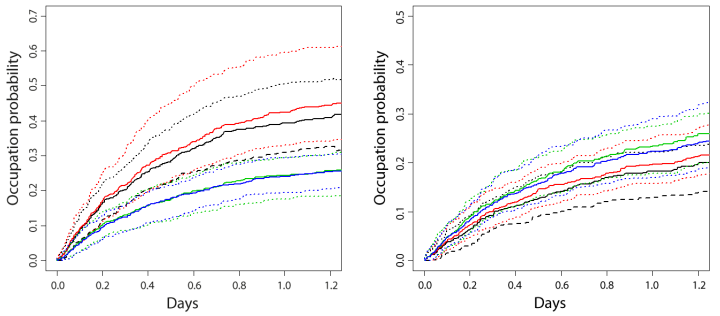
The pseudo values approach did not reveal any difference between the groups when the effect was simulated independently on the response as in Scenario I or when the effect was only dependent on the history in Scenario III (results not shown).

Figure 3.3 illustrates the median of the Aalen-Johansen estimates for the occupation probability for state Transplanted, Death and Removed for all simulated data sets for the Scenarios II and IV. As can be observed, the group effect of different transition intensities q_{iTT} on the probability of occupying state TT is reflected on occupation probabilities for all final states when the process is Markov (Scenario II) or not Markov (Scenario IV). For the state TT we could observe slightly stronger group effect for the Markov process with somewhat larger variability. This is related to lack of the history effect preventing from the transition to state TT in Markov model. The pointwise confidence bounds are based on the simulations. In contrary, as noted by Glidden [74], the pointwise confidence bounds using the recursive formula derived by Andersen et al. [82] for the occupation probabilities do not depend on covariance structure $\sqrt{n}\{(p_h(t) - \hat{F}_h(t)\}$ and depend only on $\text{var}\{p_h(t) - \hat{F}_h(t)\}$, which is the same for Markov and non-Markov processes.

We did not observe any substantial effect on the occupation probabilities of any non-absorbing state, neither for the Markov nor the non-Markov scenario.

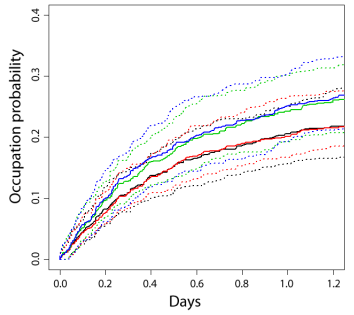
Table 3.4: *Bias and RMSE of the group effect on occupation probability of state Transplanted, Removal and Death for a Markov Model (Scenario II) and non-Markov model (Scenario IV). Results from univariate and multivariate pseudo-values approach for n=500 and n=1000, k=10 time points*

Markov Model			**Non-Markov model**	
	n=500			
	Bias	**RMSE**	**Bias**	**RMSE**
Transplanted				
Univariate	-0.012	0.037	-0.015	0.021
Multivariate	-0.025	0.017	-0.013	0.008
Death				
Univariate	-0.003	0.006	-0.004	0.005
Multivariate	-0.001	0.003	-0.008	0.002
Removal				
Univariate	-0.003	0.018	-0.009	0.019
Multivariate	-0.016	0.011	-0.018	0.017
	n=1000			
	Bias	**RMSE**	**Bias**	**RMSE**
Transplanted				
Univariate	-0.004	0.009	0.014	0.005
Multivariate	-0.013	0.006	0.022	0.004
Death				
Univariate	-0.002	0.003	-0.012	0.006
Multivariate	-0.001	0.002	-0.001	0.001
Removal				
Univariate	-0.014	0.012	-0.008	0.010
Multivariate	-0.030	0.008	-0.016	0.007



(a)

(b)



(c)

Figure 3.3: *Median of Aalen-Johansen estimators for occupation probability for state Transplanted (a), Death (b) and Removal (c) for the data simulated from Markov model from Scenario II (group 1-red, group 0-blue) and non-Markov Model from Scenario IV (group 1-black, group 0-green). Dotted lines represent pointwise confidence bounds.*

When the comparing univariate and multivariate pseudo-values approaches, the latter seems to lead to smaller RMSE. When investigating the impact of the number of time points in the GEE model for the different sample sizes for Scenarios II and IV we observe that increasing the number of time points

k and sample size n leads to smaller RMSE (Table 3.4 and 3.5). We observed rather smaller RMSE in non-Markov model, especially for the multivariate pseudo-value approach and more time points.

Table 3.5: *Bias and RMSE of the group effect on occupation probability of state Transplanted, Removal and Death for Markov Model (Scenario II) and non-Markov model (Scenario IV). Results from univariate and multivariate pseudo-values approach for n=500 and n=1000, k=5 time points.*

Markov Model			**Non-Markov model**	
	n=500			
	Bias	**RMSE**	**Bias**	**RMSE**
Transplanted				
Univariate	-0.015	0.055	-0.033	0.065
Multivariate	-0.019	0.023	-0.021	0.034
Death				
Univariate	0.007	0.015	0.014	0.066
Multivariate	0.008	0.005	0.009	0.028
Removal				
Univariate	0.006	0.047	0.017	0.032
Multivariate	0.045	0.014	0.027	0.020
	n=1000			
	Bias	**RMSE**	**Bias**	**RMSE**
Transplanted				
Univariate	-0.011	0.035	-0.013	0.010
Multivariate	-0.009	0.020	-0.003	0.008
Death				
Univariate	0.006	0.008	0.009	0.015
Multivariate	0.007	0.010	0.003	0.010
Removal				
Univariate	0.004	0.024	0.012	0.021
Multivariate	0.020	0.016	0.019	0.013

Table 3.6: *Bias and RMSE of the group effect on the probability of transition to state Transplanted (TT) and the effect on history in multinomial models for Markov Model (Scenario II) and non-Markov model (Scenario IV), adjusted for time and current state. Baseline category is probability of death.*

Markov Model

	n=500		**n=1000**	
	Bias	**RMSE**	**Bias**	**RMSE**
P(TT)/P(D)	-0.054	-0.209	0.042	0.163
Previous=T				
P(TT)/P(D)	-0.068	-0.255	0.045	0.223

Non-Markov Model (Scenario III)

	n=500		**n=1000**	
	Bias	**RMSE**	**Bias**	**RMSE**
P(TT)/P(D)	-0.055	0.174	-0.018	0.157
Previous=T				
P(TT)/P(D)	-0.072	0.249	-0.036	0.208
P(T)/P(D)	-0.292	0.366	-0.201	0.352

Non-Markov Model (Scenario IV)

	n=500		**n=1000**	
	Bias	**RMSE**	**Bias**	**RMSE**
Previous=T				
P(T)/P(D)	-0.194	0.351	-0.172	0.277

Table 3.6 illustrates the group effect from the multinomial model conditional on the previous state Transplantable adjusted for time and current state. For the data simulated from two non-Markov models we observe smaller bias and RMSE for the effect on the history $P(T)/P(D)$ for the Scenario III when there is no effect on the probability of the transition to TT, namely on $P(TT)/P(D)$. For $P(TT)/P(D)$ in non-Markov model from Scenario IV obtained with no conditioning on the previous state we observed comparable bias and RMSE as in Markov model. Conditioning on the previous state when estimating $P(TT)/P(D)$ results in larger bias and RMSE which is something to be expected since we take only subset of all transitions. Even when conditioning on the previous state the RMSE and bias for $P(TT)/P(D)$ are smaller than for the effect on the history $P(T)/P(D)$.

In an analogous manner a group effect on the transition intensity to non-final state was simulated. For this scenario the pseudo-value approach detected only the group effect on the occupational probability of absorbing states in non-Markov model. The multinomial model was able to capture the simulated group effect on transition to T both in the Markov and non-Markov models (results not shown).

We have simulated the dependence on the history through the last state since that was suspected for the analyzed real data. However, the true dependence of the process could be more complicated. The simulation study demonstrated that the pseudo-values approach can capture the group effect of the transition intensity to one final state on the occupational probability of all final states of such a non-Markov model. Additionally multinomial model was able to capture also the effect related to the history.

3.5 Discussion

Motivated by study on patients on the waiting list for heart transplantation, in this paper we presented two approaches for analyzing multi-state data for which the Markov assumption is not defensible. In particular, we use the pseudo-values approach to estimate the effects of baseline covariates on the occupation probabilities when the history dependence is a nuisance. When the interest is in a dependence history the additional information can be derived from the multinomial model approach. Both approaches are simple in principle and can be easily implemented for any general non-Markov multi-state model. We observed less efficiency in the multinomial approach in estimating the history effects than for the effects on the transition to the final state. We should consider rather series of multinomial models with the fixed time point in each model and for each of these models we should consider the subset of individuals that have made a transition in that fixed time point. Since in our case the transition times are very dense and change rather continuously, it is very hard to find the reasonable large group of individuals with the same transition time. The pseudo-values approach requires independent censoring. One can however relax that assumption to conditional independence of censoring and event times given covariates and then apply an inverse probability of censoring weighting techniques [83]. The method was recently extended for the case of clustered time-to-event data in competing risk model by Logan et al. [84]. So far pseudo-values approach was mainly evaluated using the simulation results [45, 64, 68]. Based on the simulation studies the choice of the number of time points, their location and also the choice of the correlation matrix in a GEE model was shown to have

an impact on the results. The impact of the number of time points was also demonstrated in our simulation study. We have observed an increased efficiency for the multivariate pseudo-values approach comparing to univariate model. The efficiency also increased for larger sample size and more time points in GEE model . Adjusting for the optimal number of time points could be more problematic for the occupation probabilities of non-absorbing states since they usually exhibit a non monotonic behavior. These aspects need to be further studied as well as the assessment of the goodness-of-fit and the choice of link function in GEE models.

3.6 Supplementary Materials

Supplementary materials are available online at http://smj.sagepub.com.

Acknowledgements

We thank Jacqueline M. Smits from Eurotransplant International Foundation in Leiden for providing the data set analyzed in the paper and for the medical consult.

Chapter 4

Nominal Longitudinal Response

In this chapter we present Bayesian approach to jointly model the performance of patients described by their categorical status that changes in time while waiting for the new organ together with the survival time on the waiting list. The model accounts also for the presence of competing risks due to the fact that patients are delisted from the list because of death, after transplantation or because of other reasons. Bayesian model constitutes the submodel for longitudinal categorical response being multinomial logit mixed-effects model and the cause-specific hazard model that shares the same random effects with each logit of the multinomial logit mixed-effects model. We illustrate how the fitted joint model can be used for the dynamic prediction of the cumulative incidence functions as well as the categorical response based on their available longitudinal measurements of that response. We also investigate the impact of different parameterizations of the

joint model on the dynamic predictions in a simulation study.

4.1 Introduction

In transplantation studies longitudinal measurements are often collected for patients waiting for an organ transplant. For such patients it is often of interest to investigate whether available longitudinal measurements can be used for predicting patient survival as well as further performance on the waiting list. The motivation for this work comes from the Eurotransplant heart recipients registry with 2921 recipients entering the waiting list in a period of three years. Each recipient was classified to one of the following states: Transplantable (T), Non-Transplantable (NT), Urgent (U) and High Urgent (HU). A graph depicting all possible transitions between the states is presented in Figure 4.1. The evaluation time was different for each patient and depended on the previous classification meaning that more severe patients were evaluated more frequently. The first evaluation took place at entry and additional evaluations were performed while the patient remained on the waiting list. Upon the censoring date, 528 patients had died (D) without receiving a transplant, 1565 patients received a transplant (TT) and 239 patients had been removed (R) because of other reasons. The purpose of the study is to predict the state of the patient and to estimate the risk of any of the 3 competing events (TT, R, D) based on the history of states on the waiting list with adjusting for baseline covariates.

The majority of prognostic models in the medical literature utilize only a

This chapter has been submitted for publication as "Dynamic prediction based on joint model for categorical response and time-to-event with application in transplantation studies" in *Biostatistics*

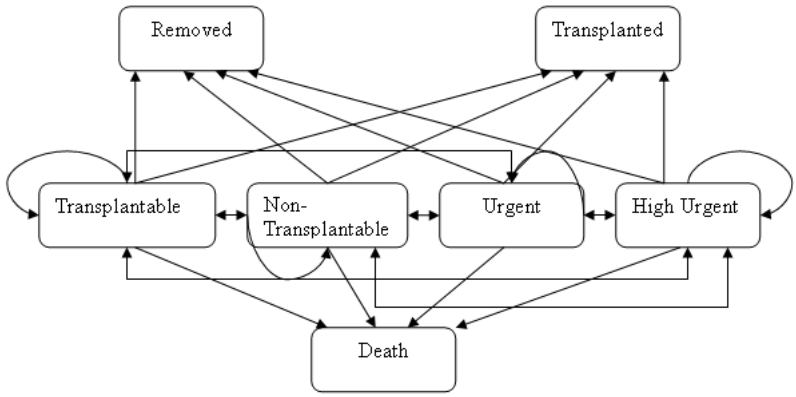


Figure 4.1: *Graph of all possible transitions between the states.*

small fraction of the available biomarker information. Even though biomarkers are measured repeatedly over time, risk scores are typically based on the last available biomarker measurement. Such an approach discards valuable information because it does not take into account that the rate of change in the biomarker levels is not only different from patient to patient but also dynamically changes over time for the same patient. Hence, it is medically relevant to investigate whether repeated measurements of a biomarker can provide a better understanding of disease progression. In particular, for the Eurotransplant clinicians the possibility of updating prediction of the risk of death based on the changing urgency status for a specific patient allows to make an intervention such as putting her at the top of the waiting list, in order to improve the chance of survival of that patient. Motivated by these arguments in this work we propose a joint modeling approach for producing dynamic predictions of the future status of a patient in the list and

her chance for getting a transplant, being removed from the list or dying. In particular, we model the urgency status as a categorical longitudinal response variable , which is assumed to be associated with the competing risks process of leaving the list. We follow a Bayesian approach for estimating the joint model based on which we derive posterior predictive distributions for the longitudinal and event time outcomes. Additionally, we also examine the impact of the different parameterizations of the joint model on the obtained predictions. By the different parameterizations we mean different functional relationship between the longitudinal and time-to-event outcomes.

Dynamic predictions have been so far studied only for continuous longitudinal outcome ([85], [86]). Categorical longitudinal responses have received much less attention in the joint modeling framework and the proposed approaches mainly handle nonrandom dropouts for discrete longitudinal data, also in Bayesian approach ([16, 18, 27, 28, 86–89]). Similarly in the context of competing risks the proposed joint models focus mainly on joint analysis of survival and repeated continuous or ordinal biomarkers, using either a likelihood ([33, 34]) or a Bayesian approach [36]. We propose a Bayesian model for joint modeling of categorical longitudinal data and time-to-event response in presence of competing risks. Following [86] we use the Monte Carlo approach to obtain dynamic subject-specific predictions based on the fitted joint model. The predictions of the cumulative incidence functions and the categorical longitudinal response are updated as additional measurements of the longitudinal response become available. Our approach can be applied using standard software (OpenBUGS/WinBUGS) with additional programming to compute the dynamic predictions.

The rest of the paper is organized as follows. Section 4.2 provides the

general modeling framework with the definition of the two submodels for the longitudinal and survival data, respectively. Section 4.3 describes the estimation methods for the proposed approach. Section 4.4 provides more details regarding the sampling procedure used for the dynamic prediction for the cumulative incidence function and the longitudinal response. In Section 4.5 we apply the proposed approach to the heart transplant data. Section 4.6 contains the setup and the results for the simulation study. Finally, in Section 4.7 we discuss the proposed methodology.

4.2 Joint Modeling Framework

4.2.1 Submodels specification

Let $\{Y_i(t) = r, r = 1, 2, \ldots, R\}$ denote the response category for individual i $(i = 1, \ldots, N)$ at time t, and let t_{ij} $(j = 1, \ldots, m_i)$ denote the time points at which measurements are taken for this subject. To model $Y_i(t)$ we postulate a multinomial mixed effects model. In particular, the probability that at time t the longitudinal response $Y_i(t)$ is equal to r conditional on the random effects $\boldsymbol{b}_i$ is given by:

$$\begin{aligned} p_{ir} = \Pr\left(Y_i(t) = r \mid \boldsymbol{b}_i\right) = &\quad \exp\{w_{ir}(t)\} \Big/ \Big[1 + \sum_{h=2}^{R} \exp\{w_{ih}(t)\}\Big], \text{ and} \\ p_{i1} = \Pr\left(Y_i(t) = 1 \mid \boldsymbol{b}_i\right) = &\quad 1 \Big/ \Big[1 + \sum_{h=2}^{R} \exp\{w_{ih}(t)\}\Big], \end{aligned} \tag{4.1}$$

where $w_{ir}(t) = \boldsymbol{x}_i^T(t)\boldsymbol{\alpha}_r + \boldsymbol{z}_i^T(t)\boldsymbol{b}_{ir}$, $\boldsymbol{x}_i(t)$ is the design vector of fixed effects and $\boldsymbol{z}_i(t)$ is the design vector for the random effects $\boldsymbol{b}_i, \boldsymbol{b}_i^T = (\boldsymbol{b}_{i1}^T, ..., \boldsymbol{b}_{iR}^T)$.

Therefore for each category r we postulate a different random vector $\boldsymbol{b}_{ir}$. In such type of models usually the random effects $\boldsymbol{b}_{ir}$ are treated as independent multivariate normal variables. We opt for using an arbitrary covariance matrix $\boldsymbol{D}$ and assume that $\boldsymbol{b}_i \sim N(0, \boldsymbol{D})$ for a complete vector $\boldsymbol{b}_i^T = (\boldsymbol{b}_{i1}^T, ..., \boldsymbol{b}_{iR}^T)$. For the survival process we consider K different causes of failure, with $T_{i1}^*, T_{i2}^*, \ldots, T_{iK}^*$ denoting the true failure times for individual i. Since the failure times are right censored we observe only $T_i = \min(T_{i1}^*, T_{i2}^*, \ldots, T_{iK}^*, C_i)$, where C_i is the censoring time with the failure indicator $\Delta_i \in \{0, 1, \ldots, K\}$, which equals 0 if the subject was censored and $1, \ldots, K$ for the corresponding competing event. We assume that censoring is at time t is independent of future failures as well as of future longitudinal responses. For each of the K causes we postulate a standard hazard model that shares all random effects $\boldsymbol{b}_{ir}$ with the multinomial logit model:

$$\begin{aligned} \lambda_{ik}(t) &= \lim_{s \to 0} \Pr(t \leq T_i^* < t + s, \Delta_i = k \mid T_i^* \geq t)/s \\ &= \lambda_{0k}(t) \exp(\boldsymbol{\gamma}_k^T \boldsymbol{f}_i(t, \boldsymbol{b}_i) + \boldsymbol{\beta}_k^T \boldsymbol{v}_i), \quad k = 1, \ldots, K, \end{aligned} \tag{4.2}$$

where $\boldsymbol{b}_i^T = (\boldsymbol{b}_{i2}^T, \ldots, \boldsymbol{b}_{iR}^T)$, $\boldsymbol{v}_i$ is a vector of baseline covariates, and $\lambda_{0k}(t)$ is the baseline hazard that can be modeled parametrically or left unspecified. Parameters $\boldsymbol{\gamma}^T = (\boldsymbol{\gamma}_1^T, \ldots, \boldsymbol{\gamma}_k^T)$ measure the strength of the association between the longitudinal and survival processes .

We will consider two options for the function $f(\cdot)$. First, we name the Random-Effects Model (R-E) the model (B.2.1) with $\boldsymbol{f}_i(t, \boldsymbol{b}_i) = \boldsymbol{b}_i$ in which case the submodels (B.2.1) and (B.2.1) for the longitudinal and survival processes share only random effects. Second, we name Time-Dependent

Model (T-D) the model (B.2.1) with $\boldsymbol{f}_i(t, \boldsymbol{b}_i) = \boldsymbol{w}_i(t)$ for which we will allow both submodels to share also linear time-dependent terms from the multinomial mixed model, i.e.:

$$\text{(R-E)}: \quad \lambda_{ik}(t) = \lambda_{0k}(t) \exp(\boldsymbol{\gamma}_k^T \boldsymbol{b}_i + \boldsymbol{\beta}_k^T \boldsymbol{v}_i), \quad k = 1, 2, \ldots, K, \tag{4.3}$$

$$\text{(T-D)}: \quad \lambda_{ik}(t) = \lambda_{0k}(t) \exp(\boldsymbol{\gamma}_k^T \boldsymbol{w}_i(t) + \boldsymbol{\beta}_k^T \boldsymbol{v}_i), \quad k = 1, 2, \ldots, K, \tag{4.4}$$

where $\boldsymbol{b}_i^T = (\boldsymbol{b}_{i2}^T, \boldsymbol{b}_{i3}^T, \ldots, \boldsymbol{b}_{iR}^T)$ and $\boldsymbol{w}_i(t) = \{w_{i2}(t), \ldots, w_{iR}(t)\}$.
Within the (T-D) formulation there are more than one parameterizations possible depending on which terms $w_{ir}(t)$ are shared by each of the hazard relative risk submodels. In particular, one may consider the parametrization, for which all $R-1$ linear predictors $w_{ir}(t), r = 2, \ldots, R$, from longitudinal part are shared with survival model:

$$\text{(T-D)}: \quad \lambda_{ik}(t) = \lambda_{0k}(t) \exp(\textstyle\sum_{r=2}^{R} \gamma_{kr} w_{ir}(t) + \boldsymbol{\beta}_k^T \boldsymbol{v}_i), k = 1, 2, \ldots, K, \tag{4.5}$$

In principle one may choose any subset of $R-1$ linear predictors $w_{ir}(t)$ to be shared in (4.4) leading to $2^{R-1} - 1$ possible parameterizations of a survival submodel for a given failure cause k. Some of the linear predictors can be reduced to only random effect shared. That results in total $3^{R-1} - 1$ possible submodels for a given k. The number of possible parameterizations increases if one would consider also parameterizations, for which a particular function of a linear predictor is shared. Within any of these formulations the same random effects account for both the association between the longitudinal and

event outcomes, and the correlation between the repeated measurements in the longitudinal process .

4.3 Estimation

4.3.1 Full likelihood framework: Bayesian approach

In the standard joint modeling framework the estimation is typically based on maximum likelihood or Bayesian methods ([9, 16, 25, 33, 89, 90]). This proceeds under the following set of conditional independence assumptions:

$$p(T_i, \Delta_i, \boldsymbol{Y}_i \mid \boldsymbol{b}_i; \boldsymbol{\theta}) = p(T_i, \Delta_i \mid \boldsymbol{b}_i; \boldsymbol{\theta}_t) p(\boldsymbol{Y}_i \mid \boldsymbol{b}_i; \boldsymbol{\theta}_y) \tag{4.6}$$

$$p(\boldsymbol{Y}_i \mid \boldsymbol{b}_i; \boldsymbol{\theta}_y) = \prod_{j=1}^{m_i} p(Y_i(t_{ij}) \mid \boldsymbol{b}_i; \boldsymbol{\theta}_y), \tag{4.7}$$

where $\boldsymbol{\theta}^T = (\boldsymbol{\theta}_y^T, \boldsymbol{\theta}_t^T, \boldsymbol{\theta}_b^T)$ is a vector of parameters from the longitudinal and survival submodels and the vector of the random effects, respectively. In particular, we assume that given the random effects the longitudinal process is independent from the event times, and moreover, the longitudinal measurements are independent from each other.

Here we will proceed under the Bayesian paradigm to estimate the model parameters. The likelihood contribution for the ith subject conditionally on the random terms is given by:

$$p(\boldsymbol{Y_i}, T_i, \Delta_i \mid \boldsymbol{b}_i; \boldsymbol{\theta}) = p(\boldsymbol{Y_i} \mid \boldsymbol{b}_i; \boldsymbol{\theta}_y) p(T_i, \Delta_i \mid \boldsymbol{b}_i; \boldsymbol{\theta}_t)$$

$$= \prod_{k=1}^{K} \left[\lambda_{0k}(T_i) \exp\{\boldsymbol{\gamma}_{\boldsymbol{k}}^T \boldsymbol{f}_i(\boldsymbol{b}_i, T_i)\}\right]^{I(\Delta_i = k)} \exp\left[-\sum_{k=1}^{K} \int_0^{T_i} \lambda_{0k}(t) \exp\{\boldsymbol{\gamma}_{\boldsymbol{k}}^T \boldsymbol{f}_i(\boldsymbol{b}_i, t)\} dt\right]$$

$$\times \left[\frac{1 + \sum_{r=2}^{R} I(y_{ir} = r) \exp\{w_{ir}(t_{ij})\}}{1 + \sum_{r=2}^{R} \exp\{w_{ir}(t_{ij})\}}\right]^{m_i}. \tag{4.8}$$

Under the conditional independence assumptions (4.6) and (4.7) the posterior distribution of the parameters and the latent terms, conditional on the observed data, are derived as:

$$p(\boldsymbol{\theta}, \boldsymbol{b}_i \mid T_i; \Delta_i; \boldsymbol{Y}_i) \propto \prod_{i=1}^{N} \left\{\prod_{j=1}^{m_i} p(Y_i(t_{ij}) \mid \boldsymbol{b}_i; \boldsymbol{\theta}_y)\right\} p(T_i, \Delta_i \mid \boldsymbol{b}_i; \boldsymbol{\theta}_t) \times p(\boldsymbol{b}_i; \boldsymbol{\theta}_b) p(\boldsymbol{\theta}_y, \boldsymbol{\theta}_t, \boldsymbol{\theta}_b), \tag{4.9}$$

where $p(\cdot)$ denotes the appropriate probability density function.

To model each cause-specific risk from (B.2.1) a piecewise constant hazard model was chosen with a step function baseline hazard with $L = 5$ quantiles and change points $t_1, t_2, \ldots, t_L$ assuring the same number of events of any cause between those time points. Let t_0 denote the start of the follow up, t_5 the maximum censoring time and $\lambda_{0k}(t)$ baseline hazard function for the time interval $[t_\iota, t_{\iota+1}], \iota = 0, 1, \ldots, L$. Then for different intervals for a given cause k we specify a separate prior hazard mean $\lambda^*_{0k\iota}(t)$ and $\lambda_{0k\iota}(t) \sim \text{Gamma}(c\lambda^*_{0k\iota}(t), c)$, where $\lambda^*_{0k\iota}(t)$ is a prior mean hazard for cause k with c being a scaling parameter. For the priors of the model parameters we make standard assumptions following [25]. In particular, for the regression

coefficients $\boldsymbol{\alpha}^T = (\boldsymbol{\alpha}_1^T, \boldsymbol{\alpha}_2^T, \ldots, \boldsymbol{\alpha}_R^T)$ of the longitudinal submodel and for the coefficients $\boldsymbol{\gamma}^T = (\boldsymbol{\gamma}_1^T, \boldsymbol{\gamma}_2^T, \ldots, \boldsymbol{\gamma}_K^T)$ of survival submodel we used multivariate normal priors. For variance-covariance matrices we assumed an inverse Wishart distribution and for the variance-covariance parameters we took an inverse-gamma prior. For all parameters vague priors have been chosen. The results were not sensitive with respect to the choice of the hyperparameters as long as the priors were sufficiently diffuse. When the longitudinal and survival submodels share only the random effects $\boldsymbol{b}_i$ then the integral over time in the likelihood expression(4.8) has of a closed-form solution. When the two submodels (B.2.1) and (B.2.1) share also time-dependent terms, the integral:

$$\int_0^{T_i} \lambda_{0k}(t) \exp\{\boldsymbol{\gamma}_k^T \boldsymbol{f}_i(\boldsymbol{b}_i, t)\} dt. \tag{4.10}$$

is approximated numerically using the Gauss-Kronrod rule with Q=15 quadrature points [91]. Therefore we obtain:

$$\int_0^{T_i} \lambda_{0k}(t) \exp\{\boldsymbol{\gamma}_k^T \boldsymbol{f}_i(\boldsymbol{b}_i, t)\} dt \approx \sum_{q=1}^{Q} \omega_q \lambda_{0k}(\xi_q) \exp\{\boldsymbol{\gamma}_k^T \boldsymbol{f}_i(\boldsymbol{b}_i, \xi_q)\}, \tag{4.11}$$

where ξ_q and ω_q, $q = 1, \ldots, Q$ denote the Gauss-Kronrod quadrature points and weights, respectively.

4.4 Dynamic Prediction of Longitudinal Trajectories and CIF

Based on the fitted model, and for a new subject l for the same population for whom we have a set of longitudinal measurements $Y_l(t) = \{Y_l(s); 0 \leq s \leq t\}$ available, our aim is to obtain predictions of conditional cumulative incidence probabilities CIF_{kl} for each cause of failure k and predictions of her future longitudinal responses . In the context of our motivating example, we are interested in predicting the future status of the patient in the waiting list, i.e., T, NT, U and HU, and the chance of the patient leaving the waiting list due to TT, D or R. We first focus on predictions of cumulative incidence probabilities. Let $\boldsymbol{\theta}$ denote the vector of parameters from the joint model and S_n the sample of size n on which the joint model was fitted, $S_n = \{T_i, \boldsymbol{Y}_i, \Delta_i; i = 1, 2, \ldots, N\}$. For a specific cause k we are interested in the conditional probability of experiencing event k before time $u > t$ given that the subject has not experienced any event up to t:

$$CIF_{kl}(u \mid t) = \Pr(T^*_{lk} < u \mid \mathcal{T}^*_l(t), Y_l(t)), \mathcal{T}^*_l(t) = \{T^*_{l1} > t, \ldots, T^*_{lk} > t\}. \tag{4.12}$$

Note that $CIF_{kl}(u \mid t)$ has a dynamic nature because when new information is recorded for a patient l at time $t' > t$, we can update these predictions and obtain $CIF_{kl}(u \mid t')$ for $u > t'$. Following [86] and [41], we derive the

posterior predictive distribution of $CIF_{kl}(u \mid t)$ as:

$$CIF_{kl}(u \mid t) = \int \Pr\left(T_{lk}^{*} < u \mid \mathcal{T}_{l}^{*}(t), Y_{l}(t); \theta\right) p(\boldsymbol{\theta} \mid S_{n}) d\boldsymbol{\theta}. \tag{4.13}$$

The first term of the integrand in (4.13) can be written as:

$$\begin{aligned} &\Pr\left(T_{lk}^{*} < u \mid \mathcal{T}_{\iota}^{*}(t), Y_{l}(t); \boldsymbol{\theta}\right) \\ &= \int \Pr\left(T_{lk}^{*} < u \mid \mathcal{T}_{l}^{*}(t), \boldsymbol{b}_{l}; \boldsymbol{\theta}\right) p\left(\boldsymbol{b}_{l} \mid \mathcal{T}_{l}^{*}(t), Y_{l}(t), \boldsymbol{\theta}\right) d\boldsymbol{b}_{l}. \end{aligned} \tag{4.14}$$

Therefore when combining (4.13) and (4.14) a Monte Carlo approach can be used to compute $CIF_{kl}(u \mid t)$ for each patient and $CIF_{kl}(u \mid t')$ can be updated for every time point $t' > t$. To derive an estimate of CIF we use the following Monte Carlo sampling scheme:

Step 1: sample $\boldsymbol{\theta}^{(\ell)}$ from the posterior $p\left(\boldsymbol{\theta} \mid S_{n}\right)$
Step 2: sample $\boldsymbol{b}_{l}^{(\ell)}$ from the posterior $p\left(\boldsymbol{b}_{l} \mid \mathcal{T}_{l}^{*}(t), Y_{l}(t); \boldsymbol{\theta}_{y}^{(\ell)}\right)$
Step 3: compute $CIF_{kl}^{(\ell)}(u \mid t, \boldsymbol{b}_{l}^{(\ell)}; \boldsymbol{\theta}^{(\ell)})$
Step 4: repeat Steps 1-3, $\ell = 1, \ldots, Ł$.

As a final prediction of the $CIF_{kl}(u \mid t, \boldsymbol{b}_{l}; \boldsymbol{\theta})$ we use the median over the Ł Monte Carlo samples together with the Monte Carlo percentiles as confidence intervals.

Step 1 takes into account the variability in Bayesian estimates of $\boldsymbol{\theta}$ obtained from the fitted joint model and $\boldsymbol{Y}_{l}$ denotes the whole observed longi-

tudinal profile for individual l. In Step 2 we utilized a Metropolis-Hastings algorithm to sample from $\{\boldsymbol{b}_l \mid \mathcal{T}_l^*(t), Y_l(t); \boldsymbol{\theta}^{(\ell)}\}$ with proposals from a multivariate t distribution and df=4 centered at the empirical Bayes (EB) estimates $\hat{\boldsymbol{b}}_l$. These EB estimates were obtained by maximizing the expression:

$$\hat{\boldsymbol{b}}_l = \arg\max_{\boldsymbol{b}} \left[\log \Pr\left(\mathcal{T}_\iota^*(t) > t \mid \hat{\boldsymbol{\theta}}_y\right) + \log \Pr\left(Y_l(t) \mid \boldsymbol{\theta}_y\right) + \log p(\boldsymbol{b} \mid \hat{\boldsymbol{\theta}}_y)\right], \tag{4.15}$$

where $\hat{\boldsymbol{\theta}}_y$ are the parameters estimates obtained from the joint model for the longitudinal part.

The low number of degrees of freedom was chosen to ensure that the proposal density has heavy tails to provide sufficient coverage of the target density $\{\boldsymbol{b}_l \mid \boldsymbol{Y}_l, \boldsymbol{\theta}\}$. The variance-covariance matrix from the joint model was additionally scaled by some tuning parameter allowing to control the acceptance rate through the range of the proposed distribution. Usually a small number of iterations (100-500) is sufficient for the purpose of calibration. More details about the Metropolis-Hastings acceptance-rejection procedure can be found in the supplementary material (part C). In order to compute $CIF_{kl}^{(\ell)}(u \mid t, \boldsymbol{b}_l^{(\ell)}; \boldsymbol{\theta}^{(\ell)})$ in Step 3 using (4.13) and (4.14) we need to calculate:

$$\begin{gathered}
\Pr\left(T_{lk}^* < u \mid \mathcal{T}_l^*(t), Y_l(t), \boldsymbol{b}_l^{(\ell)}; \boldsymbol{\theta}^{(\ell)}\right) = \mathcal{A}_{kl}(u \mid t)/S(t), \\
\mathcal{A}_{kl}(u \mid t) = \int_t^u \lambda_{0k}(s) \exp\{\boldsymbol{\gamma}_k^T \boldsymbol{f}_l(\boldsymbol{b}_l^{(\ell)}, s)\} \exp\left[-\textstyle\sum_{\kappa=1}^K \int_0^t \lambda_{0\kappa}(s) \exp\{\boldsymbol{\gamma}_\kappa^T \boldsymbol{f}_l(\boldsymbol{b}_l^{(\ell)}, s)\} ds\right], \\
S_l(t) = \exp\left[-\textstyle\sum_{k=1}^K \int_0^t \lambda_{0k}(s) \exp\{\boldsymbol{\gamma}_k^T \boldsymbol{f}_l(\boldsymbol{b}_l, s)\} ds\right].
\end{gathered} \tag{4.16}$$

In the simple case with only random effects shared between the longitudinal and risk submodels , i.e. $\boldsymbol{f}_i(\boldsymbol{b}_i, t) = \boldsymbol{b}_i$ the integral in (4.17) is of a closed form. When also time-dependent terms are shared, the integral has to be computed numerically and we use again a 15-point Gauss-Kronrod rule. In an analogous manner dynamic predictions for a future longitudinal response were calculated applying a similar sampling procedure:

Step 1: sample $\boldsymbol{\theta}_y^{(\ell)}$ from the posterior $p\left(\boldsymbol{\theta}_y \mid S_n\right)$
Step 2: sample $\boldsymbol{b}_l^{(\ell)}$ from the posterior $p\left(\boldsymbol{b}_l \mid \mathcal{T}_l^*(t), Y_l(t); \boldsymbol{\theta}_y^{(\ell)}\right)$
Step 3: compute $p(\boldsymbol{Y}_l \mid \boldsymbol{b}_l^{(\ell)}; \boldsymbol{\theta}_y^{(\ell)})$
Step 4: Repeat Steps 1-3, $\ell = 1, \ldots,$ Ł

Here in Step 3 in order to obtain a prediction for $\Pr(\boldsymbol{Y}_l \mid \boldsymbol{b}_l^{(\ell)}; \boldsymbol{\theta}_y^{(\ell)})$ we need to calculate:

$$\Pr\left(Y_\iota(t) = r \mid \boldsymbol{\theta}_y^{(\ell)}\right) = \exp(w_{lr}(t)) \Big/ \Big\{1 + \sum_{h=2}^{R} \exp(w_{ih}(t))\Big\}. \tag{4.17}$$

4.5 Analysis of the Heart Data

We return to the analysis of the heart data introduced in Section 4.1. This data were taken from an international data base of the Eurotransplant Heart recipient waiting list. A total of 2921 recipients entered the waiting list from 01.01.2006 until 31.12.2008. Recipients observation was censored at 31.03.2010. Baseline information was also collected for the patients when entering the list, namely: age, country of origin (7 countries) and blood

group. We will first consider the simple version of the longitudinal submodel (B.2.1) with time as fixed effect and random intercepts only. Therefore for each transient state Transplantable (T), Non-Transplantable (NT), Urgent (U) and High Urgent (HU) we have:

$$w_{ir}(t) = \alpha_{0r} + \alpha_{1r}t + b_{ir}, \qquad r = 2, 3, 4, \tag{4.18}$$

where for the baseline category (r=1) we choose state NT.

For the survival process we have considered first the simple parametrization of the joint model, namely when only random effects are shared (R-E):

$$\text{(R-E)}: \quad \lambda_{ik}(t) = \lambda_{0k}(t) \exp(\boldsymbol{\gamma}_k^T \boldsymbol{b}_i + \boldsymbol{\beta}_k^T \boldsymbol{v}_i), \quad k = 1, 2, 3, \tag{4.19}$$

where $\boldsymbol{b}_i^T = (b_{i2}, b_{i3}, b_{i4})$.

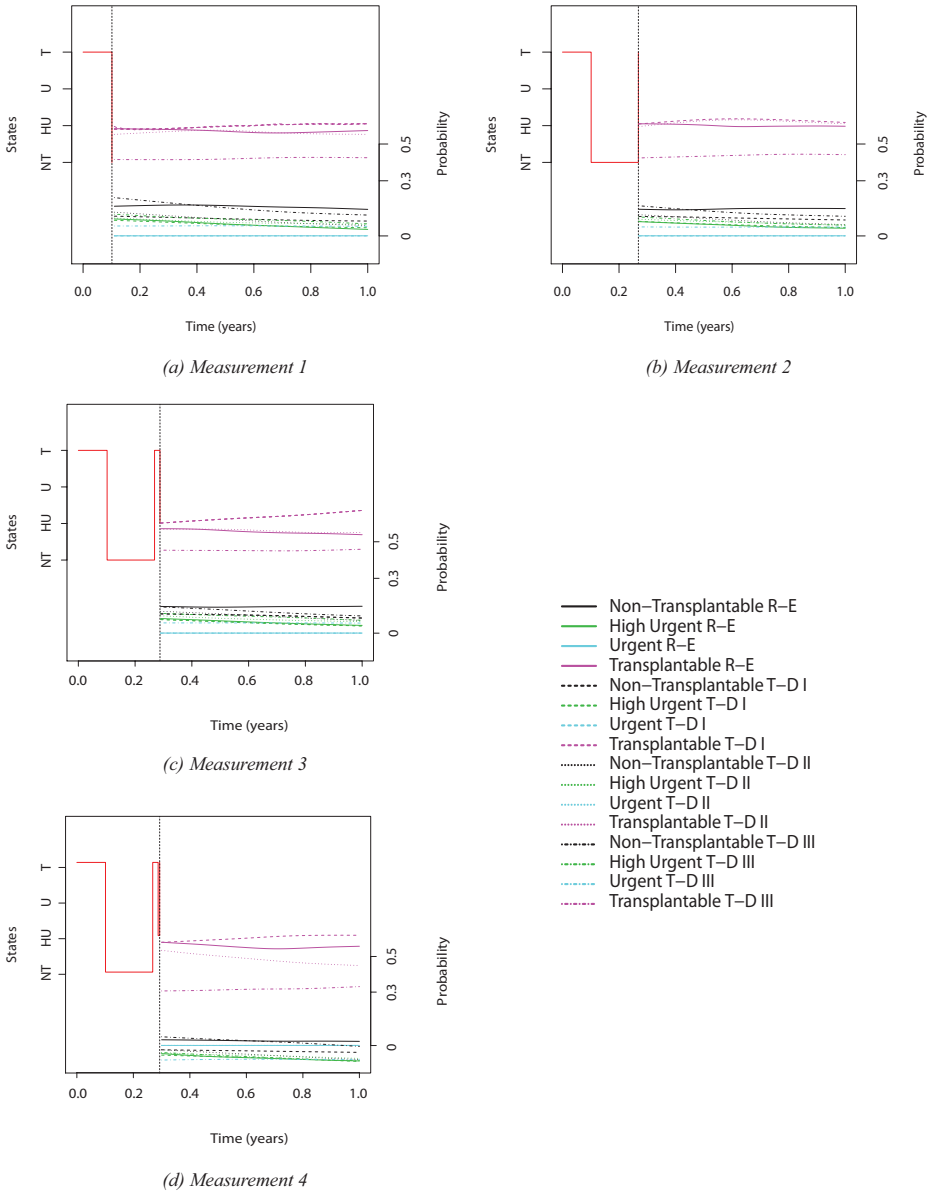
Then we fitted more complicated models when also the time dependent terms are shared (T-D). The choice of the shared terms in those models was medically motivated. For the illustration purposes we present here only three versions of (T-D) parameterizations with the best fit assessed based on Deviance Information Criterion (DIC). Therefore for each of the three competing events: Death (D), Transplantation (TT) and Removal (R) we have the following model for the survival part:

$$\text{(T-D)}: \quad \lambda_{ik}(t) = \lambda_{0k}(t) \exp(\boldsymbol{\gamma}_k^T \boldsymbol{f}_{ik}(t) + \boldsymbol{\beta}_k^T \boldsymbol{v}_i), \quad k = 1, 2, 3. \tag{4.20}$$

where $\boldsymbol{b}_i^T = (b_{i2}, b_{i3}, b_{i4})$.

For the first parametrization (T-D I) only the cause-specific submodel for Death event shares the time-dependent term, which is the log probability of being in state NT. Therefore we have: $\boldsymbol{f}_{i1}(t) = (w_{i2}(t), b_{i3}, b_{i4})$ and $\boldsymbol{f}_{ik}(t) = (b_{i2}, b_{i3}, b_{i4})$ for $k = 2, 3$. For the second parametrization (T-D II) additionally the cause-specific submodel for Removal shares the time-dependent term which is the log probability of being in state HU. Therefore, we have $\boldsymbol{f}_{i1}(t) = (w_{i2}(t), b_{i3}, b_{i4})$, $\boldsymbol{f}_{i2}(t) = (b_{i1}, w_{i3}(t), b_{i4})$ and $\boldsymbol{f}_{i4}(t) = (b_{i2}, b_{i3}, b_{i4})$. Finally for the last parametrization (T-D III) the cause-specific submodel for Transplantation shares the time-dependent term which is the log probability of being in state U, i.e: $\boldsymbol{f}_{i3}(t) = (b_{i2}, b_{i3}, w_{i4}(t))$ and $\boldsymbol{f}_{ik}(t) = (b_{i2}(t), b_{i3}, b_{i4})$ for $k = 1, 2$.

The vector $\boldsymbol{v}_i$ of baseline covariates consists of: age, blood group and IC binary covariate indication whether patients comes from the country with informed consent required (IC=1) or with the presumed consent for donation. The analysis has been performed using the R Statistical Software and the joint Bayesian model was fitted using OpenBUGS . Separate code was written by the first author for the dynamic predictions. The specific specification of the priors we used in this analysis was as follows. For the $p \times p$ variance-covariance matrices of multivariate normal priors we used inverse Wishart distribution with p degrees of freedom. For the piecewise constant baseline risk parameters in the cause-specific survival submodels a Gamma(10^{-3},10^{-3}) prior was used. We chose the scaling parameter c in the gamma prior to be equal 0.001 and a prior mean $\lambda^*_{0k\iota}(t) = 0.1$. We did not observe any substantial difference for the different values of parameter c as long as c was small enough to keep the prior noninformative. Too small values of the scaling parameter c are not recommended as they can lead to the



(a) Measurement 1

(b) Measurement 2

(c) Measurement 3

(d) Measurement 4

Figure 4.2: *Dynamic prediction of probabilities for the three longitudinal urgency categories from model with only random effects shared (solid) and different parameterizations with time-dependent terms (dashed: T-D I, dotted: T-D II, dash-dotted: T-D III) for an arbitrary individual with blood group B and non-IC.*

computation problems. In order to assess convergence for the joint Bayesian model standard MCMC diagnostic plots were used. The burn-in size was set to 10,000 iterations, which was chosen based on the visual inspection of the trace plots, and confirmed by the Raftery and Lewis diagnostics. The same number of iterations were used for constructing the summary statistics. Based on the autocorrelation plots we have chosen every 30th iteration. Therefore in total to obtain 10,000 iterations for the final inference 30,0000 iterations were required after the burn-in part. Additionally we run a second parallel chain and used Gelman and Rubin diagnostic plots to assess the convergence ([92]).

Based on the fitted models we have constructed dynamic predictions for the parameterizations (B.14) and (4.20) according to the Monte Carlo sampling procedure described in Section 4.5. We generated 1000 Monte Carlo samples for each update of the prediction. The estimates of the joint model obtained for some of the subsets considered for the Heart Data for the parametrization (B.14) and the three versions of parametrisation (4.20) are given in the Supplementary material C. The dynamic prediction plots allowed clinicians to conclude that the chance of transplantation from HU was the highest at the early times on the waiting list and decreased with time, especially as patient experienced long stable period in T before HU. Figures 4.2 and 4.4 illustrate the comparison of the dynamic prediction of longitudinal response and CIF for different parameterizations of the joint model. The prediction is constructed for an arbitrary individual with blood group B and non-informed consent (IC=0). As can be observed the CIF prediction is strongly affected by the parametrization of the joint model. This discrepancy does not diminish as more longitudinal measurements are

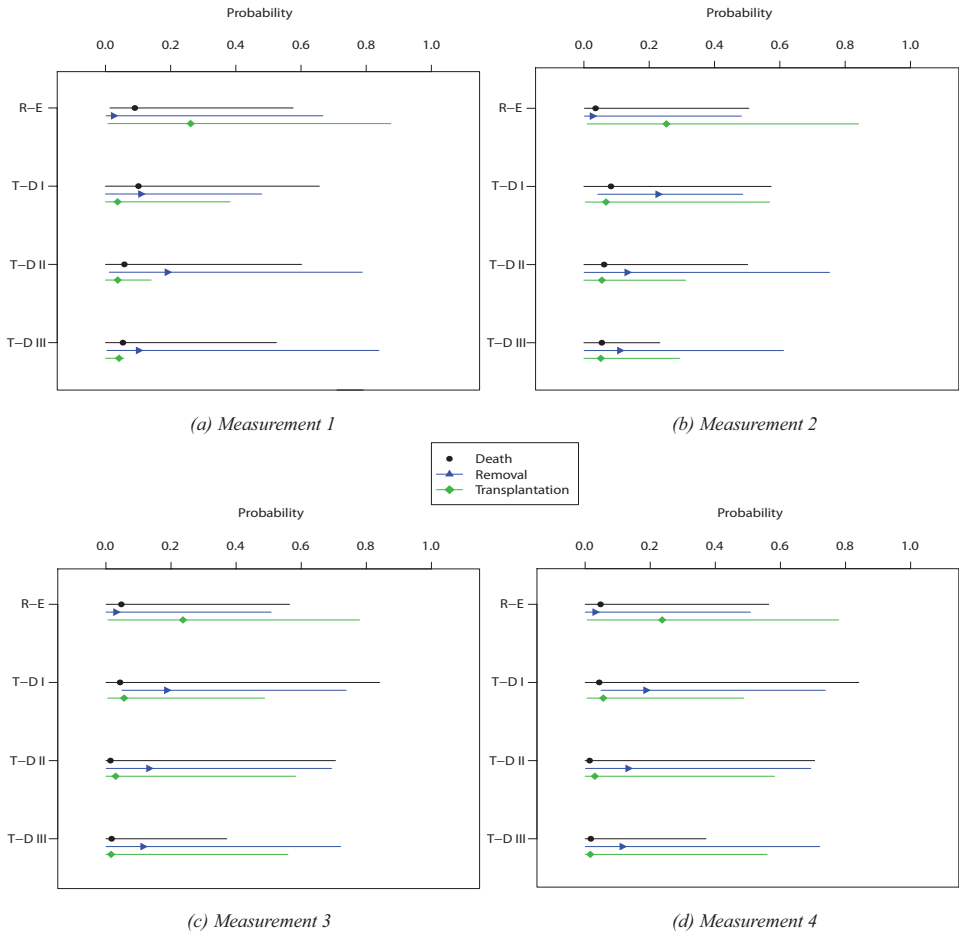
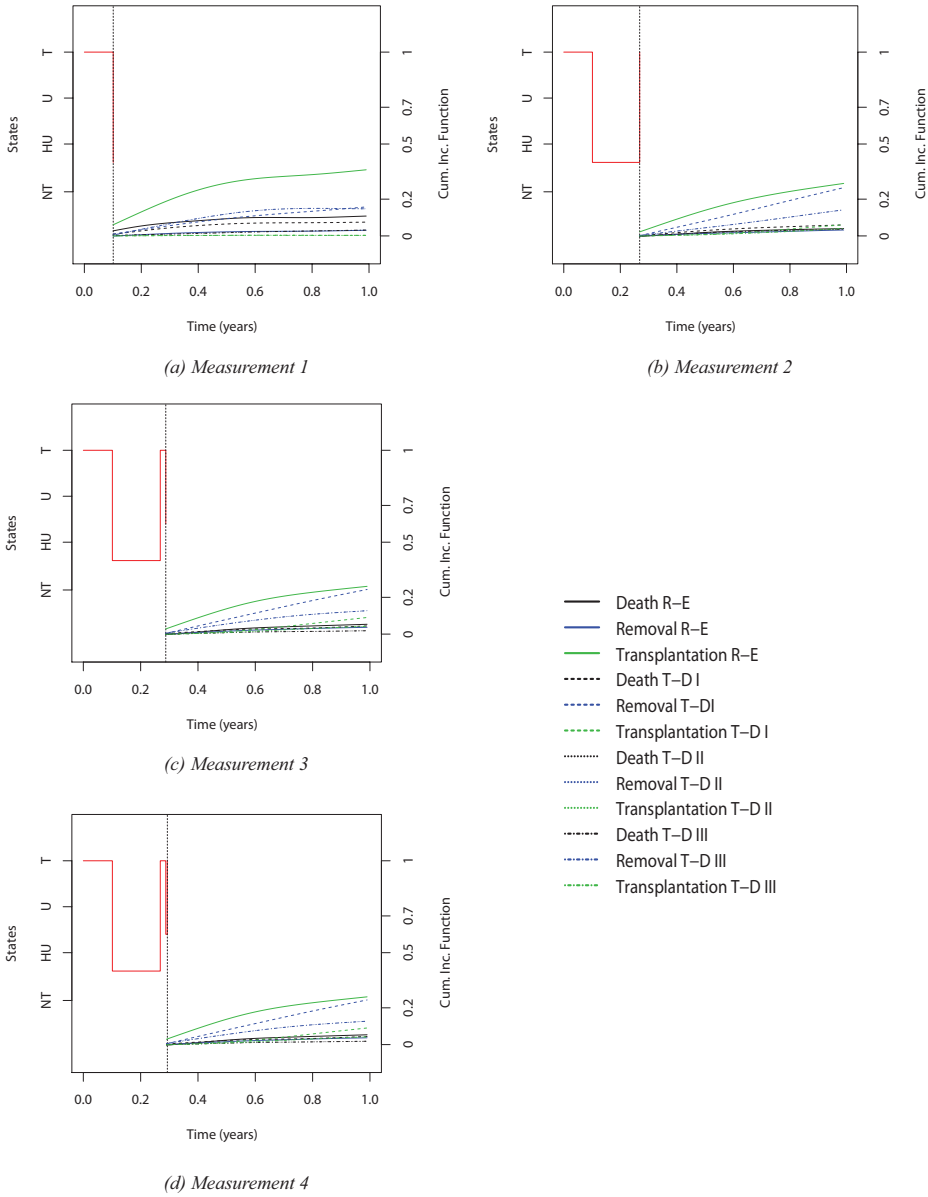


Figure 4.3: *Dynamic prediction of cumulative incidence functions for half year ahead from model with only random effects shared (R-E) and different parameterizations with time-dependent terms (TD I, TD II, TD III) for an arbitrary individual with blood group B and non-IC.*



(a) Measurement 1

(b) Measurement 2

(c) Measurement 3

(d) Measurement 4

Figure 4.4: *Dynamic prediction of cumulative incidence functions from model with only random effects shared (solid) and different parameterizations with time-dependent terms (dashed: T-D I, dotted: T-D II, dash-dotted: T-D III) for an arbitrary individual with blood group B and non-IC.*

available. It is also depicted on Figure 4.3 which presents the predicted CIF depending on the prediction time and different parameterizations. DIC criterion implied the choice of T-D I model as the best one. For most of the parameterizations we did not observe substantial impact on the dynamic predictions for the longitudinal response (Figure 4.2).

4.6 Simulations

4.6.1 Design

We performed a simulation study to evaluate the misspecification of the association structure between the longitudinal and survival process considering different association strength in presence of that misspecification. A number of simulations have been conducted to investigate the impact of these two aspects on the accuracy of (dynamic) predictions, especially for the CIF. In particular we performed a limited simulation study of two different scenarios with 200 samples per each. As comparing to the real data set we simplified the simulated model to three categories for the categorical longitudinal response and only two competing events in the survival part. In both longitudinal and survival part we have simulated a group effect of a binary factor.

For the first scenario (I) we simulated data from the model with only random effects shared corresponding to the parametrization (R-E), for the heart data set. In the second scenario (II) we simulated data from the joint model with time-dependent terms shared with one different logit shared by each survival submodel that corresponds to the parametrization (T-D II) used for the heart data.

For the longitudinal part the data were simulated for 100 patients from model (B.17) with $\alpha_{0r} = -1$, $\alpha_{1r} = 0.5$ and $\alpha_{2r} = -2$, $r = 2, 3$. The sample size corresponds to the smallest subset from the real data for which the model was fitted in order to study the interactions. The variance-covariance matrix $\boldsymbol{D}$ of the random effects was chosen to be non-diagonal with $\boldsymbol{D_{11}} = 0.11, \boldsymbol{D_{12}} = -0.08, \boldsymbol{D_{33}} = 0.27$ corresponding to the variance-covariance matrix obtained for the state HU and U from the real data. We simulated a maximum follow-up time of 4 years. The residual variance σ^2 for the longitudinal response on the logit scale was chosen to be 0.25, corresponding to the results obtained for the real data set.

Survival times were simulated from bivariate time-to-event distributions using the Clayton copula $C(\tau_1, \tau_2) = \left(\tau_1^{-\phi} + \tau_2^{-\phi} - 1\right)^{-1/\phi}, 0 \leq \tau_1, \tau_2 \leq 1, \phi > 0$ with $\phi = 15$. The aim was to reflect the correlation between the two survival responses even if in the competing risk setting we cannot asses this correlation since only one survival response is observed for each subject. The Clayton copula exhibits strong left tail dependence and weak right tail dependence. Therefore it allows early survival probabilities to be strongly correlated reflecting the higher correlation after the long time on the waiting list in the transplantation setting. We used baseline Weibull hazards with shape parameters equal 2.02 and 3.07 for the first and second survival response and scale parameters equal 1. For each marginal survival we simulated the group effect $\beta_1 = 0.03$ and $\beta_2 = 0.03$. For Scenario A the association parameters were chosen to be $\boldsymbol{\gamma}_1 = (0.8, 0.5)$ for the first survival submodel and $\boldsymbol{\gamma}_2 = (0.5, 0.8)$ for the second cause-specific survival submodel. Additionally for scenario B we simulated data with larger association parameters $\boldsymbol{\gamma}_1 = (1.8, 1.5)$ and $\boldsymbol{\gamma}_2 = (1.5, 1.8)$. The censoring mechanism was simulated

independently using an exponential distribution $Exp(\lambda_C)$. Parameter λ_C was changed in order to control proportion of censored observations. We kept 20% of censoring as in original data set and simulated 70% and 10% of events which corresponds to the smallest percentage of events in the real data set observed for the incidence of removal.

The simulation settings are summarized below:

Scenario I	Scenario II
A : $\boldsymbol{\gamma}_1 = (0.8, 0.5), \boldsymbol{\gamma}_2 = (0.5, 0.8)$,	A : $\boldsymbol{\gamma}_1 = (0.8, 0.5), \boldsymbol{\gamma}_2 = (0.5, 0.8)$,
B : $\boldsymbol{\gamma}_1 = (1.8, 1.5), \boldsymbol{\gamma}_2 = (1.5, 1.8)$,	B : $\boldsymbol{\gamma}_1 = (1.8, 1.5), \boldsymbol{\gamma}_2 = (1.5, 1.8)$,
$w_{ir}(t) = \alpha_{0r} + \alpha_{1r}u + \alpha_{2r}Group + b_{ir}$,	$w_{ir}(t) = \alpha_{0r} + \alpha_{1r}u + \alpha_{2r}Group + b_{ir}$,
$\lambda_{ik}(t) = \lambda_{0k}(t)\exp(\boldsymbol{\gamma}_k^T \boldsymbol{b}_i + \beta_k Group)$,	$\lambda_{ik}(t) = \lambda_{0k}(t)\exp(\boldsymbol{\gamma}_k^T \boldsymbol{f}_i(t) + \beta_k Group)$,
$k = 1, 2,\ r = 1, 2$,	$\boldsymbol{f}_i(t) = (w_{i2}(t), w_{i3}(t)), k = 1, 2, r = 1, 2$.

(4.21)

For each simulated data set we have fitted the three versions of the joint model, namely (R-E), (T-D II) and additionally a model with one different logit shared by each cause-specific survival submodel that corresponds to the (T-D I) model fitted for the heart data:

$$\begin{aligned}
\text{(R-E):} \quad & \lambda_{ik}(t) = \lambda_{0k}(t)\exp(\boldsymbol{\gamma}_k^T \boldsymbol{b}_i + \beta_k Group), \quad k = 1, 2, \\
\text{(T-D I):} \quad & \lambda_{ik}(t) = \lambda_{0k}(t)\exp(\boldsymbol{\gamma}_k^T \boldsymbol{f}_i(t) + \beta_k Group), \\
& \boldsymbol{f}_{i1}(t) = (w_{i2}(t), b_{i3}), \quad \boldsymbol{f}_{i2}(t) = (b_{i2}, w_{i3}(t))
\end{aligned} \tag{4.22}$$

For both scenarios we compared the prediction of cumulative incidence functions from different parameterizations with the prediction based on the gold standard model with the true parametrization and true values for the random effects and the parameters. In order to quantify the differences between the predicted and true CIFs we have considered prediction for 10 arbitrary chosen individuals with minimum 4 measurements of the longitudinal response that were randomly removed from each simulated data set before fitting any model. For each of these individuals we calculated the root mean squared error between the assumed model under each scenario and the gold standard predictions using 10 equally spaced time points in the interval from the next 6 months till 12 months ahead. We investigated how the distance from the true CIF changes as the prediction was updated. Since the simulated time points were fixed each individual could have the updated prediction in the same time points as long as he did not experience any event. Within every scenario we plot separate box plots for a given time point, in which the prediction for the chosen individuals was updated, averaged over all simulated data sets, for the two competing events separately.

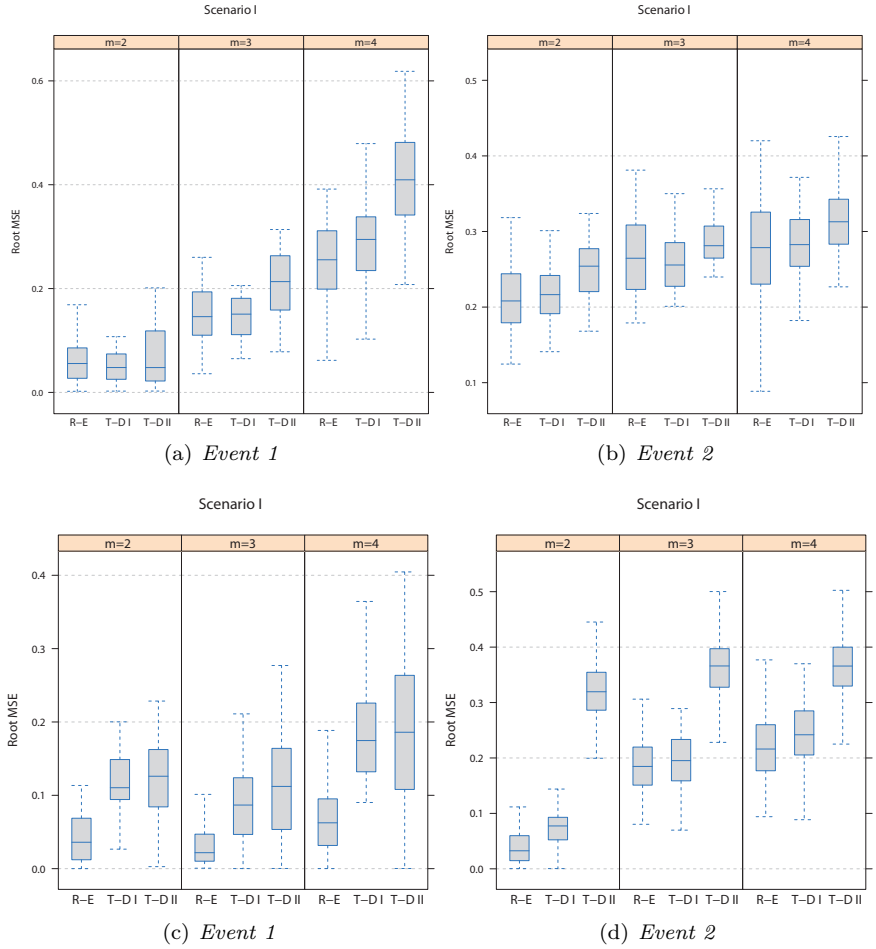
All the prior settings, size of burn-in, number of M-H iterations etc. were the same as for the real data analysis. Computations have been performed in cluster with 20 nodes with AMD Quad-Core Opteron 835X, 4 x 2GHz

and 16GB RAM per node.

4.6.2 Results

In Figures 4.5 and 4.6 the results for 200 simulated data sets of scenario IA, IB, IIA and IIB are presented. They depict the box plots for RMSE between prediction from the assumed model under each scenario and the gold standard predictions for the 10 chosen individuals, averaged over the simulated data sets. The RMSE are updated for m=2, 3 and 4 longitudinal measurements available.

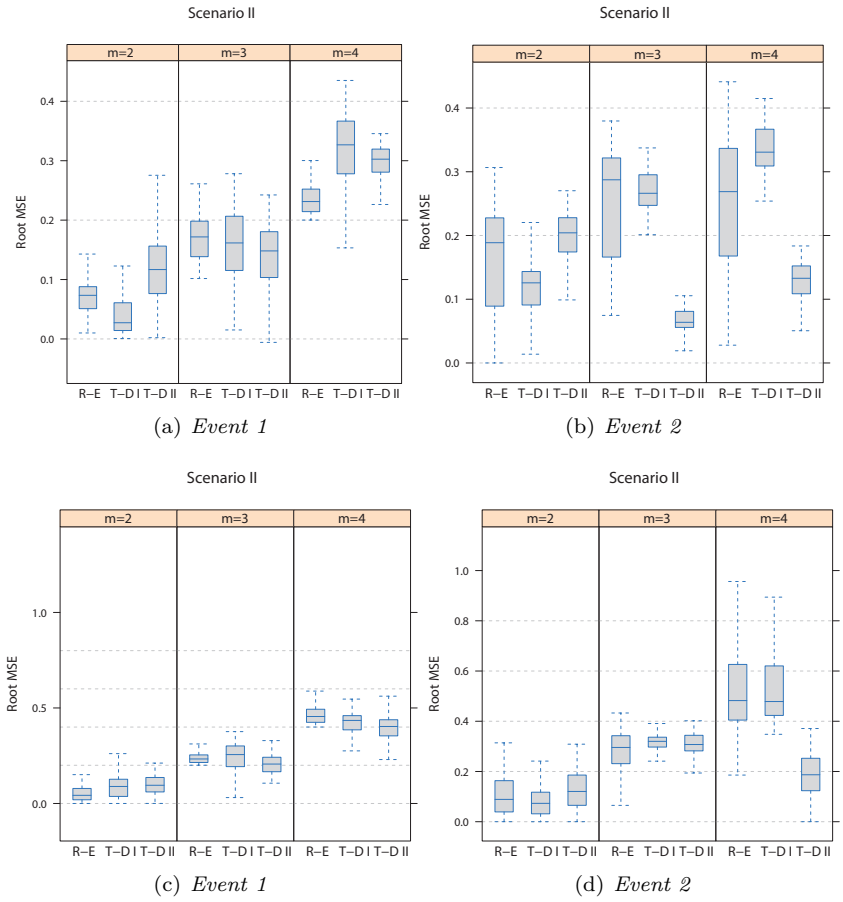
It can be observed that RMSE for the CIF predicted from the models with misspecified parametrizations increases as more measurements are recorded. For the CIF with true parametrization RMSE can also increase if there is not enough number of events to estimate the survival submodel correctly (Event 1 in Scenario A). RMSE for Event I from model T-D I and T-D II are similar since the survival submodel for event 1 has the same parametrization in both models. Analogically for Event 2 RMSE for R-E model are close to RMSE from T-D I model. For Scenarios IB and IIB with larger association parameters the discrepancies from the true model are more severe for the misspecified survival submodels from the Scenario II. When the true model shares only the random effects the increasing association parameters between the longitudinal and survival part will not affect strongly the prediction from misspecified parameterizations (T-D I) and (T-D II) as the prediction is more affected by the increasing time-dependent term than by the updated random effect. Contrary when the true model is simulated from parametrization (T-D II) the CIF from misspecified models is much more affected for larger parameters $\boldsymbol{\gamma}$ as they are associated with the omit-



(a) *Event 1* (b) *Event 2*

(c) *Event 1* (d) *Event 2*

Figure 4.5: *RMSE of the distance between the prediction from the gold standard model and the three fitted joint models (R-E, T-D I, T-D II) for the simulated scenario IA (a,b) and IB (c,d) .*



(a) *Event 1* (b) *Event 2*

(c) *Event 1* (d) *Event 2*

Figure 4.6: *RMSE of the distance between the prediction from the gold standard model and the three fitted joint models (R-E, T-D I, T-D II) for the simulated scenarios IIA (a,b) and IIB (c,d).*

ting time-dependent terms. The overestimation of the random effects does not influence seriously the dynamic prediction in any scenario since the simulated variance is small, especially for the random effect from the first logit of the longitudinal submodel. The distance from the true model was larger for more distant prediction (results not shown).

4.7 Discussion

We have presented a method to calculate individual prediction of survival and longitudinal responses that can be updated as more measurements of the longitudinal response is available for a particular patient. Predictions are based on the fitted joint model for the longitudinal and survival responses and requires specification of two separate submodels for the longitudinal and time to event data part that share only random effects and also possibly time-dependent terms.

The presented method allows to handle the variability of the subject-specific longitudinal profiles when modeling survival using the whole history of the observed longitudinal response. This has a great advantage as compared to other statistical framework such as non-Markov models where the history of the process is if an issue and no standard approaches are available. The method generalizes for any type of longitudinal responses and non- longitudinal responses such as continuous, ordinal, binary. Therefore dynamic predictions are particularly useful in all observational studies where patients are observed along time and different kind of measurements reflecting their current health status are collected. The presented approach allows to handle more than one survival response, such as in competing risks setting. This

is a typical setting in transplantation studies where patients waiting for an organ can be delisted due to the death or transplantation. Using the available *shiny* package an interface for the R code can be written resulting in a user-friendly tool for producing dynamic predictions. That tool can be used for everyday general clinical practice.

The construction of dynamic predictions requires a single fit of the joint model with a bit of additional code for Metropolis-Hastings sampling. Fitting joint model can be performed using Bayesian methods and standard software. That could be time-consuming, especially for many submodels that share time-dependent terms. For the model with only random effects shared an alternative two-stage methods could be considered.

As it was demonstrated, the chosen parametrization for joint model influences mainly the prediction for the survival part and much less the longitudinal response. Since we have many repeated measures per individual there is a lot more information about the longitudinal process than for the survival process in that setting. Even with non-random dropout the longitudinal part is therefore not seriously affected by the survival. The simulation results indicate that the misspecification of the joint model omitting the time-dependent terms is most severe for the strong association between the survival and longitudinal process. In practice one does not have a gold standard model and would like to choose the best model from the set of models with different parameterizations. When Bayesian methods are used for the joint model estimation, a DIC or other Bayesian criterion could be considered. However, due to the well-known limitations of such criteria, we are currently working on developing more general measures that would allow to choose the best model based on the quality of the produced predictions in

terms of calibration and discrimination, regardless the estimation method.

4.8 Software

The code for the Bayesian joint models with different parameterizations as well as the R code for the sampling procedure is available upon request from the first author.

4.9 Supplementary Material

Supplementary materials are available online at:
http://biostatistics.oxfordjournals.org.

Acknowledgements

The authors thank Jacqueline M. Smits from Eurotransplant International Foundation in Leiden for providing the data set analyzed in the paper and for the medical consult regarding the application of the proposed method. We also thank Interdisciplinary Centre for Mathematical and Computational Modeling from University of Warsaw for the possibility of using the Halo2 ICM supercomputer system.

Conflict of Interest: None declared.

Chapter 5

Joint Modeling vs Landmarking

A key question in clinical practice is accurate prediction of patient prognosis. To this end, nowadays, physicians have at their disposal a variety of tests and biomarkers to aid them in optimizing medical care. These tests are often performed on a regular basis in order to closely follow the progression of the disease. In this setting it is of medical interest to optimally utilize the recorded information and provide medically-relevant summary measures, such as survival probabilities, that will aid in decision making. In this chapter we present and compare two statistical techniques that provide dynamically-updated estimates of survival probabilities, namely landmark analysis and joint models for longitudinal and time-to-event data. Special attention is given to the functional form linking the longitudinal and event time processes, and to measures of discrimination and calibration in the context of

dynamic prediction.

5.1 Introduction

Nowadays there is great interest in accurate risk assessment for prevention and treatment of disease. Physicians use risk scores to reach appropriate decisions, such as prescribing treatment, or extra medical tests or suggesting alternative therapies. Patients who are informed about their health risk often decide to adjust their lifestyles to mitigate it. Risk scores are typically based on several factors that describe the patients' physical condition, such as age, BMI, smoking, genetic predisposition, and the results of medical tests. In this work we focus on the use of the results of such tests and more specifically on biomarkers. The majority of prognostic models in the medical literature utilize only a small fraction of the available biomarker information. In particular, even though biomarkers are measured repeatedly over time, risk scores are typically based on the last available biomarker measurement. It is evident that such an approach discards valuable information because it does not take into account that the rate of change in the biomarker levels is not only different from patient to patient but also dynamically changes over time for the same patient. Hence, it is medically relevant to investigate whether repeated measurements of a biomarker can provide a better understanding of disease progression and a better prediction of the risk for the event of interest than a single biomarker measurement.

In line with the previous arguments, the motivation for this research

This chapter has been submitted for publication as "Dynamic Predictions with Time-Dependent Covariates in Survival Analysis using joint modeling and Landmarking" in *Biometrical Journal* [46].

comes from a study conducted by the Department of Cardio-Thoracic Surgery of the Erasmus Medical Center in the Netherlands. This study includes 285 patients who received a human tissue valve in the aortic position in the hospital from 1987 until 2008 [44]. Aortic allograft implantation has been widely used for a variety of aortic valve or aortic root diseases. Major advantages ascribed to allografts are the excellent hemodynamic characteristics as a valve substitute; the low rate of thrombo-embolic complications, and, therefore, absence of the need for anticoagulant treatment; and the resistance to endocarditis. A major disadvantage of using human tissue valves, however is the susceptibility to degeneration and the concomitant need for re-interventions. The durability of a cryopreserved aortic allograft is age-dependent, leading to a high lifetime risk of re-operation, especially for young patients. Re-operations on the aortic root are complex, with substantial operative risks, and mortality rates in the range 4–12%. It is therefore of great interest for cardiologists and cardio-thoracic surgeons to have at their disposal an accurate prognostic tool that will inform them about the future prospect of a patient with a human tissue valve in order to optimize medical care, carefully plan re-operation and minimize valve-relate morbidity and mortality.

From the statistical analysis viewpoint the challenge is to utilize a technique capable of updating estimates of survival probabilities for a new patient as additional longitudinal information is recorded. An early approach in solving this problem has been landmarking [93–95]. The basic idea behind landmarking is to obtain survival probabilities from a Cox model fitted to the patients from the original dataset who are still at risk at the time point of interest (e.g., the last time point we know that the new patient was still alive). A relatively newer method for producing dynamic predictions

of survival probabilities is based on the class of joint models for longitudinal and time-to-event data [41, 85, 86, 96, 97]. In these models we have a complete specification of the joint distribution of the longitudinal response and the event times based on which the predictions in question can be derived. The main aim of this paper is to further study and contrast these two approaches. In particular, we show how survival probabilities are obtained under each method and what the differences are in the underlying assumptions. In addition, we focus on the functional relationship between the two processes and how this may affect predictions. We surpass the standard formulation, which only includes the current value of the marker, and we postulate functional forms that allow the rate of increase/decrease of the longitudinal outcome or a suitable summary of the whole longitudinal trajectory to determine the risk for an event. To assess the quality of the derived predictions from the two approaches we present different measures of discrimination and calibration, suitably adjusted to the context of longitudinal biomarkers.

The rest of the paper is organized as follows. Section 5.2 describes formally the context of dynamic predictions and presents the landmarking and joint modeling approaches. Section 5.3 shows different options for the functional form of the association structure between the longitudinal and event time processes. Section 5.4 presents measures of discrimination and calibration adapted to the dynamic predictions setting. Section 5.5 illustrates the use of joint modeling and landmarking in the Aortic Valve dataset and Section 5.6 refers to the results of a simulation study. Finally, Section 5.7 concludes the paper.

5.2 Dynamic Individualized Predictions

Following the discussion in Section 5.1 and the motivation from the Aortic Valve dataset, we present here the two frameworks for deriving dynamic individualized predictions. Let $\mathcal{D}_n = \{T_i, \delta_i, \boldsymbol{y}_i; i = 1, \ldots, n\}$ denote a sample from the target population, where T_i^* denotes the true event time for the i-th subject ($i = 1, \ldots, n$), C_i the censoring time, $T_i = \min(T_i^*, C_i)$ the corresponding observed event time, and $\delta_i = I(T_i^* \leq C_i)$ the event indicator, with $I(\cdot)$ being the indicator function that takes the value 1 when $T_i^* \leq C_i$, and 0 otherwise. In addition, we let $\boldsymbol{y}_i$ denote the $n_i \times 1$ longitudinal response vector for the i-th subject, with element y_{il} denoting the value of the longitudinal outcome taken at time point t_{il}, $l = 1, \ldots, n_i$.

We are interested in deriving predictions for a new subject j from the same population that has provided a set of longitudinal measurements $\mathcal{Y}_j(t) = \{y_j(t_{jl}); 0 \leq t_{jl} \leq t, l = 1, \ldots, n_j\}$, and has a vector of baseline covariates $\boldsymbol{w}_j$. The fact that biomarker measurements have been recorded up to t, implies survival of this subject up to this time point, meaning that it is more relevant to focus on the conditional subject-specific predictions, given survival up to t. In particular, for any time $u > t$ we are interested in the probability that this new subject j will survive at least up to u, i.e.,

$$\pi_j(u \mid t) = \Pr(T_j^* \geq u \mid T_j^* > t, \mathcal{Y}_j(t), \boldsymbol{w}_j, \mathcal{D}_n).$$

The time-dynamic nature of $\pi_j(u \mid t)$ is evident because when new information is recorded for patient j at time $t' > t$, we can update these predictions to obtain $\pi_j(u \mid t')$, and therefore proceed in a time-dynamic manner.

5.2.1 Landmarking

The landmarking approach provides an estimate of $\pi_j(u \mid t)$ by selecting the subjects at risk at t from the original dataset $\mathcal{D}_n$, and using these to derive predictions. More formally, let $\mathcal{R}(t) = \{i : T_i > t\}$ denote the adjusted risk set, including all subjects who were not censored or dead by the landmark time t. Then a Cox model is fitted to these subjects by resetting time with zero being the landmark time, i.e.,

$$\begin{aligned} h_i(u-t) &= \lim_{\Delta t \to 0} \frac{1}{\Delta t} \Pr\big\{u - t \leq T_i^* < u - t + \Delta t \mid T_i^* > u - t, \mathcal{Y}_i(t)\big\} \\ &= h_0(u-t) \exp\big\{\boldsymbol{\gamma}^\top \boldsymbol{w}_i + \alpha \tilde{y}_i(t)\big\}, u > t, \end{aligned}$$

where the baseline hazard function $h_0(\cdot)$ is assumed completely unspecified, $\boldsymbol{w}_i$ denotes a vector of baseline covariates, and the last available longitudinal response $\tilde{y}_i(t)$ also enters into the model as an ordinary baseline covariate. Having fitted this Cox model, an estimate of $\pi_j(u \mid t)$ is simply obtained by means of the Breslow estimator:

$$\hat{\pi}_j^{LM}(u \mid t) = \exp\Big[-\widehat{H}_0(u) \exp\{\widehat{\boldsymbol{\gamma}}^\top \boldsymbol{w}_j + \hat{\alpha} \tilde{y}_j(t)\}\Big], \tag{5.1}$$

where

$$\widehat{H}_0(u) = \sum_{i \in \mathcal{R}(t)} \frac{I(T_i \leq u)\delta_i}{\sum_{\ell \in \mathcal{R}(u)} \exp\{\widehat{\boldsymbol{\gamma}}^\top \boldsymbol{w}_\ell + \hat{\alpha} \tilde{y}_\ell(t)\}}.$$

[95] and [94] discuss several extensions of this approach that have greater flexibility by allowing the regression coefficient α to depend on time, i.e.,

$$h_i(u-t) = h_0(u-t) \exp\big\{\boldsymbol{\gamma}^\top \boldsymbol{w}_i + \alpha(u-t) \tilde{y}_i(t)\big\},$$

and also, possibly, a baseline hazard that is not only a function of the time since the last measurement $u - t$, but also a function of the measurement time t, relaxing thus the proportional hazards assumption. An advantage of landmarking is that it can be very easily applied in practice, because it only requires fitting a simple Cox model each time a new measurement has been recorded for the subject for whom predictions are of interest.

5.2.2 Joint modeling

Contrary to the landmark approach, in the framework of joint models for longitudinal and time-to-event data we have a complete specification of the joint distribution of the two outcomes [2, 3, 41, 48, 90]. For the longitudinal biomarker measurements mixed-effects models are typically employed to describe the subject-specific longitudinal trajectories. For simplicity of exposition and because the marker that we are going to use for the Aortic Valve dataset, namely the aortic gradient, is a continuous one, we focus here on linear mixed-effects models,

$$\begin{aligned} y_i(t) &= m_i(t) + \varepsilon_i(t) = \boldsymbol{x}_i^\top(t)\boldsymbol{\beta} + \boldsymbol{z}_i^\top(t)\boldsymbol{b}_i + \varepsilon_i(t), \\ \boldsymbol{b}_i &\sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{D}), \quad \varepsilon_i(t) \sim \mathcal{N}(0, \sigma^2), \end{aligned} \tag{5.2}$$

where $y_i(t)$ denotes the observed value of the longitudinal outcome at any particular time point t, $\boldsymbol{x}_i(t)$ and $\boldsymbol{z}_i(t)$ denote the time-dependent design vectors for the fixed-effects $\boldsymbol{\beta}$ and for the random effects $\boldsymbol{b}_i$, respectively, and $\varepsilon_i(t)$ the corresponding error terms that are assumed independent of the random effects, and $\text{cov}\{\varepsilon_i(t), \varepsilon_i(t')\} = 0$ for $t' \neq t$. For the survival process, we assume that the risk for an event depends on the 'true' and unobserved

value of the marker at time t (i.e., excluding the measurement error), denoted by $m_i(t)$ in (5.2). More specifically, we have

$$\begin{aligned} h_i(t \mid \mathcal{M}_i(t), \boldsymbol{w}_i) &= \lim_{\Delta t \to 0} \frac{1}{\Delta t} \Pr\{t \leq T_i^* < t + \Delta t \mid T_i^* \geq t, \mathcal{M}_i(t), \boldsymbol{w}_i\} \\ &= h_0(t) \exp\{\boldsymbol{\gamma}^\top \boldsymbol{w}_i + \alpha m_i(t)\}, \quad t > 0, \end{aligned} \tag{5.3}$$

where $\mathcal{M}_i(t) = \{m_i(s), 0 \leq s < t\}$ denotes the history of the true unobserved longitudinal process up to t, $h_0(\cdot)$ denotes the baseline hazard function, and, as before, $\boldsymbol{w}_i$ is a vector of baseline covariates with corresponding regression coefficients $\boldsymbol{\gamma}$. Parameter α quantifies the association between the true value of the marker at t and the hazard for an event at the same time point. Estimation of joint model's parameters can be based either on maximum likelihood or a Bayesian approach using Markov chain Monte Carlo algorithms. The likelihood of the model is derived under the assumptions that given the random effects, both the longitudinal and event time process are assumed independent, and the longitudinal responses of each subject are assumed independent. Formally we have,

$$p(\boldsymbol{y}_i, T_i, \delta_i \mid \boldsymbol{b}_i, \boldsymbol{\theta}) = p(\boldsymbol{y}_i \mid \boldsymbol{b}_i, \boldsymbol{\theta})\, p(T_i, \delta_i \mid \boldsymbol{b}_i, \boldsymbol{\theta}), \tag{5.4}$$

$$p(\boldsymbol{y}_i \mid \boldsymbol{b}_i, \boldsymbol{\theta}) = \prod_l p(y_{il} \mid \boldsymbol{b}_i, \boldsymbol{\theta}), \tag{5.5}$$

where $\boldsymbol{\theta}^\top = (\boldsymbol{\theta}_t^\top, \boldsymbol{\theta}_y^\top, \boldsymbol{\theta}_b^\top)$ denotes the full parameter vector, with $\boldsymbol{\theta}_t$ denoting the parameters for the event time outcome, $\boldsymbol{\theta}_y$ the parameters for the longitudinal outcomes, and $\boldsymbol{\theta}_b$ the unique parameters of the random-effects covariance matrix, and $p(\cdot)$ denotes an appropriate probability density function. More details regarding the estimation and properties of joint models

can be found in [41] and [89].

Under this framework, estimation of $\pi_j(u \mid t)$ can be based on (asymptotic) Bayesian arguments and the corresponding posterior predictive distribution:

$$\pi_j(u \mid t) = \int \Pr(T_j^* \geq u \mid T_j^* > t, \mathcal{Y}_j(t), \boldsymbol{\theta})\, p(\boldsymbol{\theta} \mid \mathcal{D}_n)\, d\boldsymbol{\theta}.$$

The calculation of the first part of each integrand takes full advantage of the conditional independence assumptions (5.4) and (5.5). In particular, we observe that the first term of the integrand of $\pi_j(u \mid t)$ can be rewritten by noting that:

$$\begin{aligned}
&\Pr(T_j^* \geq u \mid T_j^* > t, \mathcal{Y}_j(t), \boldsymbol{\theta}) \\
&= \int \Pr(T_j^* \geq u \mid T_j^* > t, \boldsymbol{b}_j, \boldsymbol{\theta})\, p(\boldsymbol{b}_j \mid T_j^* > t, \mathcal{Y}_j(t), \boldsymbol{\theta})\, d\boldsymbol{b}_j \\
&= \int \frac{S_j\{u \mid \mathcal{M}_j(u, \boldsymbol{b}_j), \boldsymbol{\theta}\}}{S_j\{t \mid \mathcal{M}_j(t, \boldsymbol{b}_j), \boldsymbol{\theta}\}}\, p(\boldsymbol{b}_j \mid T_j^* > t, \mathcal{Y}_j(t), \boldsymbol{\theta})\, d\boldsymbol{b}_j,
\end{aligned}$$

where

$$S_j\{t \mid \mathcal{M}_j(t, \boldsymbol{b}_j), \boldsymbol{\theta}\} = \exp\left\{ \int_0^t h_0(s) \exp\{\boldsymbol{\gamma}^\top \boldsymbol{w}_i + \alpha m_i(s)\} \right\} ds,$$

denotes the subject-specific survival function.

Combining these equations with the maximum likelihood estimates or with the MCMC sample from the posterior distribution of the parameters for the original data $\mathcal{D}_n$, we can devise a simple simulation scheme to obtain a Monte Carlo estimate of $\pi_j(u \mid t)$. More specifically, this is comprised of the following steps:

Step 1. Take K samples of $\{\boldsymbol{\theta}^{(k)}, k = 1, \ldots, K\}$ from either the MCMC sample

of $p(\boldsymbol{\theta} \mid \mathcal{D}_n)$ or the asymptotic normal posterior distribution $\mathcal{N}(\widehat{\boldsymbol{\theta}}, \mathcal{H}_n)$, where $\widehat{\boldsymbol{\theta}}$ denotes the maximum likelihood estimates and $\mathcal{H}_n$ the observed information matrix

$$\mathcal{H}_n = \left\{ -\sum_{i=1}^{n} \frac{\partial^2 \log p(\boldsymbol{y}_i, T_i, \delta_i, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^\top \partial \boldsymbol{\theta}} \Big|_{\theta=\hat{\theta}} \right\}^{-1}.$$

Step 2. Draw K realizations $\{\boldsymbol{b}_j^{(k)}, k = 1, \ldots, K\}$ for the random effects of the new subject j from the posterior distribution of the random effects

$$p\big(\boldsymbol{b}_j \mid T_j^* > t, \mathcal{Y}_j(t), \boldsymbol{\theta}^{(k)}\big) \propto \left\{ \prod_{l=1}^{n_j(t)} p\big(y_{jl} \mid \boldsymbol{b}_j, \boldsymbol{\theta}^{(k)}\big) \right\} S_j\big\{t \mid \mathcal{M}_j(t, \boldsymbol{b}_j), \boldsymbol{\theta}^{(k)}\big\}$$

$$\times p\big(\boldsymbol{b}_j, \boldsymbol{\theta}^{(k)}\big),$$

where $n_j(t)$ denotes the number of available measurements for subject j by time t.

Step 3. Based on these realizations an estimate of $\pi_j(u \mid t)$ is derived as

$$\hat{\pi}_j^{JM}(u \mid t) = \frac{1}{K} \sum_{k=1}^{K} \frac{S_j\big\{u \mid \mathcal{M}_j(u, \boldsymbol{b}_j^{(k)}), \boldsymbol{\theta}^{(k)}\big\}}{S_j\big\{t \mid \mathcal{M}_j(t, \boldsymbol{b}_j^{(k)}), \boldsymbol{\theta}^{(k)}\big\}}. \tag{5.6}$$

More details can be found in [85] and [41, 86].

5.2.3 Heuristic comparison between landmarking and joint modeling

The previous two sections illustrated that both landmarking and joint modeling can be utilized to derive dynamically updated estimates of conditional

survival probabilities $\pi_j(u \mid t)$. The landmark approach can be more easily implemented in practice because it only requires fitting a standard Cox model, whereas joint models require specialized software [40,41]. In addition, joint models seem to make more modeling assumptions than the landmark approach, which poses a concern regarding how a misspecification of these assumptions may affect predictions. On the other hand, the landmark approach uses less information than joint modeling (i.e., only the last observed longitudinal response), and hence is less optimal. The following points provide a more detailed exposition of the underlying differences between the two approaches.

- **Extrapolation:** The main differences in how landmarking and joint modeling tackle the problem of prediction can best be explained by Figure 5.1. This shows the longitudinal responses of a hypothetical subject who was alive up to year five and for whom we would like to obtain a predicted survival function. To produce estimates of the conditional survival probabilities both landmarking and joint modeling require a value for the longitudinal response at $t = 5$ (vertical dotted line). Since this subject provided her last longitudinal measurement at year three, some sort of extrapolation is taking place. In particular, landmarking is based on a 'last value carried forward' approach and uses as the value of the longitudinal response at year five the last available measurement of the subject at year three (horizontal dashed line). Even though this approach is conceptually simple and easy to perform in practice, unfortunately, it may lead to biased and misleading inference on the Cox model parameters [98]. Joint modeling on the other hand uses the subject-specific fitted value of the longitudinal profile from the linear

mixed model extrapolated at year 5, i.e., $m_j(5) = \boldsymbol{x}_j^\top(5)\boldsymbol{\beta} + \boldsymbol{z}_j^\top(5)\boldsymbol{b}_j$ (solid line). This approach uses all available information, because the estimate of $m_j(5)$ is based on both all past values of this subject and on the responses of other subjects. To explain how the borrowing of information between subjects is taking place, assume, hypothetically, that there was another patient, who during the first three years had exactly the same longitudinal measurements as the patient depicted in Figure 5.1, but also she had extra measurements up to year five. The joint model would make use of this patient and say that the profile of the patient in Figure 5.1 would be similar to the one of the patient with the extra measurements. From a biological point of view the joint modeling approach seems more logical than landmarking because we indeed expect the biomarker levels of a patient to continuously change over time rather than to remain constant between visits.

Note that in general even if we had observed the longitudinal response at $t = 5$, i.e., $y_j(5)$ this will not be equal to $m_j(5)$. The joint model assumes that the realizations of the longitudinal marker are the output of a stochastic process generated by the subject, and it is the underlying signal in the process, represented by $m_j(t)$, that is associated with the hazard for an event. The observed data $y_j(t)$ are a contaminated with measurement error version of the underlying signal $m_j(t)$. This measurement error most often stems from biological variation, but some times may also be attributed to the medical test/examination used to measure the marker.

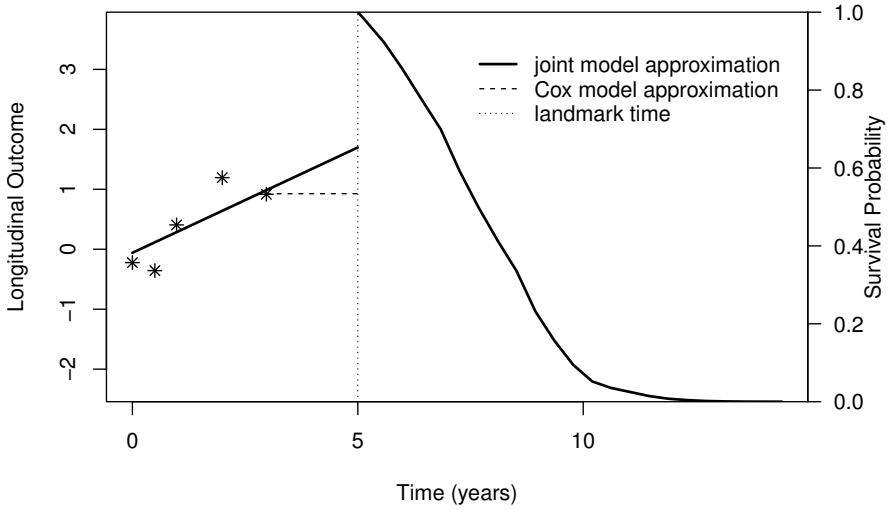


Figure 5.1: Graphical comparison on how landmarking and joint modeling use the available longitudinal measurements to provide an estimate of the longitudinal outcome at the last time point the patients was still alive. The left side of the plot shows the observed longitudinal responses, and the fitted longitudinal profile from the joint model. The right side shows the corresponding survival probability.

- **Assumptions related to the longitudinal process:** The landmark approach assumes that the visiting process, which is the process, stochastic or deterministic, that generates the visit times at which subjects provide measurements is independent of the longitudinal marker process and the survival time T_j^*. The joint modeling approach also assumes that a visit scheduled at time t is independent of a future event occurring at $T_j^* > t$ and of future longitudinal responses $\{y_j(s), t \leq s \leq T_j^*\}$, but it does allow visit times to depend on the observed longitudinal responses $\mathcal{Y}_j(t)$. This is a more realistic assumption because what we expect to happen in practice is that physicians will ask a patient to come back more often if they observe a worsening of her condition

based on her observed responses. In addition, subjects may have missing marker measurements during follow-up. The landmark approach assumes that any such missingness is completely at random [99]. On the other hand, due to the fact that joint modeling is based on a complete specification of the joint likelihood function of the longitudinal and event time processes, it allows incomplete longitudinal data to be missing at random. Hence, joint modeling is capable of providing valid inferences under less stringent assumptions than landmarking. Though, it should be mentioned that these advantageous features require the joint model to be roughly correctly specified.

- **Assumptions related to the event process:** Similarly to the assumptions for the longitudinal process, landmarking makes more stringent assumptions for the censoring process. In particular, under the landmark approach censoring is assumed independent of past longitudinal responses $\{y_j(s); 0 \leq s < t\}$, whereas under joint modeling and again because we use a complete specification of the joint likelihood function, censoring is allowed to depend in a general way on $\{y_j(s); 0 \leq s < t\}$.

5.3 Functional form

The assessment of the predictive value of baseline covariates is to a degree simple, in the sense that these covariates are typically included in a prognostic model as is or under a suitable transformation (e.g., log-scale, polynomials, splines, etc.). However, in our setting, where we have multi-

ple longitudinal measurements available per subject there could be different features of the longitudinal process that are most predictive for the event of interest. For example, in ordinary proportional hazards models, it has been long recognized that the functional form of time-varying covariates influences the derived inferences; see, for instance, [100] and references therein. In the joint modeling framework however, where the longitudinal outcome plays the role of a time-dependent covariate for the survival process, this topic has received less attention. The two main functional forms that have been primarily used so far in joint models include in the linear predictor of the relative risk model (5.3) either the subject-specific means $m_i(t)$ from the longitudinal submodel or just the random effects $\boldsymbol{b}_i$ [28, 48]. However, as argued above, there could be other characteristics of the patients' longitudinal profiles that are more predictive for the risk of an event, such as the rate of increase/decrease of the biomarker's levels or a suitable summary of the whole longitudinal trajectory. Here we present a few examples of alternative formulations for the association structure between the longitudinal outcome and the risk for an event:

$$h_i(t) = h_0(t)\exp\big\{\boldsymbol{\gamma}^\top \boldsymbol{w}_i + \alpha_1 m_i(t) + \alpha_2 m_i'(t)\big\}, \quad m_i'(t) = \frac{dm_i(t)}{dt}, \tag{5.7}$$

$$h_i(t) = h_0(t)\exp\Big\{\boldsymbol{\gamma}^\top \boldsymbol{w}_i + \alpha \int_0^t m_i(s)\, ds\Big\}, \tag{5.8}$$

$$h_i(t) = h_0(t)\exp\Big\{\boldsymbol{\gamma}^\top \boldsymbol{w}_i + \alpha \int_0^t \varrho(t-s) m_i(s)\, ds\Big\}, \tag{5.9}$$

$$h_i(t) = h_0(t)\exp(\boldsymbol{\gamma}^\top \boldsymbol{w}_i + \boldsymbol{\alpha}^\top \boldsymbol{b}_i). \tag{5.10}$$

It is evident that these parameterizations have different sets of association parameters $\boldsymbol{\alpha}$, and in addition that the interpretation of these parameters is different for each formulation. In particular, parameterization (5.7) postulates that the risk for an event at a particular time point t depends not only

on the level of the marker at this time point but also on its rate of change, captured by the slope term $m_i'(t)$. This could be of importance when two patients at a specific time point have equal marker levels, but one patient having an increasing trajectory and the other a decreasing one. Parameterization (5.8) posits that the risk for an event at time t is associated with the area under the longitudinal trajectory up to this point. This can be considered as a summary of the whole marker history up to t and contrary to the previous formulations it allows the risk to the depend on the whole history $\mathcal{M}_i(t) = \{m_i(s), 0 \leq s < t\}$ and not only on features of the marker at t. Parameterization (5.9) extends (5.8) by assigning to the past values of the longitudinal trajectory different weights, using a function $\varrho(\cdot)$. For instance, setting $\varrho(t-s) = \phi(t-s)/\{\Phi(t) - 0.5\}$, where $0 < s < t$, and $\phi(\cdot)$ and $\Phi(\cdot)$ denote the probability density and cumulative distribution functions of the standard normal distribution, respectively, we assume that the risk at t only depends on the marker levels in the interval $(t-3, t)$ with values closer to t having higher weight, because when $t - s > 3$ then $\varrho(t-s)$ is practically zero. Finally, parameterization (5.10) is time-independent and assumes that the hazard for an event is related to the random effects from the longitudinal process. This formulation shares similarities with the time-dependent slopes parameterization (5.7) when a simple random-intercepts and random-slopes structure is assumed for the longitudinal submodel.

Under the landmarking approach, and in order to improve predictive performance, we could also make better use of the observed longitudinal history than just using the last available measurement. Mimicking the formulations presented above for joint modeling, we can define Cox models fitted to the patients at risk at the landmark time t, which include $\tilde{y}_i'(t)$ that denotes the

slope calculated from the last two available measurements of each subject, and $\sum_{0\leq s\leq t} y_i(s)\Delta s$ that denotes the area under the step function defined from the observed longitudinal measurements up to t:

$$\begin{aligned} h_i(u-t) &= h_0(u-t)\exp\{\boldsymbol{\gamma}^\top \boldsymbol{w}_i + \alpha_1 \tilde{y}_i(t) + \alpha_2 \tilde{y}_i'(t)\}, \\ h_i(u-t) &= h_0(u-t)\exp\Big\{\boldsymbol{\gamma}^\top \boldsymbol{w}_i + \alpha \sum_{s=0}^{t} y_i(s)\Delta s\Big\}, \\ h_i(u-t) &= h_0(u-t)\exp\Big\{\boldsymbol{\gamma}^\top \boldsymbol{w}_i + \alpha \sum_{s=0}^{t} \varrho(t-s) y_i(s)\Delta s\Big\}, \end{aligned}$$

where, as before, $\varrho(t-s)$ is a potential weight function. Note that we do not have an analogous functional form to (5.10) under landmarking.

5.4 Measuring predictive performance

The assessment of the predictive performance of time-to-event models has received a lot of attention in the statistical literature. In general, the developed methodology has focused on calibration, i.e., how well the model predicts the observed data [101,102] or discrimination, i.e., how well can the model discriminate between patients that had the event from patients that did not [103,104]. In the following we present discrimination and calibration measures suitably adapted to the dynamic prediction setting. It should be noted that these measures require in their essence an estimate of $\pi_j(u \mid t)$, and therefore they are applicable under both landmarking and joint modeling. In the following we will use the term $\hat{\pi}_j(u \mid t)$ to generically denote either (5.1) or (5.6).

5.4.1 Discrimination

To take into account the dynamic nature of the longitudinal marker in discriminating between subjects, we focus on a time interval of medical relevance within which the occurrence of events is of interest. In this setting, a useful property of the model would be to successfully discriminate between patients who are going to experience the event within this time frame from patients who will not. To put this formally, as before, we assume that we have collected longitudinal measurements $\mathcal{Y}_j(t) = \{y_j(t_{jl}); 0 \leq t_{jl} \leq t, l = 1, \ldots, n_j\}$ up to time point t for subject j. We are interested in events occurring in the medically-relevant time frame $(t, t + \Delta t]$ within which the physician can take an action to improve the survival chance of the patient. Under the assumed model and the methodology presented in Section 5.2, we can define a prediction rule using $\pi_j(t + \Delta t \mid t)$ that takes into account the available longitudinal measurements $\mathcal{Y}_j(t)$. In particular, for any value c in $[0, 1]$ we can term subject j as a case if $\pi_j(t+\Delta t \mid t) \leq c$ (i.e., occurrence of the event) and analogously as a control if $\pi_j(t + \Delta t \mid t) > c$. Thus, in this context, we define sensitivity and specificity as

$$\Pr\big\{\pi_j(t + \Delta t \mid t) \leq c \mid T_j^* \in (t, t + \Delta t]\big\},$$

and

$$\Pr\big\{\pi_j(t + \Delta t \mid t) > c \mid T_j^* > t + \Delta t\big\},$$

respectively. For a randomly chosen pair of subjects $\{i, j\}$, in which both subjects have provided measurements up to time t, the discriminative capability of the assumed model can be assessed by the area under the receiver

operating characteristic curve (AUC), which is obtained for varying c and equals,

$$\mathrm{AUC}(t, \Delta t) = \Pr\big[\pi_i(t+\Delta t \mid t) < \pi_j(t+\Delta t \mid t) \mid \{T_i^* \in (t, t+\Delta t]\} \cap \{T_j^* > t+\Delta t\}\big],$$

that is, if subject i experiences the event within the relevant time frame whereas subject j does not, then we would expect the assumed model to assign higher probability of surviving longer than $t + \Delta t$ for the subject who did not experience the event. To summarize the discriminative power of the assumed model over the whole follow-up period, we need to take into account that the number of subjects contributing to the comparison of the fitted $\pi_i(t + \Delta t \mid t)$ with the observed data is not the same for all time points t. Following an approach similar to [105] and [106], we propose the use of a weighted average of AUCs

$$\mathrm{C}_{dyn}^{\Delta t} = \int_0^\infty \mathrm{AUC}(t, \Delta t)\ \Pr\{\mathcal{E}(t)\}\ dt \Big/ \int_0^\infty \Pr\{\mathcal{E}(t)\}\ dt, \tag{5.11}$$

where $\mathcal{E}(t) = \big[\{T_i^* \in (t, t + \Delta t]\} \cap \{T_j^* > t + \Delta t\}\big]$, and $\Pr\{\mathcal{E}(t)\}$ denotes the probability that a random pair is comparable at t. We call $\mathrm{C}_{dyn}^{\Delta t}$ the dynamic concordance index since it summarizes the concordance probabilities over the follow-up period. Note also that $\mathrm{C}_{dyn}^{\Delta t}$ depends on the length Δt of the time interval of interest, which implies that different models may exhibit different discrimination power for different Δt.

For the estimation of $\mathrm{C}_{dyn}^{\Delta t}$ we need to take care of two issues, namely, the calculation of the integrals in the definition of (5.11) and censoring. For the former we use a 15-point Gauss-Kronrod quadrature rule [107]. To take into account the fact that the number of subjects decreases over time due

to the occurrence of events and censoring, for any time point t we define as comparable pairs the pairs that satisfy the relation

$$\begin{aligned}\Omega_{ij}(t) &= \big[\{T_i \in (t, t+\Delta t]\} \cap \{\delta_i = 1\}\big] \cap \{T_j > t + \Delta t\} \text{ or} \\ &\quad \big[\{T_i \in (t, t+\Delta t]\} \cap \{\delta_i = 1\}\big] \cap \big[\{T_j = t + \Delta t\} \cap \{\delta_j = 0\}\big],\end{aligned}$$

where $i, j = 1, \ldots, n$ with $i \neq j$. For two comparable subjects i and j, we can estimate and compare their survival probabilities $\pi_i(t + \Delta t \mid t)$ and $\pi_j(t + \Delta t \mid t)$, based on the methodology presented in Section 5.2. This leads to a natural estimator for $\text{AUC}(t, \Delta t)$ as the proportion of concordant subjects out of the set of comparable subjects for time t:

$$\widehat{\text{AUC}}(t, \Delta t) = \frac{\sum_{i=1}^{n} \sum_{j=1; j\neq i}^{n} I\{\hat{\pi}_i(t + \Delta t \mid t) < \hat{\pi}_j(t + \Delta t \mid t)\} \times I\{\Omega_{ij}(t)\}}{\sum_{i=1}^{n} \sum_{j=1; j\neq i}^{n} I\{\Omega_{ij}(t)\}},$$

where $I(\cdot)$ denotes the indicator function. Having estimated $\text{AUC}(t, \Delta t)$, the next step in estimating $\text{C}_{dyn}^{\Delta t}$ is to obtain estimates for the weights $\Pr\{\mathcal{E}(t)\}$. We observe that these can be rewritten as

$$\begin{aligned}\Pr\{\mathcal{E}(t)\} &= \Pr\big[\{T_i^* \in (t, t+\Delta t]\} \cap \{T_j^* > t + \Delta t\}\big] \\ &= \Pr(T_i^* \in (t, t+\Delta t]) \times \Pr(T_j^* > t + \Delta t) \\ &= \{S(t) - S(t+\Delta t)\} S(t + \Delta t),\end{aligned}$$

where the simplification in the second line comes from the independence of subjects i and j, and $S(\cdot)$ here denotes the marginal survival function.

In practice the calculation of $\text{C}_{dyn}^{\Delta t}$ is restricted into a follow-up interval $[0, t_{max}]$ where we have information. Let $t_1, \ldots, t_{15}$ denote the re-scaled ab-

scissas of the Gauss-Kronrod rule in the interval $[0, t_{max}]$ with corresponding weights $\varpi_1, \ldots, \varpi_{15}$. We combine the estimates $\widehat{\mathrm{AUC}}(t_k, \Delta t)$, $k = 1, \ldots, 15$ with the estimates of the weights $\Pr\{\mathcal{E}(t)\}$ to obtain

$$\widehat{\mathrm{C}}^{\Delta t}_{dyn} = \frac{\sum_{k=1}^{15} \varpi_k \widehat{\mathrm{AUC}}(t_k, \Delta t) \times \widehat{\Pr}\{\mathcal{E}(t_k)\}}{\sum_{k=1}^{15} \varpi_k \widehat{\Pr}\{\mathcal{E}(t_k)\}},$$

where $\widehat{\Pr}\{\mathcal{E}(t_k)\} = \{\widehat{S}(t_k) - \widehat{S}(t_k + \Delta t)\}\widehat{S}(t_k + \Delta t)$, with $\widehat{S}(\cdot)$ denoting the Kaplan-Meier estimate of the marginal survival function $S(\cdot)$.

5.4.2 Calibration

The assessment of the accuracy of predictions of survival models is typically based on the expected error of predicting future events. In our setting, and again taking into account the dynamic nature of the longitudinal outcome, it is of interest to predict the occurrence of events at $u > t$ given the information we have recorded up to time t. This gives rise to expected prediction error:

$$\mathrm{PE}(u \mid t) = E\big[L\{N_i(u) - \pi_i(u \mid t)\}\big],$$

where $N_i(t) = I(T_i^* > t)$ is the event status at time t, $L(\cdot)$ denotes a loss function, such as the absolute or square loss, and the expectation is taken with respect to the distribution of the event times. An estimate of $\mathrm{PE}(u \mid t)$ that accounts for censoring has been proposed by [96]:

$$\widehat{\mathrm{PE}}(u \mid t) = \{n(t)\}^{-1} \sum_{i:T_i \geq t} I(T_i \geq u) L\{1 - \hat{\pi}_i(u \mid t)\} + \delta_i I(T_i < u) L\{0 - \hat{\pi}_i(u \mid t)\}$$

$$+ (1 - \delta_i) I(T_i < u) \Big[\hat{\pi}_i(u \mid T_i) L\{1 - \hat{\pi}_i(u \mid t)\} + \{1 - \hat{\pi}_i(u \mid T_i)\} L\{0 - \hat{\pi}_i(u \mid t)\}\Big],$$

where $n(t)$ denotes the number of subjects at risk at time t. The first two terms in the sum correspond to patients who were alive after time u and dead before u, respectively; the third term corresponds to patients who were censored in the interval $[t, u]$. Using the longitudinal information up to time t, $\mathrm{PE}(u \mid t)$ measures the predictive accuracy at the specific time point u. Alternatively, we could summarize the error of prediction in a specific interval of interest, say $[t, u]$, by calculating a weighted average of $\{\mathrm{PE}(s \mid t), t < s < u\}$ that corrects for censoring, similarly to $\mathrm{C}^{\Delta t}_{dyn}$. An estimator of this type for the integrated prediction error has been suggested by [101], which adapted to our time-dynamic setting takes the form

$$\widehat{\mathrm{IPE}}(u \mid t) = \frac{\sum_{i:t \le T_i \le u} \delta_i \big\{\widehat{S}_C(t)/\widehat{S}_C(T_i)\big\} \widehat{\mathrm{PE}}(u \mid t)}{\sum_{i:t \le T_i \le u} \delta_i \big\{\widehat{S}_C(t)/\widehat{S}_C(T_i)\big\}},$$

where $\widehat{S}_C(\cdot)$ denotes the Kaplan-Meier estimator of the censoring time distribution.

Both $\widehat{\mathrm{IPE}}(u \mid t)$ and $\widehat{\mathrm{PE}}(u \mid t)$ can be used to provide a measure of explained variation between nested models. Assuming model M_1 is nested in model M_2, we can compute how much the extra structure in M_2 improves accuracy by

$$R^2_{PE}(u \mid t; M_1, M_2) = 1 - \widehat{\mathrm{PE}}_{M_2}(u \mid t) \Big/ \widehat{\mathrm{PE}}_{M_1}(u \mid t)$$

or

$$R^2_{IPE}(u \mid t; M_1, M_2) = 1 - \widehat{\mathrm{IPE}}_{M_2}(u \mid t) \Big/ \widehat{\mathrm{IPE}}_{M_1}(u \mid t).$$

5.5 Analysis of the Aortic Valve dataset

We return to the Aortic Valve dataset introduced in Section 5.1. Our aim is to use the existing data and provide accurate predictions of re-operation-free survival for future patients from the same population, utilizing their baseline information, namely age, gender, BMI and the type of operation they underwent, and their recorded aortic gradient levels. In our study, a total of 77 (27%) patients received a sub-coronary implantation (SI) and the remaining 208 patients a root replacement (RR). These patients were followed prospectively over time with annual telephone interviews and biennial standardized echocardiographic assessment of valve function until July 8, 2010. Echo examinations were scheduled at 6 months and 1 year postoperatively, and biennially thereafter, and at each examination, echocardiographic measurements of aortic gradient (mmHg) were taken. By the end of follow-up, 1262 aortic gradient measurements were recorded with an average of 4.3 measurements per patient (s.d. 2.4 measurements), 59 (20.7%) patients had died, and 73 (25.6%) patients required a re-operation on the allograft. The composite event, re-operation or death, was observed for 125 (43.9%) patients, and the corresponding Kaplan-Meier estimator for the two intervention groups is shown in Figure 5.2. We can observe minimal differences in the re-operation-free survival rates between sub-coronary implantation and root replacement, with only a slight advantage of sub-coronary implantation towards the end of the follow-up. For the longitudinal process and because aortic gradient exhibits right skewness, we will proceed in our analysis using the square root transform of this outcome. Figure 5.3 depicts the subject-specific longitudinal profiles of the square root aortic gradient for the two

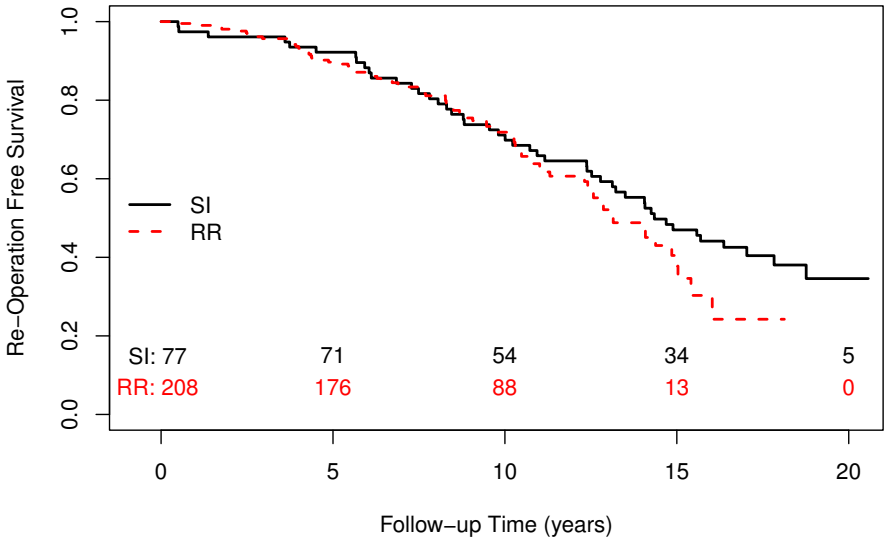


Figure 5.2: Kaplan-Meier estimates of the survival functions for re-operation-free survival for the sub-coronary implantation (SI) and root replacement (RR) groups.

intervention groups. We observe considerable variability in the shapes of these trajectories, but there are no systematic differences apparent between the two groups.

We start by defining a set of joint models based on which predictions will be calculated. For the longitudinal process we allow a flexible specification of the subject-specific square root aortic gradient trajectories using natural cubic splines of time. More specifically, the linear mixed model takes the

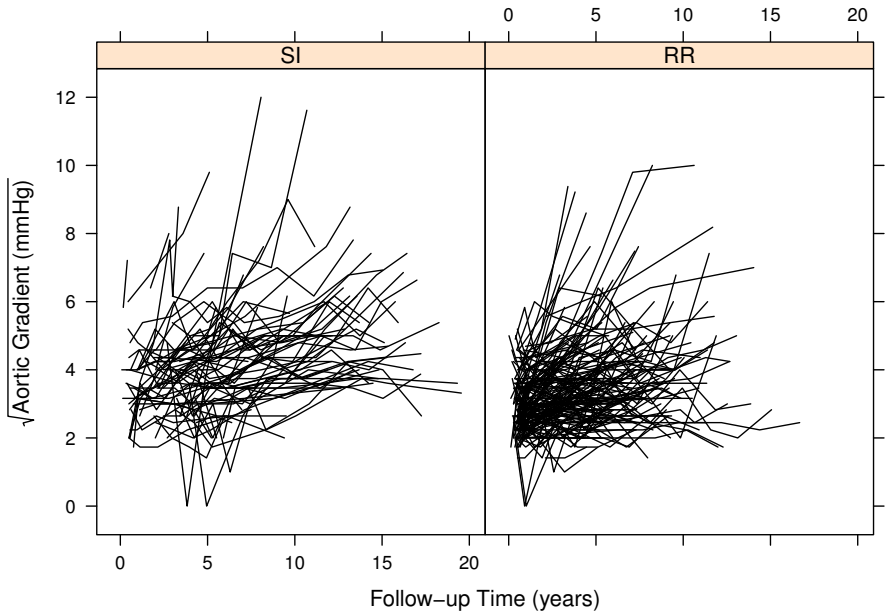


Figure 5.3: Subject-specific profiles for the square root aortic gradient separately for the sub-coronary implantation (SI) and root replacement (RR) groups.

form

$$\begin{aligned} y_i(t) \;=\; & \beta_1 \mathtt{SI}_i + \beta_2 \mathtt{RR}_i + \beta_3\{B_1(t,\lambda) \times \mathtt{SI}_i\} + \beta_4\{B_1(t,\lambda) \times \mathtt{RR}_i\} \\ & + \beta_5\{B_2(t,\lambda) \times \mathtt{SI}_i\} + \beta_6\{B_2(t,\lambda) \times \mathtt{RR}_i\} \\ & + \beta_7\{B_3(t,\lambda) \times \mathtt{SI}_i\} + \beta_8\{B_3(t,\lambda) \times \mathtt{RR}_i\} \\ & + b_{i0} + b_{i1}B_1(t,\lambda) + b_{i2}B_2(t,\lambda) + b_{i3}B_3(t,\lambda) + \varepsilon_i(t), \end{aligned}$$

where $B_n(t,\lambda)$ denotes the B-spline basis for a natural cubic spline with boundary knots at baseline and 19 years and internal knots at 2.1 and 5.5 years (i.e., the 33.3% and 66.6% percentiles of the observed follow-up times), `SI` and `RR` are the dummy variables for the sub-coronary implantation and root replacement groups, respectively, $\varepsilon_i(t) \sim \mathcal{N}(0,\sigma^2)$ and $\boldsymbol{b}_i \sim \mathcal{N}(\mathbf{0},\boldsymbol{D})$.

For the survival process we consider four relative risk models, each positing a different association structure between the two processes, namely:

$$
\begin{aligned}
M_1: \quad h_i(t) &= h_0(t)\exp\Big\{\gamma_1 \texttt{RR}_i + \gamma_2 \texttt{Age}_i + \gamma_3 \texttt{Female}_i + \gamma_4 \texttt{BMI}_i + \alpha_1 m_i(t)\Big\},\\
M_2: \quad h_i(t) &= h_0(t)\exp\Big\{\gamma_1 \texttt{RR}_i + \gamma_2 \texttt{Age}_i + \gamma_3 \texttt{Female}_i + \gamma_4 \texttt{BMI}_i + \alpha_1 m_i(t) + \alpha_2 m_i'(t)\Big\},\\
M_3: \quad h_i(t) &= h_0(t)\exp\Big\{\gamma_1 \texttt{RR}_i + \gamma_2 \texttt{Age}_i + \gamma_3 \texttt{Female}_i + \gamma_4 \texttt{BMI}_i + \alpha_1 \int_0^t m_i(s)ds\Big\},
\end{aligned}
$$

$$
M_4: \; h_i(t) = h_0(t)\exp\big(\gamma_1 \texttt{RR}_i + \gamma_2 \texttt{Age}_i + \gamma_3 \texttt{Female}_i + \gamma_4 \texttt{BMI}_i + \alpha_1 b_{i0} + \alpha_2 b_{i1} + \alpha_3 b_{i2} + \alpha_4 b_{i3}\big),
$$

where the baseline hazard is approximated with B-splines, i.e.,

$$
\log h_0(t) = \gamma_{h_0,0} + \sum_{q=1}^{Q} \gamma_{h_0,q} B_q(t, \boldsymbol{v}),
$$

with five internal knots placed at the corresponding percentiles of the observed event times, and `Female` denotes the dummy variable for females. The estimation of these models was based on a Bayesian approach and an MCMC algorithm with a single chain of 115,000 iterations from which we discarded the first 15,000 samples as burn-in. For all parameters we took standard prior distributions [89, 92]. In particular, for the vector of fixed effects of the longitudinal submodel $\boldsymbol{\beta}$, the regression parameters of the survival model $\boldsymbol{\gamma}$, the vector of spline coefficients for the baseline hazard $\boldsymbol{\gamma}_{h_0}$, and for the association parameter α we used independent univariate diffuse normal priors. For the variance of the error terms σ^2 we take an inverse-Gamma prior, while for covariance matrices we assumed an inverse Wishart prior. All computations have been performed in R (version 3.0.1) using package **JMbayes** [108] and WinBUGS (version 1.4.3) . Trace plots did not

show any alarming indications of convergence failure while auto-correlation plots showed relatively good mixing of the chains. Tables 5.1 and 5.2 show estimates and the corresponding 95% credible intervals for the parameters in the longitudinal and survival submodels , respectively. We observe that the parameter estimates in the relative risk models show greater variability between the posited association structures (in particular between the time-dependent (M_1, M_2, and M_3) and the time-independent parameterizations (M_4)) than the parameters in the linear mixed models. However, we should note that the interpretation of the regression coefficients $\boldsymbol{\gamma}$ is not the same in the four survival submodels because we condition on different components of the longitudinal process.

Table 5.1: Estimated coefficients and 95% credible intervals for the parameters of the longitudinal submodels based on the four joint models fitted to the Aortic Valve dataset.

	Value (M_1)		**Value+Slope (M_2)**		**Area (M_3)**		**Shared RE (M_4)**	
	Est.	**95% CI**	**Est.**	**95% CI**	**Est.**	**95% CI**	**Est.**	**95% CI**
`SI`	3.41	(3.055; 3.772)	3.38	(3.030; 3.718)	3.41	(3.047; 3.747)	3.38	(3.020; 3.733)
`RR`	2.86	(2.687; 3.028)	2.84	(2.671; 3.005)	2.87	(2.707; 3.038)	2.85	(2.682; 3.017)
`SI`:B-spln1	1.42	(0.885; 1.997)	1.51	(0.978; 2.048)	1.37	(0.840; 1.898)	1.59	(1.065; 2.163)
`RR`:B-spln1	1.38	(0.933; 1.815)	1.38	(0.927; 1.839)	1.42	(0.970; 1.859)	1.54	(1.096; 2.005)
`SI`:B-spln2	2.94	(1.867; 4.087)	3.19	(2.149; 4.113)	2.79	(1.874; 3.836)	3.62	(2.563; 4.745)
`RR`:B-spln2	2.57	(1.723; 3.510)	2.87	(1.978; 3.796)	2.32	(1.458; 3.158)	2.97	(2.089; 3.911)
`SI`:B-spln3	3.56	(2.585; 4.742)	3.81	(2.804; 4.787)	3.31	(2.417; 4.353)	4.44	(3.389; 5.672)
`RR`:B-spln3	2.36	(1.024; 3.736)	2.81	(1.433; 4.269)	1.94	(0.654; 3.211)	2.92	(1.574; 4.319)
σ	0.57	(0.542; 0.608)	0.58	(0.545; 0.613)	0.58	(0.545; 0.611)	0.58	(0.549; 0.617)

Table 5.2: Estimated coefficients and 95% credible intervals for the parameters of the survival submodels (parameters γ_{h_0} of the baseline hazard have been omitted) based on the four joint models fitted to the Aortic Valve dataset.

	Value (M_1)		**Value+Slope** (M_2)		**Area** (M_3)		**Shared RE** (M_4)	
	Est.	**95% CI**	**Est.**	**95% CI**	**Est.**	**95% CI**	**Est.**	**95% CI**
`RR`	0.34	(−0.040; 0.739)	0.36	(−0.056; 0.790)	0.29	(−0.104; 0.688)	−0.01	(−1.217; 1.112)
`Age`	0.01	(−0.001; 0.028)	0.02	(0.002; 0.034)	0.00	(−0.010; 0.018)	0.06	(0.022; 0.106)
`Female`	−0.15	(−0.548; 0.225)	−0.12	(−0.543; 0.275)	−0.13	(−0.509; 0.243)	−0.45	(−1.624; 0.567)
`BMI`	−0.07	(−0.130; −0.019)	−0.08	(−0.140; −0.023)	−0.06	(−0.111; −0.002)	−0.21	(−0.349; −0.095)
α_1	0.37	(0.235; 0.496)	0.28	(0.106; 0.433)	0.03	(0.006; 0.044)	−0.47	(−2.414; 1.543)
α_2			1.47	(−0.261; 3.205)			−4.14	(−7.931; −1.302)
α_3							0.97	(−0.630; 3.130)
α_4							2.53	(0.972; 4.383)

To assess the predictive ability of the four joint models and compare them with the landmark approach we consider the time interval $[t = 7.5, u = 9.5]$ years. The reason for choosing this interval is twofold. First, by time $t = 7.5$ years 75% of aortic gradient measurements have been recorded, and hence we have sufficient longitudinal information, and second, a two-year interval is considered a medically relevant time frame within which we would like to obtain accurate predictions of prognosis. For the 207 patients still at risk at 7.5 years we fitted three Cox models with corresponding association structures to the joint models defined above (except from the random effects association structure), i.e.,

$$M_5 : \; h_i(u-7.5) = h_0(u-7.5)\exp\big\{\gamma_1 \texttt{RR}_i + \gamma_2 \texttt{Age}_i + \gamma_3 \texttt{Female}_i + \gamma_4 \texttt{BMI}_i + \alpha_1 \tilde{y}_i(7.5)\big\},$$

$$M_6 : \; h_i(u-7.5) = h_0(u-7.5)\exp\big\{\gamma_1 \texttt{RR}_i + \gamma_2 \texttt{Age}_i + \gamma_3 \texttt{Female}_i + \gamma_4 \texttt{BMI}_i + \alpha_1 \tilde{y}_i(7.5) + \alpha_2 \tilde{y}'_i(7.5)\big\},$$

$$M_7 : \; h_i(u-7.5) = h_0(u-7.5)\exp\Big\{\gamma_1 \texttt{RR}_i + \gamma_2 \texttt{Age}_i + \gamma_3 \texttt{Female}_i + \gamma_4 \texttt{BMI}_i + \alpha_1 \sum_{s=0}^{7.5} y_i(s)\Delta s\Big\},$$

where $u > 7.5$, variable $\tilde{y}_i(7.5)$ denotes the last available square root aortic gradient value of each patient before year 7.5, $\tilde{y}'_i(7.5)$ denotes the slope defined from the last two available measurements, and $\sum_{0 \le s \le 7.5} y_i(s)\Delta s$ denotes the area under the step function defined from the observed square root aortic gradient measurements up to 7.5 years. The parameter estimates and confidence intervals of these Cox models are presented in Table 5.3.

We evaluate both discrimination and calibration using the predictive accu-

Table 5.3: Estimated coefficients and 95% confidence intervals for the parameters in the Cox models fitted to the patients at risk at $t = 7.5$ years.

	Value (M_5)		**Value+Slope** (M_6)		**Area** (M_7)	
	Est.	**95% CI**	**Est.**	**95% CI**	**Est.**	**95% CI**
`RR`	0.42	(−0.087; 0.930)	0.42	(−0.085; 0.927)	0.39	(−0.136; 0.907)
`Age`	−0.01	(−0.025; 0.012)	−0.01	(−0.024; 0.014)	−0.01	(−0.026; 0.011)
`Female`	−0.17	(−0.678; 0.347)	−0.16	(−0.672; 0.352)	−0.15	(−0.669; 0.363)
`BMI`	0.02	(−0.046; 0.093)	0.03	(−0.042; 0.097)	0.03	(−0.044; 0.094)
α_1	0.02	(−0.187; 0.224)	−0.01	(−0.221; 0.199)	−0.01	(−0.047; 0.031)
α_2			0.25	(−0.164; 0.669)		

racy measures presented in Section 5.4, namely $\widehat{\text{PE}}(9.5|7.5)$, $\widehat{\text{IPE}}(9.5|7.5)$, $\widehat{\text{AUC}}(9.5|7.5)$ and $\text{C}_{dyn}^{\Delta t=2}$. For the first two the absolute loss function was used, and the calculation of $\widehat{\text{C}}_{dyn}^{\Delta t=2}$ was based on the interval $[0, 15]$ years, with upper limit marking the 60% percentile of the event times distribution. The estimates of these measures are presented in Table 5.4. With respect to accuracy we observe that joint model M_4 with the shared random-effects parameterization has the smallest prediction error, followed by the other three joint models and the three Cox models using the landmark approach. This is in terms of both accuracy of prediction at year 9.5 and the weighted average of prediction errors in the interval $[7.5, 9.5]$. With respect to discriminative capability we observe that joint models M_1 and M_2 can best discriminate between patients followed by the landmark approach and the other two joint models. The overall winner could be deemed joint model M_4, which has the best accuracy and respectable discriminative capability compared to the models that offer the best discrimination. A comparison between the landmark approach and joint modeling in this particular dataset,

Table 5.4: Predictive performance measures for the Aortic Valve dataset under the four joint models and the landmark approach based on Cox models with the analogous functional forms. For $\widehat{\mathsf{PE}}(9.5|7.5)$ and $\widehat{\mathsf{IPE}}(9.5|7.5)$ the absolute loss function was used. $\widehat{\mathsf{C}}_{dyn}^{\Delta t=2}$ has been calculated in the interval $[0, 15]$ years.

	$\widehat{\mathbf{PE}}(9.5\|7.5)$	$\widehat{\mathbf{IPE}}(9.5\|7.5)$	$\widehat{\mathbf{AUC}}(9.5\|7.5)$	$\widehat{\mathbf{C}}_{dyn}^{\Delta t=2}$
M_1: JM value	0.1732	0.0904	0.6106	0.6433
M_2: JM value+slope	0.1647	0.0855	0.5958	0.6592
M_3: JM area	0.1525	0.0802	0.6090	0.5419
M_4: JM shared RE	0.1133	0.0586	0.5755	0.6201
M_5 : Cox_{LM} value	0.1888	0.1032	0.5587	0.6338
M_6 : Cox_{LM} value+slope	0.1877	0.1025	0.5300	0.6238
M_7 : Cox_{LM} area	0.1885	0.1031	0.5739	0.5930

and in particular when we compare the same parameterization (i.e., models M_1 vs. M_5, M_2 vs. M_6 and M_3 vs. M_7), reveals that the joint models perform better in terms of both accuracy and discrimination.

5.6 Simulations

5.6.1 Design

We performed a series of simulations to landmarking with joint models in the context of dynamic predictions. The design of our simulation study is motivated by the set of joint models fitted to the Aortic Valve dataset in Section 5.5. In particular, we assume 300 patients who have been followed-up for a period of 19 years, and were planned to provide longitudinal measurements at baseline and afterwards at nine random follow-up times. For the longitudinal process, and similarly to the model fitted in the Aortic Valve

dataset, we used natural cubic splines of time with two internals knots placed at 2.1 and 5.5 years, and boundary knots placed at baseline and 19 years, i.e., the form of the model is as follows

$$\begin{aligned} y_i(t) \;=\; & \beta_1 \texttt{Trt0}_i + \beta_2 \texttt{Trt1}_i + \beta_3\{B_1(t,\boldsymbol{\lambda}) \times \texttt{Trt0}_i\} + \beta_4\{B_1(t,\boldsymbol{\lambda}) \times \texttt{Trt1}_i\} \\ & + \beta_5\{B_2(t,\boldsymbol{\lambda}) \times \texttt{Trt0}_i\} + \beta_6\{B_2(t,\boldsymbol{\lambda}) \times \texttt{Trt1}_i\} \\ & + \beta_7\{B_3(t,\boldsymbol{\lambda}) \times \texttt{Trt0}_i\} + \beta_8\{B_3(t,\boldsymbol{\lambda}) \times \texttt{Trt1}_i\} \\ & + b_{i0} + b_{i1}B_1(t,\boldsymbol{\lambda}) + b_{i2}B_2(t,\boldsymbol{\lambda}) + b_{i3}B_3(t,\boldsymbol{\lambda}) + \varepsilon_i(t), \end{aligned}$$

where $B_n(t,\boldsymbol{\lambda})$ denotes the B-spline basis for a natural cubic spline with $\boldsymbol{\lambda} = (0, 2.1, 5.5, 19)$, `Trt0` and `Trt1` are the dummy variables for the two treatment groups, $\varepsilon_i(t) \sim \mathcal{N}(0,\sigma^2)$ and $\boldsymbol{b}_i \sim \mathcal{N}(\boldsymbol{0},\boldsymbol{D})$ with $\boldsymbol{D}$ taken to be diagonal.

For the survival process, we have assumed four scenarios, each one corresponding to a different functional form for the association structure between the two processes. Motivated by the arguments set forth in Section 5.2.3, we simulated survival data under the joint modeling framework (i.e., not assuming that the biomarker's levels are constant between the visit times). More specifically,

$$\begin{aligned} \text{Scenario I:} \quad & h_i(t) = h_0(t)\exp\big\{\gamma_0 + \gamma_1 \texttt{Trt1}_i + \alpha_1 m_i(t)\big\}, \\ \text{Scenario II:} \quad & h_i(t) = h_0(t)\exp\big\{\gamma_0 + \gamma_1 \texttt{Trt1}_i + \alpha_1 m_i(t) + \alpha_2 m_i'(t)\big\}, \\ \text{Scenario III:} \quad & h_i(t) = h_0(t)\exp\Big\{\gamma_0 + \gamma_1 \texttt{Trt1}_i + \alpha_1 \int_0^t m_i(s)ds\Big\}, \\ \text{Scenario IV:} \quad & h_i(t) = h_0(t)\exp\big(\gamma_0 + \gamma_1 \texttt{Trt1}_i + \alpha_1 b_{i0} + \alpha_2 b_{i1} + \alpha_3 b_{i2} + \alpha_4 b_{i3}\big), \end{aligned}$$

with $h_0(t) = \sigma_t t^{\sigma_t - 1}$, i.e., the Weibull baseline hazard. The values for the

regression coefficients in the longitudinal and survival submodels, the variance of the error terms of the mixed model, the covariance matrix for the random effects, and the scale of the Weibull baseline risk function are given in Appendix D.1, and have been chosen such that the distribution of the event times and the distribution of the follow-up longitudinal measurements were comparable across scenarios. Censoring times were simulated from a uniform distribution in $(0, t_C)$ with t_C set to result in about 45% censoring in each scenario. For each scenario we simulated 200 datasets.

5.6.2 Results

Mimicking the real-life use of a prognostic model, and to assess any potential overfitting issues, the comparison between the landmark and joint modeling approaches is based on subjects who were not used in fitting the corresponding models. More specifically, under each scenario and for each simulated dataset, we randomly excluded ten subjects whose event times were censored. For these subjects we set as landmark time the time point of their last longitudinal measurement, and produce survival probabilities from that point onwards to the end of the follow-up. Under the landmark approach these probabilities are based on the following relative risk models fitted to the remaining subjects:

$$
\begin{aligned}
LM_1: &\quad h_i(u - t_{LM}) = h_0(u - t_{LM}) \exp\big\{\gamma_0 + \gamma_1 \texttt{Trt1}_i + \alpha_1 \tilde{y}_i(t_{LM})\big\},\\
LM_2: &\quad h_i(u - t_{LM}) = h_0(u - t_{LM}) \exp\big\{\gamma_0 + \gamma_1 \texttt{Trt1}_i + \alpha_1 \tilde{y}_i(t_{LM}) + \alpha_2 \tilde{y}'_i(t_{LM})\big\},\\
LM_3: &\quad h_i(u - t_{LM}) = h_0(u - 7.5) \exp\Big\{\gamma_0 + \gamma_1 \texttt{Trt1}_i + \alpha_1 \sum_{s=0}^{t_{LM}} y_i(s)\Delta s\Big\},
\end{aligned}
$$

where t_{LM} denotes the landmark time, and as before, $\tilde{y}_i(t_{LM})$ denotes the last available measurement of subject i before t_{LM}, $\tilde{y}'_i(t_{LM})$ denotes the slope

defined from the last two available measurements, and $\sum_{s=0}^{t_{LM}} y_i(s)\Delta s$ the area under the step function defined from the observed longitudinal responses up to t_{LM}. Similarly, we also fitted four joint models to the remaining 290 subjects, with the same longitudinal submodel as the one we simulated from, and survival submodels:

$$
\begin{aligned}
JM_1: &\quad h_i(t) = h_0(t)\exp\big\{\gamma_0 + \gamma_1 \texttt{Trt1}_i + \alpha_1 m_i(t)\big\},\\
JM_2: &\quad h_i(t) = h_0(t)\exp\big\{\gamma_0 + \gamma_1 \texttt{Trt1}_i + \alpha_1 m_i(t) + \alpha_2 m_i'(t)\big\},\\
JM_3: &\quad h_i(t) = h_0(t)\exp\Big\{\gamma_0 + \gamma_1 \texttt{Trt1}_i + \alpha_1 \int_0^t m_i(s)\,ds\Big\},\\
JM_4: &\quad h_i(t) = h_0(t)\exp\big(\gamma_0 + \gamma_1 \texttt{Trt1}_i + \alpha_1 b_{i0} + \alpha_2 b_{i1} + \alpha_3 b_{i2} + \alpha_4 b_{i3}\big),
\end{aligned}
$$

based on which survival probabilities were derived. Due to the fact that our aim here is to investigate the impact on predictions of the underlying differences between landmarking and joint modeling, as explained in Section 5.2.3, in both approaches the baseline hazard is assumed of the Weibull form, i.e., $h_0(t) = \sigma_t t^{\sigma_t - 1}$ with σ_t denoting the shape parameter and the intercept term γ_0 the log scale parameter.

Based on the seven models, predictions were calculated for each of the ten subjects we have originally excluded, at ten equidistant time points between their last available longitudinal measurement and the end of follow-up. To evaluate the accuracy of these predicted survival probabilities we compared them with the gold standard survival probabilities, which are calculated as $S_j\big\{u \mid \mathcal{M}_j(u, \boldsymbol{b}_j), \boldsymbol{\theta}\big\}/S_j\big\{t \mid \mathcal{M}_j(t, \boldsymbol{b}_j), \boldsymbol{\theta}\big\}$, using the true parameter values and the true values of the random effects for the subjects we excluded. Hence, in each simulated dataset and for each of the ten subjects, we calculated root

mean squared prediction errors (RMSEs) between the gold standard survival probabilities and the predictions under the seven models. The RMSEs over all the subjects from the 200 datasets are shown in Figure 5.4. The results

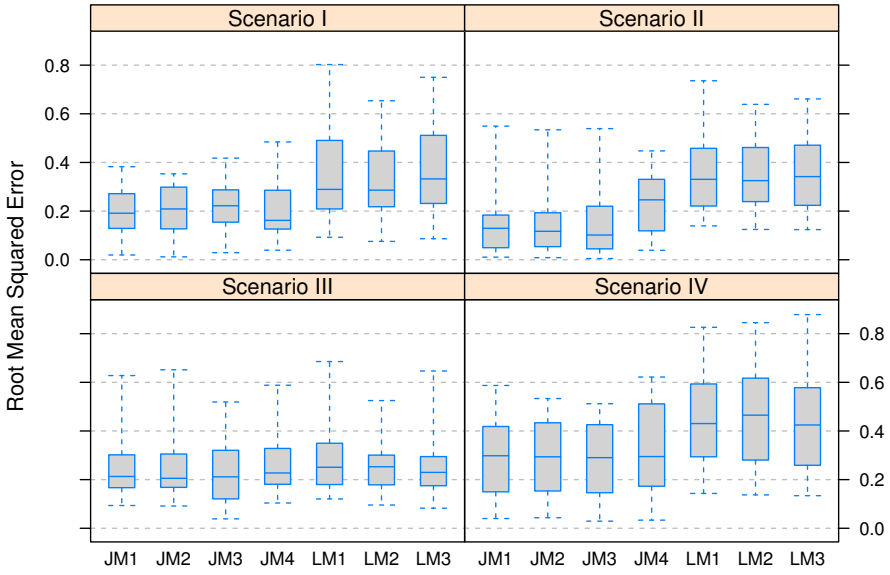


Figure 5.4: Simulation results under the four scenarios based on 200 datasets. Each boxplot shows the distribution of the root mean squared predictions error of the corresponding model to compute predictions versus the gold standard.

suggest that the joint modeling approach seems to give more accurate predictions than landmarking. More noticeable are the differences in Scenarios I, II and IV while in Scenario III both approaches gave similarly accurate results.

5.7 Discussion

In this work we have contrasted and compared two popular approaches, namely landmarking and joint modeling, for producing dynamically-updated

predictions of survival probabilities with time-dependent covariates. Landmarking can effortlessly be implemented in practice but it makes strong assumptions regarding the path of the time-dependent covariates, which may be unrealistic for longitudinal biomarker measurements. On the other hand, joint modeling allows for greater flexibility in the attributes of time-dependent covariate process, but requires more modeling assumptions to achieve this and is generally more computationally intensive. Our simulation study and the analysis of the motivating Aortic Valve dataset have shown that, in general, there is a gain from considering the joint modeling approach instead of landmarking.

In our developments we have only focused on a single continuous longitudinal biomarker. However, often in practice and in order to obtain a more complete picture of the progression of a patient, several biomarkers are recorded, which could be of either continuous or categorical nature. In this more complex setting landmarking is advantageous because it is straightforward to include extra markers as baseline covariates in the linear predictor of the Cox model fitted to the patients at risk at the landmark point. On the contrary, the joint modeling approach requires a model specification for each marker. Mathematically and under the conditional independence assumptions (5.4) and (5.5) this relatively easily achieved by considering the framework of generalized linear mixed effects models [109]. From the practical side, however, the dimensionality of the random effects may increase considerably, making joint models harder to fit. Previous and recent work by the first author is focused on resolving this problem by making use of Laplace approximations and efficient Gaussian quadrature rules [41, 47] . In addition, in our analysis of the Aortic Valve dataset we have considered the

composite event re-operation or death (whatever comes first), but for the treating physicians it could be of interest to have risk estimates separately for the two events. In this case we can extend both landmarking and joint modeling to the competing risks setting and derive estimates of the corresponding cumulative incidence functions. A general challenge when either or both of the two extensions (i.e., multiple longitudinal outcomes or multiple event times) are considered is the number of possible models. In particular, following the discussion in Section 5.3 and the different possibilities we have in building the functional relationship between the longitudinal and time-to-event outcomes, it is evident that when we move to the multivariate setting, the choice of the appropriate parameterization for each longitudinal outcome and eventually for each competing risk becomes a demanding model-selection exercise.

Regarding the software implementation of the methodology presented in the paper, the landmark approach is readily available in all statistical software that fit Cox models. The fitting of joint models, the derivation of dynamic predictions (for the survival and longitudinal outcomes) and the calculation of the calibration and discrimination measures presented in Section 5.4 are implemented in the freely available R packages **JM** [40, 41] and **JMbayes** [108], which can be downloaded from CRAN at:
`http://cran.r-project.org/package=JM` and:
`http://cran.r-project.org/package=JMbayes`, respectively.

Chapter 6

General Discussion

In this thesis we have studied several extensions of joint models for longitudinal and time-to-event data. In particular, Chapters 2 and 3 presented methods that can be used as an alternative to joint modeling approach in special settings. In Chapter 2 we proposed a two-stage procedure that can be applied in case when the longitudinal measurements are collected before the start of follow-up for survival response. The method can be applied for any type of longitudinal responses, such as continuous, ordinal, binary. In Chapter 3 we utilized multistate models and the pseudo-values approach, which allows to account for the variability in the longitudinal response when modeling survival using the whole history of this response. This has a great advantage compared to other methods for non-Markov models where the history of the process is of interest. It is also simpler and less computationally intensive than the Bayesian joint model approach presented in Chapter 4.

Dynamic predictions are becoming popular in all observational studies where patients are observed along time and there might be a need for an in-

termittent intervention. Using the freely available *shiny* package an interface for the R code was written based on the joint models presented in Chapter 4 resulting in a user-friendly tool for producing dynamic predictions. That tool can be helpful in everyday clinical practice in decision making about potential intervention. In addition, the fully Bayesian joint model of Chapter 4 allows to handle more than one survival responses, as in competing risks settings. Unlike the traditional landmark approach for dynamic predictions, presented in Chapter 5, different association structures between longitudinal and survival processes can be specified and implemented using the joint modeling approach.

6.1 Underlying Assumptions

In all joint models in this thesis the visiting and censoring processes were assumed noninformative. Nevertheless, in real data problems, we are often faced with so-called doctor's-care visiting scheme, i.e., when the physician decides about the date of next visit based on the current health status of the patient. Examinations can also take place at regular intervals, i.e., all patients are examined or observed at preassigned intervals. This situation occurs frequently in medical studies since examination times are often preassigned. Finally, we can be faced with random visiting schemes under which all subjects are examined at random times, independent of their disease history. Grüger et al. [69] showed that these three schemes are not informative when likelihood methods are used, because the likelihood given one of these examination schemes is proportional to the likelihood obtained when the examination scheme is fixed in advance.

In Chapter 3, when dealing with categorical longitudinal responses in the context of multi-state modeling we show that interval-censoring is a much more common problem than the nonignorable visiting process. This raises from the fact that even with ignorable visiting times, the exact times of change in the categorical status of a patient are often only observed for the final but not for the intermediate states. We showed that standard methods used in multi-state modeling may be problematic when dealing with interval-censoring. In contrast, using the joint modeling approach this can be easily modeled under the Missing At Random (MAR) assumption. The eventual misclassification in a categorical response could be handled similar as the measurement error problem for continuous longitudinal response through the latent formulation, both in the context of shared parameter or latent class models [110]. Therefore, under MAR, the joint modeling approach allows for ignorable visiting and censoring processes. As mentioned in Section 1, for the real data application, it implies that the two processes depend only on the observed history but neither on the event times, nor on future longitudinal measurements. In particular, for the heart transplant data, as discussed in Chapter 3, we had the ignorable doctor's-care visiting scheme since given a current state of the patient the evaluation took place every fixed number of days and the length of this period depended on the particular current state. Additionally for this data set we had to assume that the censoring was also noninformative, i.e. that it did not depend on unobserved true event times and/or future urgency status.

Recently Gueorguieva et al. [111] considered a joint model for longitudinal outcome and competing risks with cause-specific dropouts that are interval censored. Based on their likelihood formulation a subject remains at risk

of all causes of dropout until an unknown time within the interval when dropout for the given reason actually occurs. To simplify the problem it was assumed that, if a subject is interval censored for a particular reason, then she is right censored for the other competing reasons at the beginning of the interval when dropout occurred choosing the beginning of the interval. As noted by the authors, that approximation can lead to reasonable bias when the gaps between observation times are large and the majority of the data are interval censored.

MAR is closely related to another assumption in joint models, namely the conditional independence. Under that assumption the longitudinal and survival processes are independent conditional on the shared latent terms. When this is violated the dropout process is not MAR. Unfortunately, the MAR assumption is not testable using the available data, and sensitivity analysis should be performed.

Furthermore, in the joint models that we considered in this thesis we have assumed the normal distribution for the random effects. As shown by Rizopoulos et al. [87] and Huang et al. [112] , for large number of measurements per subject m_i, an eventual misspecification of the random effects distribution does not influence the results since it influences only the prior distribution. In particular, as m_i increases the role of the prior of random effect b_i in the posterior distribution of the random effects diminishes. However for categorical responses there is perhaps some concern as larger values of m_i are required.

Nevertheless, in general, we recommend, that a sensitivity analysis with respect to any aspects of the modeling step, that is the chosen parametrization, MAR and/or conditional independence assumptions as well as the dis-

tribution of the random effects, should be performed. In this thesis, for the real data applications, as presented in Chapter 4 and 5, we focused only on the impact of different model parameterizations on the results with respect to the obtained predictions.

6.2 Model Selection Problem and Goodness of Fit

As it was demonstrated in Chapter 4 the chosen parametrization for a joint model may influence predictions, mainly for the survival part but much less for the longitudinal response. Since we have many repeated measures per individual there is a lot more information in the longitudinal process than in the survival process in that setting. The simulations results presented in Chapter 4, indicate that the misspecification of the joint model omitting the time-dependent terms is most severe when the association between the survival and longitudinal process is strong. In any case, one would like to choose the best model from the set of models with different parameterizations. An alternative approach would be to average over models with different parameterizations using Bayesian model averaging [113] . Another alternative would be to use the measures proposed in Chapter 5, to choose the best model, based on the quality of the derived predictions in terms of calibration and discrimination.

Finally, when it comes to assessing modelling assumptions, a traditional way is to examine the residuals. However, in joint models residual plots can be misleading due to dropout. Patients that dropped out may have different longitudinal evolutions than patients who do not. As a result the reference

distribution of the residuals is not certain. To resolve this issue Rizopoulos [114] proposed to use multiple imputation idea to impute missing patterns using the predictive distribution $p(y_i^m \mid y_i^o, T_i, \delta_i, \theta)$ where y_i^m and y_i^o denote the missing and observed response vectors for ith subject, respectively, T_i and δ_i is her failure time and failure indicator and θ denotes the vector of model parameters. After the multiple imputation step a standard model diagnostics for mixed effects and survival models can be used as for complete data set. In particular, this procedure can be easily applied for the fully Bayesian model from Chapter 4.

6.3 Estimation Methods and Computational Issues

Under the likelihood approach in order to estimate the parameters of the shared parameter models, one needs to use numerical integration over the random effects and over the time variable. The dimension of the first integrand is more problematic since it is related to the dimension of the longitudinal response. This motivated us, in Chapter 4, to opt for a Bayesian approach that avoids the integration over the random effects space via sampling. However, even within the Bayesian framework we still have to approximate the integrand over the time variable. In our models we used the Gauss-Kronrod quadrature rule, but other rules could also be applied. This step of the estimation process is the most time-consuming, especially, in the competing risks setting when we have separate submodels for different causes of failure. As an alternative, instead of numerical integration, one can use the approximation of the integral over time followed by Ibrahim et al. [25].

This type of approximation was earlier proposed by Tsiatis, DeGruttola and Wulfson [10]. For the model with only random effects shared the integrand over time has a closed-form solution and the numerical methods are not needed.

6.4 Directions for Future Work

The majority of settings in joint modeling literature present models with a single longitudinal outcome and time-to-event. However, in practice patients are repeatedly measured for a number of outcomes that are potentially predictive for the time-to-event. The separate analysis per longitudinal outcome was shown to be less efficient than a joint analysis of all the markers simultaneously ([115], [116]). The issue of multiple correlated biomarkers has been raised by many authors within different settings. Gueorguieva and Sanacora [117] considered multivariate correlated probit models for a combination of ordinal and continuous biomarkers. Proust-Lima [38] formulated a latent class model approach for many correlated biomarkers and a binary outcome being the probability of occurrence of the clinical event according to the latent classes. Later a latent class model for the multivariate longitudinal data and time-to-event was proposed [39]. Bayesian joint models for multivariate longitudinal and survival data have been discussed by several authors ([26], [25], [27], [28]). In practice extending the joint model from single to multiple continuous longitudinal outcome creates similar problems as considering more categories in a model with a categorical outcome. In the context of the proposed Bayesian joint model, in Chapter 4, this leads to highly time- and memory consuming estimation that often suffers from

convergence problems. Therefore, it is clear that we need alternative, less computationally intensive methods that could be used in real data problems. Such an alternative could be the conditional score approach proposed by Tsiatis and Davidian [20] and extended by Song et al. [21] for the multivariate longitudinal data. This method is based on estimating equations and makes no distributional assumption on the underlying random effects, treating them as "nuisance". As noted by authors, especially for the multivariate longitudinal data, the cognitional score approach reduces considerably the computational complexity compared to likelihood or Bayesian approaches.

With respect to the survival outcome a multivariate extension is to consider multiple failure times per subject, such as recurrent event. Joint models with recurrent events processes have been discussed by Liu and Huang [118] and Rizopoulos [9]. However, this type of models require an additional submodel for the recurrent events, increasing the computational complexity.

As noted in Section 6.2 the issue of model selection in joint modeling is still under investigation and up to now not many solutions have been proposed. When Bayesian methods are used for the estimation, a DIC or other Bayesian criteria could be considered. However, due to the well-known limitations of such criteria, future work could focus on developing more general measures that would allow to choose the best model based on the quality of the produced predictions in terms of calibration and discrimination, regardless the estimation method. In particular, discrimination measures that could be applied in a competing risk setting using joint models are of a special interest. In the context of time-dependent ROC curves Heagerty et al. [119] proposed several definitions of cases and controls. Saha and Heagerty [120] and Zheng et al. [121] extended this definition for the competing risks setting.

Depending on the particular setting we could consider different methods of classifying subjects and use similar sampling procedure as Rizopoulos [86] to estimate ROC in the joint modeling framework. This extension could be applied to the fully Bayesian model from Chapter 4.

Chapter 7

Summary/Samenvatting

Summary

Many medical studies involve analyzing longitudinal responses together with event history data collected for each patient. A well-known and broadly studied example can be found in AIDS research, where CD4 cell counts taken at different time points are related to the time to death. In mainly two occasions such data need to be jointly analyzed in order to properly account for their association. First, when focus is on the longitudinal outcome, events cause nonrandom dropout that needs to be accounted for in order to obtain valid inferences. When focus is on the event times, the longitudinal responses cannot be simply included in a relative risk model because they represent the output of an internal time-dependent covariate [1].

In the framework of joint models, postulated by Faucett and Thomas [2] and Wulfson and Tsiatis [3], the longitudinal responses are considered realizations of an endogenous time-dependent covariate (Kabfleisch and Pren-

tice [1]), which is measured with error and for which we do not have the complete history of past values available. To account for these features we estimate the joint distribution of the survival and longitudinal processes.

Our research extend this standard approach in joint modeling in several ways. In particular, in Chapter 2 we present a two-stage procedure that can be used as an alternative to joint modeling approach in case when the longitudinal measurements are collected before the start of follow-up for survival response. This setting is often encountered in transplantation studies, where patients provide a series of longitudinal outcomes that are related to events occurring after transplantation. In contrast with the standard joint modeling setting, the longitudinal responses do not constitute an endogenous time dependent variable measured at the same period as the time to event. Nevertheless, the problem of measurement error still remains. The proposed two-stage procedure handles the problem of measurement error via Monte Carlo sampling from the posterior distribution of the random effects. We apply this approach for nonlinear longitudinal response and compare the results with the “naive" plug-in approach when the uncertainty about the estimates from the first step is not taken into account, as well as with the full Bayesian approach.

In Chapter 3 we consider categorical longitudinal responses in the presence of competing risks. We show how this problem can be handled using multi-state models techniques. In particular we use the pseudo-values approach introduced by Andersen et al. [45] and apply it for the Aalen-Johansen estimator of the state occupation probabilities since the transition probabilities were found to depend on the history. This approach allows to study the impact of baseline information on the occupation probabilities treating the

dependence on the history as a nuisance. To address the problem of those competing events a multinomial approach is used for the next state given the previous state observed. This has a great advantage compared to other methods for non-Markov models where the history of the process is of interest and no standard approaches are available. In Chapter 4 we formulate the problem in the joint modeling framework and propose a Bayesian model for joint modeling of categorical longitudinal data and time-to-event response taking into account the presence of competing risks.

The majority of prognostic models in the medical literature utilize only a small fraction of the available biomarker information not taking into account that the rate of change in the biomarker levels is not only different from patient to patient but also dynamically changes over time for the same patient. In Chapter 4, we present how the joint modeling approach can be used for updating the patients' predictions in the presence of competing risks. In particular, for the developed Bayesian joint model we derive posterior predictive distributions for the longitudinal and event time outcomes. Additionally, we also examine the impact of different parameterizations of the joint model on the obtained predictions. In contrary with the multi-state approach from Chapter 3 the interval-censoring problem is handled under the Missing At Random (MAR) assumption.

In Chapter 5 we compare the joint modeling technique for making dynamic predictions with an older method for producing such predictions, called landmarking . We show how survival probabilities are obtained under each method and what the differences are in the underlying assumptions. In addition, we show how the functional relationship between the two processes may affect predictions. To assess the quality of the derived predictions

from the two approaches we present different measures of discrimination and calibration , suitably adjusted to the context of longitudinal biomarkers.

Samenvatting

In vele medische studies worden longitudinale- en overlevingsgegevens verzameld voor elke patiënt. Een gekend voorbeeld van zulke data is te vinden in het AIDS onderzoek. Hierbij wordt het aantal CD4-cellen op verschillende tijdstippen gemeten dewelke gerelateerd zijn aan het tijdstip van overlijden. Voor zulke data zijn er in het algemeen twee types van vraagstellingen. Ten eerste, wanneer men geïnteresseerd is in de evolutie van de longitudinal metingen, kan men een (veralgemeend) lineair model gebruiken. Echter, dit model onderstelt dat de uitval gebeurt volgens een specifiek missing data patroon, genaamd "missing at random" en een uitbreiding van dit model dat rekening houdt met het "nonrandom dropout" gedrag van de uitvallers. Het gezamenlijk modelleren van de longitudinale- en overlevingsgegevens laat een meer complex missing data mechanisme toe. Wanneer de interesse ligt in het overlevingsproces, levert het gezamenlijk modelleren betere resultaten op dan de klassieke techniek met tijdsafhankelijke covariaten [1].

In de standaard aanpak van gemengde modellen door Faucett & Thomas [2] en Wulfsohn & Tsiatis [3]) worden de longitudinale data gezien als realisaties van een endogeen tijdsafhankelijke covariaat ([1]) gemeten met fout en waarvoor de historische gegevens maar gedeeltelijk gekend zijn. Met deze techniek wordt de gezamenlijke verdeling van de overleving en longitudinale processen bepaald. Dit wordt behandeld in hoofdstuk 1. In dit hoofdstuk geven we ook een beschrijving van de datasets die ons onderzoek hebben gemotiveerd. Tot slot worden in dit hoofdstuk de doelstellingen van dit proefschrift uiteengezet.

In deze thesis breiden we de bovenstaande standaard aanpak uit op ver-

schillende manieren. In Hoofdstuk 2 presenteren we een twee-fasen procedure als een alternatief voor de hierboven vermeldde gemengde modellen benadering in het geval dat longitudinale metingen voorafgaand aan de eigenlijke longitudinale studie werden verzameld. Dit gebeurt vaak bij transplantatie studies, waarbij longitudinale metingen van de patiënten gekend zijn dewelke kunnen gerelateerd zijn aan verwikkelingen na transplantatie. In tegenstelling tot de standaard gemengde modellen benadering, is de longitudinale respons nu geen endogene tijdafhankelijke variabele en wordt deze niet op hetzelfde tijdstip gemeten als het gebeuren. Echter, ook nu moeten we rekening houden met meetfouten. Deze worden in de twee-fasen procedure in rekening gebracht via Monte Carlo sampling van de a posteriori verdeling van de random effecten. Deze techniek hebben we toegepast op niet-lineaire longitudinale data. We vergelijken onze techniek met de " naïeve " benadering waarbij geen rekening wordt gehouden met de inherente meetfout. Verder vergelijken we onze benadering ook met een pure Bayesiaanse aanpak.

In hoofdstuk 3 hebben we categorische longitudinale data geanalyseerd in combinatie met verscheidene oorzaken voor falen (competing risks). We tonen aan hoe dit probleem met behulp van modellen voor meerdere stadia (multi-state modellen) kan aangepakt worden. Hiervoor gebruikten we pseudo - observaties, zoals voorgesteld door Andersenet al. [45]. Deze techniek hebben we toegepast op de Aalen-Johansen schatter van de kans dat een individu zich in een bepaald stadium bevindt. Deze schatter laat immers toe dat de transitiekansen (kans dat men verandert van stadium) afhankelijk zijn van voorafgaandelijke gebeurtenissen. Deze techniek laat ook toe om het effect van start informatie te gebruiken daarbij abstractie makende van de gebeurtenissen in de tijd. Een multinomiaal model werd gebruikt om de con-

ditionele kans te modelleren om naar een volgend stadium over te stappen in de aanwezigheid van meervoudig falen. Deze aanpak heeft voordelen tov andere niet-Markov modellen waarbij men ge ïnteresseerd is in de geschiedenis van het proces en waarvoor geen standaard benaderingen beschikbaar zijn. In hoofdstuk 4 gebruiken we een Bayesiaanse gemengd model voor categorische longitudinale data gecombineerd met overlevingstijden daarbij rekening houdend met mogelijk meervoudig falen.

De meerderheid van prognostische modellen in de medische literatuur gebruiken slechts een klein deel van de beschikbare biomarkers en houden geen rekening met hun longitudinale evolutie tussen en binnen pati ënten. In hoofdstuk 4 illustreren we hoe de gemengde modellen gebruikt kunnen worden om voorspellingen te doen in de context van meervoudig falen. In de context van het Bayesiaans gemengd model leiden we een posterior predictieve verdelingen af voor de longitudinale response en voor het tijdstip van het falen. Daarnaast onderzochten we ook de invloed van de verschillende parameterisaties van de gemengde modellen op de voorspellingen. In tegenstelling met de multi- state benadering van hoofdstuk 3, nemen we nu het interval-gecensureerd karakter van de response in rekening onder de Missing At Random (MAR) aanname.

In hoofdstuk 5 vergelijken we de dynamische voorspelling van onze gemengde modellen benadering met een concurrerende techniek, in het Engels genaamd 'landmarking'. We tonen hoe overlevingskansen verkregen worden onder beide methoden en wat de verschillen in de onderliggende aannames zijn. Verder tonen we hoe de functionele relatie tussen de twee processen de voorspellingen kan be ï nvloeden. Om de kwaliteit van de afgeleide voorspellingen van de twee benaderingen te beoordelen, stellen we verschillende aangepaste

maten (aangepast aan de longitudinale biomarkers) voor discriminatie voor.

Appendix A

This appendix is contains supplementary material for the paper presented in Chapter 2).

A.1 Renal Resistance Data. Descriptives

The individual profiles for different types of donor and different regions are presented in Figure A.2. It can be observed that there is some variability in the RR level at time zero, in asymptotes and also in the "slopes". After putting the kidney into the machine there is a rise of RR level for both H-B and N-H-B. This later stabilizes and appears to be almost constant. The same behavior is visible for the three donor regions. The mean profiles for different donor types and regions are presented in Figure A.1. N-H-B have higher RR level as compared to the N-H-B donors. Donors from different regions have more less similar mean RR profiles with a bit higher RR mean initial value for Region 2 and the lowest asymptote for Region 1. There were not N-H-B donors from Region 2 present in the data set. Figure A.3 presents the results from a nonlinear mixed model for the renal data.

A.2 Metropolis-Hastings algorithm

Metropolis-Hastings algorithm is a Markov chain Monte Carlo method for obtaining a sequence of random samples from a probability distribution for which direct sampling is not straightforward. The M-H algorithm is defined by two steps: a first step in which a proposal value is drawn from the candidate generating density and a second step in which the proposal value is accepted as the next iterate in the Markov chain according to the defined probability or rejected and then the next sampled value is taken to be the current value.

In order to sample form the posterior distribution for random effects for a particular individual i according to the algorithm described in Section 3.2 of the article in each step k we propose the density q for $\boldsymbol{\alpha}_i$. The proposal density was chosen to be a multivariate t distribution with 4 df, mean equal the Empirical Bayes estimate obtained from the nonlinear mixed model and variance-covariance matrix $\boldsymbol{D}$ also estimated from the nonlinear mixed model, additionally scaled by some parameter $Scale$. Before the start of the analysis the tuning parameter $Scale$ was calibrated in order to achieve the acceptance rate equal 0.5. We run 300 iterations for the calibration.
The procedure can be described as below:
Step 0 :

$$\boldsymbol{\alpha}_i^0 = EB(\boldsymbol{\alpha}_i)$$

Step K:

$$\boldsymbol{\alpha}_i^* \sim q(EB(\boldsymbol{\alpha}_i), Scale * \boldsymbol{D})(\text{proposition for } \boldsymbol{\alpha}_i)$$

Calculate acceptance-rejection criterion:

$$r_i = \frac{p(\boldsymbol{\alpha}_i^* \mid \boldsymbol{Y}_i, \boldsymbol{\theta}_y) q(\boldsymbol{\alpha}_i^{k-1}, Scale * \boldsymbol{D})}{p(\boldsymbol{\alpha}_i^{k-1} \mid \boldsymbol{Y}_i, \boldsymbol{\theta}_y) q(\boldsymbol{\alpha}_i^*, Scale * \boldsymbol{D})},$$

where $\boldsymbol{\theta}_y$ is the vector of fixed effects sampled using the estimates from the nonlinear mixed model.

Since the posterior distribution $p(\boldsymbol{\alpha}_i \mid \boldsymbol{Y}_i, \boldsymbol{\theta}_y)$ is unknown, we use the fact that:

$$p(\boldsymbol{\alpha}_i \mid \boldsymbol{Y}_i, \boldsymbol{\theta}_y) \propto p(\boldsymbol{Y}_i \mid \boldsymbol{\alpha}_i, \boldsymbol{\theta}_y) p(\boldsymbol{\alpha}_i),$$

and therefore the acceptance-rejection criterion takes the form:

$$r_i = \frac{p(\boldsymbol{Y}_i \mid \boldsymbol{\alpha}_i^*, \boldsymbol{\theta}_y) p(\boldsymbol{\alpha}_i^*) q(\boldsymbol{\alpha}_i^{k-1}, Scale * \boldsymbol{D})}{p(\boldsymbol{Y}_i \mid \boldsymbol{\alpha}_i^{k-1}, \boldsymbol{\theta}_y) p(\boldsymbol{\alpha}_i^{k-1}) q(\boldsymbol{\alpha}_i^*, Scale * \boldsymbol{D})}.$$

In above criterion $p(\boldsymbol{Y}_i \mid \boldsymbol{\alpha}_i, \boldsymbol{\theta}_y)$ is a Gaussian density for individual i with a mean being a nonlinear function of random and fixed effects $f(\boldsymbol{\alpha}_i, \boldsymbol{\theta}_y)$ and variance σ^2 estimated from the nonlinear mixed model. $p(\boldsymbol{\alpha}_i)$ is the multivariate normal distribution with mean zero and a scaled variance-covariance matrix $\boldsymbol{D}$.

Draw from uniform distribution:

$$u \sim U(0, 1).$$

If $r_i \leq u$ (accept proposition $\boldsymbol{\alpha}_i^*$ for $\boldsymbol{\alpha}_i$):

$$\boldsymbol{\alpha}_i^k = \boldsymbol{\alpha}_i^*,$$

otherwise reject $\boldsymbol{\alpha}_i^*$:

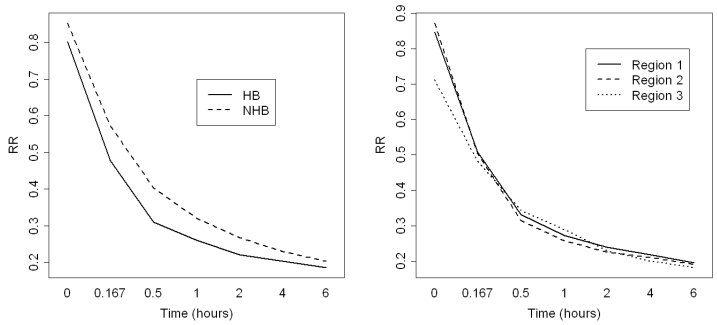
$$\boldsymbol{\alpha}_i^k = \boldsymbol{\alpha}_i^{k-1}.$$

Since we always propose the q density around the EB estimates for $\boldsymbol{\alpha}_i$ this procedure is an independence version of Metropolis-Hastings algorithm.

A.3 Tables and Figures

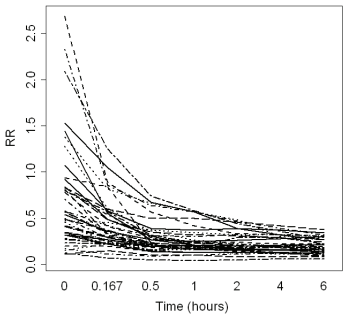
Table A.1: *Parameter estimates, standard errors and 95% credibility intervals from the longitudinal part of the joint fully Bayesian model with Weibull survival submodel*

		RR		
Effect	**Parameter**	**Estimate**	**SE**	**(95%HPD)**
ϕ_1				
`Constant`	β_{10}	2.862	0.143	(2.582; 3.142)
`Donor Age`	β_{11}	0.011	0.004	(0.003; 0.018)
`Donor Type (HB vs NHB)`	β_{12}	-0.106	0.118	(-0.337; 0.125)
`Donor Region 1 vs 3`	β_{13}	-0.09	0.096	(-0.278; 0.098)
`Donor Region 2 vs 3`	β_{14}	-0.08	0.100	(-0.276; 0.116)
ϕ_2				
`Constant`	β_{20}	3.540	0.341	(2.872; 4.208)
`Donor Age`	β_{21}	0.004	0.008	(-0.011; 0.020)
`Donor Type (HB vs NHB)`	β_{22}	-0.080	0.196	(-0.464; 0.304)
`Donor Region 1 vs 3`	β_{23}	-0.094	0.314	(-0.709; 0.521)
`Donor Region 2 vs 3`	β_{24}	0.045	0.216	(-0.378; 0.468)
ϕ_3				
`Constant`	β_{30}	1.335	0.257	(0.831; 1.839)
`Donor Age`	β_{31}	0.009	0.009	(-0.009; 0.027)
`Donor Type (HB vs NHB)`	β_{32}	0.540	0.133	(0.279; 0.801)
`Donor Region 1 vs 3`	β_{33}	-0.234	0.127	(-0.483; 0.015)
`Donor Region 2 vs 3`	β_{34}	-0.070	0.147	(-0.358; 0.218)

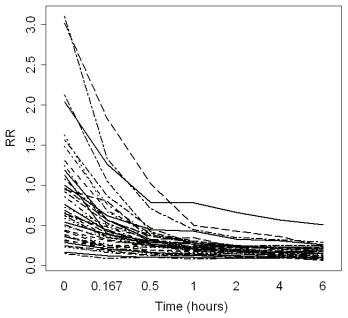


(a) (b)

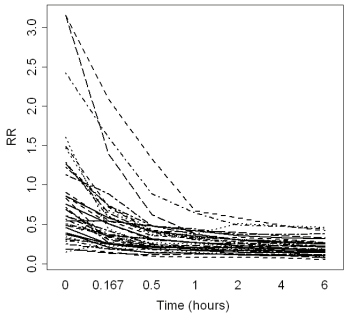
Figure A.1: *Mean profiles of Renal Resistance for the two type of donors: Heart-Beating (H-B) and Non-Heart-Beating (N-H-B) a) and 3 donor regions b)*



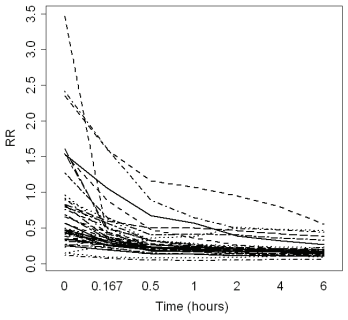
(a) *Region 1*



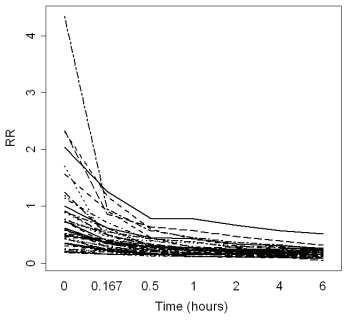
(b) *Region 2*



(c) *Region 3*

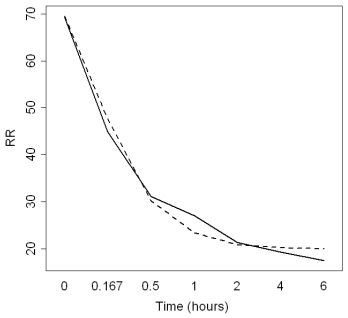


(d) *Heart Beating*

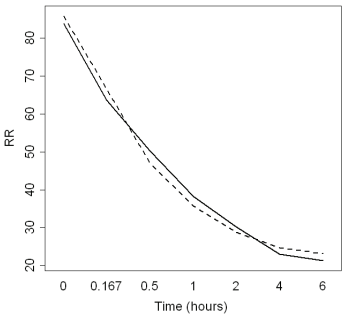


(e) *Non Heart Beating*

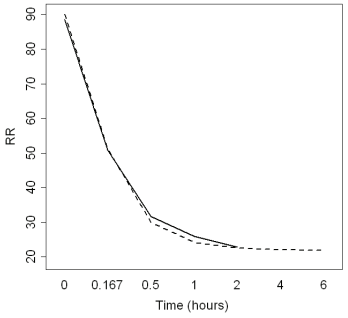
Figure A.2: *Individual profiles of renal resistance depending on donor region and donor type (50 sampled ind. in each subgroup)*



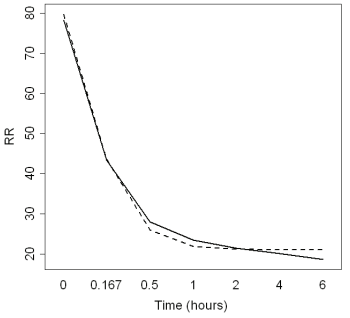
(a) *HB Region 1*



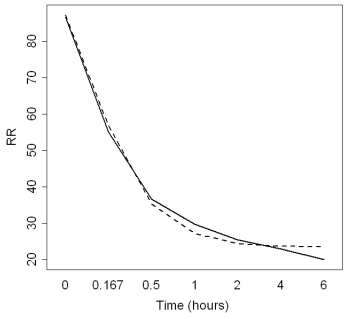
(b) *NHB Region 1*



(c) *HB Region 2*



(d) *HB Region 3*



(e) *NHB Region 3*

Figure A.3: *Mean original (solid) and fitted (dashed) RR profiles for both donor types and all donor regions*

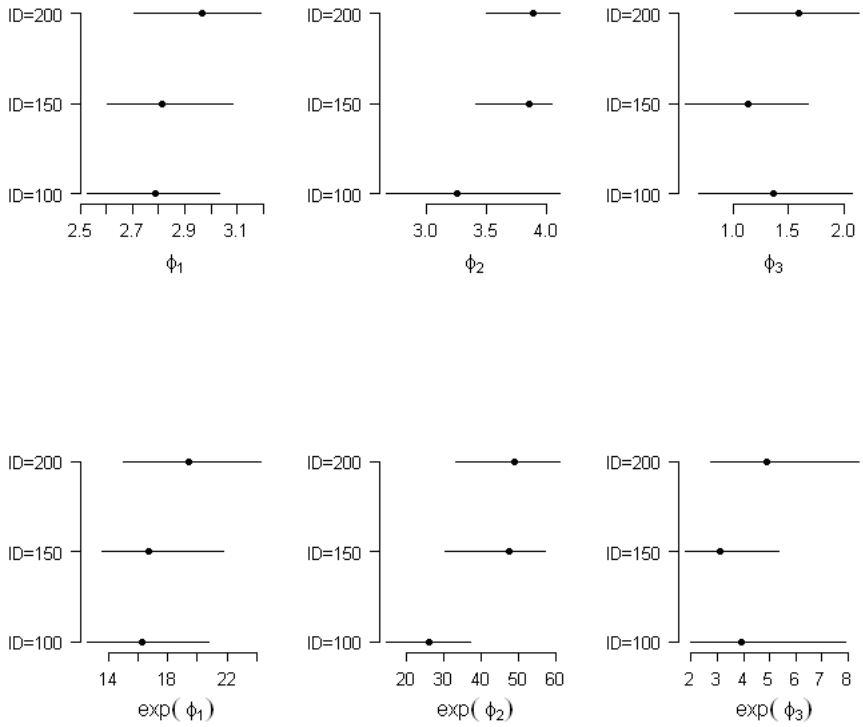


Figure A.4: *RR Parameter estimates from the nonlinear mixed model together with 25 % and 95% quartiles calculated using Monte Carlo method for arbitrary chosen individuals.*

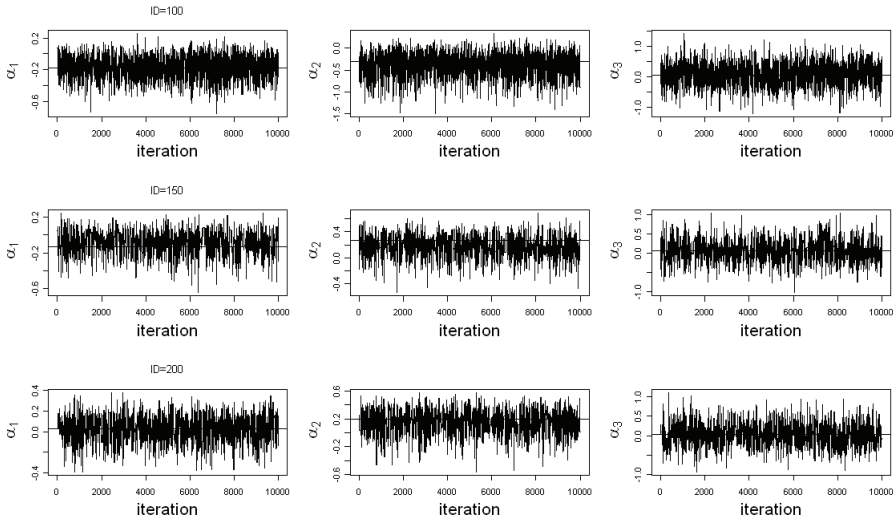


Figure A.5: *Random effects estimates together with EB (horizontal lines) estimates obtained from Monte Carlo sampling procedure for arbitrary chosen individuals.*

Appendix B

This appendix is contains supplementary material for the paper presented in Chapter 3).

B.1 Generalized Estimating Equations (GEE)

In GEE (Generalized Estimating Equations) a generalized linear model is considered:

$$g(\hat{\theta}_i) = \beta^T Z_i,$$

where g is a link function and i is individual index.

Typically a column Z_{i0} to Z_i is added to allow for intercept β_0. The estimates of β are based on unbiased estimating equations:

$$\sum_i \{(\frac{d}{d\beta} g^{-1}(\beta^T Z_i)\}^T V^{-1} \{\hat{\theta}_i - g(\beta^T Z_i)\} = \sum_i U_i(\beta) = U(\beta) = 0, \quad \text{(B.1)}$$

where V_{-1} is a working covariance matrix.

A sandwich estimator is used to estimate the variance of β.

Let

$$I(\hat{\beta}) = \sum_i \{\frac{d}{d\beta} g^{-1}(\beta^T Z_i)\}^T V^{-1} \{\frac{d}{d\beta} g^{-1}(\beta^T Z_i)\},$$
$$\text{and } \hat{Var}(\hat{\beta}) = \sum_i U_i(\hat{\beta})^T U_i(\hat{\beta}). \tag{B.2}$$

Then

$$\hat{Var}(\hat{\beta}) \approx I(\hat{\beta})^{-1} \hat{Var}(U(\hat{\beta})) I(\hat{\beta})^{-1}. \tag{B.3}$$

The estimators of β can be shown to be asymptotically normal [80] and the sandwich estimator converges in probability to the true variance.

B.2 Multi-state models

B.2.1 General framework

A multi-state process is defined as a stochastic process with a finite state space $K = \{1, 2, \ldots, N\}$ in time interval $\Gamma = [0, \tau], \tau < \infty$. For any time t, $X(t)$ denotes the state occupied at that time. The process is fully characterized by the transition probabilities between states h and r, defined as:

$$p_{hr}(s,t) = \Pr(X(t) = r \mid X(s) = h, H_{s^-}), h, r \in K; s, t \in \Gamma; s \leq t, \tag{B.4}$$

where $H_{s^-} = X(t), 0 \leq t < s$, denotes the history of the process before time s.

A multistate process can be also characterized through the transition intensities:

$$q_{hr}(s) = \lim_{\Delta s \to 0} \frac{p_{hr}(s, s + \Delta s)}{\Delta s}, \tag{B.5}$$

where q_{hr} is interpreted as the instantaneous hazard of progression from state h to state r, conditionally on occupying state h. These intensities are analogous to the standard hazard function in the Cox model. Both p_{hr} and q_{hr} in principle may depend on the history H_{s^-} and time. In Markov models we assume that the dependence on the history is only through the current state, and in the special case of time homogenous Markov models the transition probabilities/intensities are assumed constant in time. Another class of models are the semi-Markov models where the transition intensity q_{hr} is allowed to depend on the duration in state h. A model where there is only duration dependence is called homogenous semi-Markov model.

When the Markov assumption is violated, it may be difficult to obtain unbiased estimates of the transition probabilities. In that case the state occupation probabilities $p_h(t)$ are often considered, that express the probability of occupying state h at time t:

$$p_h(t) = \sum_{k=1}^{N} p_k(0) p_{kh}(0, t), \tag{B.6}$$

where $p_k(0)$ is initial distribution at time 0. The corresponding Aalen-Johansen estimates of occupation probabilities are unbiased even when the Markov assumption is violated.

B.2.2 Likelihood for multi-state Markov models

Let $i = 1, 2, \ldots, M$ denote the individuals. The data for individual i consist of a series of times $(t_{i1}, \ldots, t_{iM})$ and corresponding states $(X(t_{i1}), \ldots, X(t_{iM}))$. Consider a model where we observe a pair of successive observed states $X(t_j)$ and $X(t_{j+1})$ at times t_j and t_{j+1}. Then the contribution to the likelihood from this pair of states is:

$$L_{ij} = p_{X(t_j)X(t_{j+1})}(t_{j+1} - t_j). \tag{B.7}$$

This is entry of transition matrix $P(t)$ at $X(t_j)th$ row and $X(t_{j+1})th$ column, evaluated at time $t = t_{j+1} - t_j$.

For exact transition times the likelihood contributions are:

$$L_{ij} = p_{X(t_j)X(t_{j+1})}(t_{j+1} - t_j)q_{X(t_j)X(t_{j+1})}, \tag{B.8}$$

since the state is assumed to be $X(t_j)$ thorough the interval between t_j and t_{j+1} with a known transition to state $X(t_{j+1})$ at t_{j+1}.

Transition times to final states are always exact. Let D-death state. If $X(t_{j+1}) = D$ then the contribution to likelihood is summed over the

unknown states k on the day before death:

$$L_{ij} = \sum_{k \neq D} p_{X(t_j)k}(t_{j+1} - t_j) q_{kD}, \tag{B.9}$$

The sum is taken over all possible states k which can be visited between $X(t_j)$ and D.

B.2.3 Parametric approach for Markov models

For a general multi-state Markov model the transition probability matrix P is the unique solution to $P(s, s) = I$, and the Kolmogorov forward differential equations:

$$\frac{d}{dt} p_{hr}(s, t) = \sum_j p_{hj}(s, t) q_{jr}(t), \tag{B.10}$$

where q_{jr} are the elements of the transition-intensity matrix Q. If Q is constant over the interval (s, t), as in time homogenous Markov models, then $P(s, t) = P(t - s) = P(d)$, and the Kolmogorov equations (B.10) can be solved by the matrix of exponential of Q scaled by the time interval d:

$$P(d) = \exp(dQ). \tag{B.11}$$

For simpler models analytic expressions for each element of $P(d)$ can be calculated in terms of Q. The solution is based on an eigensystem decomposition of the intensity matrix Q.

Since for the Markov processes homogenous in time the transitions probabilities can be derived from the transition intensities, one can use standard hazard-based models to model the intensities using baseline covariates.

These may include both multiplicative hazard models and additive hazard models. The most popular is the multiplicative semi-parametric Cox regression model, which we use to model the intensity of transition from a given state h to any other state r:

$$q_{ihr}(t) = q_{0hr}(t) \exp(\beta_{hr}^T Z_i). \tag{B.12}$$

Here $q_{0hr}(t)$ denotes the baseline hazard function of time t and β is the regression coefficients vector for the covariates vector Z_i for each individual $i, i = 1, \ldots, M$. In a Cox model $q_{0hr}(t)$ we considered a fully parametric model, with for common piecewise constant hazard model where for chosen cut-points $0 = \tau_0 < \tau_1 < \ldots < \tau_L = \tau$ the hazard function $q_{0hr}(t) = q_0(t)$ is assumed to be constant, $q_0 = q_l, \tau_{l-1} \leq t < \tau_l, l = 1, \ldots, L$.

(a)

$from/to$	D	NT	HU	U	T	R	TT
D	0	0	0	0	0	0	0
NT	175	103	14	**1**	618	149	0
HU	74	166	11	203	187	**6**	915
U	**5**	12	293	1	106	**1**	46
T	273	887	1249	262	2809	83	605
R	0	0	0	0	0	0	0
TT	0	0	0	0	0	0	0

(b)

$from/to$	D	NT	HU	U	T	R	TT
D	1	0	0	0	0	0	0
NT	$1.02 * 10^{-3}$	0.965	$7 * 10^{-3}$	$3 * 10^{-4}$	0.025	$3.18 * 10^{-4}$	$1.54 * 10^{5}$
HU	$1.06 * 10^{-5}$	0.02	0.908	0.067	0.002	$3.32 * 10^{-6}$	$1.57 * 10^{-3}$
U	$1.52 * 10^{-6}$	0.003	0.188	0.774	0.031	$7.32 * 10^{-7}$	$2.4 * 10^{-3}$
T	$7.97 * 10^{-6}$	0.015	0.016	0.006	0.961	$2.2 * 10^{-5}$	$7 * 10^{-4}$
R	0	0	0	0	0	1	0
TT	0	0	0	0	0	0	1

Figure B.1: *Matrix of the observed transitions a) and estimates from Markov model b). Elements in bold are constrained. Diagonal elements cannot be constrained (intensities for diagonal elements are estimated as a sum of all remaining elements in the row).*

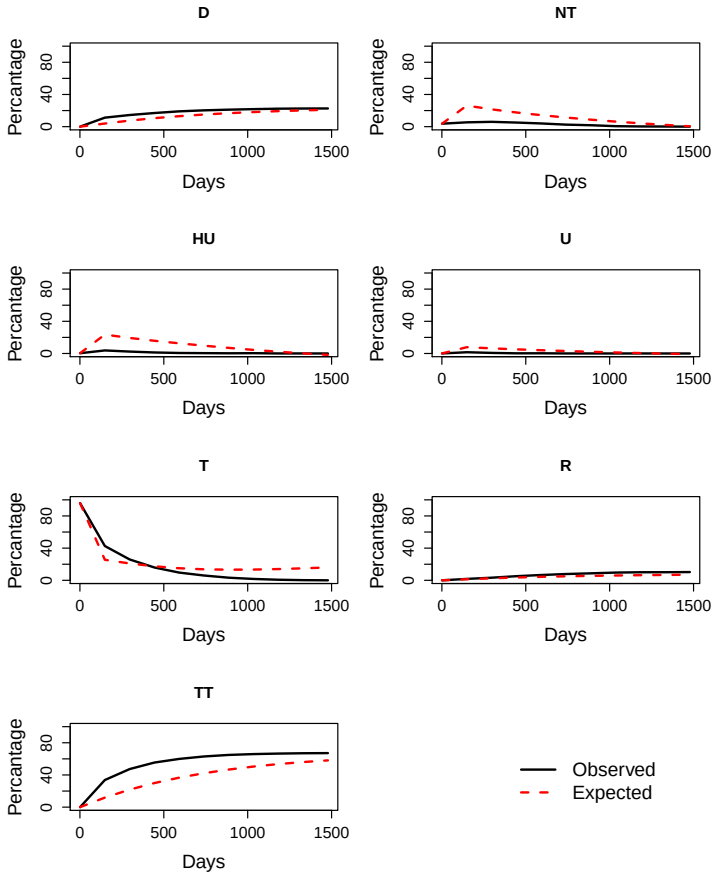


Figure B.2: *Expected (dashed) and observed (solid) prevalence assuming Markov model for each state.*

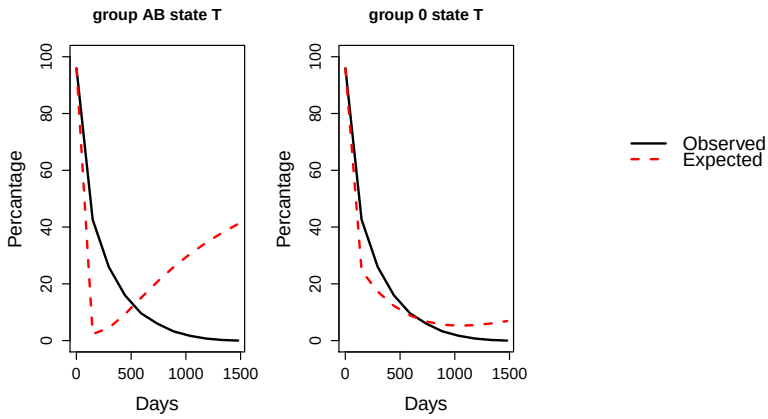


Figure B.3: *Expected (dashed) and observed (solid) prevalence for state T assuming Markov model after adjustment for blood group.*

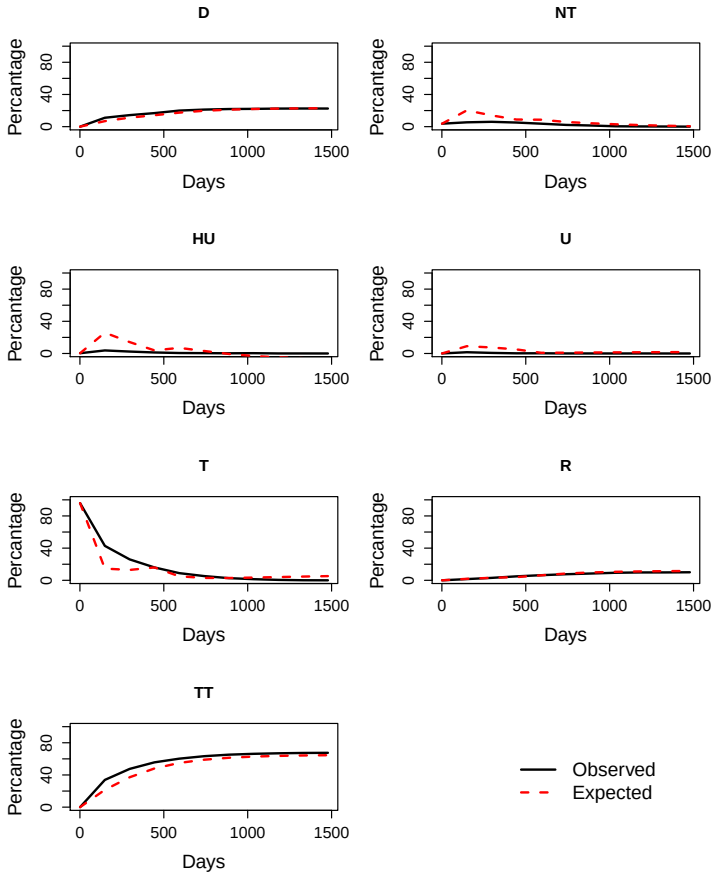


Figure B.4: *Observed and expected prevalence assuming PCI Markov Model for each state with cut point t=500 days.*

B.2.4 Non-parametric approach for Markov Models

We will now consider a multi-state Markov process defined in Section 3.2.1 for which the transition times are assumed to be observed exactly. In addition let N_{hri} denotes a multivariate counting process for individual i, $(i = 1, 2, \ldots, M)$ that counts the number of direct transitions from h to r observed for that subject in the time interval [0,t]. Let also Y_{hi} denote the indicator variable taking the value 1 when individual i is in state h at time $t-$, and 0 otherwise. A nonparametric estimator for the transition probability matrix $P(s,t)$ (B.10) can be obtained by employing the Nelson-Aalen estimator. In particular, the cumulative transition intensity:

$$A_{hr}(t) = \int_0^t q_{hr}(u)du. \tag{B.13}$$

is estimated by:

$$\hat{A}_{hr}(t) = \int_0^t \frac{dN_{hr}(u)}{Y_h(u)}, \tag{B.14}$$

where $N_{hr} = \sum_i N_{hri}$ and $Y_h = \sum_i Y_{hi}$.

Let $A(t)$ denote the (absolutely continuous) cumulative transition intensity matrix. One can show that the solution of the Kolmogorov Equations (B.10) is the matrix product-integral:

$$P(s,t) = \prod_{(s,t]} (I + q(u)du), \tag{B.15}$$

defined as:

$$\lim_{\max|s_i - s_{i-1}|\to 0} \prod (I + A(s_i) - A(s_{i-1}))), \tag{B.16}$$

where $s = s_1 < s_2 < \ldots < s_{i-1} < s_i < \ldots = t$.

By plugging in the Nelson-Aalen estimator for the cumulative transition intensity $\hat{A}_{hr}(t)$ (with $\hat{A}_{hh}(t) = -\sum_{j \neq h} \hat{A}_{hj}(t)$) into the product integral we obtain:

$$\hat{P}(s,t) = \prod_{(s,t]} (I + d\hat{A}(u)), \tag{B.17}$$

This estimator has been proposed by Aalen and Johansen and its large sample properties can be derived using the martingale theory ([79]; [82]).

B.3 Results from pseudo-values approach for Heart Data

Table B.1: *Estimates of the effect of baseline covariates on occupation probability for state Death. Results from the regression model on pseudo-values.*

	Estimate	**SE**	**p-value**
`Intercept`	-2.915	0.455	< 0.001
`Time (Days)`	-0.309	$8.2^{*}10^{-6}$	< 0.001
`Age`	$9.8^{*}10^{-4}$	$4.5^{*}10^{-4}$	0.030
`Blood B vs A`	-0.029	0.022	0.184
`Blood AB vs A`	-0.080	0.026	0.003
`Blood 0 vs A`	0.008	0.016	0.613
`DCM vs CAD`	-0.040	0.038	0.297
`Other Disease vs CAD`	-0.028	0.04	0.476
`IConsent (Y vs No)`	0.058	0.02	< 0.001

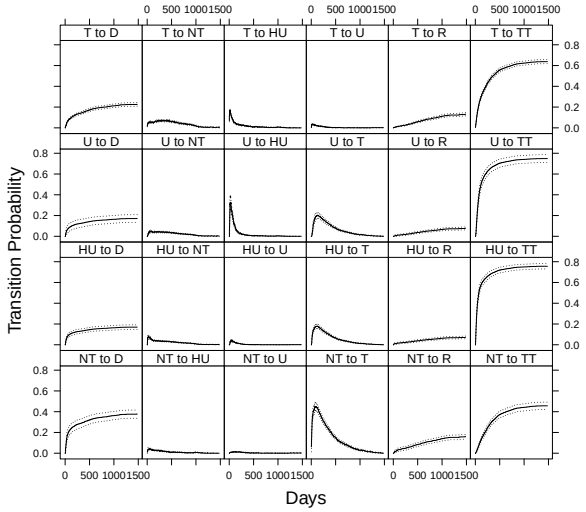
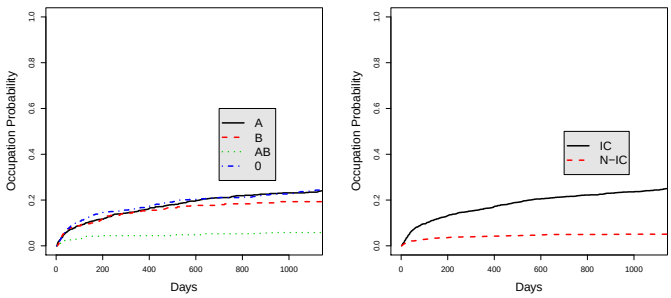


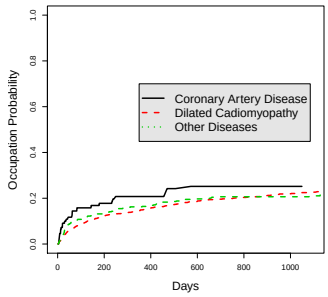
Figure B.5: *Aalen-Johansen estimators for transition probabilities* $P_{hr}(0,t)$.

Table B.2: *Estimates of the effect of baseline covariates on occupation probability for state Non-Transplantable. Results from the regression model on univariate pseudo-values.*

	Estimate	**SE**	**p-value**
`Intercept`	-1.98	0.151	< 0.001
`Time (Days)`	0.18	$2.2^{*}10^{-6}$	< 0.001
`Age`	$-1.5 * 10^{-4}$	$1.7 * 10^{-4}$	0.365
`Blood B vs A`	$7.2 * 10^{-4}$	0.009	0.937
`Blood AB vs A`	0.005	0.010	0.729
`Blood O vs A`	0.006	0.007	0.341
`DCM vs CAD`	0.04	0.007	< 0.001
`Other Disease vs CAD`	0.01	0.007	0.054
`IConsent (Y vs No)`	0.01	0.006	0.004

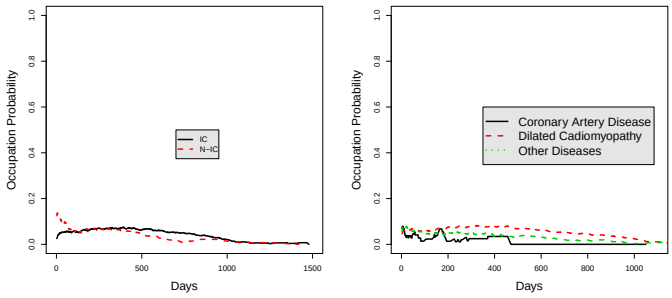


(a) (b)



(c)

Figure B.6: *Aalen-Johansen estimators for occupation probability for state Death for patients with different baseline characteristics.*

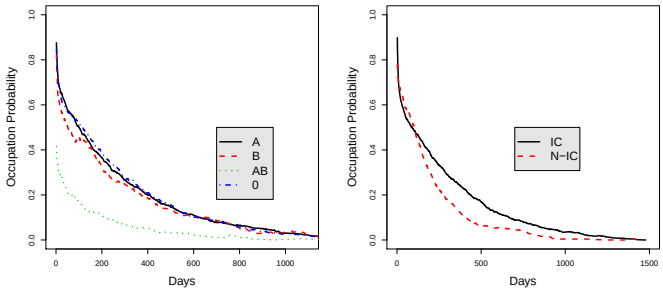


(a) (b)

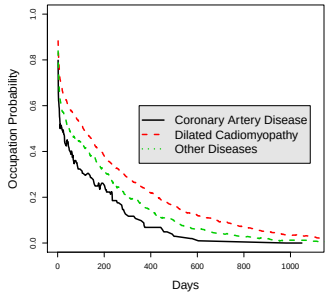
Figure B.7: *Aalen-Johansen estimators for occupation probability for state Non-Transplantable for patients with different baseline characteristics.*

Table B.3: *Estimates of the effect of baseline covariates on occupation probability for state High Urgent. Results from the regression model on univariate pseudo-values.*

	Estimate	**SE**	**p-value**
`Intercept`	-6.653	0.054	< 0.001
`Time`	0.073	$3.4 * 10^{-6}$	< 0.001
`Age`	$3.2 * 10^{-4}$	$7.6 * 10^{-5}$	< 0.001
`Blood B vs A`	-0.008	0.002	< 0.001
`Blood AB vs A`	-0.009	0.002	< 0.001
`Blood 0 vs A`	0.003	0.002	< 0.001
`DCM vs CAD`	0.004	0.003	0.281
`Other Disease vs CAD`	0.004	0.004	0.291
`IConsent (Y vs No)`	0.013	0.001	< 0.001



(a) (b)



(c)

Figure B.8: *Aalen-Johansen estimators for occupation probability for state Transplantable for patients with different baseline characteristics.*

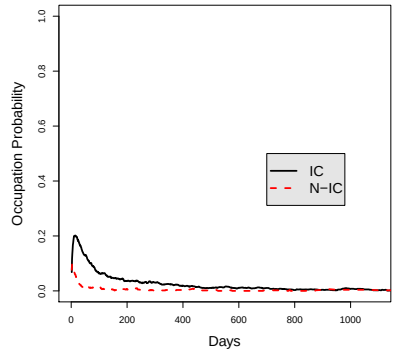
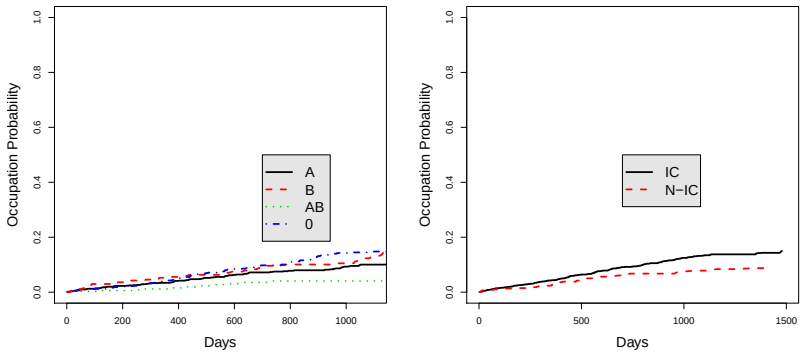


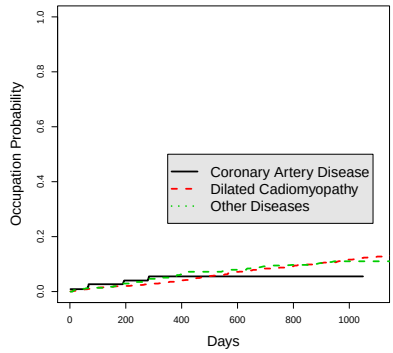
Figure B.9: *Aalen-Johansen estimators for occupation probability for state High Urgent for patients from countries with(I-C) and without(N-IC) informed consent law.*

Table B.4: *Estimates of the effect of baseline covariates on occupation probability for state Urgent. Results from the regression model on univariate pseudo-values.*

	Estimate	**SE**	**p-value**
`Intercept`	-8.309	0.004	< 0.001
`Time`	0.013	$2.9 * 10^{-6}$	< 0.001
`Age`	$3.8 * 10^{-5}$	$4.4 * 10^{-5}$	0.379
`Blood B vs A`	$7.1 * 10^{-5}$	0.003	0.979
`Blood AB vs A`	-0.002	0.002	0.314
`Blood O vs A`	0.001	0.002	0.390
`DCM vs CAD`	-0.001	0.003	0.691
`Other Disease vs CAD`	-0.002	0.003	0.521
`IConsent (Y vs No)`	0.009	0.001	< 0.001



(a) (b)



(c)

Figure B.10: *Aalen-Johansen estimators for occupation probability for state Removed for patients with different baseline characteristics.*

Table B.5: *Estimates of the effect of baseline covariates on occupation probability for state Transplantable. Results from the regression model on univariate pseudo-values.*

	Estimate	**SE**	**p-value**
`Intercept`	-9.090	0.223	< 0.001
`Time`	1.06	0.001	< 0.001
`Age`	0.002	$2.5 * 10^{-4}$	< 0.001
`Blood B vs A`	-0.015	0.015	0.323
`Blood AB vs A`	-0.067	0.017	< 0.001
`Blood 0 vs A`	0.009	0.018	0.379
`DCM vs CAD`	0.085	0.017	< 0.001
`Other Disease vs CAD`	0.045	0.018	0.014
`IConsent (Y vs No)`	0.068	0.010	< 0.001

Table B.6: *Estimates of the effect of baseline covariates on occupation probability for state Removed. Results from the regression model on univariate pseudo-values.*

	Estimate	**SE**	**p-value**
`Intercept`	-2.080	0.317	< 0.001
`Time`	-0.321	0.001	< 0.001
`Age`	$-7.6 * 10^{-5}$	$4.4 * 10^{-4}$	0.861
`Blood B vs A`	0.015	0.019	0.422
`Blood AB vs A`	-0.010	0.024	0.679
`Blood 0 vs A`	0.039	0.014	0.006
`DCM vs CAD`	0.051	0.022	0.018
`Other Disease vs CAD`	0.041	0.024	0.084
`IConsent (Y vs No)`	0.048	0.012	< 0.001

Table B.7: *Estimates of the effect of baseline covariates on occupation probabilities for all states. Results from the regression model on multivariate pseudo-values.*

	Estimate	**SE**	**p-value**
`Intercept (Death)`	-2.813	0.442	< 0.001
`NT`	8.835	0.454	< 0.001
`HU`	1.893	0.459	< 0.001
`T`	-7.534	0.492	< 0.001
`U`	2.837	0.443	< 0.001
`RR`	2.568	0.504	< 0.001
`TT`	-6.539	0.668	< 0.001
`Time`	-0.328	$8.0 * 10^{-6}$	< 0.001
`Age`	$8.5 * 10^{-4}$	$4.3 * 10^{-4}$	0.048
`Blood B vs A`	-0.021	0.021	0.330
`Blood AB vs A`	-0.072	0.025	0.004
`Blood 0 vs A`	0.009	0.015	0.056
`DCM vs CAD`	-0.048	0.037	0.197
`Other Disease vs CAD`	-0.034	0.039	0.388
`IConsent (Y vs No)`	0.052	0.015	< 0.001
`Time:NT`	0.491	$9.9 * 10^{-6}$	< 0.001
`Time:HU`	0.439	$8.6 * 10^{-6}$	< 0.001
`Time:U`	0.326	$8.3 * 10^{-6}$	< 0.001
`Time:T`	1.581	$1.3 * 10^{-5}$	< 0.001
`Time:RR`	1.298	$1.2 * 10^{-5}$	< 0.001
`Time:TT`	-0.525	$1.3 * 10^{-5}$	< 0.001
`Age:NT`	-0.001	$4.6 * 10^{-4}$	0.015
`Age:HU`	-0.001	$4.4 * 10^{-4}$	0.004
`Age:U`	$-7.9 * 10^{-4}$	$4.3 * 10^{-4}$	0.068
`Age:T`	$9.1 * 10^{-4}$	$4.9 * 10^{-4}$	0.066
`Age:RR`	-0.001	$5.4 * 10^{-4}$	0.049
`Age:TT`	-0.003	$6.8 * 10^{-4}$	< 0.001

Table B.7: *Estimates of the effect of baseline covariates on occupation probabilities for all states. Results from the regression model on multivariate pseudo-values cont.*

	Estimate	**SE**	**p-value**
`Blood B vs A:NT`	0.020	0.022	0.368
`Blood B vs A:HU`	0.009	0.021	0.659
`Blood B vs A:U`	0.021	0.021	0.326
`Blood B vs A:T`	$4.9 * 10^{-4}$	0.026	0.984
`Blood B vs A:RR`	0.036	0.025	0.149
`Blood B vs A:TT`	0.057	0.034	0.097
`Blood AB vs A:NT`	0.077	0.028	0.007
`Blood AB vs A:HU`	0.060	0.025	0.017
`Blood AB vs A:U`	0.069	0.025	0.006
`Blood AB vs A:T`	$7.1 * 10^{-4}$	0.030	0.981
`Blood AB vs A:RR`	0.067	0.031	0.023
`Blood AB vs A:TT`	0.226	0.043	< 0.001
`Blood 0 vs A:NT`	-0.004	0.017	0.822
`Blood 0 vs A:HU`	-0.006	0.016	0.725
`Blood 0 vs A:U`	-0.007	0.016	0.640
`Blood 0 vs A:T`	$-3.4 * 10^{-4}$	0.019	0.985
`Blood 0 vs A:RR`	0.010	0.018	0.583
`Blood 0 vs A:TT`	-0.057	0.024	0.019
`IConsent (Y vs No):NT`	-0.042	0.017	0.012
`IConsent (Y vs No):HU`	-0.035	0.015	0.026
`IConsent (Y vs No):U`	-0.044	0.015	0.005
`IConsent (Y vs No):T`	0.009	0.018	0.608
`IConsent (Y vs No):RR`	-0.026	0.018	0.145
`IConsent (Y vs No):TT`	-0.227	0.024	< 0.001
`DCM vs CAD:NT`	0.089	0.038	0.019
`DCM vs CAD:HU`	0.052	0.038	0.168
`DCM vs CAD:U`	0.049	0.038	0.193
`DCM vs CAD:T`	0.142	0.041	< 0.001
`DCM vs CAD:RR`	0.063	0.042	0.014
`DCM vs CAD:TT`	-0.053	0.056	0.340
`Other Disease vs CAD:NT`	0.050	0.040	0.212
`Other Disease vs CAD:HU`	0.039	0.039	0.322
`Other Disease vs CAD:U`	0.033	0.039	0.407
`Other Disease vs CAD:T`	0.085	0.043	0.049
`Other Disease vs CAD:RR`	0.052	0.044	0.244
`Other Disease vs CAD:TT`	-0.018	0.059	0.754

B.4 Results from multinomial model approach for Heart Data

Table B.8: *Estimates of log(OR) for the effect of current state and time on the probability for the next transition from the current state HU and T based on the multinomial model with the previous state Urgent, adjusted for time. Baseline category is probability of death. Jackknife SE are calculated. The estimates for the time effect are not listed.*

Previous=U

	Current=HU	**Current=T**
	Intercept	Intercept
P(NT)/P(D)	0.22(0.68)	1.38(0.98)
P(HU)/P(D)	-	2.18(0.96)
P(U)/P(D)	2.10(0.61)	2.70(1.21)
P(T)/P(D)	0.86(0.77)	-
P(R)/P(D)	-2.30(2.35)	0.38(1.29)
P(TT)/P(D)	2.53(0.60)	0.49(1.05)

Table B.9: *Estimates of log(OR) for the effect of current state and time on the probability for the next transition from the current state T, NT and U based on the multinomial model with the previous state High Urgent, adjusted for time. Baseline category is probability of death (Current=T or NT) or High Urgent. Jackknife SE are calculated. The estimates for the time effect are not listed.*

Previous=HU

	Current=T	**Current=NT**		**Current=U**
	Intercept	Intercept		Intercept
P(NT)/P(D)	-0.78(0.48)	-		
P(HU)/P(D)	0.69(0.37)	-1.67(0.52)		
P(U)/P(D)	-1.81(0.58)	-		
P(T)/P(D)	-	0.55(0.25)	P(T)/P(HU)	-1.32(0.32)
P(R)/P(D)	-1.07(0.50)	-2.24(0.58)		
P(TT)/P(D)	-2.02(0.61)	-	P(TT)/P(HU)	-2.41(0.33)

Table B.10: *Estimates of log(OR) for the effect of current state and time on the probability for the next transition from the current state Transplantable based on the multinomial model with the previous state Non Transplantable adjusted for time. Baseline category is probability of death. Jackknife SE are calculated. The estimates for the time effect are not listed.*

Previous=NT

	Current=T
	Intercept
P(NT)/P(D)	2.39(0.46)
P(HU)/P(D)	2.43(0.39)
P(U)/P(D)	0.49(0.50)
P(T)/P(D)	-
P(R)/P(D)	-
P(TT)/P(D)	2.07(0.37)

B.5 Results from simulation study

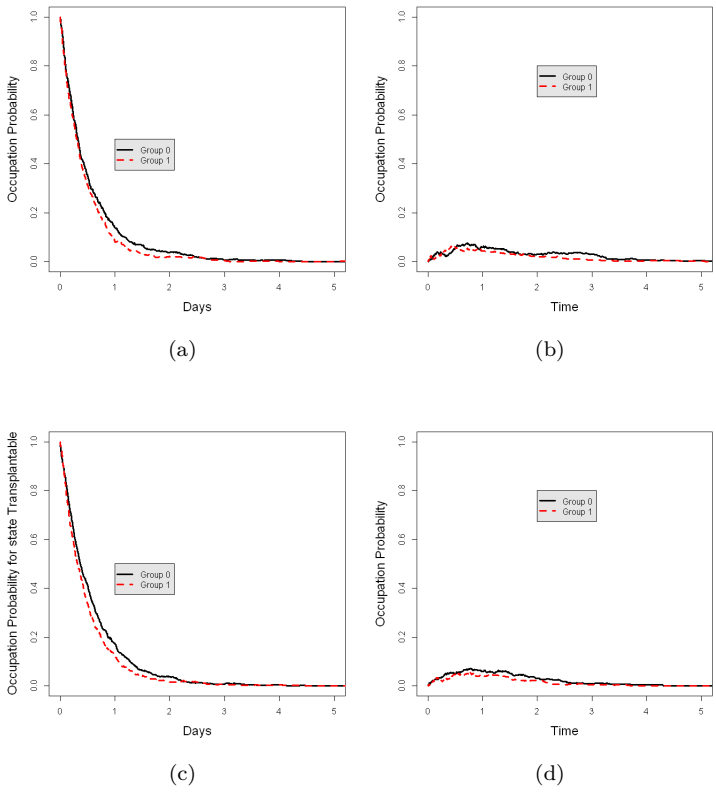


Figure B.11: *Median of Aalen-Johansen estimates for occupation probability for state Transplanted (T) (a,c) and High-Urgent (b,d) for the data simulated from Markov model with group effect on q_{iTT} (a,b) and non-Markov Model with group effect on q_{iTT} and on the history (c,d).*

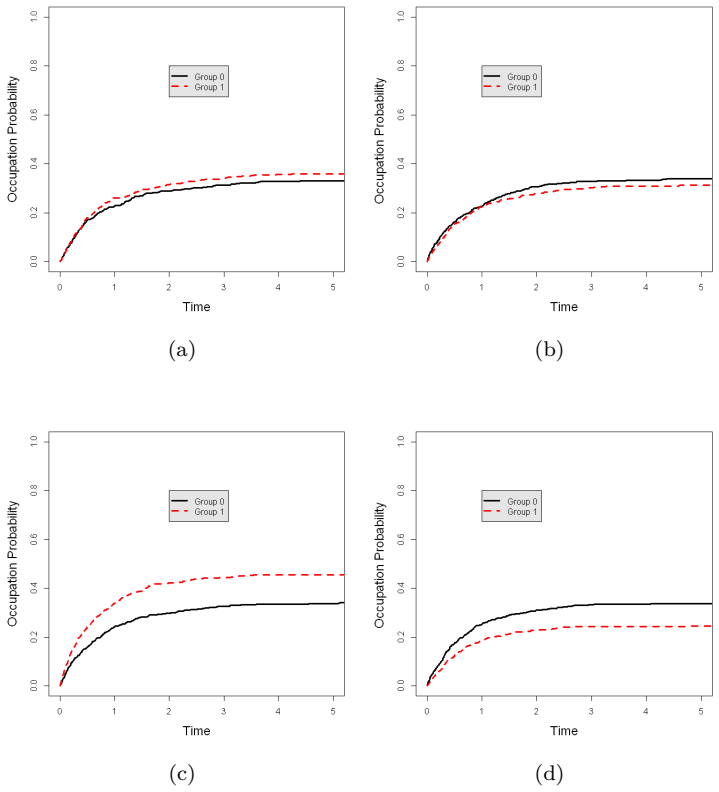


Figure B.12: *Median of Aalen-Johansen estimates for occupation probability for state Transplanted (TT) (a,c) and Death (b,d) (for the data simulated from Markov model with group effect on q_{iT} (a,b) and non-Markov Model with group effect on q_{iT} and on the history (c,d).*

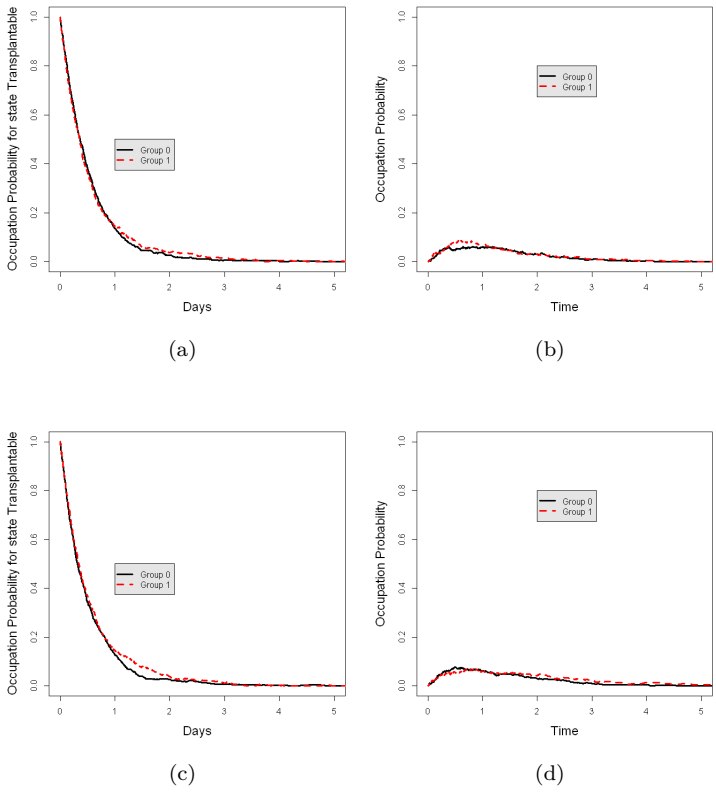


Figure B.13: *Median of Aalen-Johansen estimates for occupation probability for state Transplantable (T) (a,c) and High-Urgent (b,d) for the data simulated from Markov model with group effect on q_{iT} (a,b) and non-Markov Model with group effect on q_{iT} and on the history (c,d).*

Appendix C

This appendix is contains supplementary material for the paper presented in Chapter 4).

C.1 Tables

Table C.1: *Parameter estimates and standard errors from the longitudinal part of the joint models fitted for the heart data.*

		R-E	**T-D I**	
	Intercept	**Time**	**Intercept**	**Time**
`Pr(NT)/Pr(T)`	-1.12(0.07)	0.01(0.16)	-1.37(0.27)	-0.39(0.38)
`Pr(HU)/Pr(T)`	-1.40(0.02)	-0.55(0.31)	-1.44(0.21)	-0.73(0.57)
`Pr(U)/Pr(T)`	-6.39(0.89)	-0.02(0.01)	-8.04(3.33)	-0.39(0.12)
		T-D II	**T-D III**	
	Intercept	**Time**	**Intercept**	**Time**
`Pr(NT)/Pr(T)`	-1.39(0.29)	-0.69(0.37)	-0.85(0.25)	-0.56(0.29)
`Pr(HU)/Pr(T)`	-1.49(0.17)	-0.24(0.62)	-1.42(0.18)	-0.74(0.35)
`Pr(U)/Pr(T)`	-5.79(1.74)	-0.39(0.83)	-4.16(0.63)	0.05(0.40)

Table C.2: *Parameter estimates and standard errors from the survival part of the joint models fitted for the heart data.*

Death	**Shared**	**R-E: Estimate (SE)**	**Shared**	**T-D I: Estimate(SE)**
Pr(NT)	b_1	-0.83(0.13)	$a_{11} + a_{12}t + b_1$	0.79(0.43)
Pr(HU)	b_2	-0.59(0.09)	b_2	0.49(0.69)
Pr(U)	b_3	-3.13(0.10)	b_3	-1.14(0.26)
Age		0.006(0.002)		0.007 (0.003)
Blood A vs B		0.12 (0.10)		0.16 (0.14)
Blood AB vs B		-0.43 (0.21)		-0.48 (0.25)
Blood 0 vs B		0.18 (0.05)		0.22 (0.09)
IConsent		0.47 (0.10)		0.53 (0.12)
Removal	**Shared**	**R-E: Estimate(SE)**	**Shared**	**T-D I: Estimate(SE)**
Pr(NT)	b_1	1.68(0.02)	b_1	-1.44(0.45)
Pr(HU)	b_2	0.27(0.07)	b_2	-0.31(1.14)
Pr(U)	b_3	-1.25(0.02)	b_3	-1.10(0.47)
Age		-0.004(0.003)		-0.004(0.005)
Blood A vs B		-0.25 (0.21)		-0.31 (0.20)
Blood AB vs B		-0.29 (0.30)		-0.35 (0.32)
Blood 0 vs B		0.12 (0.13)		0.06 (0.14)
IConsent		0.61 (0.15)		0.65 (0.17)
Transplantation	**Shared**	**R-E: Estimate(SE)**	**Shared**	**T-D I: Estimate(SE)**
Pr(NT)	b_1	-0.47(0.06)	b_1	-1.82(0.24)
Pr(HU)	b_2	0.19(0.10)	b_2	0.34(0.95)
Pr(U)	b_3	-3.32(0.03)	b_3	-0.68(0.18)
Age		-0.005(0.001)		-0.006(0.002)
Blood A vs B		-0.20 (0.06)		-0.21 (0.08)
Blood AB vs B		0.30 (0.09)		0.32 (0.12)
Blood 0 vs B		-0.34 (0.05)		-0.36 (0.06)
IConsent		-0.79 (0.05)		-0.81 (0.07)

Table C.3: *Parameter estimates and standard errors from the survival part of the joint models fitted for the heart data cont.*

Death	**Shared**	**T-D II: Estimate (SE)**	**Shared**	**T-D III: Estimate(SE)**
Pr(NT)	$a_{11} + a_{12}t + b_1$	0.74(0.44)	$a_{11} + a_{12}t + b_1$	2.14(0.46)
Pr(HU)	b_2	0.42(0.09)	b_2	0.55(0.62)
Pr(U)	b_3	-0.34(0.10)	b_3	-2.16(0.28)
Age		0.008 (0.02)		0.008 (0.01)
Blood A vs B		0.12 (0.18)		0.16 (0.19)
Blood AB vs B		-0.50 (0.25)		-0.58 (0.29)
Blood 0 vs B		0.28 (0.19)		0.22 (0.22)
IConsent		0.57 (0.12)		0.55 (0.12)
Removal	**Shared**	**T-D II: Estimate(SE)**	**Shared**	**T-D III: Estimate(SE)**
(NT)	b_1	-0.04(0.35)	b_1	0.56(0.38)
(HU)	$a_{21} + a_{22}t + b_2$	-0.09(0.46)	$a_{21} + a_{22}t + b_2$	0.57(0.13)
(U)	b_3	-0.19(0.11)	b_3	-1.85(0.32)
Age		-0.005(0.003)		-0.005(0.005)
Blood A vs B		-0.26 (0.21)		-0.31 (0.20)
Blood AB vs B		-0.28 (0.30)		-0.35 (0.30)
Blood 0 vs B		0.13 (0.13)		0.15 (0.16)
IConsent		0.62 (0.15)		0.61 (0.17)
Transplantation	**Shared**	**T-D II: Estimate(SE)**	**Shared**	**T-D III: Estimate(SE)**
Pr(NT)	b_1	-2.26(0.25)	b_1	-1.99(0.30)
Pr(HU)	b_2	1.11(0.19)	b_2	-0.31(0.42)
Pr(U)	b_3	-1.72(0.35)	$a_{31} + a_{32}t + b_3$	-0.39(0.09)
Age		-0.006(0.001)		-0.007(0.002)
Blood A vs B		-0.20 (0.06)		-0.21 (0.08)
Blood AB vs B		0.32 (0.10)		0.33 (0.11)
Blood 0 vs B		-0.34 (0.05)		-0.35 (0.06)
IConsent		-0.78 (0.06)		-0.80 (0.08)

Appendix D

This appendix is related to simulation study presented in Chapter 5 (see Section 5.6.1).

D.1 Simulation Settings

For all simulation scenarios the parameter values that were used for the longitudinal submodel were

Fixed effects: $\beta_1 = 0.93$, $\beta_2 = -0.6$, $\beta_3 = 0.63$, $\beta_4 = 0.42$, $\beta_5 = 1.1$, $\beta_6 = 0.54$, $\beta_7 = 0.54$, and $\beta_8 = 0.55$;

Random effects diagonal covariance matrix: $\boldsymbol{D}_{11} = 0.49$, $\boldsymbol{D}_{22} = 4.52$, $\boldsymbol{D}_{33} = 2.33$, and $\boldsymbol{D}_{44} = 1.52$;

Measurement error standard deviation: $\sigma = 2$.

For the survival submodels the parameters that were used to simulate from each scenario are given in Table D.1.

	Scenario			
	I	II	III	IV
γ_0	-6.73	-6.73	-6.73	-6.73
γ_1	0.41	0.41	0.41	0.41
α_1	0.7	0.05	0.08	-0.3
α_2		3.3		-0.8
α_3				0.3
α_4				0.8
σ_t	1.65	1.65	1.65	1.60

Table D.1: Parameter values for the survival submodels under the four simulation scenarios.

Appendix E

Publications

International Journals

1. **Murawska, M.**, Rizopoulos, D. and Lesaffre, E. "A Two-Stage Joint Model for Nonlinear Longitudinal Response and a Time-to-Event with Application in Transplantation Studies", vol. in press, 2012. (J Probability and Statistics) [DOI:10.1155/2012/194194]

Submitted

1. **Murawska, M**, Rizopoulos, D. and Lesaffre, E. "Dynamic Predictions for Categorical Longitudinal Responses and Event Data with Application in Transplantation Studies", *Biostatistics*, 2013.

2. **Murawska, M.**, Rizopoulos D. and Lesaffre, E. "Simple analysis of non-Markov models: A case study on heart transplant data", *Statistical Modeling*, 2013. (Submitted after major revision)

3. Rizopoulos, D., Murawska, M., Andrinopoulou,E-R., Molenberghs, G,

Takkenberg, J.J.M and Lesaffre, E. "A comparison between landmarking and joint modeling for producing predictions using longitudinal outcomes", *Biometrical Journal*, 2013.

International Conferences and Workshops

1. **Murawska, M.**, Shkedy Z."Evaluation of Surrogate Markers in Early Drug Development Microarray Experiments", in *ISCB Conference*, Alexandropolis (Greece), 2007.

2. **Murawska, M.**, Rizopoulos, D. and Lesaffre, E."Measuring the Effect of Repeated Measures on an Event Time", in *ISCB Conference*, Montpelier (France), 2010.

3. **Murawska, M.**, Rizopoulos, D. and Lesaffre, E."Multi-state models for non Markov process: an application to the heart transplant data", in *IWSM Conference*, Valencia (Spain), 2011.

4. **Murawska, M.**, Rizopoulos, D. and Lesaffre, E."Dynamic Predictions Based on Joint Model for Categorical Response and Time-to-Event", in *ISCB Conference*, Bergen (Norway), 2012. **(Best paper award)**

5. **Murawska, M.**, Rizopoulos, D. and Lesaffre, E. "A comparison between landmarking and joint modeling for producing predictions using longitudinal outcomes", in *IWSM Conference*, Palermo (Italy), 2013.

National Conferences

1. **Murawska, M.**, "Bayesian methods in multivariate linear regression models in the context of localization of QTL(Quantitative Trait Loci)",

in *ICM Conference*, Nowy Zyzdroj, Poland, 2009.

2. **Murawska, M.**, "Porownanie prognozowania dynamicznego metoda joint modeling i landmarking", in *ICM Conference*, Plock, Poland, 2013.

Technical Reports

1. Frommlet F., Chakrabarti, A., **Murawska M.** , and Bogdan M, "Asymptotic Bayes optimality under sparsity for general priors under the alternative Technical Report, http://arxiv.org/abs/1005.4753v2"

Appendix F

CV

Personal Details

Name: MAGDALENA MURAWSKA

Nationality: Polish

Date of birth: 23 May 1979

Address: Rododendronstraat 106, 3053 LR Rotterdam, Netherlands

Contact: e-mail: m.murawska@erasmusmc.nl

EDUCATION

2009 - 2013 - PhD in Biostatistics, Erasmus Medical Center, Rotterdam, the Netherlands

2007 - 2009 - working on Bayesian methods in genetics, cooperation with Vienna University

2005 - 2006- MSc in Biostatistics with distinction, University Limburg-School for Informational Technology, Belgium

1998 - 2003 - MSc in Mathematics, University of Warsaw, Poland

AWARDS

ISCB 2012 Student Award, First Prize, Bergen, Norway

WORK EXPERIENCE

since XI 2009- PhD student at Erasmus Medical Center (Rotterdam) in cooperation with Eurotransplant International Foundation (Leiden), the Netherlands

II 2004 - XII 2011 - assistant researcher, Biostatistics and Clinical Trials Office, Warsaw Center of Oncology-Institute, Poland

Design and analysis of clinical trials (phase III/IV), statistical analyses of retrospective data, statistical consultations and educational courses for physicians

2008 -The Medical Centre of Postgraduate Education

Statistical analyses

- Poznan University of Medical Sciences

Statistical analyses, consultations

- Softword

Consultations, translating of user manual for statistical software

2007/2008 - statistician (freelancer) , LABoratory & Co

Testing and developing of statistical methods for the analysis of EEG data in neuromarketing

COURSES/CONFERENCES/WORKSHOPS

ISCB (The International Society for Clinical Biostatistics) Annual Conference:

2004-Leiden, Netherlands

2007-Alexandropolis, Greece

2010- Montpelier, France

2012- Bergen, Norway

IWSM (International Workshop on Statistical Modeling) Annual Conference:

2011-Valencia, Spain

2013- Palermo, Italy

ICM (Interdisciplinary Center of Mathematical Modeling) Annual Conference:

2008 -Warsaw, Poland

2009 -Nowy Zyzdroj, Poland

2013 -Plock, Poland

Internal courses in Erasmus MC

MEMBERSHIP

since 2004 - The International Society for Clinical Biostatistics

ADDITIONAL INFORMATION

2003-2004- pedagogical course, Civic Educational Association (STO), Poland

COMPUTING SKILLS

statistical software: SAS, S-Plus, R, Stata, MATLAB, STATISTICA, StatXact, Statgraphics, SPSS, Win(Open)BUGS, INLA, JAGS, C.A.MAN (mixture of population analysis);

genetic software: PBAT;

Windows, Linux, MS Office

programming in: R (including web interface and computations on clusters), Delphi/Pascal, C++ (basic level), VBA

LANGUAGES

Polish - native

English - fluent

Dutch - good

Bibliography

[1] Kabfleisch JD, Prentice RL. *The Statistical Analysis of Failure Time Data*. John Wiley & Sons: New York, 2002.

[2] Faucett C, Thomas D. Simultaneously modelling censored survival data and repeatedly measured covariates. *Statistics in Medicine* 1996; **15**:1460–1461.

[3] Wulfsohn MS, Tsiatis AA. A joint model for survival and longitudinal data measured with error. *Biometrics* 1997; **53**:330–339.

[4] Little RJA, Rubin DB. *Statistical analysis with missing data*. Wiley: New York, US, 1987.

[5] Little RJA. Modeling the dropout mechanism in repeated-measures studies. *J. A. Statist. Assoc.* 1995; **90**:1112–1121.

[6] Hogan JW, Laird NM. Model-based approaches to analyzing incomplete longitudinal and failure time data. *Statistics in Medicine* 1997; **16**:239–257.

[7] G M, G KM. *Missing Data in Clinical Studies*. John Wiley & Sons: New York, US, 2007.

[8] Davidian M, Giltman DM. *Nonlinear Models for Repeated Measurement Data*. Chapman and Hall: New York, US, 1995.

[9] Rizopoulos D. *Joint Models for Longitudinal and Time-to-Event Data: With Applications in R*. Boca Raton: Chapman and Hall/CRC, 2012.

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

[11] Prentice RL. Covariate measurement errors and parameter estimation in a failure time regression model. *Biometrika* 1982; **69**:331–342.

[12] Dafni UG, Tsiatis AA. Evaluating surrogate markers of clinical outcome measured with error. *Biometrics* 1998; **54**:1445–1462.

[13] Bycott P, Taylor J. A comparison of smoothing techniques for CD4 data measured with error in a time-dependent cox proportional hazards model. *Statist. Medicine* 1998; **17**:2061–2077.

[14] Wu MC, Carroll RJ. Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. *Biometrics* 1988; **44**:175–188.

[15] Molenberghs G, Kenward MG, E L. The analysis of longitudinal ordinal data with nonrandom drop-out. *Biometrika* 1997; **84**:33–44.

[16] Kaciroti NA, Raghunathan T, Schork MA, Clark NM, Gong M. A Bayesian approach for clustered longitudinal ordinal outcome with nonignorable missing data: evaluation of an asthma education program. *Journal of the American Statistical Association* 2006; **101**:435–446.

[17] Albert P, Follmann D, Wang S, Suh E. A latent autoregressive model for longitudinal binary data subject to informative missingness. *Biometrics* 2002; **58**:631–642.

[18] Viviani S, Alfo M, Rizopoulos D. Generalized linear mixed joint model for longitudinal and survival outcomes. *Statistics and Computing* 2013; **DOI 10.1007/s11222-013-9378-4**.

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

[20] Tsiatis A, Davidian M. A semiparametric estimator for the proportional hazards model with longitudinal covariates measured with error. *Biometrika* 2001; **88**:447–58.

[21] Song X, Davidian M, Tsiatis A. An estimator for the proportional hazards model with multiple longitudinal covariates measured with error. *Biostatistics* 2002; **3**:511Ű528.

[22] Xu J, Zeger S. Joint analysis of longitudinal data comprising repeated measures and times to events. *Applied Statistics* 2001; **50**:375–387.

[23] Wang Y, Taylor J. Jointly modeling longitudinal and event time data with application to acquired immunodeficiency syndrome. *J. A. Statist. Assoc.* 2001; **96**:895–905.

[24] Brown ER, Ibrahim JG. A Bayesian semiparametric joint hierarchical model for longitudinal and survival data. *Biometrics* 2003; **59**:221–228.

[25] Ibrahim JG, Chen M, Sinha D. Bayesian methods for joint modeling of longitudinal and survival data with applications to cancer vaccine trials. *Statistica Sinica* 2004; **14**:863–883.

[26] Xu J, Zeger S. The evaluation of multiple surrogate endpoints. *Biometrics* 2001; **57**:81–87.

[27] Brown ER, Ibrahim JG, Gruttola V. A flexible B-spline model for multiple longitudinal biomarkers and survival. *Biometrics* 2005; **61**:64–73.

[28] Rizopoulos D, Ghosh P. A Bayesian semiparametric multivariate joint model for multiple longitudinal outcomes and a time-to-event. *Statistics in Medicine* 2011; **30**:1366–1380.

[29] Albert PS, Shih JH. On estimating the relationship between longitudinal measurments and time-to-event data using a simple two-stage procedure. *Biometrics* 2009; **66**:983–991.

[30] Wu L, Hu XJ, Wu H. Joint inference for nonlinear mixed-effects models and time to event at the presence of missing data. *Biostatistics* 2008; **9**:308–320.

[31] Hu C, Sale ME. A joint model for nonlinear longitudinal data with informative dropout. *Journal of Pharmacokinetics and Pharmacodynamics* 2008; **30**:83–103.

[32] Kaciroti NA, Raghunathan ER, Schork MA, Clark MC. A Bayesian model for longitudinal count data with non-ignorable dropout. *J R Stat Soc Ser C (Appl Stat)* 2008; **57**:521–534.

[33] Elashoff R, Gang L, Li N. A joint model for longitudinal measurments and survival data in presence of multiple failure types. *Statistics in Medicine* 2008; **26**:546–557.

[34] Li R N Elashoff, Gang L. Joint modeling of longitudinal ordinal data and competing risks survival times and analysis of the ninds rt-PA stroke trial. *Statistics in Medicine* 2010; **29**:792–802.

[35] Williamson P, Kolamunnage-Dona R, Philipson P, Marson A. Joint modelling of longitudinal and competing risk data. *Statistics in Medicine* 2008; **27**:6427–6438.

[36] Huang X, Gang L, Elashoff R. A general joint model for longitudinal measurements and competing risks survival data with heterogeneous random effects. *Lifetime Data Anal.* 2011; **17**:80–100.

[37] Lin B H Turnbull, McCullocha C. Latent class models for joint analysis of longitudinal biomarker and event process data. *Journal of the American Statistical Association* 2002; **97**:53–65.

[38] Proust-Lima C, Letenneur L, Jacqmin-Gadda, Jacqmin-Gadda H. A nonlinear latent class model for joint analysis of multivariate longitudinal data and a binary outcome. *Statistics in Medicine* 2007; **26**:2229Ű2245.

[39] Proust-Lima C, Joly P, Dartiques J, Jacqmin-Gadda H. Joint modelling of multivariate longitudinal outcomes and a time-to-event: A nonlinear latent class approach. *Computational Statistics & Data Analysis* 2009; **53**:1142–1154.

[40] Rizopoulos D. Jm: An R package for the joint modelling of longitudinal and time-to-event data. *Journal of Statistical Software* 2010; **35**:1–33.

[41] Rizopoulos D. Fast fitting of joint models for longitudinal and event time data using a pseudo-adaptive gaussian quadrature rule. *Computational Statistics & Data Analysis* 2012; **56**:491–501.

[42] Philipson P, Sousa I, Diggle PJ. joineR- Joint modelling of repeated measurements and time-to-event data. http://cran.r-project.org/web/packages/joineR/index.html. 2012.

[43] Komoda T, Drews T, Hetzer R, MD H Lehmkuhl. New prioritization of heart transplant candidates on mechanical circulatory support in an era of severe donor shortage. *Journal of Heart and Lung Transplantation* 2010; **29**:989–996.

[44] Bekkers J, Klieverik L, Raap G, Takkenberg J, , Bogers A. Re-operations for aortic allograft root failure: Experience from a 21-year single-center prospective follow-up study. *European Journal of Cardio-Thoracic Surgery* 2011; **40**:35–42.

[45] Andersen P, Klein J, Rosthoj S. Generalized linear models for correlated pseudo-observations with applications to multi-state models. *Biometrika* 2003; **90**:15–27.

[46] Murawska M, Rizopoulos D, Lesaffre E. A two-stage joint model for nonlinear longitudinal response and a time-to-event with application in transplantation studies. *J Probability and Statistics* 2012; **2012**:[DOI:10.1155/2012/194 194].

[47] Rizopoulos D, Verbeke G, Lesaffre E. Fully exponential Laplace approximations for the joint modelling of survival and longitudinal data. *Journal of the Royal Statistical Society: Series B* 2009; **71**:637–654.

[48] Henderson R, Diggle P, Dobson A. Joint modelling of longitudinal measurements and event time data. *Biostatistics* 2000; **4**:465–480.

[49] DeGruttola V, Tu X. Modeling progression of CD4 lymphocyte count and its relationship to survival time. *Biometrics* 1994; **50**:1003–14.

[50] Self S, Pawitan Y. *Modeling a marker of disease progression and onset of disease*. In AIDS Epidemiology: Methodological Issues (N.P. Jewell, K. Dietz, and V.T. Farewell, eds.: BirkhÄauser, Boston, 1992.

[51] Andersen P, Gill R. Cox's regression for counting processes: A large sample study. *Annals of Statistics* 1982; **10**:1100–1120.

[52] Clayton D. *Bayesian analysis of frailty models. Tech. report.* Medical Research Council Biostatistics Unit. Cambridge, 1994.

[53] Kalbfleisch J. Non-parametric Bayesian analysis of survival time data. *J Roy Statist Soc B* 1978; **40**:214–221.

[54] Patra K, Dey DK. A general class of change point and change curve modelling for life data. *Inst. Stat. Math.* 2002; **54**:517–530.

[55] Casellas J. Bayesian inference in a piecewise Weibull proportional hazards model with unknown change points. *J.Anim Breed. Genet.* 2007; **124**:176–184.

[56] Carlin BP, Louis TA. *Bayes and Empirical Bayes Methods for Data Analysis*. Chapman & Hall: New York, 2000.

[57] Roberts GO, Gelman A, Gilks WR. Weak convergence and optimal scaling of random walk Metropolis algorithms. *The Annals of Applied Probability* 1997; **7**:110–120.

[58] Verbeke G, Molenberghs G. *Linear Mixed Models for Longitudinal Data*. Springer-Verlag: New York, US, 2000.

[59] Nelson KP, Lipsitz SR, Fitzmaurice GM, Ibrahim J, Parzen M, Strawderman R. Using the probability integral transformation for non-normal random effects in non-linear mixed models. *J. Comput. Graph. Statist.* 2004; **15**:39Ű57.

[60] Liu L, Yu Z. A likelihood reformulation method in non-normal random effects models. *Statistics in Medicine* 2008; **27**:3105–3124.

[61] Davidian M, Giltinan DM. *Nonlinear Models for Repeated Measurment Data*. Chapman and Hall: New York, US, 1998.

[62] Murawska M, Rizopoulos D. Simple analysis of non-markov models: A case study on heart transplant data. Under revision.

[63] Datta S, Satten GA. Validity of the aalen-johansen estimators of stage occupation probabilities and nelso-aalen estimators of integrated transition hazards for non-markov models. *Statistics and Probability Letters* 2001; **55**:403–411.

[64] Klein J, Andersen P. Regression modeling of competing risks data based on pseudo-values of the cumulative incidence function. *Biometrics* 2005; **61**:223–229.

[65] Klein J. Modeling competing risks in cancer studies. *Statistics in Medicine* 2006; **25**:1015–1034.

[66] Andersen P, Hansen M, Klein J. Regression analysis of restricted mean survival time based on pseudo-observations. *Lifetime Data Anal.* 2004; **10**:335–350.

[67] Andersen P, Klein J. Regression analysis for multistate models based on pseudo-value approach with application to bone marrow tranplantation studies. *Scand. J. Statist.* 2006; **34**:3–16.

[68] Graw F, Gerds T, Schumacher M. On pseudo-values for regression analysis in competing risks models. *Lifetime Data Anal.* 2009; **15**:241–255.

[69] Grüger J, Kay R, Schumacher M. The validity of inferences based on incomplete observations in disease state models. *Biometrics* 1991; **47**:595–605.

[70] Commenges D. Likelihood for interval-censored observations from multi-state models. *SORT* 2002; **27**:1–12.

[71] Datta S, Sundaram R. Nonparametric estimation of stage occupation probabilities in a multistage model with current status data. *Biometrics* 2006; **62**:829–837.

[72] Joly P, Commenges D, Helmer C, Letenneur L. A penalized likelihood approach for an illnessŰdeath model with interval-censored data:application to age-specific incidence of dementia. *Biostatistics* 2002; **3**:433–443.

[73] Meira-Machado L, Una-Alvarez J, Cadarsao-Suarez C. Nonparametric estimation of transition probabilities in a non-markov illness-death model. *Lifetime Data Anal.* 2006; **12**:325–344.

[74] Glidden D. Robust inference fpr event probabilities with non-markov event data. *Biometrics* 2002; **58**:361–368.

[75] Datta S, Satten GA. Estimation of integrated transition hazards and stage occupation probabilities for non-markov systems under dependent censoring. *Biometrics* 2002; **58**:792–802.

[76] Cook R, Grace YY, Lee KA, Gladman D. A conditional Markov model for clustered progressive multistate processes under incomplete observation. *Biometrics* 2004; **60**:436–443.

[77] Sutradhar R, Cook R. Analysis of interval-censored data from clustered multi-state processes: application to joint damage in psoriatic arthritis. *Appl. Statist.* 2008; **57**:553–566.

[78] Shao J, Tu D. *The Jackknife and Bootstrap*. Springer-Verlag, 1995.

[79] Aalen O, Johansen J. An empirical transition matrix for non-homogenous markov chains based on censored observations. *Scand. J. Statist.* 1978; **5**:141–150.

[80] Liang K, Zeger S. Longitudinal data analysis using generalized linear models. *Biometrika* 1986; **78**:13–22.

[81] Knott M, Frangos CC. Variance estimation for the jackknife using von Mises expansions. *Biometrika* 1983; **70**:501–504.

[82] Andersen P, Borgan O, Gill R, Keiding N. *Statistical Models Based on Counting Processes*. Springer-Verlag: US, New York, 1993.

[83] Scheike H, Zhang M. Direct modelling of regression effects for transition probabilities in multistate models. *Scand. J. Statist.* 2007; **12**:325–344.

[84] Logan B, Zhang M, Klein J. Marginal models for clustered time-to-event data with competing risks using pseudo-values. *Biometrics* 2011; **67**:1–7.

[85] Yu M, Taylor J, Sandler H. Individualized prediction in prostate cancer studies using a joint longitudinal-survival-cure model. *Journal of the American Statistical Association* 2008; **103**:178–187.

[86] Rizopoulos D. Dynamic prediction and prospective accuracy in joint models for longitudinal and time-to-event data. *Biometrics* 2011; **67**:819–829.

[87] Rizopoulos D, Verbeke G, Lesaffre E, Vanrenterghem Y. A two-part joint model for the analysis of survival and longitudinal binary data with excess zeros. *Biometrics* 2008; **64**:611–619.

[88] Chi YY, Ibrahim JG. Bayesian approaches to joint longitudinal and survival models accommodating both zero and nonzero cure fractions. *Statistica Sinica* 2007; **17**:445–462.

[89] Ibrahim JG, Chen M, Lipsitz SR. Missing responses in generalised linear mixed models when the missing data mechanism is nonignorable. *Biometrika* 2001; **88**:551–564.

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

[91] Kahaner D, Moler C, Nash S. *Numerical Methods and Software*. Prentice-Hall, 1989.

[92] Lesaffre E, Lawson AB. *Bayesian Biostatistics*. John Wiley & Sons: New York, 2012.

[93] Anderson J, Cain K, , Gelber R. Analysis of survival by tumor response. *Journal of Clinical Oncology* 1983; **1**:710–719.

[94] Zheng Y, Heagerty P. Partly conditional survival models for longitudinal data. *Biometrics* 2005; **61**:379–391.

[95] van Houwelingen H. Dynamic prediction by landmarking in event history analysis. *Biometrics* 2007; **34**:70–85.

[96] Henderson R, Diggle P, Dobson A. Identification and efficacy of longitudinal markers for survival. *Biostatistics* 2002; **3**:33–50.

[97] Proust-Lima C, Taylor J. Development and validation of a dynamic prognostic tool for prostate cancer recurrence using repeated measures of post-treatment PSA: A joint modeling approach. *Biostatistics* 2009; **10**:535–549.

[98] Tsiatis A, Davidian M. A semiparametric estimator for the proportional hazards model with longitudinal covariates measured with error. *Biometrika* 2001; **88**:447–58.

[99] Little RJA, Rubin DB. *Statistical analysis with missing data*. Wiley: New York, US. 2nd edition, 1987.

[100] Fisher L, Lin DY. Time-dependent covariates in the cox proportional-hazards regression model. *Annual Review of Public Health* 1999; **20**:145–157.

[101] Schemper M, Henderson R. Predictive accuracy and explained variation in cox regression. *Biometrics* 2000; **56**:249–255.

[102] Gerds T, Schumacher M. Consistent estimation of the expected brier score in general survival models with right-censored event times. *Biometrical Journal* 2006; **48**:1029–1040.

[103] Harrell F, Kerry L, Mark D. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in Medicine* 1996; **15**:361–387.

[104] Pencina M, D'Agostino, Sr D R. Evaluating the added predictive ability of a new marker: From area under the roc curve to reclassi
cation and beyond. *Statistics in Medicine* 2008; **27**:157–172.

[105] Antolini L, Boracchi P, Biganzoli E. A time-dependent discrimination index for survival data. *Statistics in Medicine* 2005; **24**:3927–3944.

[106] Heagerty P, Zheng Y. Survival model predictive accuracy and roc curves. *Biometrics* 2005; **61**:92–105.

[107] Press W, Teukolsky S, Vetterling W, , Flannery B. *Art of Scientific Computing*. Cambridge University Press: New York, 3rd edition, 2005.

[108] Rizopoulos D. *JMbayes: Shared parameter models for the joint modeling of lon- gitudinal and time-to-event data using JAGS, WinBUGS, or OpenBUGS*. R package version 0.4-1., 1013.

[109] Breslow N, Clayton D. Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association* 1993; **88**:9–25.

[110] Larsen K. Joint analysis of time-to-event and multiple binary indicators of latent classes. *Biometrics* 2004; **60**:85–92.

[111] Gueorguieva R, Rosenheck R, Lin H. Joint modelling of longitudinal outcome and interval-censored competing risk dropout in a schizophrenia clinical trial. *J R Stat Soc Ser A Stat Soc.* 2012; **175**:417–433.

[112] Huang X, Stefanski L, Davidian M. Latent-model robustness in joint models for a primary endpoint and a longitudinal process. *Biometrics* 2009; **64**:719–727.

[113] Hoeting J, Madigan A, Raftery AE, Volinsky CT. Bayesian model averaging: A tutorial. *Statistical Science* 1999; **14**:382Ű417.

[114] Rizopoulos, Verbeke G, Molenberghs G. Multiple-imputation-based residuals and diagnostic plots for joint models of longitudinal and survival outcomes. *Biometrics* 2010; **66**:20–29.

[115] Fieuws S, Verbeke G, Maes B, Vanrenterghem Y. Predicting renal graft failure using multivariate longitudinal profiles. *Biostatistics* 2008; **9**:419Ű431.

[116] McCulloch C. Joint modelling of mixed outcome types using latent variables. *Statistical Methods in Medical Research* 2008; **17**:53Ű73.

[117] Gueorguieva R, Sanacora G. Joint analysis of repeatedly observed continuous and ordinal measures of disease severity. *Statistics in Medicine* 2006; **25**:1307Ű1322.

[118] Liu L, Huang X. Joint analysis of correlated repeated measures and recurrent events processes in the presence of death, with application to a study on acquired immune deficiency syndrome. *Journal of the Royal Statistical Society, Series C* 2009; **58**:65–81.

[119] Heagerty P. Survival model predictive accuracy and ROC curves. *Biometrics* 2005; **61**:92–105.

[120] Saha P, Heagerty P. Time-dependent predictive accuracy in the presence of competing risks. *Biometrics* 2010; **66**:999–1011.

[121] Zheng Y, Cai T, Jin Y, Feng Z. Evaluating prognostic accuracy of biomarkers under competing risk. *Biometrics* 2012; **68**:388–396.

Index

Name Index