Extended Joint Models for Longitudinal & Time-to-event Data

with applications in Cardiothoracic Surgery

Grigorios Papageorgiou

Grigorios Papageorgiou
ISBN: 978-94-6419-351-0
Printed by Gildeprint

Extended Joint Models for Longitudinal and Time-to-event Data

with applications in cardiothoracic surgery

Uitgebreide joint models
voor longitudinale en overlevingsdata
met toepassingen in cardiothoracale chirurgie

Proefschrift

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de
rector magnificus

Prof. dr. A.L. Bredenoord

en volgens besluit van het College voor Promoties.
De openbare verdediging zal plaatsvinden op
dinsdag 14 December 2021 om 13:00 uur

door

Grigorios Papageorgiou
geboren te Athene, Griekeland.

Erasmus University Rotterdam

Promotiecommissie:

Promotoren: Prof. dr. D. Rizopoulos
Prof. dr. J.J.M. Takkenberg

Overige leden: Dr. I. Kardys
Prof. dr. R. Selles
Prof. dr. H. Putter

Copromotor: dr. M.M. Mokhles

The research described in this thesis was supported by Nederlandse Organisatie voor Wetenschappelijk Onderzoek VIDI grant nr. 016.146.301, by Nederlandse Organisatie voor Wetenschappelijk Onderzoek Veni grant nr. 916.160.87 and Erasmus University Medical Center funding.

To my mother, father and sister, and to Maroula.

Contents

Contents

III Missing Data

IV Discussion

Chapter 1

General Introduction

It is quite before our birth that we unconsciously register our presence in this world with a short, tiny, barely audible, rhythmic, and most importantly, repeated sound. A sound that will constantly accompany us during our lives. Our heartbeat. As long as the beat repeats steadily and with a balanced pace (not too fast and not too slow), we are going to be okay.

Of course, this will not always be the case. Our heartbeat will often change, transitioning between different states of pace and volume. We all know that it will become faster and louder under intense emotions before returning to its normal state. However, emotions will not be the sole cause of such transitions, and these transitions will not always be temporal. Events can worsen our health and increase the risk of bringing our heartbeat closer to the absorbing state of silence.

Naturally, one then reasonably wonders: If we can deduct so much information from the repeated beat of our hearts by collective experience alone, what secrets will we be able to unravel by scientifically and repeatedly monitoring our heart, its features, and its function? Can we predict the future of our health and prevent us from the paths that lead to states bad for our health?

In this thesis, motivated by a broad range of applications in heart disease, we developed statistical methodology and models that contribute to answering such questions. In the remainder of the introduction, we will briefly introduce the clinical background of the applications that motivated our research: congenital heart disease, heart failure, and tricuspid regurgitation. We will then present the datasets that we use in our applications and their features. Furthermore, we will outline the joint modeling framework and briefly introduce the theoretical background, which forms the basis for the extensions we develop.

1.1 Heart Function & Disease

The heart is a muscular organ located behind and slightly left to the breastbone that consists of two atria and two ventricles. The right atrium receives the blood from the veins and pumps it to the right ventricle, which pumps the blood to the lungs, where it becomes oxygenated. The left atrium receives the oxygenated blood and sends it to the left ventricle. Then the oxygen-rich blood is sent to the rest of the body by the contractions of the left ventricle. Ideally, blood flows unobstructedly from the atria to the ventricles and the rest of the body during the circulation process.

This ideal flow of the blood vastly depends on the function of the heart's four valves

that ensure that blood unrestrictedly flows in the right direction through the heart. The tricuspid valve, located between the right atrium and the right ventricle, ensures that blood does not flow backward to the right atrium, from where the blood will move through the pulmonary valve to the lungs. The mitral valve is responsible for the unidirectional flow between the left atrium and the left ventricle, and finally, the aortic valve ensures that the flow between the left ventricle and the aorta is both unidirectional and unobstructed.

Congenital Heart Disease and the Pulmonary Valve

Congenital heart disease (CHD) is the most common type of congenital disability. As such, it is a major contributor to morbidity and mortality in both children and adults. Around 9 out of 1000 newborns have a structural-developmental heart defect (van der Linde et al., 2011). While CHD refers to a broad spectrum of different defects and symptoms, there have been significant advancements during the past decades in treating these patients, improving their quality of life and life expectancy.

Roughly 20% of CHD involves problems with the right ventricular outflow tract (RVOT) such as pulmonary atresia, pulmonary valve stenosis, tetralogy of fallot, double outlet right ventricle. These patients will undergo RVOT reconstruction, possibly with a pulmonary valve replacement. Pulmonary valve replacement is also required for patients who undergo the Ross procedure (Ross, 1967) during which the healthy pulmonary valve of the patient is used to replace his/her diseased aortic valve, and a valve substitute (usually a homograft, human donor valve, or pericardial valve) is implanted in the pulmonary position.

Patients' heart function is closely monitored echocardiographically after RVOT reconstruction as these patients are at risk of homograft failure. More specifically, the pulmonary valve function might deteriorate over time and suffer from stenosis or pulmonary regurgitation. The former means that the valve is narrower than it should, causing higher pressure as the blood flows through the heart, and the latter means that the valve fails to fulfill its primary task, unidirectional flow of the blood, as some of the blood leaks back to the ventricle. Other complications include endocarditis or children that may outgrow their homograft due to somatic growth. If such issues occur, the patients might need an (additional) pulmonary valve replacement with an allograft/homograft.

Such additional valve replacements are challenging to practicing physicians as the optimal timing for such re-interventions is unknown. Therefore, novel methods to model the progression of pulmonary valve function in relation with patient outcomes would be beneficial to optimize the timing of these interventions.

Left Ventricular Assist Device after Heart Failure

Heart failure is a chronic, progressive condition in which the heart cannot to fulfill the body's needs for blood and oxygen. It is amongst the most common causes of hospitalization, affecting about 40 million people worldwide. Given that it commonly affects the older population and that life expectancy is increasing, it is reasonable to assume that the number of patients with heart failure is expected to increase as well (Baman and F.S., 2020). Therefore effective prevention, treatment, and management of these patients are necessary.

There are several medical and device-based therapies for heart failure. Ventricular assist devices are rapidly emerging as a safe and durable therapy for end-stage heart failure patients (Han et al., 2018). These devices are also used as a bridge while waiting for heart transplantation. Despite the technological improvements of recent years, the burden of serious adverse events for subjects treated with LVAD is still cumbersome.

Common complications of patients treated with LVAD include thrombosis, dialysis, and kidney failure. Biomarkers associated with these complications, such as total bilirubin and creatinine, may shed more light on the inter-relationships between the processes that affect the risk for such complications in LVAD patients. (Antonides et al., 2020; Yalcin et al., 2019,0) Modeling the longitudinal evolution of these biomarkers while quantifying their association with patient outcomes can bridge the knowledge gaps in their interrelationships and improve decision making and treatment for these patients.

Tricuspid Regurgitation after Heart Transplantation

Heart transplantation (HTx) is often necessary to treat end-stage heart disease. Orthotopic heart transplantation (OHT) is considered the best treatment for advanced heart failure, with more than 4.000 performed annually in the world. Tricuspid regurgitation (TR) is considered the most common valvular complication after HTx with an incidence of 84% (Wong et al., 2008).

Causes for TR in patients with a heart transplant include endomyocardial biopsies, allograft rejection, a mismatch between the donor heart size and pericardial cavity dimensions (De Simone et al., 1995; Aziz et al., 1999; Caliskan et al., 2018). The clinical impact of TR after HTx remains still unclear. Post OHT is a dynamic disease with many potential etiologies. Repeated measurements of TR over time and dynamic predictive tools can aid in investigating its association with the risk of post-OHT failure and mortality and better assess its impact.

1.2 Motivating Datasets

Several observational studies and randomized clinical trials motivate the research presented in this thesis. While these studies come from various clinical research disciplines, they share in common that they are longitudinal. That means subjects are followed up over time, and several measures such as biomarkers and clinical endpoints are being monitored.

The datasets that we use in this thesis to motivate and subsequently demonstrate the statistical methods proposed in the thesis are briefly introduced in the following sections.

Pulmonary Gradient Dataset

This dataset comes from a study on 467 congenital heart disease (CHD) patients who underwent a right ventricular outflow tract reconstruction with a pulmonary allograft and were followed up echocardiographically after that at the Department of Cardio-Thoracic surgery of Erasmus University Medical Center. The pulmonary gradient of these patients was monitored repeatedly over time and is believed to be associated with the risk of death. Death is the primary endpoint of this study. However, during follow-up, 65 (13.92%) were reoperated and received a pulmonary allograft. The main interest in this study is to investigate the association between the pulmonary gradient and the risk of death and study the impact of reoperation on this risk and the pulmonary gradient. By doing so, predictive tools can be developed to quantify the potential benefit of reoperation for future patients. Table 1.1 gives a summary of the data.

Table 1.1: Data summary: Continuous variables are presented as median (IQR), categorical variables as counts (frequencies %).

	All	Non reoperated	Reoperated
Number of subjects	467	402 (86.9%)	65 (13.1%)
Number of repeated measurements	8 (5 - 12)	7 (4 - 11)	14 (11 - 16)
Number of Events	34 (7.3%)	32 (8.0%)	2 (3.0 %)
Sex = Male	272 (58.2%)	229 (57.0%)	43 (66.1 %)
Age	20.3 (9.5 - 30.9)	21.5 (9.4 - 31.2)	16 (10.9 - 22.4)

SPRINT Data

This dataset concerns 9361 subjects who participated in the SPRINT study (Group, 2015). Subjects with increased cardiovascular risk, systolic blood pressure of 130mm Hg or higher but without diabetes were randomized to intensive or standard treatment. The composite primary outcome was myocardial infarction, acute coronary syndromes, stroke, heart failure, or death from cardiovascular causes, while systolic blood pressure is the biomarker of interest that was repeatedly measured. During follow-up, 3424(36.6%) experienced serious adverse events (SAEs). In this study, it is of interest to assess the impact of SAEs both in the systolic blood pressure evolution and the risk for the event of interest and exploiting it to derive individualized, dynamic predictions for future patients with different scenarios for the occurrence or not of SAEs.

Women & Pregnancy after RVOT

This dataset concerns 196 women between 18 and 50 years old who underwent Right Ventricular Outflow Tract (RVOT) reconstruction between April 1986 and January 2018 with an allograft at the Cardio-Thoracic Surgery department of the Erasmus Medical Center and confirmed pregnancy during this period. Complete information of 176 (90%) allografts in 165 women were recorded, including 1395 repeated measurements of pulmonary gradient and regurgitation grade from echocardiograms. 51 (30.9%) of these women had 84 completed pregnancies at an average age of $29.1 \pm 3.9(SD)years$ at 8.1 ± 6.1 years since allograft implantation. Tetralogy of Fallot was the most common diagnosis in both groups. After a mean follow-up of 15.2 years (range 0.1 to 30), 7 (13.7%) parous women underwent valve replacement versus 20 (17.5%) nulliparous women. The interest in this study was to investigate the association of Pulmonary Gradient and Pulmonary regurgitation Grade with the risk of pulmonary valve replacement and how pregnancy influences this association.

LVAD after Heart Failure

This dataset includes 232 patients who experienced heart failure and subsequently received a Left Ventricular Assist Device (LVAD) implant during follow-up at the Erasmus University Medical Center in the Netherlands. These patients may suffer multiple complications during follow-up, such as thrombosis, embolic events, and dialysis, which are assumed to increase their risk of dying. To reduce this risk, treating physicians follow these patients closely and collect biomarker measurements for their liver (Total Bilirubin) and kidney (Creatinine) function and may decide to intervene

with an LVAD implant accordingly. LVAD is consequently expected to change their risk of experiencing a complication or dying and the trajectories of their liver and kidney biomarkers. In particular, LVAD takes over the heart function of the patient, and therefore, it is expected to increase the patient's survival chance while waiting for a heart transplant.

On the other hand, however, it might increase the risk of dialysis and/or thrombosis due to acute kidney injury and/or other complications. Muslem et al. (2018). In this study, 49 (21.1%) subjects experienced a complication, of which 23 (10.0%) subsequently died, while 80 (34.5%) subjects died without experiencing such a complication.

Tricusdpid Regurgitation after Heart Transplantation

This dataset includes 572 consecutive subjects who underwent biatrial orthotopic heart transplant (OHT) for 32 years between 1984 and 2016 in Erasmus MC. These patients were echocardiographically followed up over time, and their tricuspid regurgitation measurements were recorded as well. This study aims to investigate the evolution of TR over time and its association with the risk of death, as TR is common in patients with OHT.

1.3 Joint Models for Longitudinal and Time-to-Event Data

While the motivating datasets discussed in the previous section concern different populations and aspects of heart disease, they do have many features in common. More specifically, patients are being followed up over long periods generating repeated measurements of biomarkers of interest. During follow-up, it is often the case that important events occurred, such as changes in treatment, reintervention, serious adverse events, etc., that can change the course of the disease and the risk for the clinical endpoint of interest. There is the potential of informative dropout. The repeatedly measured outcomes are assumed to be associated with the time-to-event outcome of interest. The structure of this association and its magnitude are of primary interest.

Each of the features above poses a different challenge in terms of statistical modeling. While, as we will discuss, there are methods to address some of these challenges separately, the joint modeling framework is necessary to address these features in combination.

Analysis of Repeated Measurements

The distinguishing feature of repeatedly measured data is that multiple observations are collected from a single subject (or cluster in the general case). This is in contrast to cross-sectional data, where typically, an observation belongs to a different subject. The key in the distinction between the two comes from the fact that data points are independent of each other by default in the latter case, whereas in the former case, data points within-subjects (clusters) are not. Those measurements coming from the same subject are expected to be more correlated than measurements between subjects.

This (expected) systemic difference in the correlation between observations coming from the same subject against observations from different subjects poses an additional challenge in modeling and analyzing such data. A popular framework for dealing with this challenge is mixed-effects models. The key feature of these models is the introduction of latent effects, the so-called random effects, which are specific to each subject and quantify its deviation from the average of the population, which is described by the so-called fixed effects. In other words, random effects are unobservable to us, characteristics that capture and describe the subject-specific differences that the observable characteristics cannot explain.

Technically, mixed-effects models can be described as follows. Let $\boldsymbol{y}_{ki}$ denote the $n_{ki} \times 1$ longitudinal response vector for the k^{th} longitudinal outcome taken at time point $t_{kil}, l = 1, \ldots, n_{ki}$. Each longitudinal outcome k can be described by a generalized linear mixed model conditional on the vector of random effects $\boldsymbol{b}_{ki}$:

$$\begin{cases} g_k \left[E\{y_{ki}\left(t\right) \mid \boldsymbol{b}_{ki}\} \right] = \eta_{ki}\left(t\right) = \boldsymbol{x}_{ki}^{\top}\left(t\right)\boldsymbol{\beta_k} + \boldsymbol{z}_{ki}^{\top}\left(t\right)\boldsymbol{b}_{ki}, \\ \\ \boldsymbol{b}_{ki} = \begin{bmatrix} \boldsymbol{b}_{1i} \\ \boldsymbol{b}_{2i} \\ \vdots \\ \boldsymbol{b}_{ki} \end{bmatrix} \sim \mathcal{N}\left(\boldsymbol{0}, \boldsymbol{D}\right), \end{cases}$$

where $g_k\left(\cdot\right)$ is a general link function appropriate for the k^{th} outcome, $\boldsymbol{x}_{ki}$ and $\boldsymbol{z}_{ki}$ are design vectors for the fixed-effects' regression coefficients $\boldsymbol{\beta}_k$ and the random-effects $\boldsymbol{b}_{ki}$ respectively. The association between the K longitudinal outcomes is built via the random-effects $\boldsymbol{b}_{ki}$ and $\tilde{\boldsymbol{b}}_{ki}$ which are assumed to be normally distributed with mean zero and a $q \times q$ general variance-covariance matrix D.

The formulation discussed above covers the general case of multivariate longitudinal data of different types, such as continuous or categorical. When $k = 1$, this reduces

to the basic mixed-effects models for a single longitudinal outcome.

Missing Data & Dropout

Missing data is a common complication in longitudinal studies as well as cross-sectional studies. There is a variety of reasons for which a data point might be missing from a dataset. Depending on the reason, or using a more technical term, the mechanism that generates the missing data, and whether data are missing in the outcomes of interest or the covariates of interest, the difficulty of addressing the challenges due to missing data might differ. According to the taxonomy introduced by Rubin (1976) and Little and Rubin (2002) this may happen according to three distinct missing data mechanisms:

- **Missing Completely at Random (MCAR)**: Missingness depends on neither observed nor unobserved outcomes,
- **Missing at Random (MAR)**: Missingness is independent of the unobserved outcomes after conditioning on the observed outcomes, and
- **Missing Not at Random (MNAR)**: Conditioning on the observed outcomes cannot disentangle the dependence between the missingness process and the unobserved outcomes.

MCAR is the simplest of the above cases as the missing data can be considered a random sample of the complete dataset, and therefore analyzing the observed data will not generally lead to biased results. Under MAR, this is not the case, and the missingness generating process needs to be explicitly accounted for in the model. However, under specific conditions in (Bayesian) likelihood-based approaches, such as the mixed-effects models previously discussed, missingness can be termed ignorable and valid inferences can be derived by solely analyzing the observed data. Finally, under MNAR, missingness depends on information that is not available, and therefore the missingness process needs to be modeled explicitly and jointly with the outcome of interest. As there are infinite MNAR models and the distinction between MAR and MNAR is impossible (Molenberghs et al., 2008), this is usually conducted in a framework of sensitivity analysis where different scenarios about the missingness process are explored.

To achieve so, three general frameworks exist: pattern mixture models (PMM), selection models (SeM), and shared parameter models (SPM). Their difference comes in how the joint distribution of the outcome and missingness processes are factored

to conditionally independent components. In this thesis, we focus on SPMs under which the outcome and the missingness processes are assumed to be driven by a set of unobserved variables, the so-called random effects. The use of latent variables gives rise to the conditional independence assumption under which the outcome and the missingness processes are independent, given the set of random effects.

Analysis of Time-to-Event Data

Typically in follow-up studies, apart from longitudinal outcomes, the main interest is on a clinical endpoint which is recorded as a time-to-event outcome subject to censoring. Let us denote by T_i^* the true event time for subject i and let $T_i = \min\left(T_i^*, C_i\right)$ be the observed event time, which is defined as the minimum value between the true event time T_i^* and the censoring time C_i. Furthermore, let $\delta_i \in (0, 1, 2, 3)$ denote an event indicator which is equal to: 0 under right censoring ($T_i^* > T_i$), 1 when ($T_i^* = T_i$), 2 under left censoring ($T_i^* < T_i$) and 3 under interval censoring in which case an interval of observed event times with limits T_i^L and T_i^U are observed and the true event time lies within the interval ($T_i^L < T_i^* < T_i^U$). In this thesis we will focus on the case of right censoring and assume that $\delta_i = I(T_i^* \leq C_i)$.

Under this notation, time-to-event data may be analyzed using a relative risk model such as a Cox proportional hazards model (Cox, 1972) which is specified as follows:

$$\begin{aligned} h_i\left(t \mid \boldsymbol{w}_i\left(t\right)\right) &= \lim_{\Delta t \to 0} \frac{\Pr\left\{t \leq T_i^* < t + \Delta t \mid T_i^* \geq t, \boldsymbol{w}_i\right\}}{\Delta t} \\ &= h_0\left(t\right) \exp\left[\boldsymbol{\gamma}^\top \boldsymbol{w}_i\left(t\right)\right], \end{aligned}$$

where $h_0\left(t\right)$ is the baseline hazard, $\boldsymbol{w}_i\left(t\right)$ are baseline covariates with corresponding regression coefficients $\boldsymbol{\gamma}$.

Note that the baselines covariates $\boldsymbol{w}_i\left(t\right)$ are allowed to be time-varying which referes to the extended formulation for the time-dependent Cox model (Therneau and Grambsch, 2000; Andersen and Gill, 1982). In principle this formulation allows for repeatedly measured outcomes to enter the model as time-dependent covariates and subsequently quantify their association with the hazard for the event of interest. This formulation, however, is only appropriate for exogenous factors (such as time, for example) as it assumes that the observed values are measured without error and remain constant (value carries forward) until the next measurement. This is unlikely to be the case for endogenous covariates (such as heart function), which are measured with error, are subject to biological variation, and their path cannot be fully and deterministically specified between measurements. As a result, the time-dependent Cox model

unsuitable for investigating the association between endogenous repeatedly measured outcomes and the risk for an event of interest.

The Basic Joint Model

Joint models for longitudinal and time-to-event data, which belong to the broad family of shared-parameter models, are not subject to the limitations discussed above, namely MNAR and the endogeneity of time-dependent covariates (Rizopoulos, 2012; Tsiatis and Davidian, 2004; Wulfsohn and Tsiatis, 1997; Rizopoulos and Ghosh, 2011a; Chi and Ibrahim, 2006). The building block of the joint modeling framework is the assumption that both the longitudinal and time-to-event processes are driven by unobserved and shared random-effects conditional on which the two processes are independent. The intuition behind joint models is simple, instead of using the actual observed repeated measurements as time-dependent covariates in a relative risk model, we first estimate the true underlying longitudinal trajectory based on these measurements using a mixed-effects model and use it (or functions thereof) in the relative risk model. Figure 1.1 gives a graphical representation of such a joint model where it is seen that the smooth estimated longitudinal trajectory is associated with the hazard for an event rather than the step function connecting the observed measurements as it would be the case when using a time-dependent Cox model.

Formally, the specification of a joint model is given by:

$$\begin{cases} g_k\left[E\{y_{ki}\left(t\right) \mid \boldsymbol{b}_{ki}\}\right] = \eta_{ki}\left(t\right) = \boldsymbol{x}_{ki}^{\top}\left(t\right)\boldsymbol{\beta}_{\boldsymbol{k}} + \boldsymbol{z}_{ki}^{\top}\left(t\right)\boldsymbol{b}_{ki}, \\ h_i(t \mid \mathcal{M}_i(t), \boldsymbol{w}_i(t)) = h_0(t)\, \exp[\boldsymbol{\gamma}^{\top}\boldsymbol{w}_i(t) + \sum_{k=1}^{K}\sum_{l=1}^{L_K} f_{kl}\{\boldsymbol{\alpha}_{kl}, \boldsymbol{w}_i(t), \boldsymbol{b}_{ki}, \mathcal{M}_i(t)\}] \\ \boldsymbol{b}_{ki} = \begin{bmatrix} \boldsymbol{b}_{1i} \\ \boldsymbol{b}_{2i} \\ \vdots \\ \boldsymbol{b}_{ki} \end{bmatrix} \sim \mathcal{N}\left(\boldsymbol{0}, \boldsymbol{D}\right), \end{cases}$$

where $\mathcal{M}_i(t) = \{\mathcal{M}_{1i}(t), \ldots, \mathcal{M}_{ki}(t)\}$, and $\mathcal{M}_{ki}(t) = \{\eta_{ki}(s), 0 \leq s < t\}$ denotes the history of the true unobserved longitudinal process up to time t, and $\boldsymbol{w}_i(t)$ is a vector of exogenous, possibly time-varying, covariates with corresponding regression coefficients $\boldsymbol{\gamma}$. Functions f_{kl}, parameterized by vector $\boldsymbol{\alpha}_{kl}$, specify which components/features of each longitudinal outcome are included in the relative risk model. This for-

mulation covers the case of multiple longitudinal outcomes. In case of a single longitudinal outcome subscript k and the summation over the k outcomes can be dropped.

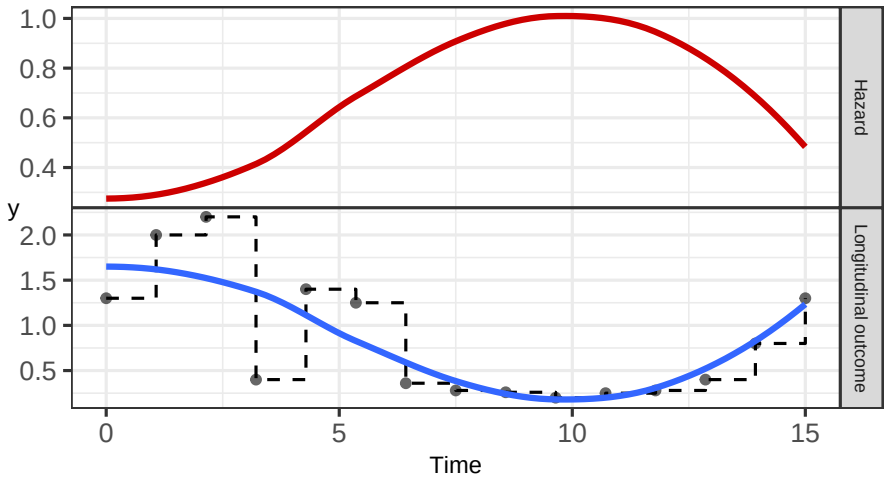

Figure 1.1: Illustration of a basic Joint Model with one longitudinal outcome and one event of interest.

Association Structures

One of the main advantages of joint modeling is that since the longitudinal marker's smooth, true underlying trajectory is used as a time-varying covariate in the relative risk model, there are no limitations in the functional form that can be used to relate the two processes. In its simplest form, function f_{kl} is

$$f\{\alpha, \boldsymbol{w}_i(t), \boldsymbol{b}_i, \mathcal{M}_i(t)\} = \alpha\eta_i(t),$$

known as the current value association. Under this structure, it is assumed that the value of the longitudinal marker(s) at time point t is associated with the hazard value for the event of interest at the same time point. Alternatively, one may assume that

$$f\{\alpha, \boldsymbol{w}_i(t), \boldsymbol{b}_i, \mathcal{M}_i(t)\} = \alpha\eta_i'(t), \quad \text{with} \quad \eta_i'(t) = \frac{d\eta_i(t)}{dt}$$

, which means that the rate of increase or decrease of the longitudinal marker at time t is associated with the hazard value for the event of interest at the same time point.

This structure is known as the current slope association. Finally, another common association structure that is useful in applications is the area under the curve association, given by:

$$f\{\alpha, \boldsymbol{w}_i(t), \boldsymbol{b}_i, \mathcal{M}_i(t)\} = \alpha \int_0^t \eta_i(s)\, ds.$$

Under this association structure, we assume that the cumulative burden of the longitudinal biomarker values up to time t is associated with the hazard value for the event of interest at the same time point. Figure 1.2 shows a graphical representation of all these three association structures.

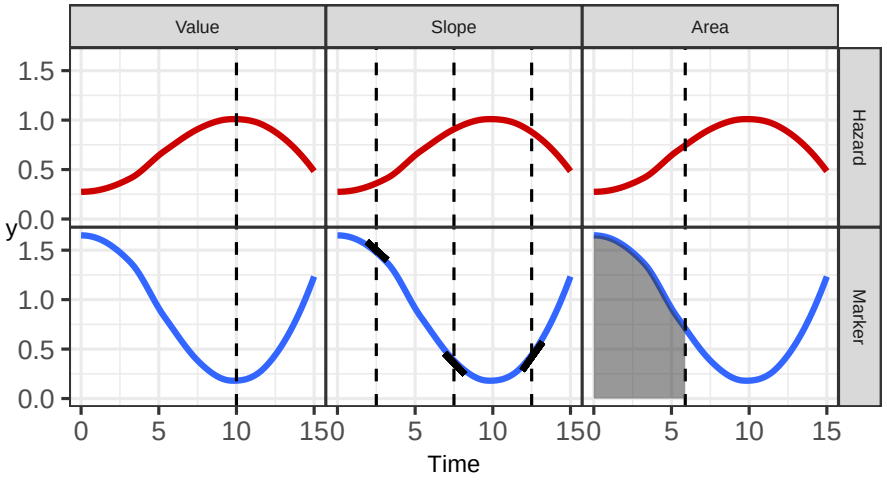

Figure 1.2: Illustration of the current value, current slope and cumulative effect association structures

1.4 Dynamic Predictions

Joint models can be used as the building block for a vital tool to aid personalized medicine in practice: dynamic predictions. The inherently dynamic nature of joint models can be exploited to build dynamic predictive tools that can be updated as new measurements for a specific subject are being gathered. This is of great importance in the field of heart disease (and overall in healthcare and medicine) as physicians need predictive tools tailored to the subject-specific characteristics of each patient to aid and improve their decision-making and facilitate communication of health risks.

More specifically and based on a joint model fitted to data, subject-specific probabilities and predictions for future biomarker levels, may derived for a new subject $i^{'}$ for whom there is a history of biomarker measurements $\mathcal{Y}_{ki'}(t) = \{y_{ki'}(t_{ki'l}); 0 \leq t_{ki'l} \leq t, k = 1, \ldots, n_k\}$ for the k^{th} longitudinal outcome and a vector of baseline covariates $w_{i'}$. The biomarker measurements are available up to a time point t and since the biomarker values are endogenous we assume that the event of interest has not occurred up to that point. n this setting, the conditional subject-specific survival probability that subject i^{new} survives at least up to a furtue time point $u > t$ is given by:

$$\pi_{i'}(u \mid t) = \Pr\left\{T_{i'}^{*} \geq u \mid T_{i'}^{*} > t, \mathcal{Y}_{1i'}(t), \ldots, \mathcal{Y}_{n_{ki'}}(t), w_{i'}\right\},$$

and the prediction for the subject's longitudinal response at u is given by:

$$\omega_{ki'}(u \mid t) = \mathrm{E}\left\{y_{ki'}(u) \mid T_{i'}^{*} > t, \mathcal{Y}_{1i'}(t), \ldots, \mathcal{Y}_{n_{ki'}}(t), w_{i'}\right\}.$$

Figure 1.3 gives a graphical representation of the concept of dynamic predictions. Both $\pi_{i'}(u \mid t)$ and $\omega_{ki'}(u \mid t)$ can be updated as new measurements for subject $i^{'}$ accrue. Estimation of both $\pi_{i'}(u \mid t)$ and $\omega_{ki'}(u \mid t)$ is based on the corresponding posterior predictive distributions (Yu et al., 2008; Proust-Lima and Taylor, 2009; Rizopoulos, 2011; Taylor et al., 2013)

1.5 Extensions

Intermediate Events & Adaptive Dynamic Predictions

The accuracy of dynamic predictions from joint models and their evaluation is a topic of active research. Recent work on the topic has revealed that the performance of dynamic predictions from joint models depends on the modeling of subject-specific longitudinal profiles and is sensitive to misspecification (Rizopoulos et al., 2017; Ferrer et al., 2019; Suresh et al., 2017). This is an important observation as longitudinal studies, due to their nature, are susceptible to intermediate events that may occur on any subject during follow-up and directly affect the shapes of their longitudinal trajectories. There are many examples of such intermediate events that may be under the investigator's control. For the pulmonary gradient dataset, some subjects received additional valves during follow-up. In the SPRINT dataset, subjects may have experienced serious adverse events. Pregnancy may or may not affect the evolution of pulmonary valve markers during follow-up. Receiving an LVAD can drastically change the subject's longitudinal trajectory and heart failure or transplantation risk.

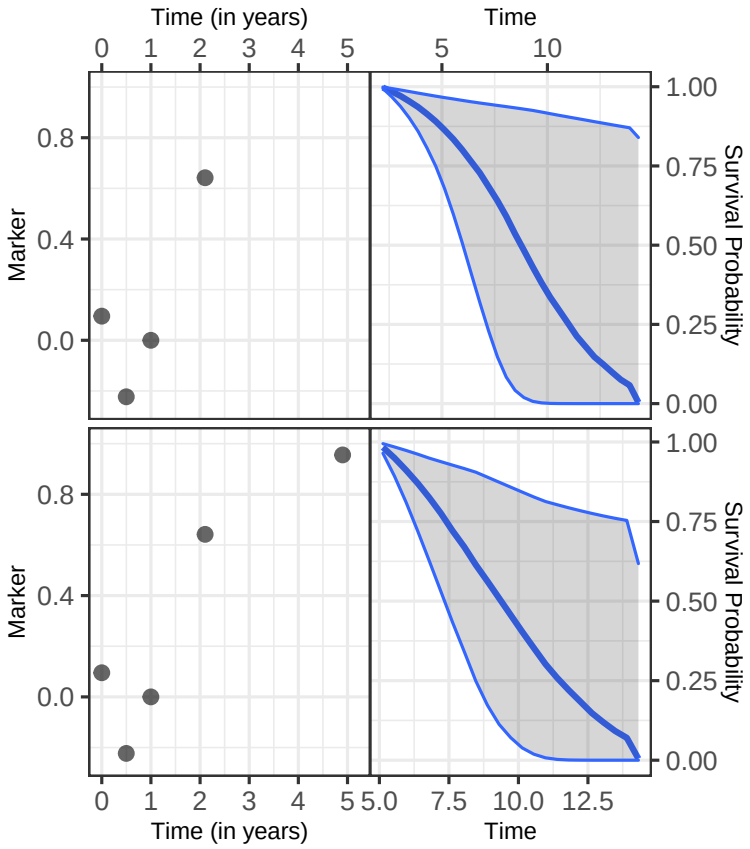

Figure 1.3: Illustration of subject-specific dynamic prediction of conditional survival probabilities after obtaining a new biomarker measurement.

Therefore, it is clear that such intermediate events are common in follow-up studies, and ignoring their occurrence might affect the quality of predictions for such cases. Despite that, there has not been much work in developing predictive tools that can incorporate intermediate events and adapt their prediction accordingly. Recent and interesting work on the topic has been produced by Sène et al. (2016) and Taylor et al. (2005) in the context of prostate cancer but only considered the longitudinal trajectories before the occurrence of the intermediate event.

In this thesis, we propose a different approach to include the occurrence of intermediate events in joint models and exploit them in adaptive dynamic prediction tools. We further propose adapted performance metrics for appropriately evaluating these tools.

Multi-state Models

The basic joint model is formulated for the case of a single longitudinal marker coupled with a single event of interest. In a previous section of this thesis, we presented a generalization that allows for multiple longitudinal markers and a single time-to-event outcome. However, it is very often the case that multiple events of interest may occur that may affect each other, as it is the case in the presence of competing risks (Andrinopoulou et al., 2014). A natural extension in this setting is the so-called multi-state models to which competing risks and single time-to-event outcomes are special cases of (Putter et al., 2007). This generalization allows for the inclusion of multiple disease states of interest and enables us to investigate the transition between such states. Such models can be advantageous in the field of heart failure and LVAD transplantation, where subjects may change between different disease states such as LVAD transplantation, thrombosis, heart failure, etc., and aid in understanding the interrelationship between biomarkers and the transition between such states (Muslem et al., 2018). Joint models for a single longitudinal marker and multi-state processes have been developed by Ferrer et al. (2016) under the frequentist paradigm with applications in the field of prostate cancer. In this thesis, we discuss extensions for the case of multiple biomarkers.

Feature Selection

The specification of the relationship between biomarkers and events of interest is not always straightforward. While clinical background, previous research, and subject-matter knowledge can aid in defining this relationship, it is often the case, especially in exploratory analyses and in the development of prediction models, that this relationship is still unknown. Previous work on the topic has revealed that different features

of longitudinal outcomes might be related to the events of interest (Rizopoulos and Ghosh, 2011b; Mauff et al., 2017). This makes selecting an appropriate association structure complicated in practice and creates a need for tools to aid in this selection when developing such models. The dimensionality of the task is also essential in this framework, as in the case of multiple longitudinal and time-to-event outcomes, the number of possible association structures increases substantially. For this purpose, Bayesian shrinkage priors have been used in the past to automates the selection process (Andrinopoulou and Rizopoulos, 2016). In this thesis, we extend the research on this topic by considering and assessing the performance of global-local shrinkage priors in the context of LVAD transplantation where multiple outcomes are of interest and information about their interrelationships is still unknown.

Missing Data & Sensitivity Analysis

Dropout is very common in longitudinal studies and can create complications in the subsequent analysis of the data depending on the causes of the dropout. As the distinction between MAR and MNAR is intractable, the standard approach in practice is to use an analysis valid under MAR and explore MNAR deviations from it as part of sensitivity analysis. Shared parameter models can theoretically be used for this purpose, but in practice, the SeM and PMM modeling frameworks are preferred because they entail a more straightforward MAR specification. There has been a lot of work recently towards achieving a MAR specification under the SPM framework with the introduction of the Generalized Shared Parameter Model (Creemers et al., 2010; Njagi et al., 2014). In this thesis, we propose an alternative specification which while simpler, can more easily be used in practice with readily available software.

1.6 Research Goals

Motivated by the studies discussed in previous sections, this thesis aims to extend the joint modeling framework in order to answer questions such as:

- **Intermediate Events & Adaptive Dynamic Predictions**:
  - Investigate the association between the evolution of pulmonary gradient and the risk of death. Quantify the impact of re-intervention on the course of heart disease for patients who underwent RVOT reconstruction with a pulmonary valve. Develop a dynamic prediction tool that is adaptive to different scenarios with respect to the decision to re-intervene or not to the patient.

  - Assess the impact of SAEs both in the systolic blood pressure evolution and the risk for the event of interest and exploiting it to derive individualized, dynamic predictions for future patients with different scenarios with respect to the occurrence or not of SAEs.
  - Investigate how and whether pregnancy as an intermediate event influences the long-term durability of allografts in RVOT.

- **Multi-state Models & Shrinkage**:
  - To investigate the impact of LVAD at different stages after heart failure and understand its association with predictive biomarkers and clinical outcomes.
  - To automate feature selection of biomarkers in the context of multivariate longitudinal and survival data by using Bayesian shrinkage priors.
  - Investigate factors are associated with tricuspid regurgitation after heart transplantation and how tricuspid regurgitation is associated with the risk of death.

- **Missing Data**:
  - To introduce an alternative framework for sensitivity analysis for missing data generating mechanisms in shared parameter joint models.
  - To propose guidelines for dealing with missing data in practice.

1.7 Outline

The rest of this thesis is structured as follows. In **PART I** we discuss the incorporation of intermediate events as time-varying covariates in joint models and their use in deriving dynamic predictions adaptive to different scenarios for the occurrence of these intermediate events. More specifically, in **Chapter** 2 we give the technical details of incorporating intermediate events such as reintervention or serious adverse events in joint models and for deriving dynamic predictions from these models. We also discuss measures for validation of the predictive performance of these models, such as the time-varying AUC and the PE. The approach is illustrated in two motivating datasets. In **Chapter** 3 we investigate the influence of pregnancy on the durability of allografts in RVOT. We consider pregnancy as an intermediate event, and we quantify its influence on biomarkers for valve replacement such as pulmonary regurgitation and the pulmonary gradient and its association with the risk of valve replacement.

In **PART II** we discuss feature selection of biomarkers in the presence of multiple longitudinal markers and multi-state processes by using Bayesian shrinkage priors. Specifically, in **Chapter** 4 we discuss specification and estimation of joint models for multivariate longitudinal outcomes and multiple processes under the Bayesian framework and how to use Bayesian shrinkage priors to automate the selection of the appropriate association structure between biomarkers and transitions between states of interest. In **Chapter** 5 we investigate the association between Tricuspid Regurgitation and the risk of dying in patients who underwent Heart Transplantation. We use bayesian shrinkage priors to select features of tricuspid regurgitation that are associated with the risk of death.

In **PART III** we discuss missing data. In **Chapter** 6 we discuss the Missing at Random characterization for the shared parameter models and propose its use for sensitivity analysis in cases with dropout. **Chapter** 7 we briefly discuss the subject of missing covariate data in practice and propose useful guidelines on how to deal with missing covariate data in practice.

1.8 References

Andersen, P. K. and Gill, R. D. (1982). Cox's regression model for counting processes: A large sample study. *The Annals of Statistics*, 10(4):1100–1120.

Andrinopoulou, E.-R. and Rizopoulos, D. (2016). Bayesian shrinkage approach for a joint model of longitudinal and survival outcomes assuming different association structures. *Statistics in Medicine*, 35(26):4813–4823.

Andrinopoulou, E.-R., Rizopoulos, D., Takkenberg, J. J. M., and Lesaffre, E. (2014). Joint modeling of two longitudinal outcomes and competing risk data. *Statistics in Medicine*, 33(18):3167–3178.

Antonides, C. F., Schoenrath, F., de By, T. M., Muslem, R., Veen, K., Yalcin, Y. C., Netuka, I., Gummert, J., Potapov, E. V., Meyns, B., Özbaran, M., Schibilsky, D., Caliskan, K., and the EUROMACS investigators (2020). Outcomes of patients after successful left ventricular assist device explantation: a euromacs study. *ESC Heart Failure*, 7(3):1085–1094.

Aziz, T. M., Burgess, M. I., Rahman, A. N., Campbell, C. S., Deiraniya, A. K., and Yonan, N. A. (1999). Risk factors for tricuspid valve regurgitation after orthotopic heart transplantation. *The Annals of Thoracic Surgery*, 68(4):1247 – 1251.

Baman, J. and F.S., A. (2020). Heart failure. *JAMA*, 324(10):1015.

Caliskan, K., Strachinaru, M., and Soliman, O. I. (2018). *Tricuspid Regurgitation in Patients with Heart Transplant*, pages 49–58. Springer International Publishing, Cham.

Chi, Y. and Ibrahim, J. (2006). Joint models for multivariate longitudinal and multivariate survival data. *Biometrics*, 62:432–445.

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

Creemers, A., Hens, N., Aerts, M., Molenberghs, G., Verbeke, G., and Kenward, M. G. (2010). A sensitivity analysis for shared-parameter models for incomplete longitudinal outcomes. *Biometrical Journal*, 52(1):111–125.

De Simone, R., Lange, R., Sack, F.-U., Mehmanesh, H., and Hagl, S. (1995). Atrioventricular valve insufficiency and atrial geometry after orthotopic heart transplantation. *The Annals of Thoracic Surgery*, 60(6):1686 – 1693. Thirty-second Annual Meeting.

Ferrer, L., Putter, H., and Proust-Lima, C. (2019). Individual dynamic predictions using landmarking and joint modelling: Validation of estimators and robustness assessment. *Statistical Methods in Medical Research*, 28(12):3649–3666. PMID: 30463497.

Ferrer, L., Rondeau, V., Dignam, J., Pickles, T., Jacqmin-Gadda, H., and Proust-Lima, C. (2016). Joint modelling of longitudinal and multi-state processes: application to clinical progressions in prostate cancer. *Statistics in Medicine*, 35(22):3933–3948.

Group, T. S. R. (2015). A randomized trial of intensive versus standard blood-pressure control. *New England Journal of Medicine*, 373(22):2103–2116. PMID: 26551272.

Han, J. J., Acker, M. A., and Atluri, P. (2018). Left ventricular assist devices. *Circulation*, 138(24):2841–2851.

Little, R. J. A. and Rubin, D. B. (2002). *Statistical Analysis with Missing Data*. Wiley, New York.

Mauff, K., Steyerberg, E. W., Nijpels, G., van der Heijden, A. A., and Rizopoulos, D. (2017). Extension of the association structure in joint models to include weighted cumulative effects: K. MAUFF *ET AL* . *Statistics in Medicine*, 36(23):3746–3759.

Molenberghs, G., Beunckens, C., Sotto, C., and Kenward, M. G. (2008). Every missingness not at random model has a missingness at random counterpart with equal fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(2):371–388.

Muslem, R., Caliskan, K., Akin, S., Sharma, K., Gilotra, N. A., Constantinescu, A. A., Houston, B., Whitman, G., Tedford, R. J., Hesselink, D. A., Bogers, A. J., Russell, S. D., and Manintveld, O. C. (2018). Acute kidney injury and 1-year mortality after left ventricular assist device implantation. *The Journal of Heart and Lung Transplantation*, 37(1):116 – 123.

Njagi, E. N., Molenberghs, G., Kenward, M. G., Verbeke, G., and Rizopoulos, D. (2014). A characterization of missingness at random in a generalized shared-parameter joint modeling framework for longitudinal and time-to-event data, and sensitivity analysis. *Biometrical Journal*, 56(6):1001–1015.

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

Putter, H., Fiocco, M., and Geskus, R. B. (2007). Tutorial in biostatistics: competing risks and multi-state models. *Statistics in Medicine*, 26(11):2389–2430.

Rizopoulos, D. (2011). Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data. *Biometrics*, 67:819–829.

Rizopoulos, D. (2012). *Joint Models for Longitudinal and Time-to-Event Data, with Applications in R*. Chapman & Hall/CRC, Boca Raton.

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

Rizopoulos, D. and Ghosh, P. (2011b). A bayesian semiparametric multivariate joint model for multiple longitudinal outcomes and a time-to-event. *Statistics in Medicine*, 30(12):1366–1380.

Rizopoulos, D., Molenberghs, G., and Lesaffre, E. M. (2017). Dynamic predictions with time-dependent covariates in survival analysis using joint modeling and landmarking. *Biometrical Journal*, 59(6):1261–1276.

Ross, D. (1967). Replacement of aortic and mitral valves with a pulmonary autograft. *The Lancet*, 290(7523):956 – 958. Originally published as Volume 2, Issue 7523.

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63(3):581–592.

Sène, M., Taylor, J. M., Dignam, J. J., Jacqmin-Gadda, H., and Proust-Lima, C. (2016). Individualized dynamic prediction of prostate cancer recurrence with and without the initiation of a second treatment: Development and validation. *Statistical Methods in Medical Research*, 25(6):2972–2991.

Suresh, K., Taylor, J. M., Spratt, D. E., Daignault, S., and Tsodikov, A. (2017). Comparison of joint modeling and landmarking for dynamic prediction under an illness-death model. *Biometrical Journal*, 59(6):1277–1300.

Taylor, J. M., Yu, M., and Sandler, H. M. (2005). Individualized predictions of disease progression following radiation therapy for prostate cancer. *Journal of Clinical Oncology*, 23(4):816–825. PMID: 15681526.

Taylor, J. M. G., Yongseok, P., Ankerst, D. P., Proust□Lima, C., Williams, S., Kestin, L., Bae, K., Pickles, T., and Sandler, H. (2013). Real□time individual predictions of prostate cancer recurrence using joint models. *Biometrics*, 69(1):206–213.

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

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

van der Linde, D., Konings, E. E., Slager, M. A., Witsenburg, M., Helbing, W. A., Takkenberg, J. J., and Roos-Hesselink, J. W. (2011). Birth prevalence of congenital heart disease worldwide: A systematic review and meta-analysis. *Journal of the American College of Cardiology*, 58(21):2241 – 2247.

Wong, R. C.-C., Abrahams, Z., Hanna, M., Pangrace, J., Gonzalez-Stawinski, G., Starling, R., and Taylor, D. (2008). Tricuspid regurgitation after cardiac transplantation: An old problem revisited. *The Journal of Heart and Lung Transplantation*, 27(3):247 – 252.

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

Yalcin, Y. C., Muslem, R., Veen, K. M., Soliman, O. I., Hesselink, D. A., Constantinescu, A. A., Brugts, J. J., Manintveld, O. C., Fudim, M., Russell, S. D., Tomashitis, B., Houston, B. A., Hsu, S., Tedford, R. J., Bogers, A. J., and Caliskan, K. (2020). Impact of continuous flow left ventricular assist device therapy on chronic kidney disease: A longitudinal multicenter study. *Journal of Cardiac Failure*, 26(4):333 – 341.

Yalcin, Y. C., Muslem, R., Veen, K. M., Soliman, O. I., Manintveld, O. C., Darwish Murad, S., Kilic, A., Constantinescu, A. A., Brugts, J. J., Alkhunaizi, F., Birim, O., Tedford, R. J., Bogers, A. J. J. C., Hsu, S., and Caliskan, K. (2019). Impact of preoperative liver dysfunction on outcomes in patients with left ventricular assist devices. *European Journal of Cardio-Thoracic Surgery*, 57(5):920–928.

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

Part I

Intermediate Events & Adaptive Dynamic Predictions

Chapter 2

Individualized dynamic prediction of survival with the presence of intermediate events

This chapter is based on:

Papageorgiou, G., Mokhles, M.M., Takkenberg, J.J.M. and Rizopoulos, D. (2019), Individualized dynamic prediction of survival with the presence of intermediate events *Statistics in Medicine*, 38: 5623–5640. doi: https://doi-org.eur.idm.oclc.org/10.1002/sim.8387

Abstract

Often, in follow-up studies, patients experience intermediate events, such as reinterventions or adverse events, which directly affect the shapes of their longitudinal profiles. Our work is motivated by two studies in which such intermediate events have been recorded during follow-up. In both studies, we are interested in the change of the longitudinal evolutions after the occurrence of the intermediate event and in utilizing this information to improve the accuracy of dynamic prediction of their risk. To achieve so, we propose a flexible joint modeling framework for longitudinal and time-to-event data, which includes features of the intermediate event as time-varying covariates in both the longitudinal and survival submodels. We consider a set of joint models that postulate different effects of the intermediate event in the longitudinal profile and the risk of the clinical endpoint, with different formulations for the association structure while allowing its functional form to change after the occurrence of the intermediate event. Based on these models, we derive dynamic predictions of conditional survival probabilities which are adaptive to different scenarios with respect to the occurrence of the intermediate event. We evaluate the predictive accuracy of these predictions with a simulation study using the time-dependent area under the receiver operating characteristic curve and the expected prediction error adjusted to our setting. The results suggest that accounting for the changes in the longitudinal profiles and the instantaneous risk for the clinical endpoint is important, and improves the accuracy of the dynamic predictions.

2.1 Introduction

Nowadays there is great interest in the medical field for predictive tools that facilitate precision medicine. In the context of follow-up studies, in which patients are monitored with several longitudinally measured parameters and biomarkers, physicians are interested in utilizing this information for predicting clinical endpoints. In this setting, joint models for longitudinal and survival outcomes provide a flexible framework to study the association between these outcomes and derive dynamic individualized predictions (Wulfsohn and Tsiatis, 1997; Tsiatis and Davidian, 2004; Rizopoulos, 2012).

The evaluation of the accuracy of these predictions obtained from joint models has gathered a lot of attention lately (Rizopoulos et al., 2017; Ferrer et al., 2019; Suresh et al., 2017). An important observation that has been made is that the accuracy of the derived predictions is influenced by the appropriate modeling of the subject-specific longitudinal profiles. In that regard, often in follow-up studies intermediate events occur in some patients that directly affect the shapes of their longitudinal evolutions. These may include events that are either directly in the control of the investigators, such as additional reinterventions, or may be not, such as adverse events that the patients may experience. While such intermediate events are common, very little work has been done in the direction of developing predictive tools that account for them and are adaptive to different scenarios with respect to their occurrence. To our knowledge, only Sène et al. (2016) and Taylor et al. (2005) investigated this topic in the context of prostate cancer recurrence and radiotherapy as an intermediate event. In their approach, however, they only considered the biomarker trajectories up to the occurrence of the intermediate event assuming extrapolation of the longitudinal profile thereafter. That is, changes in the shape of the longitudinal profile due to the occurrence of the intermediate event were not accounted for. Our goal is to show that utilizing the whole longitudinal trajectory, while capturing the changes to its shape due to the occurrence of intermediate events, can considerably improve the accuracy of such predictions.

In our work, we are motivated by two studies in which such intermediate events have been recorded during follow-up. The first study concerns 467 Congenital Heart Diseased (CHD) patients who underwent a Right Ventricular Outflow Tract (RVOT) reconstruction with a pulmonary valve and were followed-up echocardiographically thereafter, at the Department of Cardio-Thoracic surgery of Erasmus University Medical Center. Death is considered as the study endpoint while pulmonary gradient is the biomarker of interest which is believed to be related to the risk of death. During follow-up 65 (13.92%) were reoperated and received a pulmonary allograft. In Figure

2.1 the pulmonary gradient evolutions of 4 randomly selected patients, one from each combination between the reoperation and event status, are shown and it can be seen that when a patient is reoperated the pulmonary gradient drops. The interest in this study lies in the association between the pulmonary gradient and the risk of death but the main focus is to study the impact of reoperation on the risk of death both directly and indirectly (i.e. through its association with the pulmonary gradient) in order to develop predictive tools that can quantify the potential benefit of reoperation for future patients. The second study concerns 9361 subjects who participated in the SPRINT study (Group, 2015). Subjects with increased cardiovascular risk, systolic blood pressure of 130 mm Hg or higher but without diabetes, were randomized to intensive or standard treatment. The composite primary outcome was myocardial infarction, acute coronary syndromes, stroke, heart failure or death from cardiovascular causes while systolic blood pressure is the biomarker of interest which was repeatedly measured. During follow-up 3424 (36.6%) experienced serious adverse events (SAE). In Figure 2.2 the systolic blood pressure evolutions over time of the subjects who experienced SAEs and those who did not are depicted along with a loess curve for the average evolution over time for each group. The interest lies in assessing the impact of serious adverse events both in the systolic blood pressure evolution and the risk for the event of interest and exploiting it to derive individualized dynamic predictions for future patients with different scenarios with respect to the occurrence or not of SAEs. A more detailed description of the datasets is on the online supplementary material Sections S1.1 and S1.2.

In both studies, physicians are interested in obtaining predictions of the respective clinical endpoints. However, to provide predictions of adequate accuracy, it will be required to carefully model the subject-specific longitudinal trajectories. Borrowing ideas from piecewise regression models, we achieve this by explicitly introducing the occurrence of these intermediate events as binary time-varying covariates in the specification of both the longitudinal and survival submodels of the joint model. The regression coefficient associated with this covariate can then capture changes, due to the occurrence of the intermediate event, in both the biomarker trajectory and the hazard for the event of interest. Furthermore, we allow features of the biomarker trajectory, such as the rate of change to differ after the occurrence of the intermediate event. This allows us to estimate the impact of intermediate events as well as their specific features which then can be utilized in deriving dynamic predictions for a future patient under different scenarios: for example how the risk of a patient changes assuming different treatment strategies, such as no reintervention, reintervention now or reintervention at a later time-point.

The rest of the paper is structured as follows. Section 2.2 describes the formulation

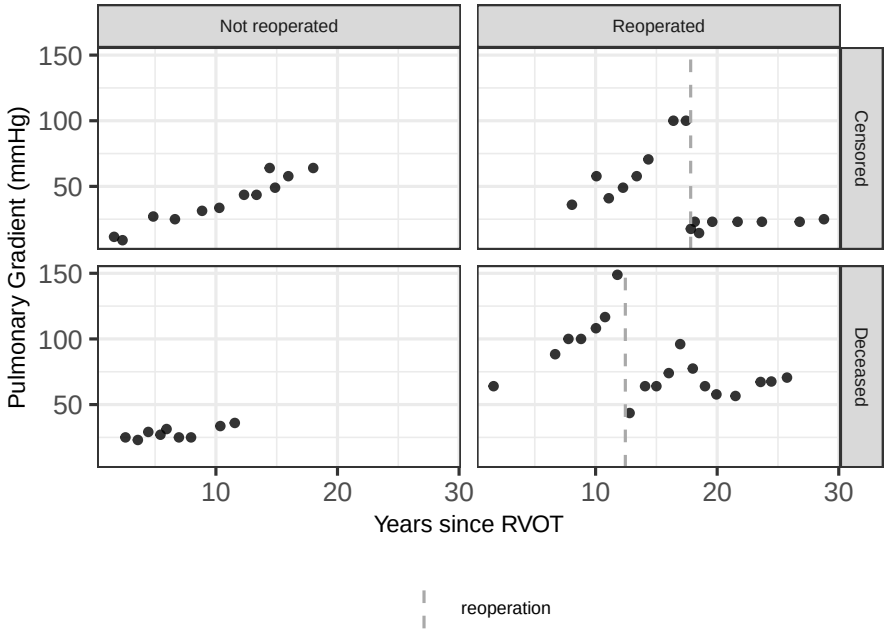

Figure 2.1: Pulmonary Gradient profile of 4 randomly selected patients, one from each of the following categories: Not Reoperated and Censored, Not Reoperated and Deceased, Reoperated and Censored, Reoperated and Deceased. The vertical red dashed lines depict the time of reoperation.

of the joint model in the presence of intermediate events. Section 2.3 presents the individualized dynamic predictions under different scenarios with respect to the occurrence of intermediate events and measures of predictive accuracy. In Section 2.4 we present the results of the analyses of the two motivating studies while in Section 2.5 we show the results of a simulation study. Finally, in Section 2.6 we close with a discussion.

2.2 Joint Model for Longitudinal and Time-to-Event Data with an Intermediate Event

Assuming n individuals under study, let $\mathcal{D}_n = \{T_i, \delta_i, \boldsymbol{y}_i, \rho_i; \quad i = 1, \ldots, n\}$ denote the sample from the target population, where $T_i = \min\left(T_i^*, C_i\right)$ denotes the observed event time, which is defined as the minimum value between the true event time T_i^* and the

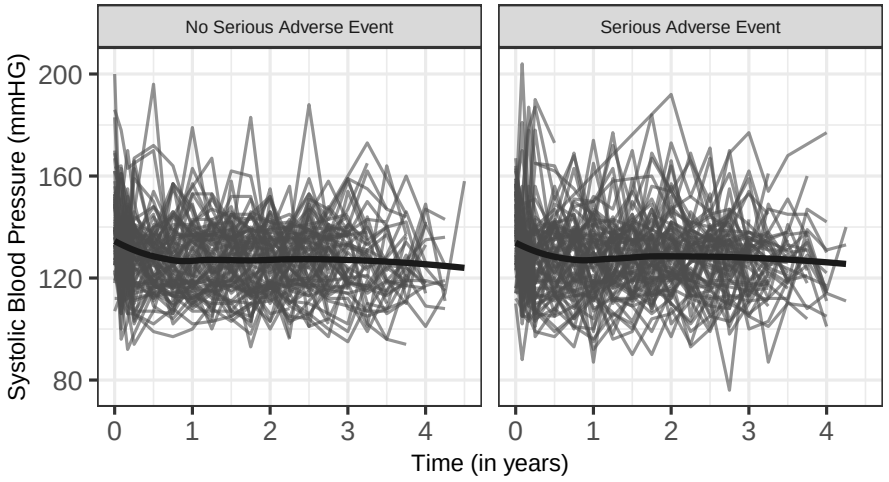

Figure 2.2: Individual evolutions and loess splines curves (solid thick lines) of systolic blood pressure for 80 randomly selected subjects who experienced a Serious Adverse Event and for 80 randomly selected subjects who did not experience a Serious Adverse Event.

censoring time C_i, and $\delta_i = I\left(T_i^* \leq C_i\right)$ the event indicator with $I(\cdot)$ being the indicator function which equals 1 if $T_i^* \leq C_i$ and 0 otherwise. Moreover, let ρ_i denote the time to the intermediate event with a corresponding indicator $R_i(t) = I(t \geq \rho_i)$ at any time t during follow-up, which takes the value 1 if a subject experienced the intermediate event and 0 otherwise. Furthermore, let $t_{i+} = \max\left(0, t_{ij} - \rho_i; j = 1, \ldots, n_i\right)$ denote the time relative to the occurrence of the intermediate event and $\boldsymbol{y}_i$ be the vector of size $n_i \times 1$ of repeated measurements for the i^{th} subject, with element y_{ij} being the observed value of the longitudinal outcome at time point $t_{ij}, \quad j = 1, \ldots, n_i$. We assume $\boldsymbol{y}_i$ to be a contaminated with measurement error version of the true and unobserved value of the longitudinal outcome at any time t: $\boldsymbol{y}_i(t) = \boldsymbol{\eta}_i(t) + \boldsymbol{\epsilon}_i(t)$ with $\boldsymbol{\eta}_i(t)$ denoting the true value of the longitudinal outcome at time t and measurement error $\boldsymbol{\epsilon_i}(\boldsymbol{t}) \sim \mathcal{N}\left(0, \sigma^2 I_{n_i}\right)$. The true level of the longitudinal outcome is then formulated as:

$$\eta_i(t) = \begin{cases} \boldsymbol{x}_i^\top(t)\boldsymbol{\beta} + \boldsymbol{z}_i^\top(t)\boldsymbol{b}_i, 0 < t < \rho_i \\ \\ \boldsymbol{x}_i^\top(t)\boldsymbol{\beta} + \boldsymbol{z}_i^\top(t)\boldsymbol{b}_i + \tilde{\boldsymbol{x}}_i^\top(t)\tilde{\boldsymbol{\beta}} + \tilde{\boldsymbol{z}}_i^\top(t)\tilde{\boldsymbol{b}}_i, t \geq \rho_i, \end{cases} \tag{2.1}$$

where $\boldsymbol{x}_i^\top(t)$ and $\boldsymbol{z}_i^\top(t)$ are design vectors for the fixed-effects regression coefficients $\boldsymbol{\beta}$ and the random-effects $\boldsymbol{b_i}$ respectively. Design vectors $\tilde{\boldsymbol{x}}_i^\top(t)$ and $\tilde{\boldsymbol{z}}_i^\top(t)$ include any function of the time-varying covariates $R_i(t)$ and t_{i+}, which describe the changes of the longitudinal trajectory after the occurrence of the intermediate event. These changes are then captured by the corresponding fixed-effects regression coefficients $\tilde{\boldsymbol{\beta}}$ allowing for subject-specific variation via the random-effects $\tilde{\boldsymbol{b}}_i$. The random-effects $\boldsymbol{b}_i$ and $\tilde{\boldsymbol{b}}_i$ are assumed to be normally distributed with mean zero and a $q \times q$ variance-covariance matrix $\boldsymbol{D}$.

Depending on how the trajectory of the biomarker changes after the occurrence of the intermediate event, the specification of $\tilde{\boldsymbol{x}}_i^\top(t)$ and $\tilde{\boldsymbol{z}}_i^\top(t)$ may vary. Let $g\{R_i(t), t_{i+}\} = \tilde{\boldsymbol{x}}_i^\top(t)\tilde{\boldsymbol{\beta}} + \tilde{\boldsymbol{z}}_i^\top(t)\tilde{\boldsymbol{b}}_i$ denote the part of (2.1) which describes the changes in the longitudinal profile after the occurrence of the intermediate event. Then in a setting as the one illustrated in Figure 2.1, for the Pulmonary Gradient dataset, where the longitudinal trajectory is characterized by a seemingly linear evolution before and after the occurrence of the intermediate event, a steep drop at the occurrence of the intermediate event and a potential change in the slope after the occurrence of the intermediate event, function $g\{R_i(t), t_{i+}\}$ could be specified as $R_i(t)\left(\tilde{\beta}_1 + \tilde{b}_{i1}\right) + t_{i+}\left(\tilde{\beta}_2 + \tilde{b}_{i2}\right)$. That is, the steep drop at the occurrence of the intermediate event will be captured by $\left(\tilde{\beta}_1 + \tilde{b}_{i1}\right)$ and the change in the slope after the occurrence of the intermediate event will be captured by $\left(\tilde{\beta}_2 + \tilde{b}_{i2}\right)$. On the other hand, in the setting of the SPRINT data where the longitudinal profiles show a nonlinear evolution over time without steep sudden changes, function $g\{R_i(t), t_{i+}\}$ could be specified as $\sum_{k=0}^{Q}\left(\tilde{\beta}_k + \tilde{b}_{ik}\right)B_k(t_{i+}, \boldsymbol{k})$ where $B_k(t_{i+}, \boldsymbol{k})$ denotes the k^{th} basis function of a B-spline with knots $k_1, \ldots, k_Q$. In that case the change in the shape of the nonlinear longitudinal trajectory after the occurrence of the intermediate event will be captured by $\sum_{k=1}^{Q}\left(\tilde{\beta}_k + \tilde{b}_{ik}\right)$ while there is no need to include $R_i(t)$. Generally, the functional form of $g\{R_i(t), t_{i+}\}$ may vary allowing for a broad range of mixed-effects models that can capture various types of changes in the longitudinal profile after the occurrence of the intermediate event.

Let $\mathcal{H}_i(t, \rho_i) = [\eta_i\{s, \rho_i(s)\},\ \ 0 \leq s \leq t]$ denote the history of the longitudinal outcome up to time t. Note that in the definition of the history of the longitudinal outcome, we explicitly indicate that the true underlying value of the longitudinal outcome is also a function of the time to the intermediate event ρ_i. That is, to highlight that the subject-specific trajectory, $\eta_i(t)$, differs from the occurrence of the intermediate event onwards. Then the effects of the longitudinal outcome and the intermediate event, while adjusting for other covariates, on the risk for an event are quantified by utilizing a relative risk model of the form:

$$h_i\{t \mid \mathcal{H}_i(t,\rho_i), \boldsymbol{w}_i\} = \begin{cases} h_0(t)\exp\left[\gamma^\top \boldsymbol{w}_i + f_{t<\rho_i}\{\mathcal{H}_i(t,\rho_i), \boldsymbol{b}_i\}^\top \boldsymbol{\alpha}\right], 0 < t < \rho_i \\ h_0(t)\exp\left[\gamma^\top \boldsymbol{w}_i + \mathbf{1}\zeta + f_{t\geq\rho_i}\{\mathcal{H}_i(t,\rho_i), \boldsymbol{b}_i\}^\top \boldsymbol{\alpha}\right], t \geq \rho_i, \end{cases} \tag{2.2}$$

where $h_0(t)$ is the baseline risk function and $\boldsymbol{w}_i$ is a vector of baseline covariates with a corresponding vector of regression coefficients γ. The effect of the intermediate event on the risk is captured by the regression coefficient ζ which quantifies the change in risk from the the occurrence of the intermediate event onwards. Furthermore, the hazard of an event for patient i at any time t is associated with the subject-specific trajectory, $\eta_i(t)$, through $f\{(\mathcal{H}_i(t,\rho_i), \boldsymbol{b}_i\}$ which is a function of the history of the longitudinal outcome up to time $\mathcal{H}_i(t,\rho_i)$ and/or the vector of subject-specific effects $\boldsymbol{b}_i$. Function $f_{(t,\rho_i)}\{(\mathcal{H}_i(t,\rho_i), \boldsymbol{b}_i\}$ determines the association structure between the longitudinal and the time-to-event processes while the corresponding vector of regression coefficients $\boldsymbol{\alpha}$ quantifies the magnitude of the association. Several functional forms for the specification of the association structure have been used in the literature, such as the current value, current slope and the cumulative effect (Mauff et al., 2017). The functional form of the association structure is an important feature of the joint model formulation, especially with regard to the accuracy of the dynamic predictions (Rizopoulos et al., 2017,0; Andrinopoulou et al., 2017). Hence, to allow for more flexibility we explicitly allow for the functional form of the association structure to differ before and after the occurrence of the intermediate event. In general, any functional form can be used for $f_{t\geq\rho_i}\{\mathcal{H}_i(t,\rho_i), \boldsymbol{b}_i\}$ and $f_{t<\rho_i}\{\mathcal{H}_i(t,\rho_i), \boldsymbol{b}_i\}$ including, of course, the case where the association structure remains the same and $f_{t\geq\rho_i}\{\mathcal{H}_i(t,\rho_i), \boldsymbol{b}_i\} = f_{t<\rho_i}\{\mathcal{H}_i(t,\rho_i), \boldsymbol{b}_i\}$.

The estimation of the parameters of the proposed joint model is achieved under the Bayesian framework using Markov Chain Monte Carlo (MCMC) algorithms. For more details regarding Bayesian estimation of joint models the reader may refer to Rizopoulos (2016), Ibrahim et al. (2001)] and Brown et al. (2005).

2.3 Individualized Dynamic Predictions with Time-varying Intermediate Events

2.3.1 Dynamic Predictions

Based on the joint model fitted in the sample $\mathcal{D}_n = \{T_i, \delta_i, \boldsymbol{y}_i, \rho_i;\ \ i = 1, \ldots, n\}$ from the target population, dynamic predictions for a new subject j from the same population can be derived up to a future time of interest $u > t$ given his/hers biomarker

history $\mathcal{H}_j(t) = [\eta_j\{s, R_j(s)\},\ 0 \le s \le t]$. Let $\mathcal{Y}_j(t) = \{y_j\left(t_{jl}\right); 0 \le t_{jl} \le t, l = 1, \ldots, n_j\}$ denote the history of observed biomarker values taken up to time t for patient j, then under the Bayesian joint model framework these predictions can be estimated using the corresponding posterior predictive distributions, namely:

$$\begin{aligned}
\pi_j(u \mid t) &= Pr\{T_j^* \ge u \mid T_j^* > t, \mathcal{Y}_j(t), \boldsymbol{\theta}\} \\
&= \int \frac{S_j\{u \mid \mathcal{H}_j\left(u, \boldsymbol{b}_j\right), \boldsymbol{\theta}\}}{S_j\{t \mid \mathcal{H}_j\left(t, \boldsymbol{b}_j\right), \boldsymbol{\theta}\}} p\{\boldsymbol{b}_j \mid T_j^* > t, \mathcal{Y}_j(t), \boldsymbol{\theta}\} d\boldsymbol{b}_j \\
&= \int \exp\left[-\int_t^u h_j\{s \mid \mathcal{H}_j\left(u, \boldsymbol{b}_j\right)\} ds\right] p\{\boldsymbol{b}_j \mid T_j^* > t, \mathcal{Y}_j(t), \boldsymbol{\theta}\} d\boldsymbol{b}_j, \\
&\quad t \le s \le u.
\end{aligned} \tag{2.3}$$

Note that in equation 2.3 and for the remainder of the text, covariates w_j and x_j are suppressed from notation for simplicity and without loss of generality. Expressing the fraction term in (2.3) as $\exp\{-\int_t^u h_i(s)\, ds\}, t \le s \le u$ has two main advantages. First, it reduces the computational time required, since the denominator part $S_j\{t \mid \mathcal{H}_j\left(t, \boldsymbol{b}_j\right), \boldsymbol{\theta}\}$ does not need to be computed anymore. Second, it improves the precision of numerical integration. The latter benefit is due to the fact that by re-expressing the fraction term in (2.3) as such, an adaptive Gauss-Kronrod scheme can be deployed for the numerical computation of smaller regions of the target interval. This improves the precision of Gaussian quadrature since the quadrature points are spent for smaller regions of the interval.

Incorporating the time-varying covariate $R_i(t)$ in both the longitudinal and relative risk submodels of the joint model allows us to evaluate how the occurrence of the intermediate event of interest at a future time point will influence the individualized risk predictions, for subjects who have not experienced the intermediate event by time t. The main difference of our approach when compared with the approach of Sène et al. (2016), is that we assume that both the instantaneous risk of the primary endpoint and the longitudinal profile change after the occurrence of the intermediate event, whereas they assumed an extrapolated longitudinal profile, instead. More specifically, by assuming an extrapolated longitudinal profile, Sène et al. (2016), were more interested in assessing how predictions change with and without a second treatment whereas we are more interested in studying how individualized risk predictions are influenced by intermediate events, such as reintervention or adverse events, by explicitly allowing for changes in both the longitudinal and survival submodels.

That is, different scenarios regarding the time of the intermediate event may lead to changes in the risk captured via different individual dynamic predictions accordingly. More specifically, for a future time of interest $u > t$ different assumptions can be made: 1) The patient experiences an intermediate event immediately $\rho_i = t$ or at a time point within the time-interval of prediction $t \leq \rho_i \leq u$. 2) The patient does not experience an intermediate event within the time-interval of prediction $\rho_i > u$. The individualized dynamic predictions in (2.4) are then further dependent on the scenario of choice:

$$\pi_j\left(u \mid t, \rho_j\right) = \begin{cases} \int & \exp\left\{-\int_t^u h_j\left(s \mid b_j, \rho_j \leq u\right) ds\right\} \times \\ & p\left\{b_j \mid T_j^* > t, \mathcal{Y}_j\left(t, t \leq \rho_j \leq u\right), \boldsymbol{\theta}\right\} d\boldsymbol{b}_j, t \leq \rho_j \leq u \\ \int & \exp\left\{-\int_t^u h_j\left(s \mid b_j, \rho_j > u\right) ds\right\} \times \\ & p\left\{b_j \mid T_j^* > t, \mathcal{Y}_j\left(t, t \leq u \leq \rho_j\right), \boldsymbol{\theta}\right\} d\boldsymbol{b}_j, t \leq u < \rho_j. \end{cases} \tag{2.4}$$

In Equation (2.4) the full conditional posterior density of the random effects can be analytically expressed as $\left\{p\left(T_j^* > t \mid \rho_j \leq u, b_j\right)\right\} \prod_l \left\{p\left(y_j \mid \rho_j \leq u, b_j\right)\right\} p\left(\boldsymbol{b}_j \mid \boldsymbol{\theta}\right)$ for $t \leq \rho_j \leq u$ and as $\left\{p\left(T_j^* > t \mid \rho_j \leq u, b_j\right)\right\} \prod_l \left\{p\left(y_j \mid \rho_j > u, b_j\right)\right\} p\left(\boldsymbol{b}_j \mid \boldsymbol{\theta}\right)$ for $t \leq u < \rho_j$, where $\prod_l \{p\left(y_j \mid \rho_j \leq u, b_j\right)\}$ and $\prod_l \{p\left(y_j \mid \rho_j > u, b_j\right)\}$ are multivariate Gaussian joint densities for the longitudinal responses with means $\boldsymbol{x}_j^\top(t)\boldsymbol{\beta} + R_j(t)\beta_R + t_+\beta_{t_+} + \boldsymbol{z}_j^\top(t)\boldsymbol{b_j} + R_j(t)b_{jR} + t_+ b_{jt_+}$ and $\boldsymbol{x}_j^\top(t)\boldsymbol{\beta} + \boldsymbol{z}_j^\top(t)\boldsymbol{b_j}$ respectively and variance-covariance matrices $\sigma^2 I_{n_j}$. $p\left(\boldsymbol{b}_j \mid \boldsymbol{\theta}\right)$ is a multivariate Gaussian density function with mean 0 and variance covariance matrix D.

To estimate $\pi_j\left(u \mid t, \rho_j\right)$ a Monte Carlo scheme is employed, for the integration over the random-effects, where a large set of $\theta^m\ (m = 1, \ldots, M)$ and $b_j^m\ (m = 1, \ldots, M)$ are sampled from their posterior distributions and subsequently used to compute $\pi_j^m\left(u \mid t, \rho_j\right)$. The median value of $\pi_j^m\left(u \mid t, \rho_j\right)$ is the point estimate and the 2.5% and 97.5% percentiles give a 95% credible interval.

2.3.2 Evaluation of predictive performance

To assess the performance of the individualized dynamic predictions described in the previous section, we will work under a similar framework as the one presented in (Rizopoulos et al., 2017). More specifically, we will use the time-dependent area under the receiver operating characteristic curve (AUC) and the expected prediction error (PE), adapted for the presence of intermediate events.

Under the framework presented in Sections 2 and 3.1 a rule can be defined using the individualized dynamic predictions $\pi_j\left(u = t + \Delta t \mid t\right)$ while utilizing the available longitudinal measurements up to t, $\mathcal{Y}_j\left(t\right)$. More specifically, a subject j can be termed as either to experience the event $\pi_j\left(u = t + \Delta t \mid t\right) \leq c$ or not $\pi_j\left(u = t + \Delta t \mid t\right) > c$ within a clinically relevant time interval $(t, \Delta t]$, with $c \in [0, 1]$. That is, for a pair of subjects which is randomly chosen $\{i, j\}$ for both of which the measurements up to t are provided, the AUC which is calculated for varying values of c is a measure of the discriminative capability of the assumed model and is given by:

$$\text{AUC}\left(t, \Delta t\right) = \Pr\left[\pi_i\left(t + \Delta t \mid t\right) < \pi_j\left(t + \Delta t \mid t\right) \mid \{T_i^* \in (t, t + \Delta t]\} \cap \{T_j^* > t + \Delta t\}\right], \tag{2.5}$$

which intuitively means that we expect the assumed model to give higher probability of surviving longer than the time interval of interest $(t + \Delta t]$ to the subject who did not experience the event (in this case subject j).

However, in the presence of intermediate events, the dynamic predictions change depending on whether a subject experienced the intermediate event or not. That is, the AUC in (2.5) changes to:

$$\text{AUC}\left(t, \Delta t\right) = \begin{cases} \Pr\Big[\pi_i\left(t + \Delta t \mid t, \rho_i > t\right) < \pi_j\left(t + \Delta t \mid t, \rho_i > t\right) \mid \\ \quad \{T_i^* \in (t, t + \Delta t]\} \cap \{T_j^* > t + \Delta t\}\Big], 0 < t < \rho_i, \rho_j \\ \Pr\Big[\pi_i\left(t + \Delta t \mid t, \rho_i \leq t\right) < \pi_j\left(t + \Delta t \mid t, \rho_i \leq t\right) \mid \\ \quad \{T_i^* \in (t, t + \Delta t]\} \cap \{T_j^* > t + \Delta t\}\Big], t \geq \rho_i, \rho_j. \end{cases}$$

Estimation of $\text{AUC}\left(t, \Delta t\right)$ is based in counting the concordant pairs of subjects by appropriately distinguishing between the comparable and the non-comparable (due to censoring) pairs of subjects at time t. More specifically, the following decomposition is used:

$$\hat{\text{AUC}}\left(t, \Delta t\right) = \hat{\text{AUC}}_1\left(t, \Delta t\right) + \hat{\text{AUC}}_2\left(t, \Delta t\right) + \hat{\text{AUC}}_3\left(t, \Delta t\right) + \hat{\text{AUC}}_4\left(t, \Delta t\right).$$

Term $\hat{\text{AUC}}_1\left(t, \Delta t\right)$ refers to the pairs of subjects who are comparable,

$$\Omega_{ij}^{(1)}(t) = \begin{cases} [\{T_i \in (t, t+\Delta t]\} \cap \{\delta_i = 1\} \cap \{0 < t < \rho_i\}] \cap \left[\{T_j > t + \Delta t\} \cap \{0 < t < \rho_j\}\right], \\ 0 < t < \rho_i, \rho_j \\ \\ [\{T_i \in (t, t+\Delta t]\} \cap \{\delta_i = 1\} \cap \{t \geq \rho_i\}] \cap \left[\{T_j > t + \Delta t\} \cap \{t \geq \rho_j\}\right], \\ t \geq \rho_i, \rho_j, \end{cases}$$

where $i, j = 1, \ldots, n$ with $i \neq j$. We can then estimate and compare the survival probabilities $\pi_i(t + \Delta t \mid t, t \geq \rho_i)$ and $\pi_j(t + \Delta t \mid t, t \geq \rho_j)$ for subjects i and j who did not experienced the intermediate event and $\pi_i(t + \Delta t \mid t, 0 < t < \rho_i)$ and $\pi_j(t + \Delta t \mid t, 0 < t < \rho_j)$ for subjects who experienced the intermediate event. Then $\hat{\text{AUC}}_1(t, \Delta t)$ is the proportion of concordant subjects out of the set of comparable subjects at time t:

$$\hat{\text{AUC}}_1(t, \Delta t) = \frac{\sum_{i \in \mathcal{A}} \sum_{j \neq i \in \mathcal{A}} I\{\hat{\pi}_i(t + \Delta t \mid t, t < \rho_i) < \hat{\pi}_j\left(t + \Delta t \mid t, t < \rho_j\right)\} \times I\{\Omega_{ij}^{(1)}(t)\}}{\sum_{i \in \mathcal{A}} \sum_{j \neq i \in \mathcal{A}} I\{\Omega_{ij}^{(1)}(t)\}} + \frac{\sum_{i \in \mathcal{B}} \sum_{j \neq i \in \mathcal{B}} I\{\hat{\pi}_i(t + \Delta t \mid t, t \geq \rho_i) < \hat{\pi}_j\left(t + \Delta t \mid t, t \geq \rho_j\right)\} \times I\{\Omega_{ij}^{(1)}(t)\}}{\sum_{i \in \mathcal{B}} \sum_{j \neq i \in \mathcal{B}} I\{\Omega_{ij}^{(1)}(t)\}},$$

where $\mathcal{A} = \{i, j : t < \rho_i; i, j = 1, \ldots, n\}$ and $\mathcal{B} = \{i, j : t \geq \rho_i; i, j = 1, \ldots, n\}$. The remaining terms, $\hat{\text{AUC}}_2(t, \Delta t)$, $\hat{\text{AUC}}_3(t, \Delta t)$ and $\hat{\text{AUC}}_4(t, \Delta t)$ refer to the pairs of subjects who due to censoring cannot be compared, namely

$$\Omega_{ij}^{(2)}(t) = \begin{cases} [\{T_i \in (t, t+\Delta t]\} \cap \{\delta_i = 0\} \cap \{0 < t < \rho_i\}] \cap \left[\{T_j > t + \Delta t\} \cap \{0 < t < \rho_j\}\right], \\ 0 < t < \rho_i, \rho_j \\ \\ [\{T_i \in (t, t+\Delta t]\} \cap \{\delta_i = 0\} \cap \{t \geq \rho_i\}] \cap \left[\{T_j > t + \Delta t\} \cap \{t \geq \rho_j\}\right], \\ t \geq \rho_i, \rho_j, \end{cases}$$

$$\Omega_{ij}^{(3)}(t) = \begin{cases} [\{T_i \in (t, t+\Delta t]\} \cap \{\delta_i = 1\} \cap \{0 < t < \rho_i\}] \cap \left[\{T_i < T_j \leq t + \Delta t\} \cap \right. \\ \left. \{0 < t < \rho_j\} \cap \{\delta_j = 0\}\right], 0 < t < \rho_i, \rho_j \\ \\ [\{T_i \in (t, t+\Delta t]\} \cap \{\delta_i = 1\} \cap \{t \geq \rho_i\}] \cap \left[\{T_i < T_j \leq t + \Delta t\} \cap \right. \\ \left. \{t \geq \rho_j\} \cap \{\delta_j = 0\}\right], t \geq \rho_i, \rho_j, \end{cases}$$

$$\Omega_{ij}^{(4)}(t) = \begin{cases} [\{T_i \in (t, t+\Delta t]\} \cap \{\delta_i = 0\} \cap \{0 < t < \rho_i\}] \cap \left[\{T_i < T_j \leq t + \Delta t\} \cap \right. \\ \left. \{0 < t < \rho_j\} \cap \{\delta_j = 0\}\right], 0 < t < \rho_i, \rho_j \\ \\ [\{T_i \in (t, t+\Delta t]\} \cap \{\delta_i = 0\} \cap \{t \geq \rho_i\}] \cap \left[\{T_i < T_j \leq t + \Delta t\} \cap \right. \\ \left. \{t \geq \rho_j\} \cap \{\delta_j = 0\}\right], t \geq \rho_i, \rho_j, \end{cases}$$

which contribute to the overall AUC after being appropriately weighted with the probability that they would be comparable:

$$\hat{\text{AUC}}_m(t, \Delta t) = \frac{\sum_{i \in \mathcal{A}} \sum_{j \neq i \in \mathcal{A}} I\{\hat{\pi}_i(t + \Delta t \mid t, t < \rho_i) < \hat{\pi}_j\left(t + \Delta t \mid t, t < \rho_j\right)\} \times I\{\Omega_{ij}^{(m)}(t)\} \times \hat{\nu}_{ij}^{(m)}}{\sum_{i \in \mathcal{A}} \sum_{j \neq i \in \mathcal{A}} I\{\Omega_{ij}^{(m)}(t)\} \times \hat{\nu_{ij}^{(m)}}}$$
$$+ \frac{\sum_{i \in \mathcal{B}} \sum_{j \neq i \in \mathcal{B}} I\{\hat{\pi}_i(t + \Delta t \mid t, t \geq \rho_i) < \hat{\pi}_j\left(t + \Delta t \mid t, t \geq \rho_j\right)\} \times I\{\Omega_{ij}^{(m)}(t)\} \times \hat{\nu}_{ij}^{(m)}}{\sum_{i \in \mathcal{B}} \sum_{j \neq i \in \mathcal{B}} I\{\Omega_{ij}^{(m)}(t)\} \times \hat{\nu}_{ij}^{(m)}},$$

with $m = 2, 3, 4$ and

$$\hat{\nu}_{ij}^{(2)} = \begin{cases} 1 - \hat{\pi}_i \left(t + \Delta t \mid T_i, t < \rho_i\right), & i \in \mathcal{A} \\ \\ 1 - \hat{\pi}_i \left(t + \Delta t \mid T_i, t \geq \rho_i\right), & i \in \mathcal{B}, \end{cases}$$

$$\hat{\nu}_{ij}^{(3)} = \begin{cases} \hat{\pi}_j \left(t + \Delta t \mid T_j, t < \rho_j\right), & j \in \mathcal{A} \\ \\ \hat{\pi}_j \left(t + \Delta t \mid T_j, t \geq \rho_j\right), & j \in \mathcal{B}, \end{cases}$$

$$\hat{\nu}_{ij}^{(4)} = \begin{cases} \left\{1 - \hat{\pi}_i \left(t + \Delta t \mid T_i, t < \rho_i\right)\right\} \times \hat{\pi}_j \left(t + \Delta t \mid T_j, t < \rho_j\right), & i, j \in \mathcal{A} \\ \\ \left\{1 - \hat{\pi}_i \left(t + \Delta t \mid T_i, t \geq \rho_i\right)\right\} \times \hat{\pi}_j \left(t + \Delta t \mid T_j, t \geq \rho_j\right), & i, j \in \mathcal{B}. \end{cases}$$

The expected error of predicting future events can be used to assess the accuracy of individualized dynamic predictions. Similarly as for the AUC, to account for the dynamic nature of the longitudinal outcome, we focus our interest in predicting events that occur at a time point $u > t$ given the information available up to time t, $\mathcal{Y}_j(t)$. Let $N_j(t) = I\left(T_i^* > t\right)$ denote the event status of subject j at time t. Using the square loss function the expected prediction error then is:

$$\text{PE}(u \mid t) = E\left[\left\{N_j(u) - \pi_j(u \mid t)\right\}^2\right], \tag{2.6}$$

where the expectation is taken with respect to the distribution of the event times. Adapting the above to the framework of intermediate events, (2.6) can be re-expressed as:

$$\text{PE}(u \mid t) = \begin{cases} E\left[\left\{N_j(u) - \pi_j(u \mid t, \rho_i > t)\right\}^2\right], & 0 < t < \rho_i \\ \\ E\left[\left\{N_j(u) - \pi_j(u \mid t, \rho_i \leq t)\right\}^2\right], & t \geq \rho_i, \end{cases}$$

where for each case the corresponding individualized dynamic predictions showed in (2.4) are used. The estimate of $\text{PE}(u \mid t)$ as proposed Henderson et al. (2002) and adjusted for the presence of intermediate events is given by:

$$\begin{aligned}\hat{\text{PE}}\left(u \mid t, \rho_i\right) = & \left\{n\left(t, t<\rho_i\right)\right\}^{-1} \sum_{i \in \mathcal{A}, T_i \geq t} I\left(T_i \geq u\right)\left\{1-\hat{\pi}_i\left(u \mid t, t<\rho_i\right)\right\}^2+ \\ & \delta_i I\left(T_i<u\right)\left\{0-\hat{\pi}_i\left(u \mid t, t<\rho_i\right)\right\}^2+ \\ & \left(1-\delta_i\right) I\left(T_i<u\right)\Big[\hat{\pi}_i\left(u \mid T_i, t<\rho_i\right)\left\{1-\hat{\pi}_i\left(u \mid t, t<\rho_i\right)\right\}^2\left\{1-\hat{\pi}_i\left(u \mid T_i, t<\rho_i\right)\right\}\times \\ & \left\{0-\hat{\pi}_i\left(u \mid t, t<\rho_i\right)\right\}^2\Big]+\left\{n\left(t, t \geq \rho_i\right)\right\}^{-1} \sum_{i \in \mathcal{B}, T_i \geq t} I\left(T_i \geq u\right)\left\{1-\hat{\pi}_i\left(u \mid t, t \geq \rho_i\right)\right\}^2+ \\ & \delta_i I\left(T_i<u\right)\left\{0-\hat{\pi}_i\left(u \mid t, t \geq \rho_i\right)\right\}^2\left(1-\delta_i\right) I\left(T_i<u\right) \\ & \Big[\hat{\pi}_i\left(u \mid T_i, t \geq \rho_i\right)\left\{1-\hat{\pi}_i\left(u \mid t, t \geq \rho_i\right)\right\}^2\times \\ & \left\{1-\hat{\pi}_i\left(u \mid T_i, t \geq \rho_i\right)\right\}\left\{0-\hat{\pi}_i\left(u \mid t, t<\rho_i\right)\right\}^2\Big],\end{aligned}$$

where $n\left(t, t<\rho_i\right)$ and $n\left(t, t \geq \rho_i\right)$ denote the number of subjects still at risk at time t, who have not/have experienced the intermediate event respectively.

2.4 Analysis of Pulmonary Gradient and SPRINT trial data

2.4.1 Pulmonary Gradient dataset

The Pulmonary Gradient dataset was introduced in Section 1. Our goal, is to investigate the association between the pulmonary gradient and the risk of death, how reoperation as an intermediate event changes the evolution of the pulmonary gradient and the instantaneous risk for death, and then to utilize this information to derive individualized dynamic predictions under different scenarios with respect to a future time of reoperation.

In Figure 2.1, the evolutions of pulmonary gradient for reoperated and non reoperated patients are depicted, where it is shown that for the case of reoperated patients the profiles are characterized by a linear increasing trend which drops at the moment of reoperation and then continues to increase whereas for the case of non reoperated patients the profiles show a linear increasing trend. Therefore, for this outcome, we assumed a linear mixed-effects submodel including a linear effect of time, a drop at the moment of reoperation and a change in slope after the occurrence of reoperation in both the fixed-effects and random-effects parts of the model while correcting for baseline differences in age and sex. A preliminary analysis suggested that assuming nonlinear effects of time did not improve the fit of the model to the data. Hence, we used the following specification for the mixed-effects model:

$$PG_i(t) = \begin{cases} (\beta_0 + b_{i0}) + (\beta_1 + b_{i1}) \times t_i + \beta_4 Age_i + \beta_5 Sex_i + \epsilon_i(t), 0 < t < \rho_i \\ (\beta_0 + b_{i0}) + (\beta_1 + b_{i1}) \times t_i + \left(\tilde{\beta}_2 + \tilde{b}_{i2}\right) \\ \times R_i(t) + \left(\tilde{\beta}_3 + \tilde{b}_{i3}\right) \times t_{i+} + \beta_4 Age_i + \beta_5 Sex_i + \epsilon_i(t), t \geq \rho_i. \end{cases}$$

where $PG_i(t)$ are the measurements of pulmonary gradient, $R_i(t)$ is a binary time-dependent indicator of reoperation and $t_{i+} = \max(0, t_{ij} - \rho_{ij}; j = 1, \ldots, n_i)$ is the time relative to reoperation.

To investigate the association between the pulmonary gradient and the risk of death, we postulated relative risk submodels with different parametrizations for the pulmonary gradient. The baseline hazard was expressed as a B-splines function. We also corrected for age and sex and assumed reoperation to have a direct effect on the hazard. Based on a preliminary analysis we assessed various formulations for the association structure. However, only functional forms including the slope of the pulmonary gradient were found to be stronger. Assuming a different functional form after the occurrence of reoperation did not improve the fit of the model to the data. Therefore, we present the joint models that include the slope of the pulmonary gradient along with the joint model that assumes an association with the current value of the pulmonary gradient for the sake of comparison, since it is the most common form of the association structure:

$$\begin{aligned} \text{M1} : h_i(t) &= h_0(t) \exp\{\gamma_1 \text{Age}_i + \gamma_2 \text{Sex}_i + \zeta R_i(t) + \alpha_1 \eta_{PGi}(t)\}, \\ \text{M2} : h_i(t) &= h_0(t) \exp\{\gamma_1 \text{Age}_i + \gamma_2 \text{Sex}_i + \zeta R_i(t) + \alpha_1 \eta_{PGi}(t) + \alpha_2 \tfrac{d}{dt} \eta_{PGi}(t)\}, \\ \text{M3} : h_i(t) &= h_0(t) \exp\{\gamma_1 \text{Age}_i + \gamma_2 \text{Sex}_i + \zeta R_i(t) + \alpha_2 \tfrac{d}{dt} \eta_{PGi}(t)\}. \end{aligned}$$

Table 2.1 summarizes the parameter estimates and the 95% credibility intervals of the longitudinal submodel that was used for the Pulmonary Gradient dataset. Table 2.2 summarizes the parameter estimates and the 95% credibility intervals of the survival submodels based on the three joint models fitted to the Pulmonary Gradient dataset. As shown in Table 2, the association of the pulmonary gradient with the instantaneous risk of death was weak regardless the functional form of the association structure. The strongest association in magnitude was found when using the slope parametrization both before and after the occurrence of the intermediate event.

Despite the weak magnitude of the association, reoperation was found to have a strong effect on the longitudinal evolution of pulmonary gradient. The fitted joint

Table 2.1: Estimated coefficients and 95% credibility intervals for the parameters of the longitudinal submodel fitted to the Pulmonary Gradient dataset.

	Est.	95% CI
β_0	23.41	(20.677; 26.068)
β_1	0.99	(0.797; 1.178)
$\tilde{\beta}_2$	-12.83	(-18.735; -6.647)
$\tilde{\beta}_3$	-0.02	(-0.994; 1.198)
β_4	-0.13	(-0.226; -0.033)
β_5	-4.01	(-6.671; -1.306)
σ	10.52	(10.262; 10.8)

Table 2.2: Estimated hazard ratios and 95% credibility intervals for the parameters of the survival submodels based on the three joint models fitted to the Pulmonary Gradient dataset. Age at baseline is measured in years, α_1 and α_2 are measured in the units of the pulmonary gradient (mmHg) when referring to the current value association and in (mmHg/time) when referring to the current slope association.

	Value		Value + Slope		Slope	
	HR	95% CI	HR	95% CI	HR	95% CI
γ_1	1.05	(1.03; 1.078)	1.06	(1.03; 1.083)	1.06	(1.03; 1.08)
γ_2	0.51	(0.223; 1.073)	0.52	(0.235; 1.099)	0.51	(0.25; 1.061)
ζ	0.34	(0.056; 1.399)	0.32	(0.042; 1.406)	0.31	(0.042; 1.255)
α_1	1.01	(0.992; 1.031)	1.01	(0.983; 1.032)	1.32	(0.783; 1.919)
α_2			1.17	(0.624; 1.987)		

models can be used to show how individualized dynamic predictions can be derived for a new subject, under different scenarios with respect to the timing of reoperation in the future. For this illustration, we will use model M3 since the slope parametrization was found to have a stronger effect on the instantaneous risk of death. In Figure 2.3 the individual prediction of survival for a subject with 7 measurements of pulmonary gradient, so far, are shown under the assumptions of no reoperation in the future versus reoperation immediately and after four years respectively. As shown in Figure 2.3, reoperation improves the survival probability for the new subject regardless its timing. When the reoperation is not assumed to be immediate, the survival curves

overlap up to the point of reoperation and then separate, depicting the improvement on the survival probability for this subject due to reoperation.

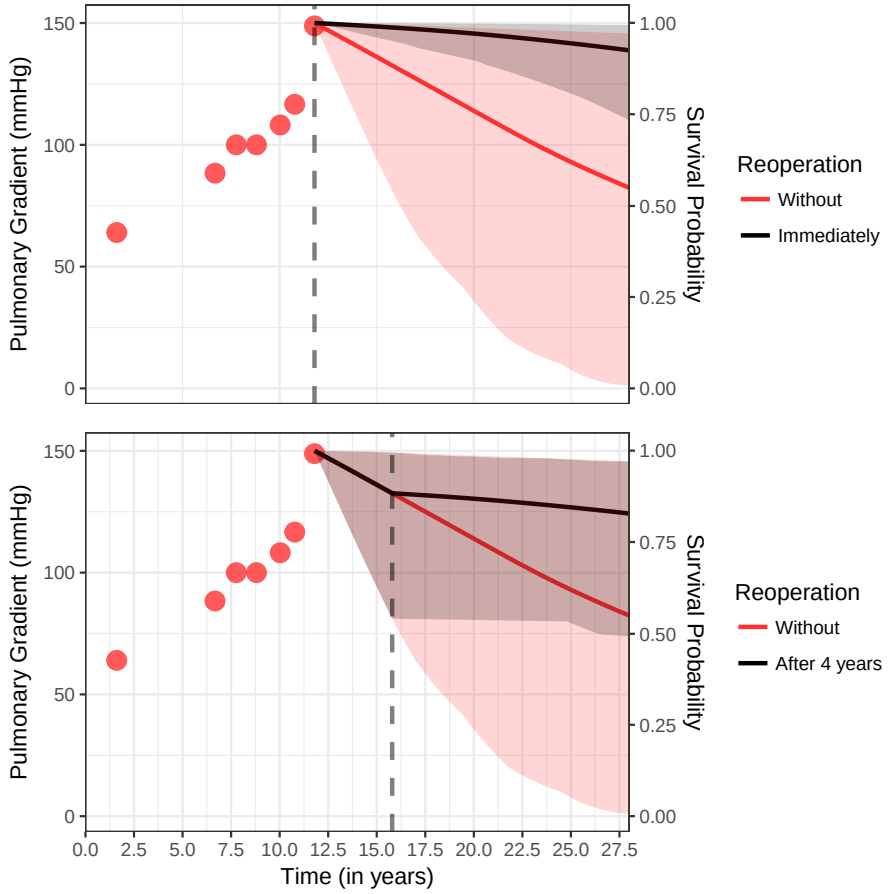

Figure 2.3: Survival Probabilities for a new subject under different scenarios with respect to the timing of reoperation. The vertical gray dashed line depicts the time of reoperation.

2.4.2 SPRINT dataset

The SPRINT dataset was also introduced in Section 1. Our goal, is to investigate how serious adverse events during follow-up change the evolution of the systolic blood pressure and the instantaneous risk for the composite endpoint, and then to utilize this

information to derive individualized dynamic predictions under different scenarios with respect to the occurrence of serious adverse events in the future.

In Figure 2.2, a random sample of the evolutions of systolic blood pressure for patients who experienced and who did not experience serious adverse events are depicted. For both sets of patients, the profiles show diverse nonlinear trends which we assume to change after the occurrence of serious adverse events. Therefore, for this outcome, we assumed a nonlinear mixed-effects submodel using natural cubic splines with 3 knots for the effect of time, and the effect of time relative to the occurrence of serious adverse events in both the fixed-effects and random-effects parts of the model while adjusting for differences in treatment. More specifically, the following specification for the mixed-effects model was used:

$$SBP_i\left(t\right) = \begin{cases} \left(\beta_0 + b_{i0}\right) + \left(\sum_{k=0}^{3}\left(\beta_{(k+1)} + b_{i(k+1)}\right) B_k\left(t, \boldsymbol{k}\right)\right) + \beta_5 \text{Treatment}_i + \\ \left(\sum_{k=0}^{3} \beta_{(k+6)} B_k\left(t, \boldsymbol{k}\right)\right) \times \text{Treatment}_i + \epsilon_i\left(t\right), 0 < t < \rho_i \\ \\ \left(\beta_0 + b_{i0}\right) + \left(\sum_{k=0}^{3}\left(\beta_{(k+1)} + b_{i(k+1)}\right) B_k\left(t, \boldsymbol{k}\right)\right) + \beta_5 \text{Treatment}_i + \\ \left(\sum_{k=0}^{3} \beta_{(k+6)} B_k\left(t, \boldsymbol{k}\right)\right) \times \text{Treatment}_i + \\ \left(\sum_{k=0}^{3}\left(\tilde{\beta}_{(k+10)} + \tilde{b}_{i(k+10)}\right) B_k\left(t_+, \boldsymbol{k}\right)\right) + \\ \left(\sum_{k=0}^{3} \tilde{\beta}_{(k+14)} B_k\left(t_+, \boldsymbol{k}\right)\right) \times \text{Treatment}_i + \epsilon_i\left(t\right), t \geq \rho_i. \end{cases}$$

where $SBP_i\left(t\right)$ are the measurements of systolic blood pressure and $t_{i+} = \max\left(0, t_{ij} - \rho_{ij}\right.$ $\left.; j = 1, \ldots, n_i\right)$ is the time relative to the occurrence of serious adverse event.

To investigate the association between the systolic blood pressure and the composite endpoint, we postulated relative risk submodels with different parametrizations for the systolic blood pressure. The baseline hazard was expressed as a B-splines function. We also corrected for treatment and assumed the occurrence of serious adverse event to have a direct effect on the hazard. The functional forms we used for the association structure were the current value, slope, area, both the current value and slope, as well as more elaborate ones assuming that the value, slope, and area all have an effect on the hazard and assuming that after the occurrence of the serious adverse event the effect of the current slope on the hazard changes. The joint models are given in detail below:

$$\begin{aligned}
\text{M4} : h_i(t) &= h_0(t)\exp\{\gamma_1 \text{Treatment}_i + \zeta R_i(t) + \alpha_1 \eta_{SBPi}(t)\},\\
\text{M5} : h_i(t) &= h_0(t)\exp\{\gamma_1 \text{Treatment}_i + \zeta R_i(t) + \alpha_2 \tfrac{d}{dt}\eta_{SBPi}(t)\},\\
\text{M6} : h_i(t) &= h_0(t)\exp\{\gamma_1 \text{Treatment}_i + \zeta R_i(t) + \alpha_1 \eta_{SBPi}(t) + \alpha_2 \tfrac{d}{dt}\eta_{SBPi}(t)\}.\\
\text{M7} : h_i(t) &= h_0(t)\exp\{\gamma_1 \text{Treatment}_i + \zeta R_i(t) + \alpha_3 \textstyle\int_0^t \eta_{SBPi}(s)\,ds\}.\\
\text{M8} : h_i(t) &= h_0(t)\exp\{\gamma_1 \text{Treatment}_i + \\
&\quad \zeta R_i(t) + \alpha_1 \eta_{SBPi}(t) + \alpha_2 \tfrac{d}{dt}\eta_{SBPi}(t) + \alpha_3 \textstyle\int_0^t \eta_{SBPi}(s)\,ds\}.\\
\text{M9} : h_i(t) &= h_0(t)\exp\{\gamma_1 \text{Treatment}_i + \\
&\quad \zeta R_i(t) + \alpha_1 \eta_{SBPi}(t) + \alpha_2 \tfrac{d}{dt}\eta_{SBPi}(t) + \alpha_3 R_i(t) \times \tfrac{d}{dt}\eta_{SBPi}(t)\}.
\end{aligned}$$

Table 2.3 summarizes the parameter estimates and the 95% credibility intervals of the longitudinal submodel that was used for the SPRINT dataset. Tables 2.4 and 2.5 summarize the parameter estimates and the 95% credibility intervals of the survival submodels based on the six joint models fitted to the SPRINT dataset. As shown in Tables 2.4 and 2.5, the association of the pulmonary gradient with the instantaneous risk of composite endpoint was weak in magnitude but significant in the cases of value and slope association and area association.

Similarly, as for the Pulmonary Gradient Dataset, in Figure 2.4 we illustrate how the individualized subject-specific predictions for the composite endpoint of interest change under different scenarios for the timing of the occurrence of a serious adverse event. More specifically, we illustrate the cases of an immediate occurrence of the adverse event and an occurrence after a year using the joint model that postulates effects of both the current value and current slope of the systolic blood pressure trajectory on the instantaneous risk of the composite endpoint. In both cases, the occurrence of the serious adverse event worsens the survival probability considerably.

Finally, as an illustration of the use of the predictive performance measures, as they were presented in Section 3, we refer the reader to the online supplementary material S2 where we present a comparison in terms of the AUC and the PE for the SPRINT data where the extrapolation method (that is ignoring the longitudinal data observed after the occurrence of the intermediate event) and the proposed model are compared.

Table 2.3: Estimated coefficients and 95% credibility intervals for the parameters of the longitudinal submodel fitted to the SPRINT dataset.

	Est.	95% CI
β_0	137.95	(137.549; 138.32)
β_1	-1.33	(-1.808; -0.851)
β_2	-0.40	(-0.873; 0.05)
β_3	-8.39	(-9.302; -7.505)
β_4	1.01	(0.574; 1.407)
β_5	-0.57	(-1.116; 0.021)
β_6	-13.45	(-14.125; -12.778)
β_7	-11.10	(-11.734; -10.456)
β_8	-25.93	(-27.184; -24.583)
β_9	-9.51	(-10.094; -8.938)
$\tilde{\beta}_{10}$	0.06	(-0.954; 1.088)
$\tilde{\beta}_{11}$	-0.06	(-1.282; 1.181)
$\tilde{\beta}_{12}$	-1.82	(-3.097; -0.389)
$\tilde{\beta}_{13}$	-0.62	(-2.402; 1.268)
$\tilde{\beta}_{14}$	1.84	(0.423; 3.333)
$\tilde{\beta}_{15}$	0.35	(-1.321; 2.107)
$\tilde{\beta}_{16}$	5.06	(2.983; 6.847)
$\tilde{\beta}_{17}$	1.48	(-1.028; 3.808)
σ	11.22	(11.164; 11.266)

2.4.3 Software

The R package **JMbayes** was extended to appropriately account for the occurrence of intermediate events and deriving individualized dynamic predictions under different scenarios with respect to the occurrence of the intermediate event. These changes are already integrated in the package both on CRAN and in the development version on GitHub. However, since the specification of such models can be very application-specific, in terms of special features which need to be accounted for due to the occurrence of the intermediate event, the datasets need to be appropriately prepared before their use with the functions of package **JMbayes**. The reader may refer to the online Supplementary Material Section S3, for a step by step tutorial on how to fit joint models with the occurrence of intermediate events and derive predictions thereafter.

Table 2.4: Estimated hazard ratios and 95% credibility intervals for the parameters of the joint models fitted to the SPRINT dataset. α_1, α_2 and α_3 are measured in the units of 20 times systolic blood pressure (20mmHg) when referring to the current value association, in (10mmHg/time) when referring to the current slope association and in (20mmHg$\times t$) when referring to the area under the curve association.

	Value		Slope		Area	
	HR	95% CI	HR	95% CI	HR	95% CI
γ_1	0.91	(0.59; 0.859)	0.72	(0.59; 0.859)	0.81	(0.69; 0.97)
ζ	1.86	(1.592; 2.379)	1.93	(1.592; 2.379)	1.88	(1.567; 2.236)
α_1	1.34	(0.940; 1.310)	1.11	(0.940; 1.310)	1.07	(1.000; 1.143)

Table 2.5: Estimated hazard ratios and 95% credibility intervals for the parameters of the joint models fitted to the SPRINT dataset. α_1, α_2 and α_3 are measured in the units of 20 times systolic blood pressure (20mmHg) when referring to the current value association, in (10mmHg/time) when referring to the current slope association and in (20mmHg$\times t$) when referring to the area under the curve association.

	Value + Slope		Value + Slope + Area		Value + Slope Interaction	
	HR	95% CI	HR	95% CI	HR	95% CI
γ_1	0.9	(0.743; 1.093)	0.9	(0.768; 1.054)	0.9	(0.737; 1.116)
ζ	1.89	(1.536; 2.294)	1.87	(1.649; 2.154)	1.58	(1.203; 2.065)
α_1	1.390	(1.166; 1.646)	1.32	(1.173; 1.471)	1.37	(1.118; 1.671)
α_2	1.100	(1.012; 1.206)	1.09	(1.049; 1.130)	1.05	(0.964; 1.158)
α_3	1.02	(0.985; 1.059)	1.84	(0.985; 2.571)		

Finally, all the analyses presented in Section 4, were performed using R version 3.5.1 and package **JMbayes**.

2.5 Simulation Study

2.5.1 Design

To evaluate the performance of the proposed models and to compare, in terms of predictive accuracy, the dynamic predictions that account for the whole biomarker trajectory against the cases where extrapolation or a simple time-dependent Cox model

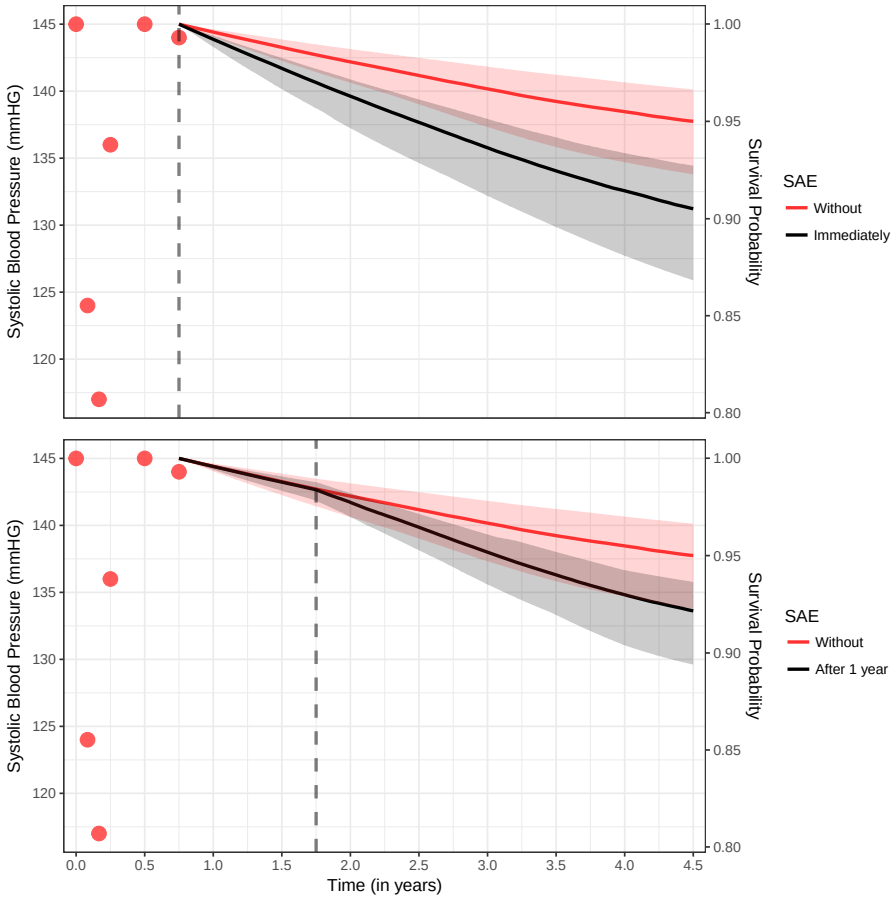

Figure 2.4: Survival Probabilities for a new subject under different scenarios with respect to the occurrence of serious adverse event (SAE). The vertical gray dashed line depicts the time of serious adverse event.

without accounting for the longitudinal data are assumed, we performed a simulation study. The main goal of the simulation study is to show the benefit in the accuracy of the individualized dynamic predictions when assuming that the intermediate event changes both the risk for the event of interest and the longitudinal trajectory against the case of assuming that the intermediate event only changes the risk for the event of interest while the longitudinal trajectory is extrapolated and the case where the longitudinal data are not taken into account. We assumed 2000 patients and then randomly selected follow-up visits. Each visit time t_{ij} was simulated from a uniform

distribution between 0 and 30. We assumed a total number of 20 measurements per subject. The final number of measurements, however, varies depending on when a subject experienced the clinical event or was censored. To mimic a realistic situation, the timing of the intermediate event was assumed to depend on the value of the biomarker trajectory. Specifically, if the biomarker exceeded a specific value then reintervention took place at the next visit. For the cases that this value was not reached, the patient was assumed to never have experienced the intermediate event. For simplicity, we assumed a linear mixed-effects model and a survival submodel without any baseline covariates.

For the continuous longitudinal outcome, we simulated the data from a linear mixed-effects models similar to the model that we used for the pulmonary gradient dataset:

$$y_i(t) = \eta_i(t) + \epsilon_i(t) = (\beta_0 + b_{i0}) + (\beta_1 + b_{i1})\, t_i + \left(\tilde{\beta}_2 + \tilde{b}_{i2}\right) R_i(t) + \left(\tilde{\beta}_3 + \tilde{b}_{i3}\right) t_{i+} + \epsilon_i(t),$$

where $\epsilon_i(t) \sim \mathcal{N}\left(0, \sigma^2 I_{n_i}\right)$ and $\boldsymbol{b}_i = \left(b_{i0}, b_{i1}, \tilde{b}_{i2}, \tilde{b}_{i3}\right)^\top \sim \mathcal{N}(0, \boldsymbol{D})$. More specifically, we adopted a linear effect for time, a "drop" effect that occurs at the time of reoperation and an effect for the change in the slope from the time of reoperation onwards for both the fixed and the random part. Time t was simulated from a uniform distribution between 0 and 30. Based on this model for the continuous outcome we assumed three different scenarios:

- Scenario 1: $\beta_1 = 20.7, \quad \beta_2 = 1.6, \quad \tilde{\beta}_3 = -15.5, \quad \tilde{\beta}_4 = -0.76,$
- Scenario 2: $\beta_1 = 20.7, \quad \beta_2 = 1.6, \quad \tilde{\beta}_3 = -15.5, \quad \tilde{\beta}_4 = 0,$
- Scenario 3: $\beta_1 = 20.7, \quad \beta_2 = 1.6, \quad \tilde{\beta}_3 = 0, \quad \tilde{\beta}_4 = -0.76.$

The assumed longitudinal trajectories for each of the three scenarios are depicted in Figure 2.5.

More specifically, in the first scenario, we assume that the longitudinal profile drops at the occurrence of the intermediate while the slope changes after its occurrence. In the second and third scenarios, the slope does not change after the occurrence of the intermediate event and the longitudinal profile does not drop respectively.

For the survival outcome we assumed a relative risk model of the form:

$$h_i(t) = h_0(t) \exp\{\gamma \mathbf{1} + R_i(t)\, \zeta + \alpha_1 \eta_i(t)\},$$

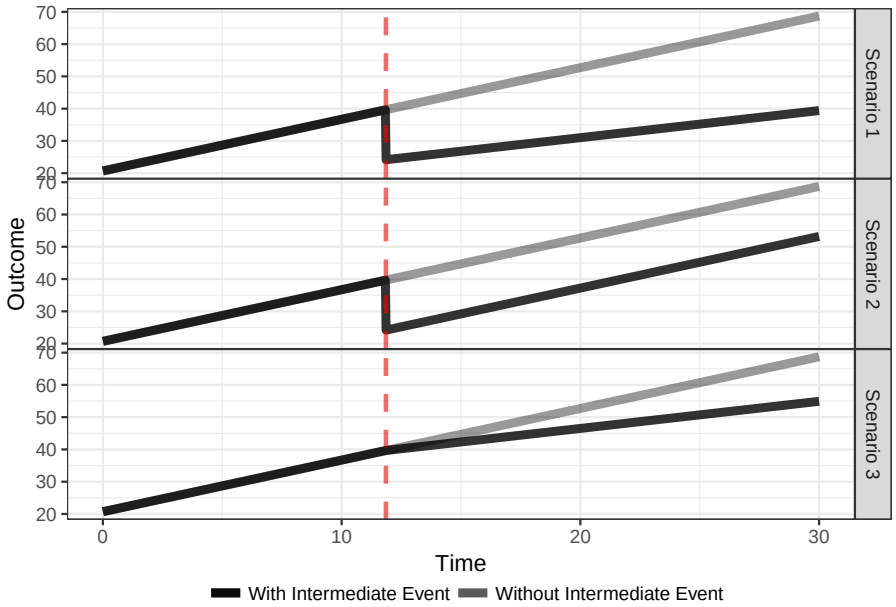

Figure 2.5: Assumed average longitudinal evolutions under the three simulation scenarios for subjects who experienced an intermediate event and for subjects who did not. The vertical red dashed line depicts the occurrence of the intermediate event.

where $\mathbf{1}$ is a vector of ones for the intercept term with a corresponding regression coefficient γ while the baseline risk was simulated from a Weibull distribution $h_0(t) = \xi t^{\xi-1}$ with $\xi = 20.4$. The censoring process was assumed to follow an exponential distribution with mean equal to 22.6. For the time-dependent Cox model, the following relative risk model was used instead:

$$h_i(t) = h_0(t) \exp\{\gamma \mathbf{1} + R_i(t)\zeta\},$$

where no association with the longitudinal outcome is assumed.

2.5.2 Results

Under the settings described in the previous Section, 500 datasets were simulated for each of the three scenarios. All the datasets were split in half to a training and test part with 1000 subjects each. For all the scenarios, to account for the whole trajectory of the biomarker, the joint model which consists of the submodels shown in Section 2.5.1 was fitted to the part of the simulated datasets which were kept for training. On the

other hand, for the extrapolation method the observations after reintervention were omitted from the analysis of the longitudinal outcome and the following mixed-effects model was fitted to the data up to reintervention time:

$$y_i(t) = \eta_i(t) + \epsilon_i(t) = (\beta_0 + b_{i0}) + (\beta_1 + b_{i1})\, t_i + \epsilon_i(t),$$

while for the survival process the same model was used. For the time-dependent Cox model, no longitudinal model was used and we only accounted for a change in the instantaneous hazard after the occurrence of the intermediate event as discussed in the previous section.

To assess the performance of the three approaches we used the models that were fitted on the training data to calculate the time-dependent AUCs and the prediction errors based on the test data. Both the time-dependent AUCs and the prediction errors were calculated at 3 different time-intervals starting at: $t = 20$, $t = 22$ and $t = 24$ respectively and assuming a clinically relevant time interval of two years $\Delta t = 2$. The time-intervals were selected on the basis of when the most events occur in the simulated data.

In Figures 2.6 and 2.7 we present the results of the simulation study depicted by boxplots. Specifically, the boxplots in each row represent different scenarios: non-zero effects, zero change in slope and zero "drop" at time of reintervention and in each column a different time-interval for prediction. In all the scenarios and time-intervals, both the AUC and PE are better when assuming that the intermediate event changes both the risk for the event of interest and the longitudinal trajectory. The simple time-dependent Cox model performs considerably worse than the other approaches. Moreover, there is a slight increase in the difference between the WT and Extrapolation methods for both predictive measures as the time-interval is set at later time-points. That is, the more information used the greater the difference between the two methods tends to become. Moreover, the relative performance of the two approaches does not differ between the three scenarios as well as for the different follow-up times. Therefore, the results support the argument that accounting for the changes in the longitudinal trajectories due to the occurrence of the intermediate event improve the predictive accuracy when compared to the approach that the longitudinal profile remains unaffected by its occurrence.

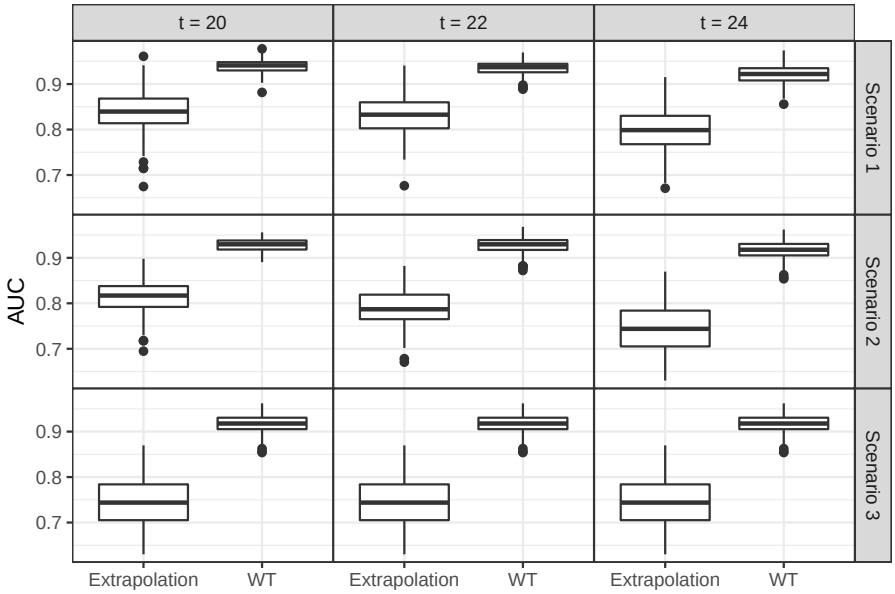

Figure 2.6: AUCs for the individualized dynamic predictions, evaluated using the testing part of the 500 datasets for two different joint models. 1) Extrapolation: Assuming that the longitudinal profile does not change after the occurrence of the intermediate event, 2) WT (Whole Trajectory): Assuming that the longitudinal profile changes after the occurrence of the intermediate event.

2.6 Discussion

Using the joint modeling framework we developed tools for deriving individualized dynamic predictions that are adaptive to different scenarios regarding intermediate events, such as treatment changes or the occurrence of adverse events. We proposed a range of joint models for longitudinal and time-to-event data which can accommodate special features due to the occurrence of intermediate events in both the longitudinal and survival submodels. That is, by incorporating features, such as the ones described in (2.1) and (2.2) a broad range of flexible joint models is sketched which accounts for the impact of an intermediate event by allowing for: 1) a direct effect of the intermediate event on the risk of the clinical endpoint through the time-varying binary covariate $R_i(t)$, 2) a direct effect of the intermediate event on the longitudinal trajectory through $g\{R_i(t), t_{i+}\}$ and 3) an indirect effect of the intermediate event on the the risk of the clinical endpoint through the association between the two outcomes

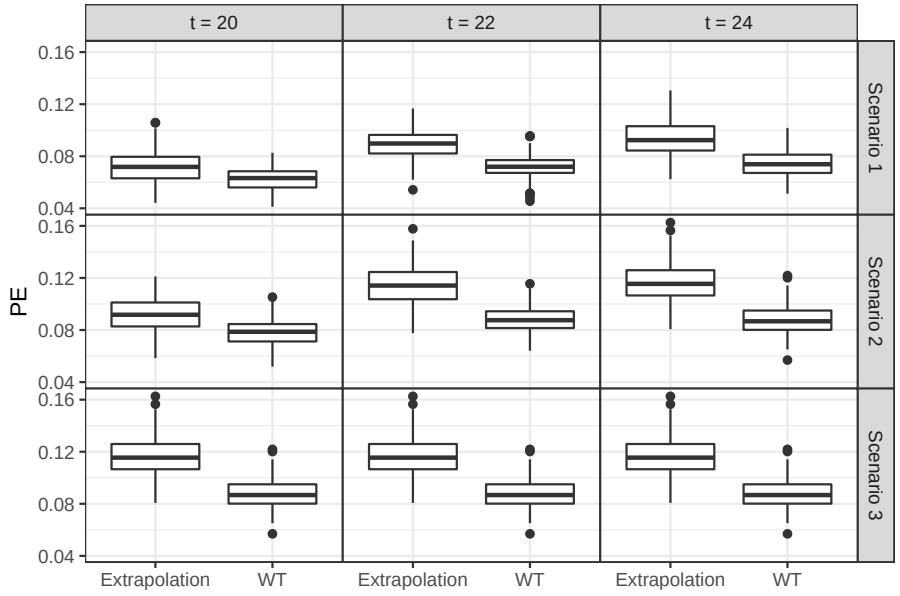

Figure 2.7: PEs for the individualized dynamic predictions, evaluated using the testing part of the 500 datasets for two different joint models. 1) Extrapolation: Assuming that the longitudinal profile does not change after the occurrence of the intermediate event, 2) WT (Whole Trajectory): Assuming that the longitudinal profile changes after the occurrence of the intermediate event.

which is defined by $f_{(t,\rho_i)}\{(\mathcal{H}_i(t,\rho_i), \boldsymbol{b}_i\}$ and is allowed to differ before and after the intermediate event. All these features allow for great flexibility in the specification of the joint model which when utilized accordingly can lead to accurate predictions.

In the same line as recent observations with regard to dynamic predictions from joint models, we have seen that the accuracy of the predictions is influenced by intermediate events occurring during follow-up. Such events will need to be appropriately modeled as time-varying covariates in both the longitudinal and survival submodels. As illustrated in our simulation study, doing so, improves the predictive accuracy of the individualized dynamic predictions.

The focus of the applications presented in this paper, was to illustrate how joint models can be utilized to provide individualized dynamic predictions under different scenarios with respect to the occurrence of intermediate events. It should be noted, however, that in the analysis of the Pulmonary Gradient dataset, presented in Section 4.1, models M1, M2 and M3 assume that reoperation has a proportional effect on the

hazard at any time point. This is a strong assumption since clinically it would make sense to assume that reoperation increases the hazard shortly after its occurrence before its beneficial effect takes place. However, model diagnostics based on Schoenfeld residuals did not point to a violation of this assumption. Since the focus of this application was to illustrate the exploitation of joint models in deriving individualized dynamic predictions which are adaptive to different scenarios with respect to the occurrence of reoperation, we did not further explore the impact of this assumption. We do believe, however, that a more flexible model, such as a joint model that considers a multi-state survival process with reoperation as a potential state would allow for such assumptions to be incorporated which sets a potential direction for further research. Furthermore, it should also be noted that potential interaction effects between time and baseline characteristics such as age and sex, were not explored in the modelling of the pulmonary gradient for parsimony, but they could be added.

It is also essential to discuss to what extent the proposed models can be used to draw inference on potential outcomes following different scenarios on the future occurrence of intermediate events. It is, therefore, crucial to note that the focus of this paper was prognostic modelling and the development of a prediction tool that can quantify potential changes in the outcome under different scenarios concerning the occurrence of future intermediate events. Consequently, the adaptive individualized prediction tools we developed can be used to answer questions such as: "What is the expected change in the survival probability of a patient if he/she gets reoperated in a year from now given the information available on the patient so far?". Such questions should not be confused with inferential statements of causal nature such as: "A reoperation on this patient a year from now will cause his/her risk to change by this quantity". For such statements to be possible, a set of essential assumptions including positivity, consistency and exchangeability should be satisfied, similar to the assumptions that any statistical model must fulfil to be used for causal inference. An informative reference on the topic, specifically for the case of longitudinal data and time-varying treatments can be found in Hernán and Robins (2019) (Chapters 19-21).

Another interesting point of discussion for such applications is whether the proposed models are susceptible to time-dependent confounding. Indeed, in the framework of a randomised clinical trial, the target treatment effect would no longer be protected if a specific selection of patients (e.g. severe cases) experienced the intermediate event post-randomization. This complicates the interpretation of the treatment effect coefficient. Therefore, it should be noted that in the applications presented, in this paper, we worked under the assumption that the intermediate event depends only on previous measurements of the marker and does not carry any further informa-

tion. Under this assumption, including the intermediate event in the specification of both the mixed-effect sub-model and the relative risk sub-model, is sufficient and the process that generates the intermediate event does not have to be explicitly modeled.

The joint model formulation we presented allows to utilize the quantification of the effects the intermediate event imposes on the risk for the clinical endpoint of interest. As such, it can be utilized to derive individualized dynamic predictions for new subjects who did not experience the intermediate event and quantify how its occurrence at any future time point will influence their risk predictions. Such a predictive tool can provide valuable information to the physicians and assist in their decision making process for potential treatment changes. Based on such predictions further prognostic tools can potentially be developed. For example in settings where the timing of a future treatment is important, having both benefits and disadvantages when applied either too late or too early, such dynamic predictions that are adaptive to the timing of the intermediate event can become the basis for methodology that can be used to predict the optimal time for the future treatment. Unfortunately, the applications at hand did not allow for exploring such a possibility, but this is a clear direction for future research. Moreover, in our paper we only consider the joint analysis of one longitudinal and one survival outcome. While the extension of the proposed models to their multivariate counterparts is straightforward, such multivariate joint models have not been explored in the literature in the context of intermediate events that may occur during follow-up and alter the course of the disease for the patient. Therefore, individualized dynamic predictions based on even more complex joint models, such as with multiple longitudinal biomarkers or with multi-state processes instead of a single time-to-event outcome might lead to improved accuracy depending on the application.

Acknowledgements The first and last author would like to acknowledge support by the Netherlands Organization for Scientific Research VIDI (grant number: 016.146.301). The first and second author would like to acknowledge support by the Netherlands Organisation for Scientific Research NWO Veni (grant number: 916.160.87).

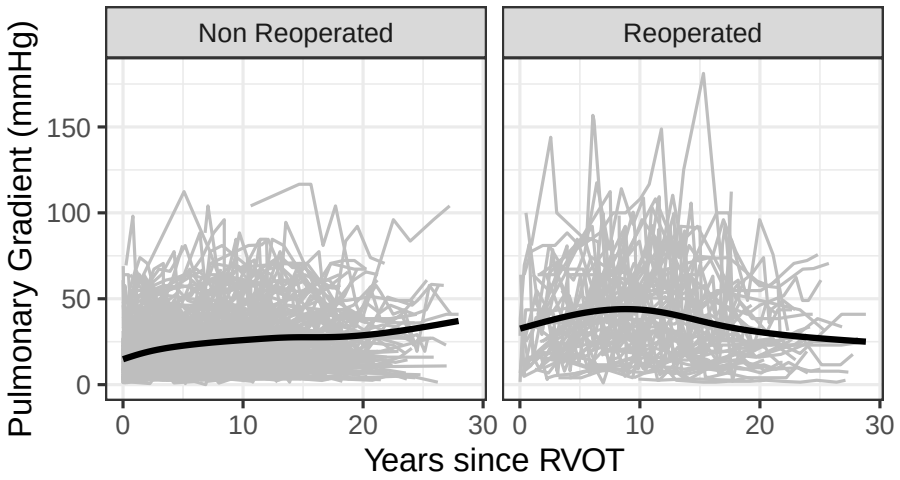

Figure 2.8: Individual trajectories of observed pulmonary gradient values over time for non reoperated subjects (left panel) and reoperated subjects (right panel) along with a smooth curve (red line).

Appendix

2.A Data Description

2.A.1 Pulmonary Gradient Data

Table 2.6: Data description: Continuous variables are presented as median (IQR), categorical variables as counts (frequencies %).

	All	Non reoperated	Reoperated
Number of subjects	467	402 (86.9%)	65 (13.1%)
Number of repeated measurements	8 (5 - 12)	7 (4 - 11)	14 (11 - 16)
Number of Events	34 (7.3%)	32 (8.0%)	2 (3.0 %)
Sex = Male	272 (58.2%)	229 (57.0%)	43 (66.1 %)
Age	20.3 (9.5 - 30.9)	21.5 (9.4 - 31.2)	16 (10.9 - 22.4)

Table 2.7: Data description: Continuous variables are presented as median (IQR), categorical variables as counts (frequencies %).

	All	No SAE	SAE
Number of subjects	9068	5644 (62.2%)	3424 (37.8%)
Number of repeated measurements	15 (13 - 17)	15 (13 - 17)	14 (12 - 17)
Number of Events	535 (5.9%)	20 (0.3%)	515 (15.0 %)
Treatment = Intensive	4552 (50.2%)	2804 (49.7%)	1748 (51.0%)

2.A.2 SPRINT Data

2.B Predictive performance comparison between extrapolation method and joint models with intermediate events in the SPRINT data

To illustrate the use of the predictive performance measures presented in Section 3, we hereby present a comparison in terms of the time-dependent area under the receiver operating characteristic curve (AUC) and the expected prediction error (PE) for the SPRINT data.

More specifically, we are interested to compare the joint model used for the analysis of the SPRINT data which postulates effects of both the current value and current slope of the systolic blood pressure trajectory on the instantaneous risk of the composite endpoint (as presented in Section 4.2) against the corresponding model that ignores the longitudinal data after the occurrence of the serious adverse event (Extrapolation method). Specifically, for the extrapolation method the following model was used for the evolution of the systolic blood pressure:

$$SBP_i(t) = (\beta_0 + b_{i0}) + \left(\sum_{k=0}^{3} \left(\beta_{(k+1)} + b_{i(k+1)} \right) B_k(t, \boldsymbol{k}) \right) + \beta_5 \text{Treatment}_i + \left(\sum_{k=0}^{3} \beta_{(k+6)} B_k(t, \boldsymbol{k}) \right) \times \text{Treatment}_i + \epsilon_i(t),$$

while for the instantaneous risk the same model was used for both methods.

To compare the two approaches, we randomly split the data in half to a training and test part. Then we fitted the models to the training part and evaluated the AUC and PE for both models using the test part. The evaluation was done at 6 time-intervals:

$t = 2.5$, $t = 2.75$, $t = 3$, $t = 3.25$, $t = 3.5$, and $t = 3.75$ assuming a clinically relevant time interval of a quarter of a year $\Delta t = 0.25$.

The results are shown in table 2.8. In the specific application the two approaches are virtually similar. This can be explained by the fact that the occurrence of a serious adverse event did not have a strong effect on the evolution of systolic blood pressure. However, even in this case both approaches achieve similar performance in terms of predictive accuracy.

Table 2.8: Evaluation of Predictive Performance between WT and Extrapolation Methods for the SPRINT data

	WT		Extrapolation	
t	AUC	PE	AUC	PE
2.50	0.66088	0.00362	0.65809	0.00363
2.75	0.79924	0.00417	0.79290	0.00418
3.00	0.54684	0.00606	0.56923	0.00607
3.25	0.42924	0.00368	0.46713	0.00369
3.50	0.62177	0.00421	0.63544	0.00423
3.75	0.68023	0.00570	0.66712	0.00571

2.C Sample R Code

In this section we provide a sample **R** code for fitting the joint models proposed in the paper and then utilizing them to derive predictions under different scenarios for the occurrence of intermediate events. As an illustrative example we use one of the 500 datasets generated for the simulation study, presented in the paper, under scenario 1. Under this scenario we assume that the trajectory of the longitudinal outcome immediately changes (drops) at the occurrence of the intermediate event while the rate of increase (slope) also changes after the occurrence of the intermediate event, implying two special features of the longitudinal outcome that need to be considered.

To start with the code, the **R** package **JMbayes** needs to be installed and loaded.

```
# For the remainder the R symbol 'dat' will be used to refer to the dataset
# Replace NAs for Interm.Evnt.Time with 1e+6
dat$Interm.Evnt.Time <- ifesle(is.na(dat$Interm.Evnt.Time),
                       1e+5, dat$Interm.Evnt.Time)
```

Here we chose a large unrealistic value of 100000 years that is impossible to occur in reality. Table 2.9 shows how the data for the 2^{nd} subject changed.

Table 2.9: Example Dataset

ID	time	Y	Interm.Evnt.Time	Tstart	Tstop	Event
1	0.000	22.043	12.846	0.000	1.179	0
1	1.179	20.319	12.846	1.179	4.793	0
1	4.793	23.387	12.846	4.793	8.202	0
1	8.202	26.034	12.846	8.202	9.105	0
1	9.105	24.857	12.846	9.105	12.846	0
1	12.846	27.046	12.846	12.846	16.821	0
1	16.821	17.251	12.846	16.821	17.200	0
1	17.200	18.261	12.846	17.200	17.223	0
1	17.223	19.069	12.846	17.223	18.753	0
2	0.000	23.424	100000.000	0.000	2.401	0
2	2.401	24.342	100000.000	2.401	4.210	0
2	4.210	26.170	100000.000	4.210	4.451	0
2	4.451	25.781	100000.000	4.451	6.419	0
2	6.419	29.663	100000.000	6.419	7.265	1

Now we can proceed to compute and include in the dataset the time-varying covariates of interest: $R_{(t)}$ and t_{i+} which correspond to the inermediate event indicator and the time relative to the occurrence of the intermediate event, respectively. To do so in **R** we run the following code:

```
# Compute the intermediate event indicator
dat$Int.Evnt.Index <- as.numeric(dat$time >= dat$Interm.Evnt.Time)
# Compute Time relative to the intermediate event
dat$time.relative <- pmax(0, dat$time - dat$Interm.Evnt.Time)
```

Table 2.10 shows the updated data on the subset of the data that includes the two subjects we use in the example. We see that the two time-varying covariates are now included in the dataset. For subject 1, for all the time points following the occurrence of the intermediate event, the two time-varying covariates change accordingly whereas for subject 2 they are zero since the occurrence of the intermediate event was not observed during follow-up.

Table 2.10: Example Dataset

ID	time	Y	Interm.Evnt.Time	Tstart	Tstop	Event	Int.Evnt.Index	time.relative
1	0.000	22.043	12.846	0.000	1.179	0	0	0.000
1	1.179	20.319	12.846	1.179	4.793	0	0	0.000
1	4.793	23.387	12.846	4.793	8.202	0	0	0.000
1	8.202	26.034	12.846	8.202	9.105	0	0	0.000
1	9.105	24.857	12.846	9.105	12.846	0	0	0.000
1	12.846	27.046	12.846	12.846	16.821	0	1	0.000
1	16.821	17.251	12.846	16.821	17.200	0	1	3.975
1	17.200	18.261	12.846	17.200	17.223	0	1	4.354
1	17.223	19.069	12.846	17.223	18.753	0	1	4.377
2	0.000	23.424	100000.000	0.000	2.401	0	0	0.000
2	2.401	24.342	100000.000	2.401	4.210	0	0	0.000
2	4.210	26.170	100000.000	4.210	4.451	0	0	0.000
2	4.451	25.781	100000.000	4.451	6.419	0	0	0.000
2	6.419	29.663	100000.000	6.419	7.265	1	0	0.000

Now that the data are ready, we can proceed to fitting the joint model. This is no different than fitting any other joint model with package **JMbayes** apart from including in the model specification the newly introduced time-varying covariates. Hence we first fit separately the Cox relative risk and mixed-effects submodels and then we pass the corresponding objects to function `jointModelBayes()` to obtain the final joint model fit:

```
# Fit time-dependent Cox model
cox.fit <- coxph(Surv(Tstart, Tstop, Event) ~ Int.Evnt.Index + cluster(ID),
                 data = dat)
# Fit the mixed-effects model
mix.fit <- lme(Y ~ time + Int.Evnt.Index + time.relative,
               random = ~ time + Int.Evnt.Index + time.relative | ID,
               data = dat)
# Fit the Joint model
joint.fit <- jointModelBayes(mix.fit, cox.fit, timeVar = "time")
```

After obtaining the joint model fit, we can proceed in deriving dynamic predictions under different scenarios for the occurrence of the intermediate event for new subjects. In Table 2.11 the data for a new subject which has not experienced the intermediate event are shown.

Table 2.11: Example Dataset

ID	time	Y	Interm.Evnt.Time	Int.Evnt.Index	time.relative
New	0.5548873	15.86804	NA	NA	NA
New	1.3547850	26.81457	NA	NA	NA
New	2.1550550	13.02823	NA	NA	NA
New	6.5422016	29.46144	NA	NA	NA
New	7.5409739	23.48326	NA	NA	NA

For the code illustration this data are assumed to be included in a `data.frame` called `newdata`. Since this subject has not experienced the intermediate event, all the relevant information regarding the occurrence of the intermediate event are shown as missing observations. Again in order to derive predictions under different scenarios regarding the occurrence of the intermediate event we need to manipulate the data accordingly. We will illustrate how the data need to be changed for investigating three different scenarios regarding the occurrence of the intermediate event:

1. No occurrence of the intermediate event,
2. immediate occurrence of the intermediate event,
3. occurrence of the intermediate event after 1 year.

For the first case (no occurrence of the intermediate event) we will use the same trick as before. That is we will specify the time of the intermediate event as an arbitrarily large number and then calculate the intermediate event indicator and time relative to the occurrence of the intermediate event accordingly.

```
# Create a copy of newdata and call it newdata 1
newdata1 <- newdata
# Specify arbitrarily large number for Interm.Evnt.Time
newdata1$Interm.Evnt.Time <- 1e+05
# Compute Int.Evnt.Index and time.relative
newdata1$Int.Evnt.Index <-  as.numeric(newdata$time >= newdata$Interm.Evnt.Time)
newdata$time.relative <- pmax(0, newdata$time - newdata$Interm.Evnt.Time)
```

The changes in `newdata1` are shown in Table 2.12.

Table 2.12: Example Dataset

ID	time	Y	Interm.Evnt.Time	Int.Evnt.Index	time.relative
New	0.5548873	15.86804	1e+05	0	0
New	1.3547850	26.81457	1e+05	0	0
New	2.1550550	13.02823	1e+05	0	0
New	6.5422016	29.46144	1e+05	0	0
New	7.5409739	23.48326	1e+05	0	0

In a similar manner we create datasets `newdata2` and `newdata3` for scenarios 2 and 3 respectively. The only notable difference here, is that we add a new row in the dataset for which we only specify the covariates associated with time and the intermediate event accordingly but we specify the longitudinal outcome as missing for the specific row, since no data are available.

```
# DATA FOR SCENARIO 2
# create copy of newdata1 and call it newdata2
newdata2 <- newdata1
# Specify the time of the intermediate event to be equal
# to the last observed time point plus
# an arbitrarily small value
newdata2$Interm.Evnt.Time <- newdata2$time[nrow(newdata2)] + 0.00001
# create additional row
new.row <- newdata2[nrow(newdata2), ]
new.row$time <- newdata2$Interm.Evnt.Time[1]
new.row$Y <- NA
# attach new row to the dataset
newdata2 <- rbind(newdata2, new.row)
# compute time-varying covariates
newdata2$Int.Evnt.Index <- as.numeric(newdata2$time >= newdata2$Interm.Evnt.Time)
newdata2$time.relative <- pmax(0, newdata2$time - newdata2$Interm.Evnt.Time)

# DATA FOR SCENARIO 3
# create copy of newdata1 and call it newdata3
newdata3 <- newdata1
# Specify the time of the intermediate event to be equal
# to the last observed time point
# plus 1
newdata3$Interm.Evnt.Time <- newdata3$time[nrow(newdata3)] + 1
# create additional row
new.row <- newdata3[nrow(newdata3), ]
new.row$time <- newdata3$Interm.Evnt.Time[1]
```

```
new.row$Y <- NA
# attach new row to the dataset
newdata3 <- rbind(newdata3, new.row)
# compute time-varying covariates
newdata3$Int.Evnt.Index <- as.numeric(newdata3$time >= newdata3$Interm.Evnt.Time)
newdata3$time.relative <- pmax(0, newdata3$time - newdata3$Interm.Evnt.Time)
```

The changes in `newdata2` and `newdata3` are shown in Tables 2.13 2.14, respectively. Finally, the predictions under the different scenarios can be obtained by applying function `survfitJM()` to the fitted models and datasets we just created:

```
# Predictions for scenario 1
preds.1 <- survfitJM(object = joint.fit, newdata = newdata1, idVar = "ID")
# Predictions for scenario 2
preds.2 <- survfitJM(object = joint.fit, newdata = newdata2, idVar = "ID")
# Predictions for scenario 3
preds.3 <- survfitJM(object = joint.fit, newdata = newdata3, idVar = "ID")
```

Table 2.13: Example Dataset

ID	time	Y	Interm.Evnt.Time	Int.Evnt.Index	time.relative
New	0.5548873	15.86804	7.540984	0	0
New	1.3547850	26.81457	7.540984	0	0
New	2.1550550	13.02823	7.540984	0	0
New	6.5422016	29.46144	7.540984	0	0
New	7.5409739	23.48326	7.540984	0	0
New	7.5409839	NA	7.540984	1	0

2.D Data Availability Statement

The data used for the Pulmonary Gradient analysis are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions. The data used for the SPRINT trial are available from BioLINCC.

Table 2.14: Example Dataset

ID	time	Y	Interm.Evnt.Time	Int.Evnt.Index	time.relative
New	0.5548873	15.86804	8.540974	0	0
New	1.3547850	26.81457	8.540974	0	0
New	2.1550550	13.02823	8.540974	0	0
New	6.5422016	29.46144	8.540974	0	0
New	7.5409739	23.48326	8.540974	0	0
New	8.5409739	NA	8.540974	1	0

2.5 References

Andrinopoulou, E.-R., Rizopoulos, D., Takkenberg, J. J., and Lesaffre, E. (2017). Combined dynamic predictions using joint models of two longitudinal outcomes and competing risk data. *Statistical Methods in Medical Research*, 26(4):1787–1801. PMID: 26059114.

Brown, E. R., Ibrahim, J. G., and DeGruttola, V. (2005). A flexible b-spline model for multiple longitudinal biomarkers and survival. *Biometrics*, 61(1):64–73.

Ferrer, L., Putter, H., and Proust-Lima, C. (2019). Individual dynamic predictions using landmarking and joint modelling: Validation of estimators and robustness assessment. *Statistical Methods in Medical Research*, 28(12):3649–3666. PMID: 30463497.

Group, T. S. R. (2015). A randomized trial of intensive versus standard blood-pressure control. *New England Journal of Medicine*, 373(22):2103–2116. PMID: 26551272.

Henderson, R., Diggle, P., and Dobson, A. (2002). Identification and efficacy of longitudinal markers for survival. *Biostatistics*, 3(1):33–50.

Hernán, M. A. and Robins, J. M. (2019). *Causal Inference*. Chapman & Hall/CRC, forthcoming, Boca Raton.

Ibrahim, J. G., Chen, M.-H., and Sinha, D. (2001). *Bayesian Survival Analysis*. Springer-Verlag, New York.

Mauff, K., Steyerberg, E. W., Nijpels, G., van der Heijden, A. A., and Rizopoulos, D. (2017). Extension of the association structure in joint models to include weighted cumulative effects: K. MAUFF *ET AL* . *Statistics in Medicine*, 36(23):3746–3759.

Rizopoulos, D. (2012). *Joint Models for Longitudinal and Time-to-Event Data*. Chapman & Hall/CRC, New York.

Rizopoulos, D. (2016). The r package jmbayes for fitting joint models for longitudinal and time-to-event data using mcmc. *Journal of Statistical Software, Articles*, 72(7):1–46.

Rizopoulos, D., Hatfield, L. A., Carlin, B. P., and Takkenberg, J. J. M. (2014). Combining dynamic predictions from joint models for longitudinal and time-to-event data using bayesian model averaging. *Journal of the American Statistical Association*, 109(508):1385–1397.

Rizopoulos, D., Molenberghs, G., and Lesaffre, E. M. (2017). Dynamic predictions with time-dependent covariates in survival analysis using joint modeling and landmarking. *Biometrical Journal*, 59(6):1261–1276.

Sène, M., Taylor, J. M., Dignam, J. J., Jacqmin-Gadda, H., and Proust-Lima, C. (2016). Individualized dynamic prediction of prostate cancer recurrence with and without the initiation of a second treatment: Development and validation. *Statistical Methods in Medical Research*, 25(6):2972–2991.

Suresh, K., Taylor, J. M., Spratt, D. E., Daignault, S., and Tsodikov, A. (2017). Comparison of joint modeling and landmarking for dynamic prediction under an illness-death model. *Biometrical Journal*, 59(6):1277–1300.

Taylor, J. M., Yu, M., and Sandler, H. M. (2005). Individualized predictions of disease progression following radiation therapy for prostate cancer. *Journal of Clinical Oncology*, 23(4):816–825. PMID: 15681526.

Tsiatis, A. A. and Davidian, M. (2004). Joint modeling of longitudinal and time-to-event data: An overview. *Statistica Sinica*, 14(3):809–834.

Wulfsohn, M. S. and Tsiatis, A. A. (1997). A joint model for survival and longitudinal data measured with error. *Biometrics*, 53(1):330–339.

Chapter 3

Influence of pregnancy on long-term durability of allografts in right ventricular outflow tract

This chapter is based on:

Romeo, J.L.R.*, Papageorgiou, G*, Takkenberg, J.J.M., Roos-Hesselink, J.W., van Leeuwen W.J., Cornette J.M.J., Rizopoulos, D., Bogers A.J.J.C. and Mokhles, M.M. (2020), Influence of pregnancy on long-term durability of allografts in right ventricular outflow tract *The Journal of Thoracic and Cardiovascular Surgery*, 159: 1508–1516.E1. doi: https://doi.org/10.1016/j.jtcvs.2019.08.083

* both authors contributed equally

Abstract

Background
There is very limited published evidence about the influence of pregnancy on allograft durability in right ventricular outflow tract (RVOT) position. We present the first study using mixed and joint modeling.
Methods
This retrospective study compared clinical and valve related outcomes of all consecutive female hospital survivors in their fertile life phase (18–50 years) based on pregnancy. Serial echocardiographic measurements of pulmonary gradient and regurgitation were analyzed for their association with valve replacement using joint models for longitudinal and time-to-event data. Occurrence of first pregnancy was included as a time-dependent intermediate event in both the longitudinal and survival analyses of the joint model in order to assess its impact on the hemodynamic and clinical outcome.
Results
In total, 196 consecutive women in their fertile life-phase with an allograft were included. Complete information of 176 (90%) allografts in 165 women was available including 1395 echocardiograms. Of these women, 51 (30.9%) women had 84 completed pregnancies at an average age of 29.1 ± 3.9 (SD) years; 8.1 ± 6.1 years since allograft implantation. Tetralogy of Fallot was the most common diagnosis in both groups. After a mean follow up of 15.2 years (range 0.1-30), 7 (13.7%) parous underwent valve replacement, versus 20 (17.5%) nulliparous. During this follow up, the mean allograft gradient in parous (24.2 mmHg) and nulliparous (21.0 mmHg) was comparable (p=.225). One mmHg increase in pulmonary gradient increased the instantaneous risk of pulmonary valve replacement (PVR) by a ratio of 1.051 (p<.001), regardless of pregnancy. Similarly, development of moderate or severe regurgitation increased the risk of PVR (p=.038), regardless of pregnancy. Pregnancy was not associated with a change in the allograft gradient (p=.258), regurgitation grade (p=.774) or hazard of PVR (p=.796) during follow-up.
Conclusions
Pregnancy is not associated with impaired allograft durability in women with good cardiac health.

3.1 Introduction

Congenital heart disease (CHD) is present in 8-10 per 10.000 live births and often predisposes towards increased morbidity and earlier mortality, compared to the normal population (van der Bom et al., 2012; Gilboa et al., 2016; O'Leary et al., 2016; Hoffman and Kaplan, 2002) [1-4]. The prevalence of CHD in young adulthood is expected to increase due to constantly improving surgical and clinical treatment options enabling patients to reach their fertile life phase (van der Bom et al., 2012; Gilboa et al., 2016; van der Linde et al., 2011)[1, 2, 5]. In this regard, clinicians are confronted with a growing population of adults with congenital heart disease (ACHD). Approximately 20% of all CHD involves the right ventricular outflow tract (RVOT). Pulmonary valve replacement (PVR) has, therefore, become the most frequently performed procedure in ACHD comprising approximately 10% of all procedures (T.S.o.T.S., 2017)[6]. Suitable valve alternatives for RVOT reconstruction, however, are still scarce. The allograft is generally preferred over biological or mechanical prostheses for its excellent hemodynamic capabilities, durability and independence from anticoagulance which could be especially beneficial for childbearing women (Mokhles et al., 2011; Wong et al., 1993; Steinberg et al., 2017; van Hagen et al., 2015)[7-10]. Structural valve deterioration (SVD) remains the main caveat of the allograft however, ultimately indicating repeated valve intervention (Romeo et al., 2018b)[11]. As pregnancy induces hormonal changes and an important hemodynamic burden by increasing circulating volume, heart rate and bloodpressure (Stout and Otto, 2007; Basquin et al., 2010)[12, 13] it could potentially affect allograft durability. During the life long care and follow up of female ACHD, the wish and possibilities of pregnancy along with potential risks and long term outcome must be considered and actively addressed; ideally prior to pregnancy (Stout, 2005)[14]. This information could also be used in pregnancy heart team meetings on women considering pregnancy and requiring valve surgery, as recommended by the most recent guidelines (Class 1 recommendation) (Regitz-Zagrosek et al., 2018)[15]. This is the first report about the influence of pregnancy on long term function and durability of allograft conduits in RVOT position in women of childbearing age, using repeated measurement analysis.

3.2 Methods

3.2.1 Study design

All consecutive patients who underwent RVOT reconstruction between April 1986 and January 2018 with an allograft at the Cardio-Thoracic Surgery department of the

Erasmus Medical Center were included and screened for pregnancies. The Erasmus Medical Center is a tertiary referral center specialized in (congenital) cardiac surgery. All female patients older than 18 years at the time of this study and younger than 50 years at moment of surgery were retrospectively included. A detailed questionnaire was sent out to all women who confirmed pregnancy and gave informed consent. A supplementary telephone interview was conducted for additional information. The Medical Ethics Commission reviewed and approved this study (MEC 17-158) after which written informed consent was acquired prior to onset of this study.

3.2.2 Operation technique

Allograft implantation was performed with standard cardiopulmonary bypass with moderate- or normothermia on a beating heart after median sternotomy. In case of concomitant intracardiac procedures the aorta was cross-clamped after which myocardial protection was achieved with crystalloid cardioplegia (St. Thomas' solution). Allografts were interposed between the right ventricle and pulmonary artery with running polypropylene 5.0 sutures. Our center exclusively uses cryopreserved allografts for RVOT reconstruction, preferring pulmonary allografts over aortic ones. Our institutional policy does not indicate any anticoagulation for homografts postoperatively after adequate mobilization.

3.2.3 Definitions

For this study, pregnancies were considered when reaching beyond >20 weeks of gestation. Parous and nulliparous women were thus defined as respectively having or having not experienced a pregnancy and delivery beyond 20 weeks of gestation. Gestational age was based on first trimester ultrasound and last menstrual period. After pulmonary valve replacement, follow up by a cardiologist was regularly performed in all women during outpatient visits after one and six months and annually thereafter. Cardiac and valve related outcomes were reported according to the guidelines for reporting mortality and morbidity after cardiac valvular interventions (Akins et al., 2008)[16]. All interventions were discussed during weekly multidisciplinary meetings involving (congenital) cardiologist, cardiac surgeons and radiologists. The primary endpoint of this study was pulmonary valve replacement (PVR) either surgically or percutaneously.

3.2.4 Echocardiographic

Allograft function was repeatedly assessed using standardized transthoracic echocardiograms during outpatient visits. Peak pulmonary gradient (mmHg) and regurgita-

tion grade were assessed from a multi-window perspective according to the guidelines for the assessment of valve stenosis and regurgitation (Baumgartner et al., 2009; Lancellotti et al., 2010; Cleuziou et al., 2015)[17-19]. Peak pulmonary gradient was calculated with the modified Bernoulli equation from the peak transvalvular speed (m/s) and pulmonary regurgitation was graded qualitatively as non-significant (i.e. no regurgitation, trace or mild) or significant (moderate or severe).

3.2.5 Statistics

Baseline characteristics and measurements are summarized using descriptive statistics. Continuous measurements are summarized using means ± standard deviations or medians with range. Qualitative measurements are summarized as absolute counts with percentages. All statistical tests were two-sided with a p-value lower than 0.05 considered as significant. Kaplan-Meier plots were used to describe cumulative freedom from valve replacement and mortality. Serial echocardiographic measurements of pulmonary gradient and regurgitation were analyzed using a bivariate non-linear mixed-effects model, with a linear mixed-effects regression sub-model for the pulmonary gradient, and a mixed-effects logistic regression sub-model for pulmonary regurgitation. Both longitudinal sub-models were adjusted for the effects of age, donor sex, allograft diameter and type of allograft (i.e. aortic or pulmonary). The effect of pregnancy was determined by including a nonlinear effect of time using natural cubic splines from the onset of pregnancy in the sub-models of both pulmonary gradient and regurgitation grade. For the random effects structure we used a nonlinear effect of time as well as a nonlinear effect of time from the onset of pregnancy using natural cubic splines. A time-dependent Cox model was used for the instantaneous risk of valve replacement. The value of both the pulmonary gradient and regurgitation, as estimated by the joint mixed-effects model, were included in the Cox model in order to determine their association with the instantaneous risk for valve replacement. The occurrence of a first pregnancy was entered in the model as a binary time-varying variate. Furthermore, interaction terms between the current value of the longitudinal trajectories of pulmonary gradient and regurgitation were included to investigate potential changes in the magnitude of the association between the hemodynamic outcomes and the instantaneous risk for valve replacement due to the occurrence of first pregnancy. The effect pregnancy on allograft function over time was thus assessed in three manners: 1. The effect of pregnancy on the longitudinal evolution of pulmonary gradient and regurgitation grade. 2. The effect of pregnancy as an independent covariate in a conventional proportional hazards model. 3. The interaction effect of pregnancy and the evolutions of both pulmonary gradient and the regurgitation grade in the joint model. I.e. does the occurrence of pregnancy alter the relationship between

both markers and PVR? Statistical analyses were performed with SPSS (IBM Corp. Released 2012. IBM SPSS Statistics for Windows, Version 24.0. Armonk, NY: IBM Corp.) and R (R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/).

3.3 Results

During the study period, 280 consecutive female patients underwent allograft implantation. Overall hospital mortality was 3.2% (n=9). We excluded 11 (3.9%) patients who were over 50 years at the time of surgery, 48 (17.4%) patients who had not reached 18 years and 16 (5.7%) patients who were lost to follow up due to return to their country of origin immediately post-surgery. Of the 196 allografts eligible for inclusion, 176 (90%) had complete clinical and echocardiographic follow-up (Table 1). None of the patients who underwent repeated PVR had multiple pregnancies during the follow-up time of different allografts. Compared to nulliparous women, parous women had less often right ventricular hypertrophy (12% vs 26%, p=.046) at time of surgery. Other baseline characteristics were comparable (Table 3.1).

3.3.1 Pregnancy outcome

The cardiac and obstetric course and fetal outcomes of all pregnancies have been extensively described previously (Romeo et al., 2018a)[20]. In short, parous women were in overall good cardiac and clinical health prior to pregnancy with an average pulmonary gradient of 19 mmHg, non-significant regurgitation in 83.3% and significant regurgitation in 16.7%. Mean age at last menstruation was 29.6± 4.3 years, which was 8.6± 6.6 years after allograft implantation. Two women were identified who became pregnant despite explicit contraindications on cardiac grounds. The first had a severely dilated and dysfunctioning left ventricle and the second had severe pulmonary hypertension. They were induced into premature labor but recovered uneventfully. Both still have their homograft in situ without signs of valve deterioration. No woman died during pregnancy or within the first year postpartum.

3.3.2 Long term clinical outcome

Median follow-up length from operation to last follow-up was 18.0 years (mean 17.4 years; range 2.0-30.0 years, total patient-years 887) and 13.7 years (mean 14.3 years; range 0.01-28.6 years, total patient-years 1784) for parous and nulliparous women,

Table 3.1: Perioperative characteristics of 176 female allograft recipients

	Total	Nulliparous	Parous	P-value
Allografts n(%)	176	125 (71)	51 (29)	
Aortic allograft	18 (10)	16 (13)	2 (4)	.078
Pulmonry allograft	158 (90)	109 (87)	49 (96)	
Pregnancies n (%)		0	84	NA
1			26	
2			20	
≥ 3			5	
Age at valve replacement (years)	22.2± 12.5	22.7± 14.1	21.4± 7.3	.467
Age at first pregnancy (years)		NA	29.1± 3.9	NA
Time of pregnancy since PVR (years)		NA	8.1± 6.1	NA
Height at valve replacement (m)	151±31	147±34	159±17	0.16
Weight at valve replacement (kg)	56±18	56 ±19	56±15	.925
Creatinine	59±18	58±19	61±15	.341
Right ventricular hypertrophy n(%)	38 (22)	32 (26)	6 (12)	.046
Left ventricular hypertrophy n(%)	27 (15)	22 (18)	5 (10)	.251
Diagnosis n(%)				
TOF	72 (41)	49 (39)	23 (45)	.235
Aortic Pathology*	42 (24)	27 (22)	15 (29)	
PA + VSD	19 (11)	16 (13)	3 (6)	
TA	12 (7)	11 (9)	1 (2)	
TGA + PA/PS	5 (3)	4 (3)	1 (2)	
PA-VSD	5 (3)	3 (2)	2 (4)	
Discordant	5 (3)	5 (4)	0 (0)	
Tricuspid Atresia	1 (1)	0 (0)	1 (2)	
Other	15 (9)	10 (8)	5 (10)	
Donor sex (male) n (%)	89 (51)	65 (52)	24 (47)	.619
Allograft size (mm)	23 (11 - 30)	23 (11 - 30)	24 (16 - 28)	.206
Perfusion time (min) (n = 161)	145 (35-465)	143 (35-465)	151 (39-295)	.691
Pulmonary valve replacement	27	20 (16)	7 (14)	.332
Severe stenosis	14	12 (60)	2 (29)	
Severe regurgitation	6	4 (20)	2 (29)	
Mixed stenosis and regurgitation	7	4 (20)	3 (43)	

NYHA: New York Heart Association, PA: Pulmonary Atresia, PS: Pulmonary Stenosis, TA: Truncus Arteriosus, TGA: Transposition of the Great Arteries, TOF:Tetralogy of Fallot, VSD:Ventricular Septal Defect.
*Primary aortic valve pathology for which a pulmonary autograft ('Ross') procedure was performed

respectively. Thirteen women died during follow up (0.73%/patient-year), all nulliparous, with 10 (0.56%/patient-year) cases of cardiac death and 3 cases (0.17%/patient-year) of non-cardiac related death. Among the cardiac related deaths there was 1 (0.06%/patient-year) case of valve related death, 4 (0.22%/patient-year) cases of sudden unexplained death and 5 (0.28%/patient-year) cases of non-valve related cardiac death all comprising end-stage heart failure. One nulliparous woman died during an attempt to percutaneously implant a Melody valve indicated by severe stenosis 23.0 years after initial allograft implantation. All other reinterventions were performed successfully. Following allograft implantation, there were 4 (3.2%) percutaneous balloon dilatation in the nulliparous, and 2 (3.9%) in the parous women. In the parous group there were 7 (0.8%/patient-year) PVR after 14.3± 4.9 years after initial surgery. Six were surgically (4 allografts, 1 mechanical valve, 1 bioprosthesis (Medtronic Freestyle® valve)) and 1 was percutaneously (Melody® valve). Among the nulliparous 20 (1.24%/patient-year) underwent PVR after 14.1± 5.5 years of which 13 surgically (all allografts) and 7 transcatheter Melody® valve implantations. All PVR were indicated by structural valve deterioration. Freedom from valve replacement after 20 years was 78± 16% for parous and 66± 14% for nulliparous women (Figures 3.1 and 3.2).

3.3.3 Echocardiographic outcome

There were 1395 serial echocardiograms available for all included patients (mean 7, range 1-26) serial echocardiograms/patient). For women with multiple pregnancies, only the effect of the first pregnancy on the allograft hemodynamics were analyzed. Measurements in those cases were censored from the onset of the second pregnancy. The median time for women to have their first pregnancy after allograft implantation was 8 years. Figure 3.3 present the longitudinal evolutions of pulmonary gradient and regurgitation for scenarios with and without pregnancy. Figure 3.4 present both longitudinal evolutions along with the influence of pregnancy on the hazard of PVR.

The analyses of the association between both gradient and regurgitation, and the hazard of PVR showed that an 1 mmHg increase in the pulmonary gradient, increases the instantaneous risk of PVR by a ratio of 1.051 (p<.001). Similarly, an increase in the probability of significant regurgitation increases the instantaneous risk of PVR by a ratio of 1.209 (p=0.038). Pregnancy was, however, not associated with an increased hazard of PVR (p=0.796, Table 3.2). The interaction between the longitudinal pulmonary gradient and regurgitation grade with pregnancy was non-significant (p=0.258 and p=0.774, respectively), indicating that pregnancy did not change the association between the hemodynamic outcomes and the hazard of valve replacement.

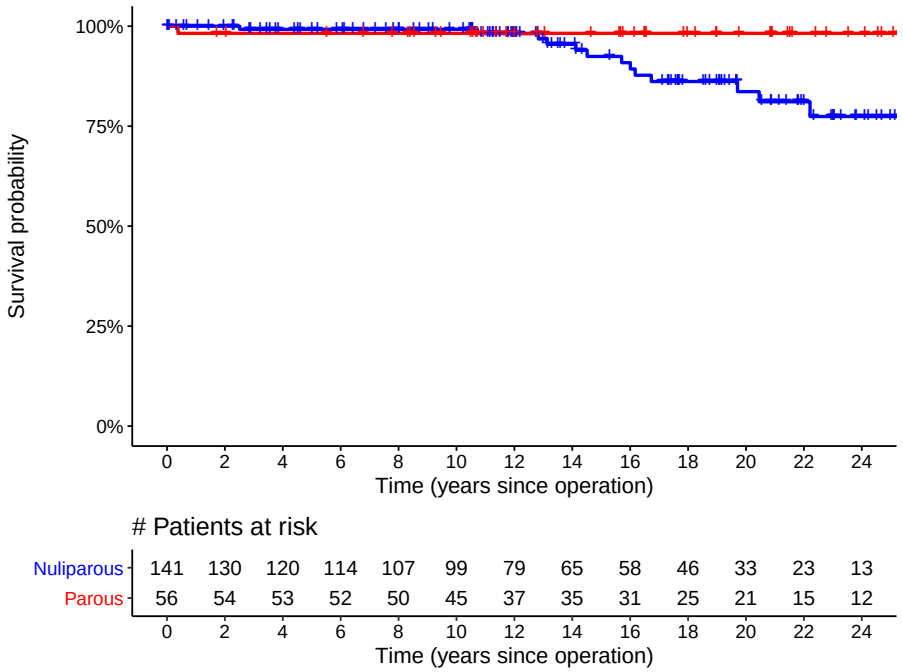

Figure 3.1: Kaplan-Meier curves of overall survival after allograft replacement

3.4 Discussion

We described the long term effect of pregnancy on the clinical and longitudinal echocardiographic outcome of allografts in RVOT position. The results show that the occurrence of pregnancy at any moment after allograft implantation is not associated with an increased hazard of PVR. Therefore, there is no evidence to support that pregnancy negatively influences allograft function over time.

Increased rates of complications during pregnancy like premature labor, fetal demise, severe bleeding and heart failure have been reported among a wide spectrum of corrected CHD, compared to the general population (Romeo et al., 2018a; Trigas et al., 2014; Balci et al., 2011; Bédard et al., 2009; Yap et al., 2010; Grewal et al., 2009)[20-25]. Much less is known about cardiac outcome late after pregnancy and impact of pregnancy on cardiac function. As life expectancy among practically all ACHD has substantially improved during the last decades, the number of pregnancies will likely increase as well (Gilboa et al., 2016; O'Leary et al., 2016)[2, 3]. Multiple studies have demonstrated mechanical prostheses to carry extensive risks of pregnancy

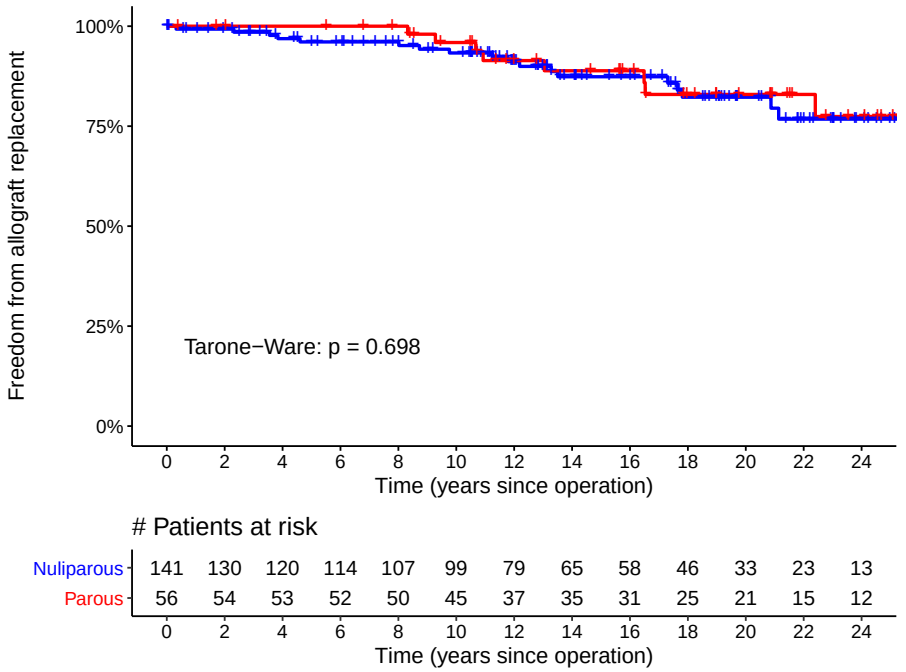

Figure 3.2: Kaplan-Meier curves of freedom from allograft replacement

related complications (Steinberg et al., 2017; van Hagen et al., 2015; Ayad et al., 2016)[9, 10, 26]. Previous studies on pregnancy have mainly focused on bioprostheses and mechanical prostheses in the aortic position (Arabkhani et al., 2012; Doer and Somerville, 1997; North et al., 1999; Basude et al., 2014)[27-30]. In their study on allografts and autografts in the aortic position, Arabkhani et al. described the influence of pregnancy on the longitudinal evolution of gradient, sino-tubular junction diameter, annulus diameter and regurgitation grade. They reported no significant influence of pregnancy on all mentioned evolutions with a comparable freedom from valve reoperation between parous and nulliparous (Arabkhani et al., 2012)[27]. In another report on allografts in aortic position North et al. reported a 10 year freedom from valve loss defined as allograft replacement or valve-related death of 72%, an no influence of pregnancy (North et al., 1999)[29].

Although allografts and bioprostheses are less thrombogenic and therefore independent from anticoagulation, they suffer from structural deterioration (Romeo et al., 2018b; Christenson et al., 2010)[11, 31]. The outcome of parous women with RVOT allograft conduits has been reported earlier in smaller (Avila et al., 2002; Cleuziou

Table 3.2: Joint model estimates of the association with pulmonary valve replacement

	HR	95% CI	P
Occurrence of first pregnancy	0.726	(0.049 - 4.482)	0.796
Age	1	(0.951 - 1.030)	0.890
Current value of gradient	1.051	(1.030 - 1.062)	<0.001
Current value of regurgitation	1.209	(1.010 - 1.537)	0.038
Interaction with occurrence of first pregnancy:			
Current value of gradient	1.030	(0.980 - 1.105)	0.258
Current value of regurgitation	1.030	(0.844 - 1.391)	0.774

CI = Credible interval, HR = Hazard Ratio

et al., 2010; Eric Jamieson et al., 1995; Sadler et al., 2000)cohorts[32-35]. However, these studies only focused on complications during or directly after pregnancy without reporting the longitudinal allograft function over time and the effect of pregnancy on this outcome. These studies have generally reported positive clinical outcomes (Avila et al., 2002; Cleuziou et al., 2010; Eric Jamieson et al., 1995; Sadler et al., 2000)[32-35]. Cleuziou et al. studied the rate of degeneration of 45 bioprostheses and 42 allografts, of which 40 in RVOT position. Pregnancy was not a risk factor for accelerated valve failure with freedom from valve replacement at 10 years of 73% and 52% (p=0.2) for patients with and without pregnancy, respectively (Cleuziou et al., 2010)[33]. Oosterhof et al. studied the risk of allograft failure and dysfunction after repair of Tetralogy of Fallot. Data from the Zahara study were included and indicated no effect of pregnancy on allograft durability (Oosterhof et al., 2006)[36]. Metz et al. were the only authors to report a negative influence of pregnancy on general cardiac health in patients after pulmonary valve surgery. Among women who underwent PVR or pulmonary repair, 33 women with 95 completed pregnancies were compared to 20 nulliparous. A higher incidence rate of the composite outcome of unanticipated cardiac surgery, heart failure and death was reported in the parous group compared to nulliparous women. However, when the authors looked closely at this group of patients the found that adverse cardiac outcome after pregnancy was more prevalent in patients with pulmonary valve surgery (not replacement) compared to those did

had undergone PVR. (Metz et al., 2013)[37]. The mixed cohort of Metz et al. including a modest proportion of patients with replaced valves can potentially explain their findings. In general however, research about the influence of pregnancy on prosthetic valve durability is limited and even though in line with our results, often lack statistically robust methodology (Arabkhani et al., 2012; Avila et al., 2002; Cleuziou et al., 2010; Eric Jamieson et al., 1995; Sadler et al., 2000; Oosterhof et al., 2006; Heuvelman et al., 2013)[27, 32-36, 38].

In our cohort, parous women differed in baseline characteristics only with regard to right ventricular hypertrophy. Other baseline characteristics were comparable. However, cardiac and general health, other than allograft function alone, might have influenced the desire, ability and advise from consulting cardiologists to become pregnant. In parous women, pregnancy did not influence allograft function and hazard of PVR. However, the results and, therefore, conclusions are restricted to parous women and are therefore more likely to be in better cardiac health. Only two women could be identified who became pregnant despite explicit contraindications on cardiac grounds.

This study shows that longitudinal allograft function over time is unaffected by the occurrence of pregnancy. Therefore, pregnancy does not influence durability of allografts as demonstrated by using a novel and advanced biostatistical framework (Papageorgiou et al., 2019)[39]. The use of mixed and joint modelling has been limited in Cardio-Thoracic research even though it has been recommended by the 2008 guidelines for reporting mortality and morbidity after cardiac valve interventions (Akins et al., 2008; Arabkhani et al., 2012; Mokhles et al., 2014)[16, 27, 40]. Combining serial measurements of any biomarker with a time dependent event is a statistically robust method to test the potential influence of the latter, by more optimally utilizing available data.

3.4.1 Strength and Limitations

We assessed the effect of pregnancy on RVOT allograft durability was longitudinally assessed using innovative statistical methods. Furthermore, follow-up completeness was high and a large number of standardized echocardiograms were available for the analyses. In contrast to earlier studies, innovative statistical methods were used to analyze the data. Because pregnancy is ethically and practically impossible to randomize, future research should combine these type of innovative statistics with data from prospectively followed cohorts.

The exact potential mechanism through which pregnancy could influence long term allograft function is largely unknown. As the change in hemodynamics are transitory

during pregnancy, it would be more clinically accurate to model a temporary effect of pregnancy which fades of after labor. However, this approach would require an extensive amount of measurements before, during and after pregnancy to reliably estimate the temporal effect and was therefore impossible in the current analysis. Furthermore, while the effect of pregnancy on pulmonary gradient and pulmonary regurgitation was not found to substantially alter their evolutions, a small change was detected. Taking into account that only 13.7% of the subjects experienced pregnancy, results should be interpreted with care since power is low.

3.5 Conclussion

The occurrence of pregnancy at any moment during at least the first 15 years after RVOT reconstruction with an allograft conduit does not influence allograft durability.

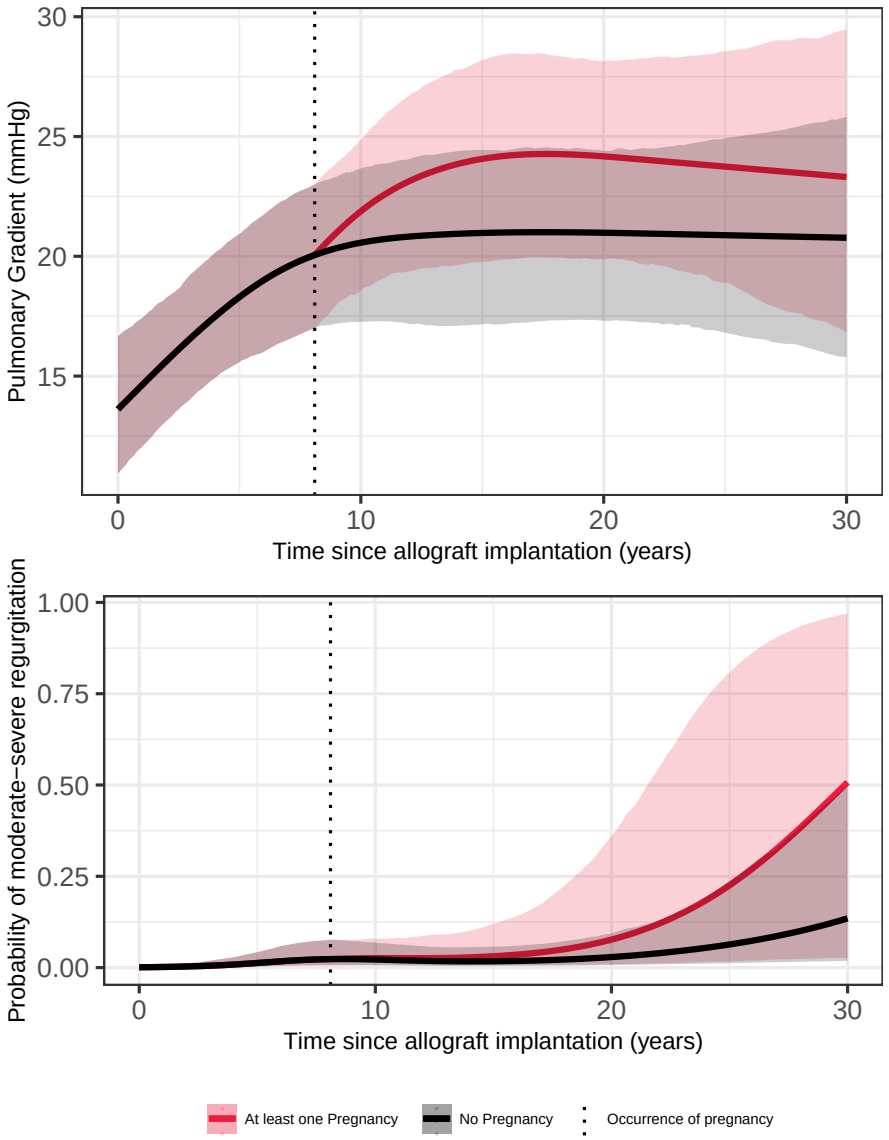

Figure 3.3: Mixed-effects models of transpulmonary peak gradient (mmHg) and moderate-severe regurgitation

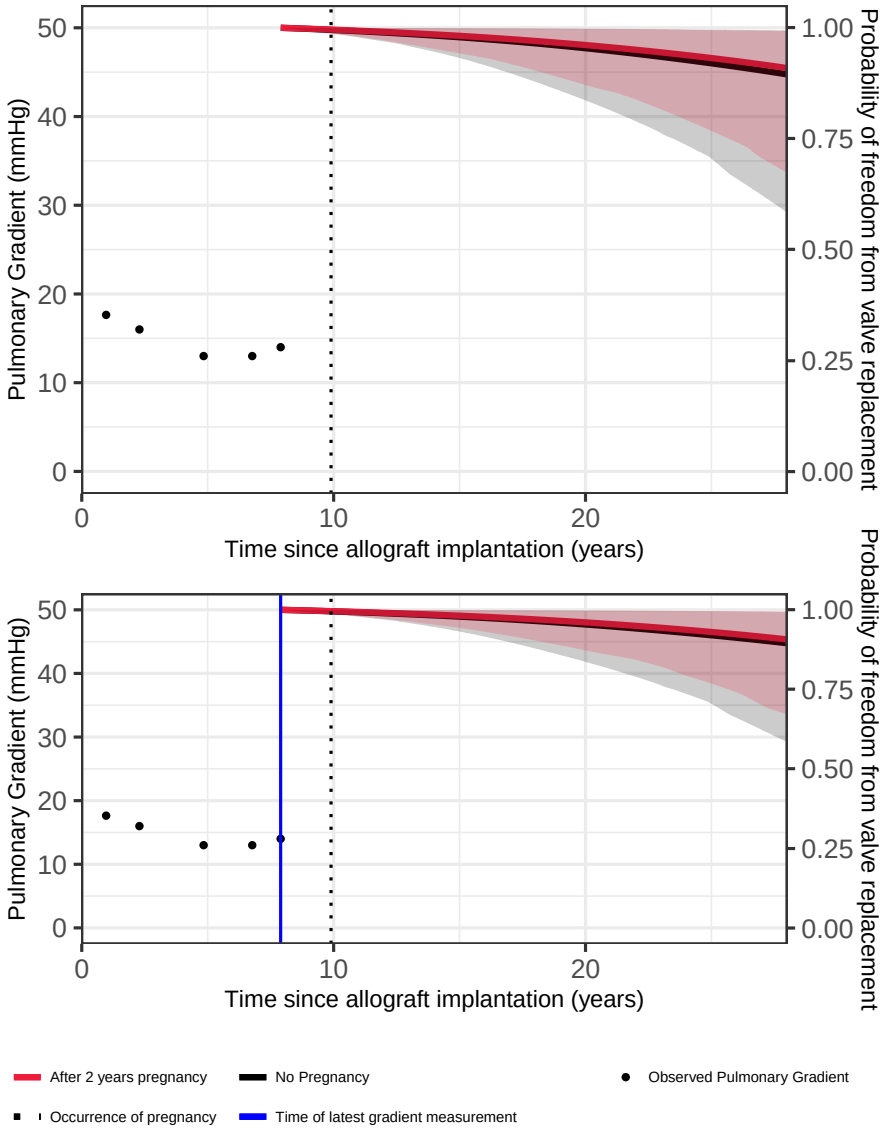

Figure 3.4: Influence of pregnancy on freedom from valve replacement

3.6 References

Akins, C. W., Miller, D. C., Turina, M. I., Kouchoukos, N. T., Blackstone, E. H., Grunkemeier, G. L., Takkenberg, J. J., David, T. E., Butchart, E. G., Adams, D. H., Shahian, D. M., Hagl, S., Mayer, J. E., and Lytle, B. W. (2008). Guidelines for reporting mortality and morbidity after cardiac valve interventions. *The Annals of Thoracic Surgery*, 85(4):1490 – 1495.

Arabkhani, B., Heuvelman, H. J., Bogers, A. J., Mokhles, M. M., Roos-Hesselink, J. W., and Takkenberg, J. J. (2012). Does pregnancy influence the durability of human aortic valve substitutes? *Journal of the American College of Cardiology*, 60(19):1991–1992.

Avila, W. S., Rossi, E. G., Grinberg, M., and Ramires, J. A. F. (2002). Influence of pregnancy after bioprosthetic valve replacement in young women: a prospective five-year study. *The Journal of heart valve disease*, 11(6):864—869.

Ayad, S. W., Hassanein, M. M., Mohamed, E. A., and Gohar, A. M. (2016). Maternal and fetal outcomes in pregnant women with a prosthetic mechanical heart valve. *Clinical Medicine Insights: Cardiology*, 10:CMC.S36740. PMID: 26884686.

Balci, A., Drenthen, W., Mulder, B. J., Roos-Hesselink, J. W., Voors, A. A., Vliegen, H. W., Moons, P., Sollie, K. M., van Dijk, A. P., van Veldhuisen, D. J., and Pieper, P. G. (2011). Pregnancy in women with corrected tetralogy of fallot: Occurrence and predictors of adverse events. *American Heart Journal*, 161(2):307 – 313.

Basquin, A., Pineau, E., Galmiche, L., Bonnet, D., Sidi, D., and Boudjemline, Y. (2010). Transcatheter valve insertion in a model of enlarged right ventricular outflow tracts. *The Journal of Thoracic and Cardiovascular Surgery*, 139(1):198 – 208.

Basude, S., Trinder, J., Caputo, M., and Curtis, S. L. (2014). Pregnancy outcome and follow-up cardiac outcome in women with aortic valve replacement. *Obstetric Medicine*, 7(1):29–33. PMID: 27512416.

Baumgartner, H., Hung, J., Bermejo, J., Chambers, J. B., Evangelista, A., Griffin, B. P., Iung, B., Otto, C. M., Pellikka, P. A., Quiñones, M., of Echocardiography, A. S., and of Echocardiography, E. A. (2009). Echocardiographic assessment of valve stenosis: Eae/ase recommendations for clinical practice. *Journal of the American Society of Echocardiography : official publication of the American Society of Echocardiography,*, 22(1):1 – 102.

Bédard, E., Dimopoulos, K., and Gatzoulis, M. A. (2009). Has there been any progress made on pregnancy outcomes among women with pulmonary arterial hypertension? *European Heart Journal*, 30(3):256–265.

Christenson, J. T., Sierra, J., Colina Manzano, N. E., Jolou, J., Beghetti, M., and Kalangos, A. (2010). Homografts and xenografts for right ventricular outflow tract reconstruction: Long-term results. *The Annals of Thoracic Surgery*, 90(4):1287 – 1293.

Cleuziou, J., Hörer, J., Kaemmerer, H., Teodorowicz, A., Kasnar-Samprec, J., Schreiber, C., and Lange, R. (2010). Pregnancy does not accelerate biological valve degeneration. *International Journal of Cardiology*, 145(3):418 – 421.

Cleuziou, J., Vitanova, K., Kasnar-Samprec, J., Hörer, J., Lange, R., and Schreiber, C. (2015). Durability of down-sized homografts for the reconstruction of the right ventricular outflow tract †. *European Journal of Cardio-Thoracic Surgery*, 49(5):1421–1425.

Doer, A. and Somerville, J. (1997). Pregnancy in patients with pulmonary autograft valve replacement. *European Heart Journal*, 18(10):1659–1662.

Eric Jamieson, W., Craig Miller, D., Akins, C. W., Ian Munro, A., Glower, D. D., Moore, K. A., and Henderson, C. (1995). Pregnancy and bioprostheses: Influence on structural valve deterioration. *The Annals of Thoracic Surgery*, 60:S282 – S287.

Gilboa, S. M., Devine, O. J., Kucik, J. E., Oster, M. E., Riehle-Colarusso, T., Nembhard, W. N., Xu, P., Correa, A., Jenkins, K., and Marelli, A. J. (2016). Congenital heart defects in the united states. *Circulation*, 134(2):101–109.

Grewal, J., Siu, S. C., Ross, H. J., Mason, J., Balint, O. H., Sermer, M., Colman, J. M., and Silversides, C. K. (2009). Pregnancy outcomes in women with dilated cardiomyopathy. *Journal of the American College of Cardiology*, 55(1):45 – 52.

Heuvelman, H. J., Arabkhani, B., Cornette, J. M., Pieper, P. G., Bogers, A. J., Takkenberg, J. J., and Roos-Hesselink, J. W. (2013). Pregnancy outcomes in women with aortic valve substitutes. *The American Journal of Cardiology*, 111(3):382 – 387.

Hoffman, J. I. E. and Kaplan, S. (2002). The incidence of congenital heart disease. *Journal of the American College of Cardiology*, 39(12):1890—1900.

Lancellotti, P., Tribouilloy, C., Hagendorff, A., Moura, L., Popescu, B. A., Agricola, E., Monin, J.-L., Pierard, L. A., Badano, L., Zamorano, J. L., on behalf of the European Association of Echocardiography, Reviewers:, D., Sicari, R., Vahanian, A., and Roelandt, J. R. (2010). European Association of Echocardiography recommendations for the assessment of valvular regurgitation. Part 1: aortic and pulmonary regurgitation (native valve disease). *European Journal of Echocardiography*, 11(3):223–244.

Metz, T. D., Hayes, S. A., Garcia, C. Y., and Yetman, A. T. (2013). Impact of pregnancy on the cardiac health of women with prior surgeries for pulmonary valve anomalies. *American Journal of Obstetrics and Gynecology*, 209(4):370.e1 – 370.e6.

Mokhles, M. M., Rajeswaran, J., Bekkers, J. A., Borsboom, G. J., Roos-Hesselink, J. W., Steyerberg, E. W., Bogers, A. J., Takkenberg, J. J., and Blackstone, E. H. (2014). Capturing echocardiographic allograft valve function over time after allograft aortic valve or root replacement. *The Journal of Thoracic and Cardiovascular Surgery*, 148(5):1921 – 1928.e3.

Mokhles, M. M., van de Woestijne, P. C., de Jong, P. L., Witsenburg, M., Roos-Hesselink, J. W., Takkenberg, J. J., and Bogers, A. J. (2011). Clinical outcome and health-related quality of life after right-ventricular-outflow-tract reconstruction with an allograft conduit. *European Journal of Cardio-Thoracic Surgery*, 40(3):571–578.

North, R. A., Sadler, L., Stewart, A. W., McCowan, L. M., Kerr, A. R., and White, H. D. (1999). Long-term survival and valve-related complications in young women with cardiac valve replacements. *Circulation*, 99(20):2669–2676.

O'Leary, J. M., Siddiqi, O. K., de Ferranti, S., Landzberg, M. J., and Opotowsky, A. R. (2016). The changing demographics of congenital heart disease hospitalizations in the united states, 1998 through 2010. *JAMA*, 309(10):984–986.

Oosterhof, T., Meijboom, F. J., Vliegen, H. W., Hazekamp, M. G., Zwinderman, A. H., Bouma, B. J., van Dijk, A. P., and Mulder, B. J. (2006). Long-term follow-up of homograft function after pulmonary valve replacement in patients with tetralogy of Fallot. *European Heart Journal*, 27(12):1478–1484.

Papageorgiou, G., Mauff, K., Tomer, A., and Rizopoulos, D. (2019). An overview of joint modeling of time-to-event and longitudinal outcomes. *Annual Review of Statistics and Its Application*, 6(1):223–240.

Regitz-Zagrosek, V., Roos-Hesselink, J. W., Bauersachs, J., Blomström-Lundqvist, C., Cífková, R., De Bonis, M., Iung, B., Johnson, M. R., Kintscher, U., Kranke, P., Lang, I. M., Morais, J., Pieper, P. G., Presbitero, P., Price, S., Rosano, G. M. C., Seeland, U., Simoncini, T., Swan, L., Warnes, C. A., and Group, E. S. D. (2018). 2018 ESC Guidelines for the management of cardiovascular diseases during pregnancy: The Task Force for the Management of Cardiovascular Diseases during Pregnancy of the European Society of Cardiology (ESC). *European Heart Journal*, 39(34):3165–3241.

Romeo, J. L., Takkenberg, J. J., Roos-Hesselink, J. W., Hanif, M., Cornette, J. M., van Leeuwen, W. J., van Dijk, A., Bogers, A. J., and Mokhles, M. M. (2018a). Outcomes of pregnancy after right ventricular outflow tract reconstruction with an allograft conduit. *Journal of the American College of Cardiology*, 71(23):2656 – 2665.

Romeo, J. L. R., Mokhles, M. M., van de Woestijne, P., de Jong, P., van den Bosch, A., van Beynum, I. M., Takkenberg, J. J. M., and Bogers, A. J. J. C. (2018b). Long-term clinical outcome and echocardiographic function of homografts in the right ventricular outflow tract†. *European Journal of Cardio-Thoracic Surgery*, 55(3):518–526.

Sadler, L., McCowan, L., White, H., Stewart, A., Bracken, M., and North, R. (2000). Pregnancy outcomes and cardiac complications in women with mechanical, bioprosthetic and homograft valves. *BJOG: An International Journal of Obstetrics & Gynaecology*, 107(2):245–253.

Steinberg, Z. L., Dominguez-Islas, C. P., Otto, C. M., Stout, K. K., and Krieger, E. V. (2017). Maternal and fetal outcomes of anticoagulation in pregnant women with mechanical heart valves. *Journal of the American College of Cardiology*, 69(22):2681 – 2691.

Stout, K. (2005). Pregnancy in women with congenital heart disease: the importance of evaluation and counselling. *Heart*, 91(6):713–714.

Stout, K. K. and Otto, C. M. (2007). Pregnancy in women with valvular heart disease. *Heart*, 93(5):552–558.

Trigas, V., Nagdyman, N., von Steinburg, S. P., Oechslin, E., Vogt, M., Berger, F., Schneider, K.-T. M., Ewert, P., Hess, J., and Kaemmerer, H. (2014). Pregnancy-related obstetric and cardiologic problems in women after atrial switch operation for transposition of the great arteries. *Circulation Journal*, 78(2):443–449.

T.S.o.T.S. (2017). Adults - spring 2017.

van der Bom, T., Bouma, B. J., Meijboom, F. J., Zwinderman, A. H., and Mulder, B. J. (2012). The prevalence of adult congenital heart disease, results from a systematic review and evidence based calculation. *American heart journal*, 164(4):568–575.

van der Linde, D., Konings, E. E., Slager, M. A., Witsenburg, M., Helbing, W. A., Takkenberg, J. J., and Roos-Hesselink, J. W. (2011). Birth prevalence of congenital heart disease worldwide. *Journal of the American College of Cardiology*, 58(21):2241–2247.

van Hagen, I. M., Roos-Hesselink, J. W., Ruys, T. P., Merz, W. M., Goland, S., Gabriel, H., Lelonek, M., Trojnarska, O., Mahmeed, W. A. A., Balint, H. O., Ashour, Z., Baumgartner, H., Boersma, E., Johnson, M. R., and Hall, R. (2015). Pregnancy in women with a mechanical heart valve. *Circulation*, 132(2):132–142.

Wong, V., Cheng, C. H., and Chan, K. C. (1993). Fetal and neonatal outcome of exposure to anticoagulants during pregnancy. *American Journal of Medical Genetics*, 45(1):17–21.

Yap, S.-C., Drenthen, W., Pieper, P., Moons, P., Mulder, B., Vliegen, H., van Dijk, A., Meijboom, F., Jaddoe, V., Steegers, E., Boersma, E., Roos-Hesselink, J., and on behalf of the ZAHARA investigators (2010). Pregnancy outcome in women with repaired versus unrepaired isolated ventricular septal defect. *BJOG: An International Journal of Obstetrics & Gynaecology*, 117(6):683–689.

Part II

Multi-state Models & Shrinkage

Chapter 4

Feature Selection of Longitudinal Biomarkers in Multivariate Joint Models for Longitudinal and Multi-State Processes

This chapter is based on:

Papageorgiou, G, Mokhles, M.M., Muslem, R., Brugts, J., Manintveld, O.C., Constantinescu, A.A., Bekkers, J., Takkenberg, J.J.M., Rizopoulos, D. (2021) Feature Selection of Longitudinal Biomarkers in Multivariate Joint Models for Longitudinal and Multi-State Processes *Manuscript in Preparation*

Abstract

This work is motivated by a study at the Erasmus University Medical Center in the Netherlands on patients who received a Left Ventricular Assist Device (LVAD) after heart failure. These patients may experience multiple and potentially successive complications during follow-up, such as thrombosis and embolism which may increase their risk of dying. To reduce these risks, treating physicians follow these patients closely and collect repeated marker measurements for their liver and kidney function, both before and after LVAD intervention. Their primary research goals are to understand better how these biomarkers are associated with the risk of each complication and how LVAD intervention changes the dynamics between these associations, to better guide their decision making with respect to LVAD intervention.

In this context, the joint modeling framework is a popular choice in addressing such objectives. However, while there has been a lot of work towards extensions of the classic joint modeling framework, such as multivariate joint models, joint models for a longitudinal outcome and a multi-state process and Bayesian shrinkage selection for the association structure in joint models, a united approach allowing for all these features simultaneously has yet to be studied. Furthermore, investigating the performance of Bayesian shrinkage methods in terms of feature selection in highly correlated settings has not received a lot of attention.

We therefore propose a flexible multivariate joint model for longitudinal and multi-state processes that includes LVAD intervention as a time-varying covariate in both the multivariate longitudinal and multi-state sub-models. We further allow for different association structures between the longitudinal markers and the transition-specifc intensities, while allowing for baseline characteristics' effects to differ for each transition. To determine the most appropriate functional forms we use informative global-local shrinkage priors on the regression coeffcients that correspond to the covariates of the transition-specifc intensities.

The results suggest that the use of the global-local shrinkage priors may improve the estimation and identifcation of association parameters and functional forms respectively. However, the impact of the correlation among features and the number of observed transtions needs to be further investigated.

4.1 Introduction

Multiple repeatedly measured outcomes and times-to-events are often available on the patient-specific level in follow-up studies which aim to investigate the association between the processes that generate these data. It is also often the case that intermediate events such as complications and re-interventions occur during follow-up with the potential to alter the dynamics of the disease (Proust-Lima and Taylor, 2009; Sène et al., 2016). Joint modelling of longitudinal and survival data is a popular choice in studying the link between such processes. There has been a large body of research proposing joint models and extensions thereof, such as joint models for multiple longitudinal outcomes and a single time-to-event, competing risks and recurrent events, multi-state processes and the use of shrinkage priors for the selection of the appropriate association structure (Faucett and Thomas, 1996; Wulfsohn and Tsiatis, 1997; Andrinopoulou et al., 2014; Andrinopoulou and Rizopoulos, 2016; Ferrer et al., 2016). While these approaches address several common challenges of joint models separately, a unified approach allowing for all these features, simultaneously, has yet to be studied. Furthermore, the impact of intermediate events on both processes and their association has not received enough attention.

This work focuses on the selection of the appropriate functional form for the association between multiple longitudinal outcomes and a multi-state process in the presence of intermediate events. It has been recognised, by previous work on the topic, that different features of a longitudinal process may be related to the risk for an event, such as the current value, the current slope or the area under the trajectory of the longitudinal profiles (Rizopoulos and Ghosh, 2011; Mauff et al., 2017). Specifically for the case of joint models for multiple longitudinal outcomes and multi-state processes, however, the selection of the most appropriate functional forms becomes even more cumbersome. In contrast with the situation of a single time-to-event process, the interest now lies in the association of the longitudinal outcomes with multiple transitions between states. Hence, it might be the case that different features of the longitudinal outcomes drive the risk for the transition between a state "A" to a state "B" than the features that drive the risk for the transition between a state "A" to a state "C". It also may be the case that multiple features of the same longitudinal profile impact a single transition. Apparently, in this case, the dimensionality of the selection problem becomes higher since the number of association parameters that describe the candidate functional forms increases multiplicatively with the number of longitudinal outcomes and state transitions under investigation. On top of that, when multiple features of the same biomarker are considered, high correlation among them is expected which might affect the performance of standard variable selection procedures.

Such challenges are present in the motivating study for our work. It concerns 232 patients who experienced heart failure and subsequently received a Left Ventricular Assist Device (LVAD) implant during follow-up, at the Erasmus University Medical Center in the Netherlands. These patients may suffer multiple complications during follow-up, such as thrombosis, embolic events and dialysis, which are assumed to increase their risk of dying. To reduce this risk, treating physicians follow these patients closely and collect biomarker measurements for their liver (Total Bilirubin) and kidney (Creatinine) function and may decide to intervene with an LVAD implant accordingly. LVAD is consequently expected to change their risk of experiencing a complication or dying as well as the trajectories of their liver and kidney biomarkers. In particular, LVAD takes over the heart function of the patient, and therefore, it is expected to increase the survival chance of the patient while waiting for a heart transplant. On the other hand, however, it might increase the risk of dialysis and/or thrombosis due to acute kidney injury and/or other complications. (Muslem et al., 2018; Antonides et al., 2020; Yalcin et al., 2020). In this study 49 (21.1%) subjects experienced a complication of which 23 (10.0%) subsequently died while 80 (34.5%) subjects died without experiencing such a complication. Figures 4.1 and 4.2 depict the log-transformed observed longitudinal profiles of Total Bilirubin (TB) and Creatinine (Cr) for 8 subjects before and after LVAD implantation.

The rest of this paper is structured as follows. Section 2 describes the formulation of the joint model in the presence of intermediate events. Section 3 presents the individualized dynamic predictions under different scenarios with respect to the occurrence of intermediate events and measures of predictive accuracy. In Section 4, we present the results of the analyses of the two motivating studies, while in Section 5, we show the results of a simulation study. Finally, in Section 6, we close with a discussion.

4.2 Multivariate Joint Models for Longitudinal and Multi-state Processes with the Occurrence of Intermediate Events

4.2.1 Formulation of the Joint Model

For each individual i in the study multiple longitudinal processes and a multi-state process are observed. Let $S = \{1, \ldots, M\}$ denote the finite state space and let $\{N_i(t), t \geq 0\}$ be a multi-state process where $N_i(t)$ represents the state occupied by subject i at time t. The multi-state process is assumed to be a non-homogeneous Markov process so that the future evolution of the process depends on the time since entry in the study T_{i0} and so that the future of the process depends only on the current state

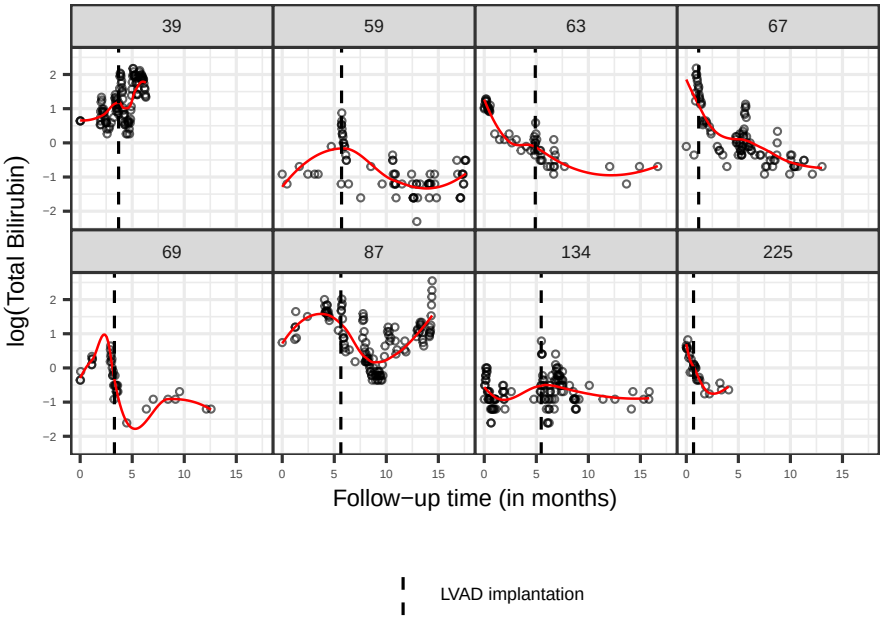

Figure 4.1: Profiles of log-transformed Total Bilirubin (mg/dl) for 8 subjects before and after LVAD implantation.

and not on the past states. Additionally, process N_i is assumed to be continuous and observed between the time of entry in the study T_{i0} and the right censoring time C_i, that is $N_i = \{N_i(t), T_{i0} \leq t \leq C_i\}$. Let $\mathcal{D}_n = \{\boldsymbol{T}_i, \boldsymbol{\delta}_i, \boldsymbol{y}_{ki}, \rho_i; i = 1, \ldots, n\}$ denote the sample from the target population with $\boldsymbol{T}_i = (T_{i1}, \ldots, T_{im_i})^\top$ denoting the vector of observed transition times with $T_{il} < T_{i(l+1)}, \forall l \in \{0, \ldots, m_i - 1\}$, $\boldsymbol{\delta}_i = (\delta_{i1}, \ldots, \delta_{im_i})^\top$ the corresponding vector of transition indicators, with $\delta_{i(l+1)}$ equal to 1 if a direct transition is observed at time $T_{i(l+1)}$ and 0 otherwise, $\forall l \in \{0, \ldots, m_i - 1\}$. Note that when $N_i(T_{im_i})$ (the last observed state for subject i) is absorbing, meaning that it is impossible to observe further transitions, then m_i direct transitions are observed whereas when $N_i(T_{im_i})$ is not absorbing, $m_i - 1$ transitions are observed and T_{im_i} equals to the right censoring time C_i. Furthermore, for each subject i we observe the vector of the K longitudinal outcomes $\boldsymbol{y}_{ki} = (y_{1i}, \ldots, y_{1n_i}, \ldots, y_{ki}, \ldots, y_{kn_i})^\top$ with y_{kij} being the observed value of the k^{th} longitudinal outcome at time $t_{ij}, j = 1, \ldots, n_i$, and ρ_i, the time to the occurrence of the intermediate event.

To incorporate the time to the occurrence of the intermediate event as a time-

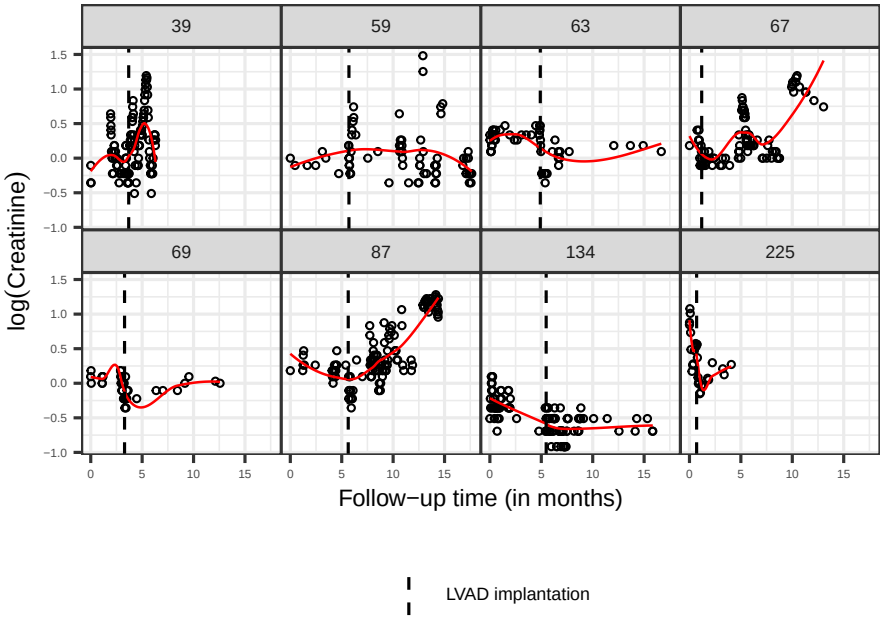

Figure 4.2: Profiles of log-transformed Creatinine (mg/dl) for 8 subjects before and after LVAD implantation.

varying covariate, we further introduce $R_i(t) = I(t > \rho_i)$ which equals 1 at any time point t after the occurrence of the intermediate event and zero otherwise, and the time relative to the occurrence of the intermediate event $t_{i+} = \max(0, t_{ij} - \rho_i; j = 1, \ldots, n_i)$. The true level, $\eta_{ki}(t)$, of the k^{th} longitudinal outcome at time t is then expressed as:

$$g_k\left[E\{y_{ki}(t) \mid \boldsymbol{b}_{ki}\}\right] = \eta_{ki}(t) = \begin{cases} \boldsymbol{x}_{ki}^{\top}(t)\,\boldsymbol{\beta_k} + \\ \boldsymbol{z}_{ki}^{\top}(t)\,\boldsymbol{b}_{ki}, 0 < t < \rho_i \\ \\ \boldsymbol{x}_{ki}^{\top}(t)\,\boldsymbol{\beta}_k + \boldsymbol{z}_{ki}^{\top}(t)\,\boldsymbol{b}_{ki} + \tilde{\boldsymbol{x}}_{ki}^{\top}(R_i, t_{i+})\,\tilde{\boldsymbol{\beta}}_k + \\ \tilde{\boldsymbol{z}}_{ki}^{\top}(R_i, t_{i+})\,\tilde{\boldsymbol{b}}_{ki}, t \geq \rho_i, \end{cases} \tag{4.1}$$

where $g_k(\cdot)$ is a general link function appropriate for the k^{th} outcome, $\boldsymbol{x}_{ki}$ and $\boldsymbol{z}_{ki}$ are design vectors for the fixed-effects' regression coefficients $\boldsymbol{\beta}_k$ and the random-effects

$\boldsymbol{b}_{ki}$ respectively. Furthermore, we explicitly introduce design vectors $\tilde{\boldsymbol{x}}_{ki}\left(R_i, t_{i+}\right)$ and $\tilde{\boldsymbol{z}}_{ki}\left(R_i, t_{i+}\right)$ which may include any function of the time-varying covariates R_i and t_{i+} for the corresponding fixed-effects' regression coefficients $\tilde{\boldsymbol{\beta}}_k$ and the random-effects $\tilde{\boldsymbol{b}}_{ki}$ respectively. The association between the K longitudinal outcomes is built via the random-effects $\boldsymbol{b}_{ki}$ and $\tilde{\boldsymbol{b}}_{ki}$ which are assumed to be normally distributed with mean zero and a $q \times q$ general variance-covariance matrix D.

The generalized multivariate mixed-effects model is then linked to a Markov multi-state model under the assumption of shared random-effects. More specifically, let $\mathcal{H}_{ki}\left(t, \rho_i\right) = \left[\eta_{ki}\left(s, \rho_i\right), 0 \leq s \leq t\right]$ denote the history of the k^{th} longitudinal outcome up to time t. Then, for the transition between state $x \in S$ to state $v \in S$, the transition intensity at time t takes the following form:

$$h_{xv}^{i}\left(t\right) = h_{xv,0}\left(t\right) \exp\left[\boldsymbol{W}_{xv,i}{}^{\top}\boldsymbol{\gamma}_{xv} + \sum_{k=1}^{K}\sum_{u=1}^{U} f_{xv,u}\left\{\eta_{ki}\left(t\right), \alpha_{xv,ku}\right\}\right], \tag{4.2}$$

where $h_{xv,0}$ is the transition-specific baseline hazard, $\boldsymbol{W}_{xv,i}^{\top}$ is a state-specific design vector for the state-specific regression coefficients $\boldsymbol{\gamma}_{xv}$. The occurrence of the intermediate event is included as a time-varying covariate in the model allowing the transition to differ before and after the occurrence of the intermediate event. The transition-specific effect of the intermediate event is captured by the corresponding regression coefficient ζ_{xv}. The association structure between the K longitudinal outcomes and the multi-state process is defined via U transition-specific multivariate functions $f_{xv,u}\left\{\eta_{ki}\left(t\right), \alpha_{xv,ku}\right\}$. Common choices are:

$$\text{Current Value Association: } f_{xv,u}\left\{\eta_{ki}\left(t\right), \alpha_{xv,ku}\right\} = \eta_{ki}\left(t\right)\alpha_{xv,ku} \tag{4.3}$$

$$\text{Current Slope Association: } f_{xv,u}\left\{\eta_{ki}\left(t\right), \alpha_{xv,ku}\right\} = \frac{\partial \eta_{ki}\left(t\right)}{\partial t}\alpha_{xv,ku} \tag{4.4}$$

$$\text{Area Under the Curve Association: } f_{xv,u}\left\{\eta_{ki}\left(t\right), \alpha_{xv,ku}\right\} = \int_0^t \eta_{ki}\left(s\right)\alpha_{xv,ku} ds, \tag{4.5}$$

or any other function of the history of the k^{th} longitudinal outcome up to time t, $\eta_{ki}\left(t\right)$ or the respective random-effects in the context of the study. The corresponding regression coefficients $\boldsymbol{\alpha}_{xv,k}$ quantify the impact of the k^{th} longitudinal outcome on the transition between states x and v.

4.3 Bayesian Shrinkage

4.3.1 Estimation

We employ a Bayesian approach with Markov chain Monte Carlo (MCMC) sampling methods to estimate the parameters of the multivariate joint model. The likelihood of the model is derived under the conditional independence assumption. Under this assumption, we assume that the longitudinal and multi-state processes are independent conditional on the unobserved random effects. Furthermore, the repeated measurements within each subject are also assumed to be independent given the random effects. The posterior distribution is given by:

$$p\left(\boldsymbol{\theta} \mid y_{ki}, N_i, \delta_i\right) \propto \prod_{k=1}^{K} \prod_{j=1}^{n_{ki}} p\left(y_{kij} \mid \boldsymbol{b}_{ki}, \boldsymbol{\theta}_{y_k}\right) p\left(N_i, \delta_i \mid \mathcal{H}_{ki}, \boldsymbol{\theta}_N\right) p\left(\boldsymbol{b}_{ki} \mid \boldsymbol{\theta}_{y_k}\right) p\left(\boldsymbol{\theta}_{y_k}\right) p\left(\boldsymbol{\theta}_N\right) \tag{4.6}$$

, where $\boldsymbol{\theta} = (\boldsymbol{\theta}_{y_k}, \boldsymbol{\theta}_N)$ is the vector of all the parameters contained in 4.1 4.2 respectively. The likelihood contribution of the longitudinal outcomes is given by:

$$p\left(\boldsymbol{y}_{ki} \mid \boldsymbol{b}_{ki}, \boldsymbol{\theta}_{y_k}\right) = \exp\left\{ \sum_{j=1}^{n_{ki}} \frac{y_{kji}\psi_{kji}\left(\boldsymbol{b}_{ki}\right) - c\left[\psi_{kji}\left(\boldsymbol{b}_{ki}\right)\right]}{\alpha_\phi} - d\left(y_{kji}, \phi\right) \right\}, \tag{4.7}$$

where $\psi_{kji}\left(\boldsymbol{b}_{ki}\right)$ and phi denote the natural and dispersion parameters of the exponential family of distributions respectively, and $c\left(\cdot\right)$, $d\left(\cdot\right)$ and $\alpha\left(\cdot\right)$ are known functions that specify the member of the exponential family.

For the likelihood contribution of the multi-state process let $P^i_{xv}\left(s,t\right)$ denote the transition probability from states x to v for subject i between times s and t: $P^i_{xv}\left(s,t\right) = P\left(N_i\left(t\right) = v \mid N_i\left(s\right) = x\right)$. Due to the continuity assumption of the Markov chain it holds that individual i remains in state $N_i\left(T_{il}\right)$ between times T_{il} and $T_{i(l+1)}$ with probability $P^i_{N_i(T_{il}),N_i(T_{il})}\left(T_{il}, T_i\left(l+1\right) \mid \boldsymbol{b}_i\right)$ and moves to state $N_i\left(T_{i(l+1)}\right)$ with transition intensity $h^i_{N_i(T_{il}),N_i(T_i(l+1))}\left(N_i\left(T_i\left(l+1\right) \mid \boldsymbol{b}_i\right)\right)$. The likelihood contribution, conditioning on the starting state is given by:

$$
\begin{aligned}
p\left(N_i \mid \boldsymbol{b}_i; \boldsymbol{\theta}_N\right) = & \prod_{l=0}^{m_i-1} \left\{ P^i_{N_i(T_{il}),N_i(T_{il})} \left(T_{il}, T_{i(l+1)} \mid \boldsymbol{b}_i\right) \right. \\
& \left. \lambda^i_{N_i(T_{il}),N_i(T_{i(l+1)})} \left(T_{i(l+1)} \mid \boldsymbol{b}_{ki}\right)^{\delta_{i(l+1)}} \right\} \\
= & \prod_{l=0}^{m_i-1} \left\{ \exp\left(\int_{T_{il}}^{T_{i(l+1)}} \lambda^i_{N_i(T_{il}),N_i(T_{il})} \left(u \mid \boldsymbol{b}_{ki}\right) du \right) \right. \\
& \left. \lambda^i_{N_i(T_{il}),N_i(T_{i(l+1)})} \left(T_{i(l+1)} \mid \boldsymbol{b}_i\right)^{\delta_{i(l+1)}} \right\}
\end{aligned}
$$

, with $h^i_{xx}\left(t\right) = -\sum_{x,x\neq v} h^i_{xv}\left(t\right)$.

4.3.2 Bayesian Shrinkage Global-local Priors

As we discussed in Section 4.2 there are different association structures that can be used to link the longitudinal and multi-state processes. While the choice of the form of these association structures can and if possible should be driven by clinical and biological background for the application at hand, this is not always possible. Furthermore, especially for the case of linking multiple longitudinal outcomes and multi-state processes the dimensionality of the vector of association parameters $\boldsymbol{\alpha}_{ku}$ can increase substantially leading to situations where the number of parameters to estimate is larger or close to the number of observed transitions. To overcome this and to automate the selection of appropriate association sructures we suggest the use of global-local shrinkage priors. In particular, we study two different shrinkage priors and how they behave in our setting.

The general specification of the prior for the association parameters is given by:

$$
\alpha_{xv,ku} \mid \tau^2_{xv,ku}, \lambda^2 \sim \mathcal{N}\left(0, \tau^2_{xv,ku}, \lambda^2\right), \tag{4.8}
$$

where $\tau^2_{xv,ku}, \lambda^2 \sim \mathcal{F}$ is the local shrinkage parameter, that applies to the corresponding single alpha parameter for transition $x \rightarrow v$ for longitudinal outcome k and association structure u and $\lambda^2 \sim \mathcal{G}$ is the global shrinkage parameter, that applies to all alpha parameters.

By changing the distributions $\mathcal{F}$ and $\mathcal{G}$ we achieve different shrinkage priors. More specifically we consider the two following hierachies:

$$\begin{aligned} \alpha_{xv,ku} \mid \tau^2_{xv,ku}, \lambda^2 &\sim \mathcal{N}\left(0, \tau^2_{xv,ku}, \lambda^2\right) \\ \tau^2_{xv,ku} \mid \nu_{xv,ku} &\sim \mathcal{IG}\left(\frac{1}{2}, \frac{1}{\nu_{xv,ku}}\right) \\ \lambda^2 \mid \xi &\sim \mathcal{IG}\left(\frac{1}{2}, \frac{1}{\xi}\right) \\ \nu_{xv,1u}, \ldots, \nu_{xv,KU}, \xi &\sim \mathcal{IG}\left(\frac{1}{2}, 1\right) \end{aligned} \tag{4.9}$$

and

$$\begin{aligned} \alpha_{xv,ku} \mid \tau^2_{xv,ku}, \lambda^2 &\sim \mathcal{N}\left(0, \tau^2_{xv,ku}, \lambda^2\right) \\ \tau^2_{xv,ku} &\sim \mathcal{IG}\left(\frac{1}{2}, 1\right) \\ \lambda^2 &\sim \mathcal{IG}\left(\frac{1}{2}, 1\right) \end{aligned} \tag{4.10}$$

, where $\nu_{xv,ku}$ and ξ are the local and global hypervariances of the inverse gamma distributions which also follow an inverse gamma distribution. The specification in 4.9 leads to the half-Cauchy $C^+(0,1)$ distribution leading to the horseshoe shrinkage prior. Alternatively, by setting variance of the global and local shrinkage parameters to be equal to 1 as shown in 4.9, we specify the ridge shrinkage prior. The two are expected to differ as the horseshoe shrinkage prior is characterized by a strong spike that is expected to cause severe shrinkage near zero in contrast to the ridge prior which has a weaker spike. Furthermore, the horseshoe shrinkage prior has more narrow tails compared to the ridge prior and therefore the ridge shrinkage will be stronger further away from zero.

4.3.3 Software

The R package **JMbayes** was extended to include multi-stage processes and shrinkage priors. These changes are already integrated in the package both on CRAN and in the development version on GitHub. The inclusion of intermediate events was already incorporated from previous work (Papageorgiou et al., 2019). The help page of function `mvJointModelBayes()` provides examples on how to fit joint models for longitudinal and multi-state processes with and without shrikage.

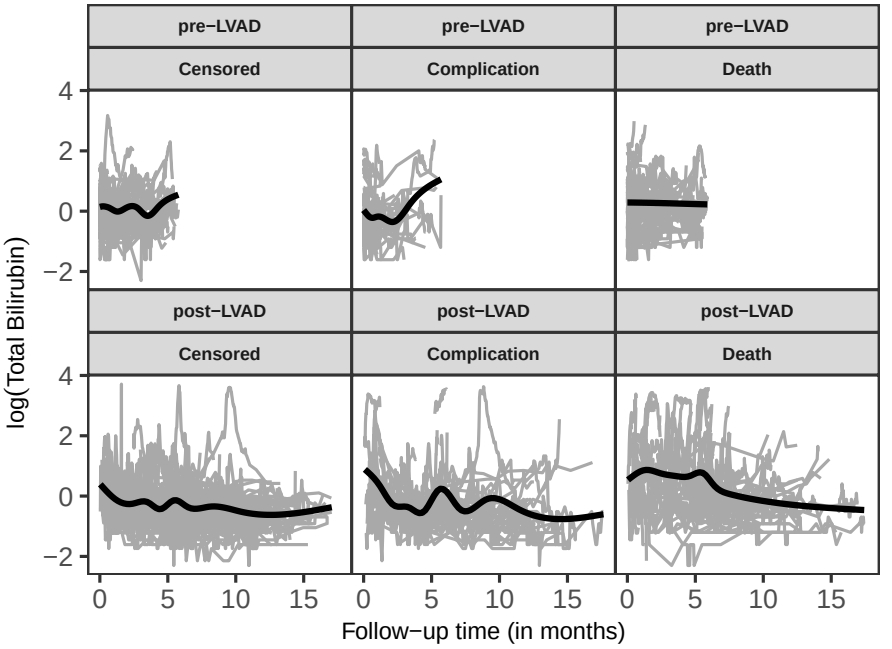

Figure 4.3: Profiles of log-transformed Total Bilirubin (mg/dl) for all subjects before (upper panel) and after (lower panel) LVAD implantation per clinical outcome.

4.4 Application to the LVAD Data

In this section we discuss the analysis of the LVAD dataset which was introduced in Section 1. Our goals are to: 1) assess the impact of LVAD on the evolution of TB and Cr, 2) to investigate the association between TB and Cr with each transition intensity between the clinical states of interest, and 3) to quantify these associations while automating the identification of the most important features that describe the association between each biomarker and each transition.

Table 4.1 summarizes some additional information about the LVAD dataset. More specifically, four baseline characteristics were considered including age at baseline, sex, BMI and a binary indicator on whether heart failure etiology is ischemic or not. On Figures 4.3 and 4.4 the profiles of all subjects are shown before and after LVAD for both $\log(\text{TB})$ and $\log(\text{Cr})$. Looking at the sample of subject-specific profiles on Figures 4.1 and 4.2 we see that the observed repeated measurements are characterized by high variation among measurements from the same subject while the loess spline

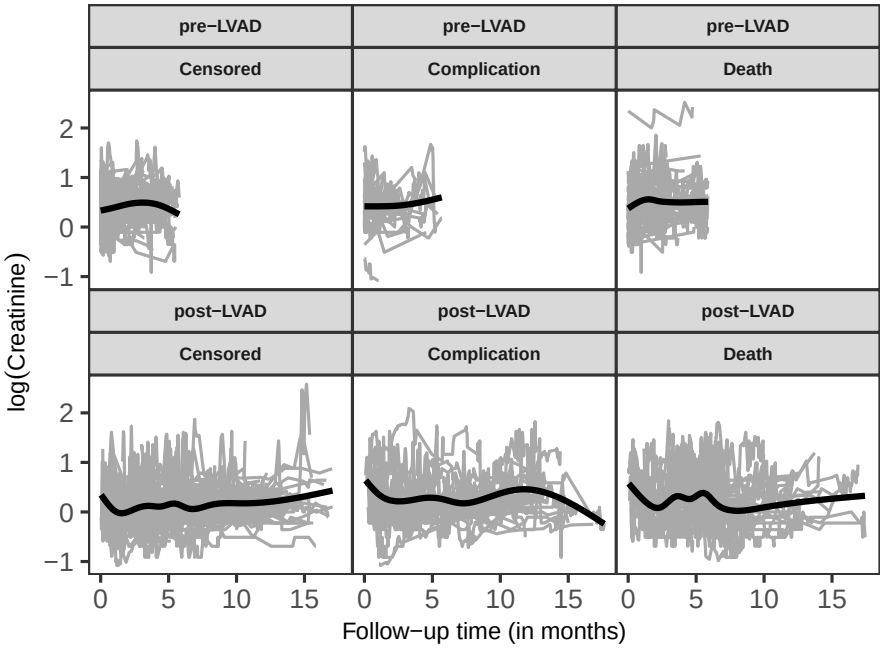

Figure 4.4: Profiles of log-transformed Creatinine (mg/dl) for all subjects before (upper panel) and after (lower panel) LVAD implantation per clinical outcome.

curves indicate considerable variation accross subjects. For the majority of the profiles, an immediate but temporal reduction on the the levels of both biomarkers after LVAD is hinted, which then is followed by increasing, stable or decreasing profiles. The shape of the trajectories is non linear for the majority of the trajectories. To accommodate these features in our analysis we used the following multivariate linear mixed-effects models:

$$\log TB_i(t) = \begin{cases} (\beta_{0_{TB}} + b_{i0_{TB}}) + \left(\sum_{k=0}^{4} \left(\beta_{(k+1)_{TB}} + b_{i(k+1)_{TB}}\right) B_k(t, \boldsymbol{k})\right) \\ +\beta_{11_{TB}}\text{Age}_i + \beta_{12_{TB}}\text{Sex}_i + \beta_{13_{TB}}\text{BMI}_i + \beta_{14_{TB}}\text{Etiology}_i + \epsilon_{i_{TB}}(t), t_+ = 0 \\ \\ (\beta_{0_{TB}} + b_{i0_{TB}}) + \left(\sum_{k=0}^{3} \left(\beta_{(k+1)_{TB}} + b_{i(k+1)_{TB}}\right) B_k(t, \boldsymbol{k})\right) + \\ \left(\sum_{k=0}^{3} \left(\tilde{\beta}_{(k+5)_{TB}} + \tilde{b}_{i(k+5)_{TB}}\right) B_k(t_{i+}, \boldsymbol{k})\right) \\ +\beta_{11_{TB}}\text{Age}_i + \beta_{12_{TB}}\text{Sex}_i + \beta_{13_{TB}}\text{BMI}_i + \beta_{14_{TB}}\text{Etiology}_i + \epsilon_{i_{TB}}(t), t_+ > 0, \end{cases} \tag{4.11}$$

$$\log Cr_i(t) = \begin{cases} (\beta_{0_{Cr}} + b_{i0_{Cr}}) + \left(\sum_{k=0}^{4} \left(\beta_{(k+1)_{Cr}} + b_{i(k+1)_{Cr}}\right) B_k(t, \boldsymbol{k})\right) \\ +\beta_{11_{Cr}}\text{Age}_i + \beta_{12_{Cr}}\text{Sex}_i + \beta_{13_{Cr}}\text{BMI}_i + \beta_{14_{Cr}}\text{Etiology}_i + \epsilon_{i_{Cr}}(t), t_{i+} = 0 \\ \\ (\beta_{0_{Cr}} + b_{i0_{Cr}}) + \left(\sum_{k=0}^{3} \left(\beta_{(k+1)_{Cr}} + b_{i(k+1)_{Cr}}\right) B_k(t, \boldsymbol{k})\right) + \\ \left(\sum_{k=0}^{3} \left(\tilde{\beta}_{(k+5)_{Cr}} + \tilde{b}_{i(k+5)_{Cr}}\right) B_k(t_{i+}, \boldsymbol{k})\right) \\ +\beta_{11_{Cr}}\text{Age}_i + \beta_{12_{Cr}}\text{Sex}_i + \beta_{13_{Cr}}\text{BMI}_i + \beta_{14_{Cr}}\text{Etiology}_i + \epsilon_{i_{Cr}}(t), t_{i+} > 0, \end{cases} \tag{4.12}$$

for the log-transformed measurements of TB and Cr respectively, where B_k denotes natural cubic splines basis matrix with 4 knots at equally spaced quantiles of time. The regression coefficients β_k corresponding to B_k capture the non-linear shapes of the trajectories. We then allow for the shape to change after LVAD by including natural cubic splines of the time relative to LVAD. The regression coefficients $\tilde{\beta}_k$ capture the changes to the biomarker trajectory after LVAD while they are zeroed out before LVAD. An unstructured variance-covariance matrix $\boldsymbol{D}$ was used for the random-effects $\boldsymbol{b} = \left(\boldsymbol{b_{0_T B}}, \boldsymbol{b_{k_T B}}, \tilde{\boldsymbol{b}}_{\boldsymbol{k_T B}}, \boldsymbol{b_{0_C r}}, \boldsymbol{b_{k_C r}}, \tilde{\boldsymbol{b}}_{\boldsymbol{k_C r}}\right)^{\top} \sim \mathcal{N}(0, \boldsymbol{D})$ which were assumed to be independent of the error terms $\boldsymbol{\epsilon}_{i_{TB}} \sim \mathcal{N}\left(0, \sigma^2_{TB}\boldsymbol{I}_n\right)$, $\boldsymbol{\epsilon}_{i_{Cr}} \sim \mathcal{N}\left(0, \sigma^2_{Cr}\boldsymbol{I}_n\right)$.

Table 4.1: Description of the LVAD Dataset. Continuous variables are presented with median (IQR). Categorical variables are presented with n (percentage).

	All
Number of Measurements	14828
Number of Measurements per Subject	53 (36; 77.25)
Follow-up time	10.27 (5.4; 12.66)
LVAD Time	2.18 (0.77; 3.88)
Age	56.98 (45; 63)
Sex = Male	178 (76.72%)
BMI	26.16 (22.78; 30.45)
Etiology = Ischemic	82 (35.34%)

The multi-state data are summarized and depicted on Figure 4.5 which also illustrates the transitions between the three clinical states of interest. To address the objectives of our analyses we allowed for the three types of association structure (value, slope, area) as described in Section 2 for each biomarker and for each transition intensity. This resulted in nine association parameters in total. We further allowed for transition specific effects for age, sex, BMI, and ischemic etiology of heart failure. We

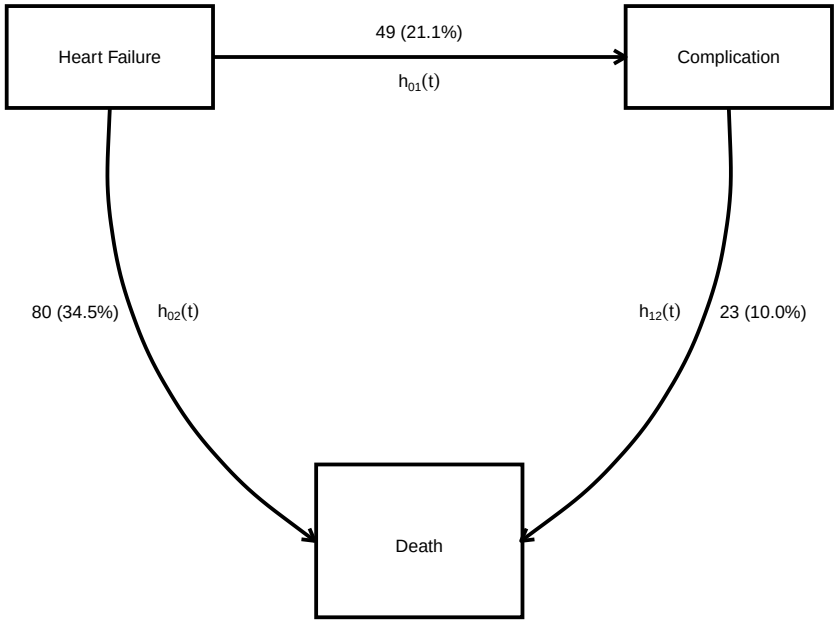

Figure 4.5: Multi-state representation of the transitions from heart failure to complication and/or death. Arrows indicate the direction of each transition. $h_{xv}(t)$ represents the transition intensity from state x to state v.

assumed transition specific baseline hazards which were approximated by P-splines. The formulation of the joint model is, therefore, given by:

$$h_{xv}^{i}\left(t \mid \boldsymbol{b}_i, \tilde{\boldsymbol{b}}_i\right) = h_{xv,0}(t) \exp\left[\boldsymbol{W}_{xv,i}{}^{\top}\boldsymbol{\gamma}_{xv} + \sum_{k=1}^{2}\sum_{u=1}^{3} f_{xv,u}\left\{\eta_{ki}(t), \alpha_{xv,ku}\right\}\right], \tag{4.13}$$

where:

- $f_{xv,1}\left\{\eta_{ki}(t), \alpha_{xv,k1}\right\} = \eta_{ki}(t)\,\alpha_{xv,k1}$,
- $f_{xv,2}\left\{\eta_{ki}(t), \alpha_{xv,k2}\right\} = \frac{\partial \eta_{ki}(t)}{\partial t}\alpha_{xv,k2}$,
- $f_{xv,3}\left\{\eta_{ki}(t), \alpha_{xv,k3}\right\} = \int_0^t \eta_{ki}(s)\,\alpha_{xv,k3} ds$,

and η_{ki} is the linear predictor of the trajectory of the i^{th} subject for the k^{th} longitudinal outcome with $k = 1$ corresponding to $\log(\text{TB}_i)$ and $k = 2$ corresponding to $\log(\text{Cr}_i)$. The vector of baseline regression coeffcients $\boldsymbol{\gamma}_{xv} = \{\gamma_{xv_1}, \gamma_{xv_2}, \gamma_{xv_3}, \gamma_{xv_4}\}^\top$ includes a separate parameter for BMI, age, sex and etiology respectively for each transition $x \to v$.

Tables 4.2 and 4.3 summarize the posterior parameter estimates and the 95% credibility intervals of the regression coefficients and the residual standard deviation of the longitudinal submodels for $\log(\text{TB}_i)$ and $\log(\text{Cr}_i)$. The baseline characteristics seems to have a quite small effect on the evolution of both $\log(\text{TB}_i)$ and $\log(\text{Cr}_i)$. The larger effect is observed for male sex $0.328(0.142; 0.513)$ and $0.175(0.093; 0.255)$ for $\log(\text{TB}_i)$ and $\log(\text{Cr}_i)$ respectively. The effect of the baseline factors is going towards the same direction for both outcomes with the exception age which however seems to have a relatively small effect in magnitude. In Figures 4.6 4.7 the fitted longitudinal models are visualized. For these plots the following values were used for the baseline variables: age = 56.98, BMI = 26.16, sex = male and etiology = Not ischemic. The LVAD implantation time was set at 91.25 days. As shown in the plot it seems that LVAD implantation had an immediate effect on the evolution of $\log(\text{TB}_i)$ and $\log(\text{Cr}_i)$ as a decrease in both biomarkers followed LVAD implantation for both biomarkers. For $\log(\text{TB}_i)$ it seems that the decrease stabilizes and a constant difference between the profiles (LVAD vs no LVAD) remains over follow-up which is not the case for $\log(\text{Cr}_i)$ as the difference seems to fade out near the end of follow-up.

In Table 4.4 the posterior estimates of hazard ratios and 95% credibility intervals for the parameters of the joint models without shrinkage (NoS), using the horseshoe prior (HS) and the ridge prior (R) are shown. The hazard ratios of the baseline effects remain similar despite using shrinkage or not as expecte since shrinkage is only applied to the association parameters. Moreover, from these results it does not seem that any of the baseline characteristics has an effect that changes between different transitions. In Figures 4.8 and 4.9 we visualize the posterior mcmc samples for all the association parameters used in the model for $\log(\text{TB}_i)$ and $\log(\text{Cr}_i)$. This graph helps us make some interesting observations. First, we see that the horseshoe and ridge shrinkage priors have a very similar behavior in the specific application. However, we can still observe the expected behavior of the priors for specific parameters. For example, in figure 4.8 we see that for the slope parametrization and for transitions HF $\to$ Complication and HF $\to$ Death the HS prior results with more density concetrated near zero when compared with the ridge prior. This is expected, as the horseshoe prior has a stronger spike near zero. Second, we see that when the effect size of a parameter is relatively strong, as for example the value association parameter for transition HF $\to$ Death in 4.8, then there is no shrinkage and the result of both priors is close to the

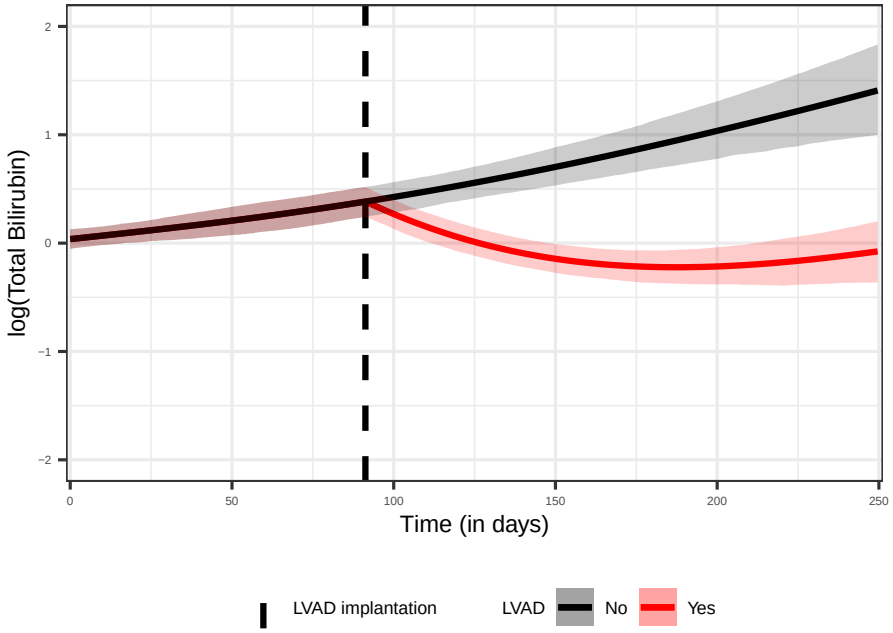

Figure 4.6: Effect plot of log-transformed Total Bilirubin (mg/dl) over time. Baseline characteristics are set to median for continuous factors and to most frequent value for categorical factors.

result when no shrinkage is applied. Finally, the results suggest that based on the data, the current value association strutcure seems to be stronger for both outcomes and for transtitions HF $\rightarrow$ Death and Complication $\rightarrow$ Death. The other association structures are shrunk towards zero.

4.5 Simulation Study

4.5.1 Design

To further investigate the behavior of the shrinkage priors in terms of feature selection and evaluate the proposed models we performed a simulation study. We assumed 250 subjects and then randomly selected follow-up visits. Each visit time t_{ij} was simulated from a uniform distribution between 0 and 12. We assumed a total number of 20 measurements per subject. The final number of measurements, however, varies

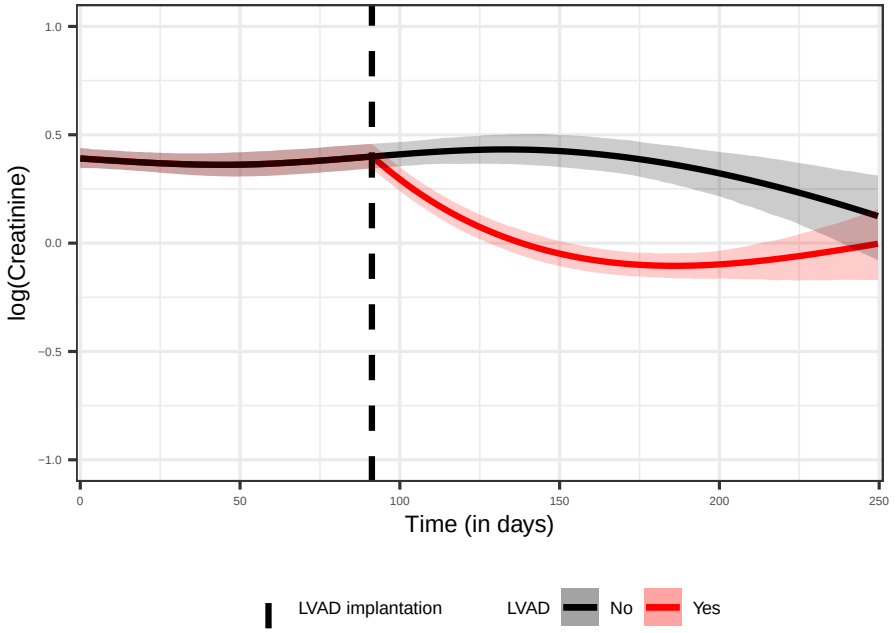

Figure 4.7: Effect plot of log-transformed Creatinine (mg/dl) over time. Baseline characteristics are set to median for continuous factors and to most frequent value for categorical factors.

depending on when a subject was censored. For simplicity we assumed one longitudinal outcome and a multi-state process with three states, similar to the one depicted in figure 4.5. For the continuous longitudinal outcome, we fitted a linear mixed-effects model given by:

$$y_{ij}(t) = (\beta_0 + b_{i0}) + (\beta_1 + b_{i1}) \times t_{ij} + \beta_2 X_i + \epsilon_{ij}(t), \tag{4.14}$$

where $\epsilon_i \sim \mathcal{N}\left(0, \sigma_y^2\right)$, with $\sigma_y = 1.242$, $\boldsymbol{b}_i \sim \mathcal{N}(0, D)$ with $D = (d_{11} = 1.35, d_{12} = 4, d_{22} = 2)^\top$ and X is a continuous baseline variable simulated from a normal distribution with mean 4.763 and standard deviation 2.8.

For the multi-state process we assumed the following scenarios:

$$h_{xv}^i(t \mid \boldsymbol{b}_i) = h_{xv,0}(t) \exp\left[\boldsymbol{X}_{xv,i}{}^\top \boldsymbol{\gamma}_{xv} + \eta_i(t)\, \alpha_{xv,1}\right], \tag{4.15}$$

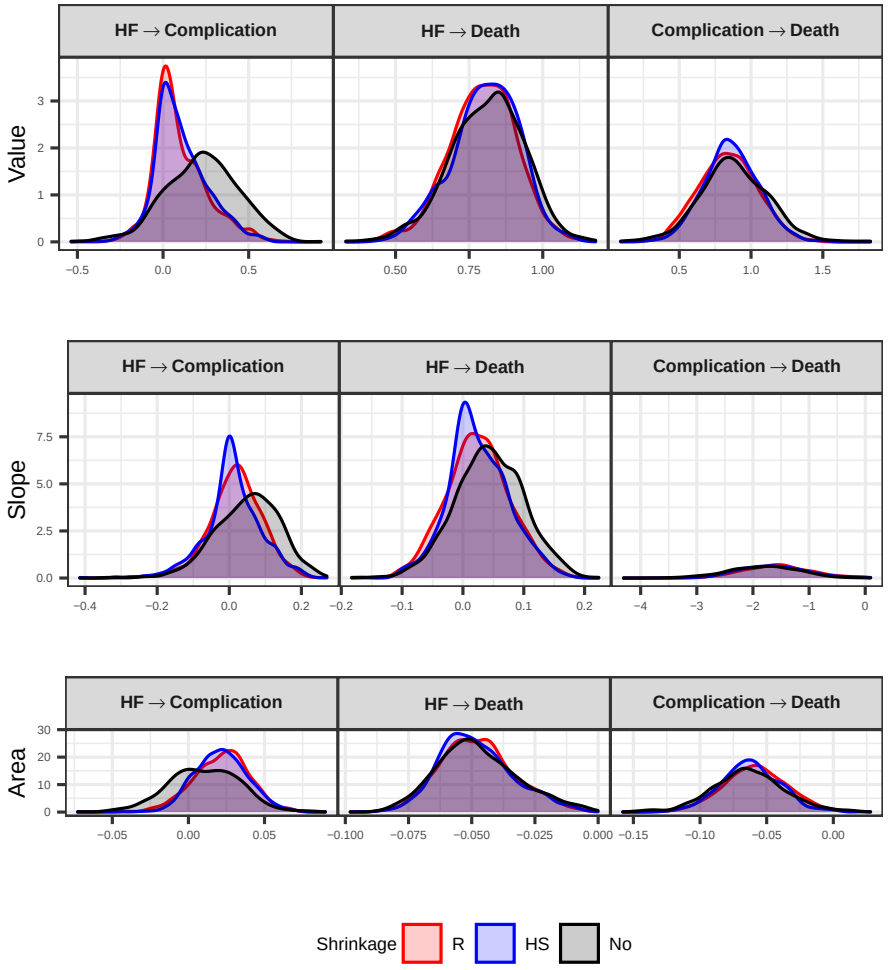

Figure 4.8: Density plots of posterior mcmc samples for the $alpha$ parameters associated with log-transformed Total Bilirubin (mg/dl) included in the joint model under different shrinkage priors.

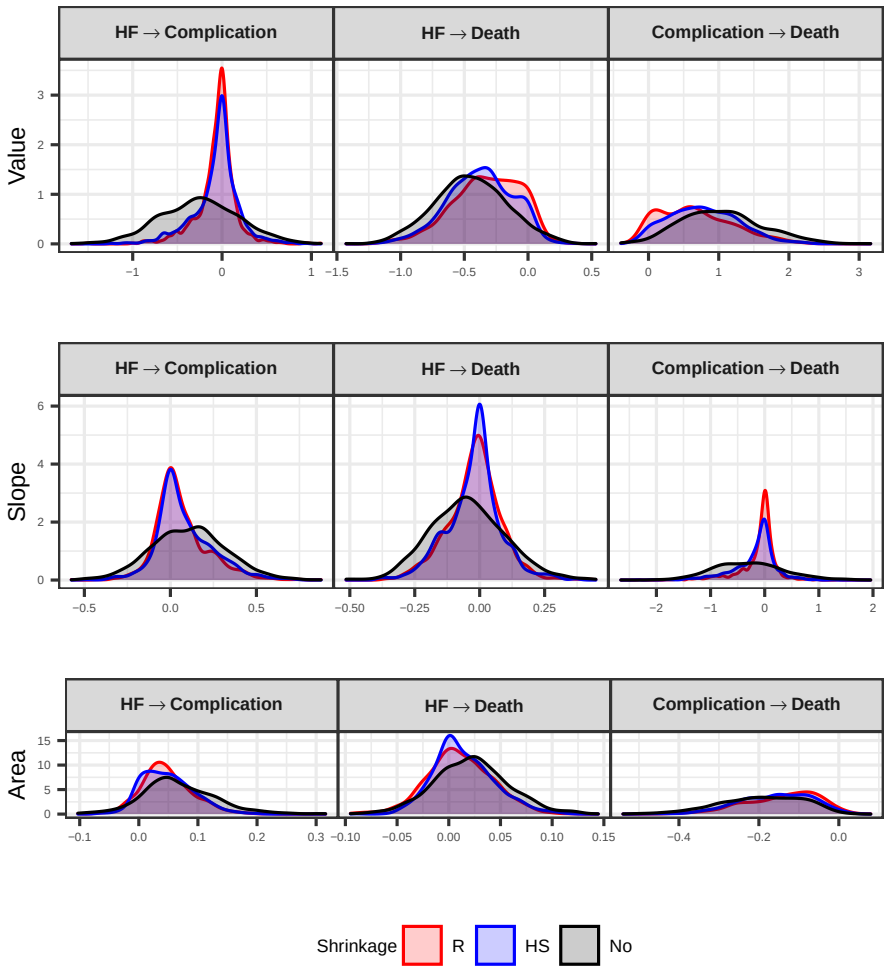

Figure 4.9: Density plots of posterior mcmc samples for the *alpha* parameters associated with log-transformed Creatinine (mg/dl) included in the joint model under different shrinkage priors.

Table 4.2: Posterior estimates and 95% credibility intervals for the parameters of the longitudinal submodel for log (TB)

Parameter	Est. (95% CI)
$\beta_{0_{TB}}$	-0.161(-2.240; 3.027)
$\beta_{1_{TB}}$	0.626(0.337; 0.937)
$\beta_{2_{TB}}$	0.620(0.367; 0.878)
$\beta_{3_{TB}}$	1.314(0.770; 1.887)
$\beta_{4_{TB}}$	1.761(0.856; 2.676)
$\beta_{5_{TB}}$	1.989(0.699; 3.351)
$\tilde{\beta}_{6_{TB}}$	0.241(-2.686; 2.077)
$\tilde{\beta}_{7_{TB}}$	-1.466(-4.641; 0.529)
$\tilde{\beta}_{8_{TB}}$	-0.877(-2.387; 0.107)
$\tilde{\beta}_{9_{TB}}$	-1.151(-7.791; 3.025)
$\tilde{\beta}_{10_{TB}}$	-2.109(-3.191; -1.101)
$\beta_{11_{TB}}$	-0.007(-0.013; -0.001)
$\beta_{12_{TB}}$	0.328(0.142; 0.513)
$\beta_{13_{TB}}$	0.004(-0.009; 0.017)
$\beta_{14_{TB}}$	-0.033(-0.206; 0.147)
σ	0.388(0.334; 0.342)

$$h^i_{xv}\left(t \mid \boldsymbol{b}_i\right) = h_{xv,0}\left(t\right) \exp\left[{\boldsymbol{X}_{xv,i}}^\top \boldsymbol{\gamma}_{xv} + \frac{\partial \eta_i\left(t\right)}{\partial t}\alpha_{xv,2}\right], \tag{4.16}$$

$$h^i_{xv}\left(t \mid \boldsymbol{b}_i\right) = h_{xv,0}\left(t\right) \exp\left[{\boldsymbol{X}_{xv,i}}^\top \boldsymbol{\gamma}_{xv} + \eta_i\left(t\right)\alpha_{xv,1} + \frac{\partial \eta_i\left(t\right)}{\partial t}\alpha_{xv,2}\right], \tag{4.17}$$

which correspond to using the current value association for all transitions, the slope association for all transitions and both association structures for all transitions. Finally, we considered an additional scenario under which the current value association was used for transitions $x = 0 \rightarrow v = 1$ and $x = 1 \rightarrow v = 2$ while the slope association was used for transition $x = 0 \rightarrow v = 2$. The baseline risk was simulated from a Weibull distribution $h_{xv,0}(t) = \xi_{xv} t^{\xi_{xv}-1}$ with $\xi_{01} = 11.325$, $\xi_{02} = 9.216$ and $\xi_{12} = 9.657$. The censoring process was assumed to follow an exponential distribution with mean equal to 9. Under these settings we simulated 100 datasets from each scenario. For the fitting part, we used the exact same model for the longitudinal outcome as the one we

Table 4.3: Posterior estimates and 95% credibility intervals for the parameters of the longitudinal submodel for log (Cr)

Parameter	Est. (95% CI)
$\beta_{0_{Cr}}$	-0.639(-2.012; 0.755)
$\beta_{1_{Cr}}$	0.042(-0.055; 0.141)
$\beta_{2_{Cr}}$	0.128(0.019; 0.256)
$\beta_{3_{Cr}}$	0.047(-0.217; 0.316)
$\beta_{4_{Cr}}$	0.303(-0.010; 0.639)
$\beta_{5_{Cr}}$	0.401(-0.047; 0.870)
$\tilde{\beta}_{6_{Cr}}$	-0.201(-1.460; 1.088)
$\tilde{\beta}_{7_{Cr}}$	-0.383(-1.736; 0.984)
$\tilde{\beta}_{8_{Cr}}$	-0.114(-0.810; 0.529)
$\tilde{\beta}_{9_{Cr}}$	0.061(-2.866; 3.012)
$\tilde{\beta}_{10_{Cr}}$	-0.285(-0.834; 0.214)
$\beta_{11_{Cr}}$	0.007(0.004; 0.009)
$\beta_{12_{Cr}}$	0.175(0.093; 0.255)
$\beta_{13_{Cr}}$	0.013(0.007; 0.019)
$\beta_{14_{Cr}}$	-0.011(-0.090; 0.066)
σ	0.224(0.221; 0.227)

simulated from while assuming both the value and slope association structures and fitted a separate model for each of the three shrinkage scenarios: no shrinkage (NoS), horseshoe (HS) and ridge (R).

4.5.2 Results

In Figures 4.10, 4.11, 4.12, 4.13 we present boxplots with the results of the simulations under the scenario of the current value association for all transitions, the current slope association for all transitions, both the current value and slope association for all transitions and the current value association for transitions 1 and 3 and the slope association for transition 2 respectively. Small shrinkage seems to apply across scenarios for nuisance effects, with the horseshoe prior consistently applying more shrinkage than the ridge prior when the coefficient value is close to zero. All three methods seem to perform equally well or poorly in identifying the actual effect estimates across all scenarios, but the horseshoe prior seems to perform best in shrinking nuisance effects

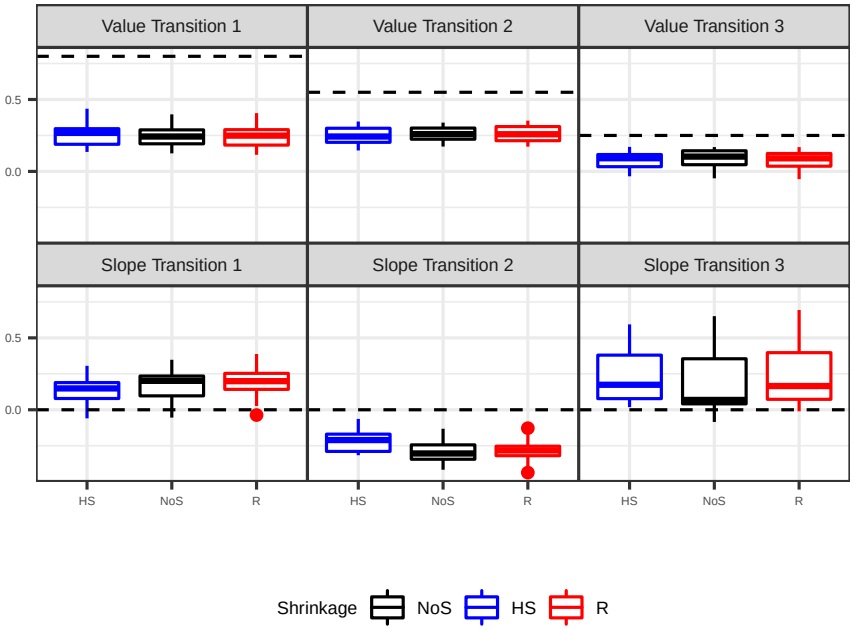

Figure 4.10: Posterior estimates for the coefficients of the association parameters under the scenario of the current value association for all transitions

towards zero. The observed bias in the results is expected as in most situations, the observed transitions were relatively small compared to the estimated parameters. This was done to mimic realistic scenarios such as the one we encountered in applying to the LVAD dataset. In cases where the number of observed transitions is relatively high, all three methods appear to perform equally well.

4.6 Disucssion

In this work, we considered an extended joint model that may include multivariate longitudinal and multi-states processes. Further, following our previous work, we incorporated intermediate events that can alter the course of the disease for the subjects under study. However, the inclusion of multiple longitudinal and multi-state processes poses some new challenges to the specification of the association structure between all these outcomes. First, the dimensionality of the parameter space becomes relatively higher than usual as the number of association parameters increases with the

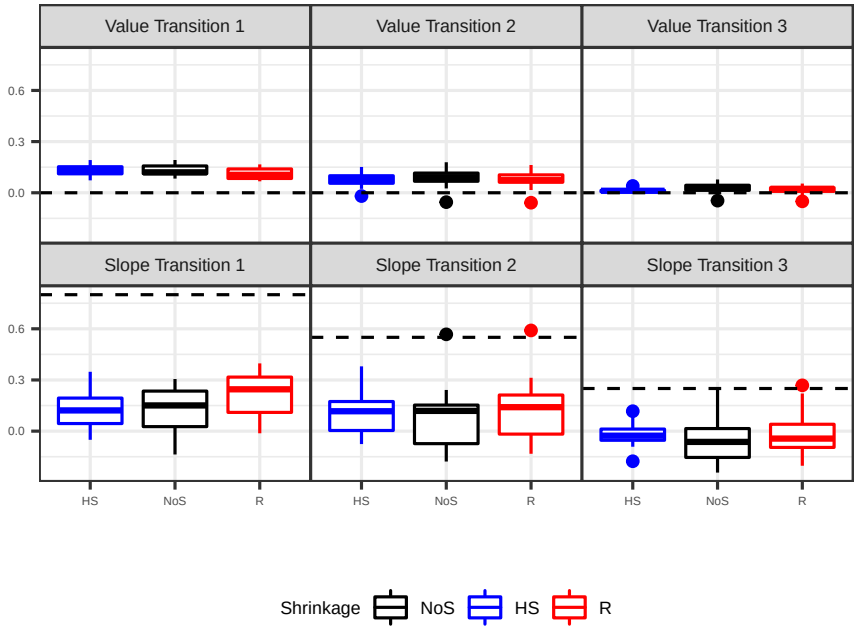

Figure 4.11: Posterior estimates for the coefficients of the association parameters under the scenario of the current slope association for all transitions

number of transitions and candidate association features. Secondly, these features are expected to be highly correlated as they come from the same longitudinal processes.

While the specification of the inter-relationships between outcomes should be driven by clinical background knowledge, to the extent possible, this is not always feasible in practice as the identification of their association structure might be the research question itself. Therefore, we proposed using Bayesian global-local shrinkage priors to aid in selecting features that better describe the association between the longitudinal outcomes and the transitions of interest. More specifically, we used the horseshoe and the ridge shrinkage priors in a setting with two longitudinal outcomes, three transitions of interest, and three candidate association features: the current value, the slope, and the area under the curve of the longitudinal outcomes. Our application to the LVAD dataset identified the current value association as the feature that better captures the association between TB and Cr and the risk for transitioning from heart failure to complications and death and from complication to death. Finally, our simulation study revealed that automating feature selection in such models is not

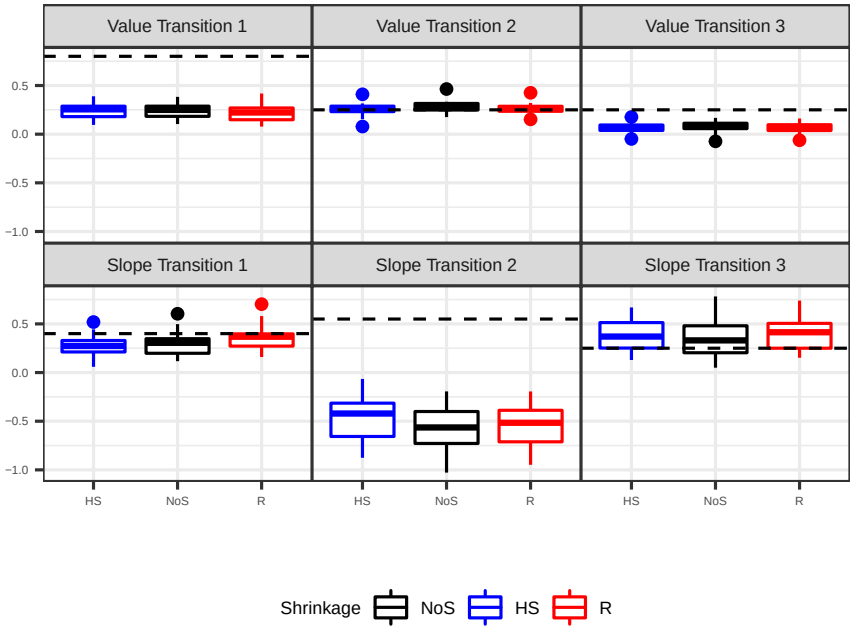

Figure 4.12: Posterior estimates for the coefficients of the association parameters under the scenario of both the current value and slope association for all transitions

a straightforward task. While global-local shrinkage priors were useful in shrinking nuisance parameters towards zero in contrast to strong effects, we observed that the actual value of the association parameters was not always captured despite a slightly better performance compared to not using shrinkage. Previous work on the topic (Andrinopoulou and Rizopoulos, 2016) identified similar issues but to a much lesser extent when using local shrinkage priors. However, using multi-state processes instead of a single survival outcome results in a more complicated setting and may cause such deviances especially when the number of observed transitions is relatively small.

While global-local shrinkage priors showed to be a promising tool that can help feature selection in joint models, many questions remain unanswered, sketching the map for future research on the topic. First, the cause of bias in the simulation results needs to be further explored to understand its causes. The correlation between the features of the longitudinal outcome(s) and the dimensionality of the parameter space need to be explored to identify their impact on feature selection. Furthermore, we only considered shrinkage for the association parameters in our work and not the

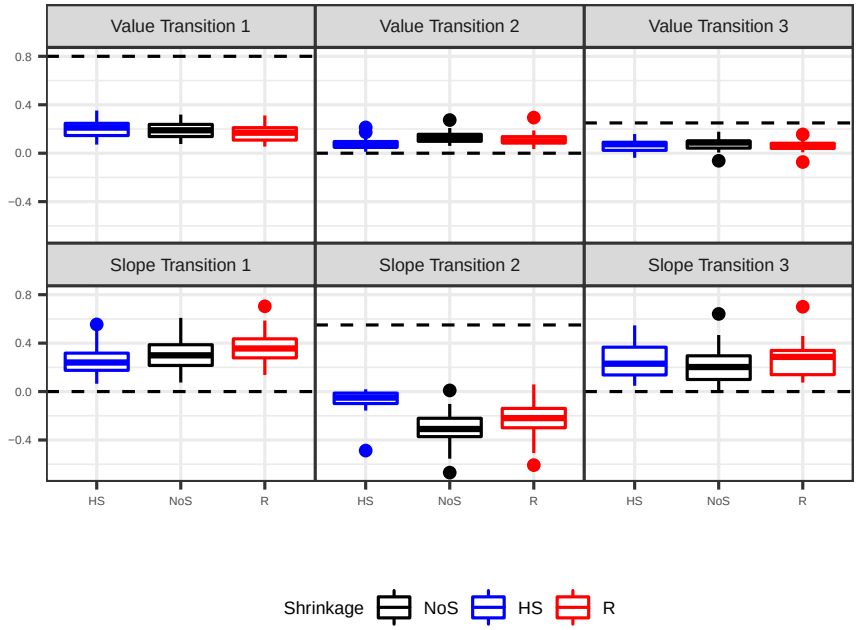

Figure 4.13: Posterior estimates for the coefficients of the association parameters under the scenario of the current value association for transitions 1 and 3 and the slope association for transition 2

baseline covariates. Applying shrinkage to all the parameters might help in the model's accuracy, especially in settings with low transition rates. Finally, further research is necessary to evaluate whether the use of global-local shrinkage priors can lead to better predictive performance in such models.

Table 4.4: Posterior estimates of hazard ratios and 95% credibility intervals for the parameters of the joint models without shrinkage (NoS), using the horseshoe prior (HS) and the ridge prior (R)

	NoS	**HS**	**R**
Parameter	HR(95% CI)	HR(95% CI)	HR(95% CI)
$\gamma 01_1$	1.039(0.999; 1.080)	1.042(1.007; 1.082)	1.041(1.000; 1.082)
$\gamma 02_1$	0.989(0.963; 1.015)	0.991(0.969; 1.014)	0.989(0.963; 1.015)
$\gamma 12_1$	1.008(0.958; 1.060)	1.012(0.964; 1.058)	1.010(0.961; 1.063)
$\gamma 01_2$	1.018(0.996; 1.041)	1.019(0.998; 1.040)	1.018(0.997; 1.040)
$\gamma 02_2$	1.005(0.991; 1.018)	1.005(0.993; 1.018)	1.005(0.992; 1.017)
$\gamma 12_2$	1.035(0.996; 1.072)	1.032(0.999; 1.065)	1.032(0.998; 1.067)
$\gamma 01_3$	0.913(0.450; 1.852)	0.870(0.462; 1.729)	0.886(0.452; 1.755)
$\gamma 02_3$	1.169(0.764; 1.871)	1.166(0.763; 1.832)	1.170(0.760; 1.822)
$\gamma 12_3$	1.180(0.400; 3.839)	1.179(0.383; 3.441)	1.223(0.428; 3.683)
$\gamma 01_4$	1.176(0.595; 2.276)	1.127(0.567; 2.140)	1.117(0.532; 2.193)
$\gamma 02_4$	1.052(0.692; 1.588)	1.021(0.681; 1.585)	1.042(0.691; 1.631)
$\gamma 12_4$	0.618(0.252; 1.542)	0.646(0.260; 1.617)	0.639(0.282; 1.527)
$\alpha_{01,TB1}$	1.257(0.810; 1.874)	1.111(0.853; 1.553)	1.104(0.851; 1.632)
$\alpha_{02,TB1}$	2.249(1.724; 2.820)	2.239(1.759; 2.740)	2.217(1.757; 2.756)
$\alpha_{12,TB1}$	2.419(1.519; 3.892)	2.372(1.649; 3.446)	2.315(1.574; 3.417)
$\alpha_{01,TB2}$	1.052(0.876; 1.236)	1.010(0.854; 1.183)	1.018(0.862; 1.172)
$\alpha_{02,TB2}$	1.044(0.932; 1.165)	1.026(0.931; 1.137)	1.025(0.928; 1.139)
$\alpha_{12,TB2}$	0.164(0.046; 0.513)	0.191(0.065; 0.683)	0.208(0.067; 0.749)
$\alpha_{01,TB3}$	1.010(0.966; 1.055)	1.022(0.989; 1.058)	1.022(0.985; 1.056)
$\alpha_{02,TB3}$	0.954(0.927; 0.988)	0.953(0.929; 0.984)	0.953(0.929; 0.983)
$\alpha_{12,TB3}$	0.938(0.891; 0.993)	0.940(0.899; 0.984)	0.943(0.901; 0.988)
$\alpha_{01,Cr1}$	0.777(0.327; 1.809)	0.941(0.515; 1.547)	0.944(0.559; 1.398)
$\alpha_{02,Cr1}$	0.639(0.365; 1.141)	0.688(0.411; 1.042)	0.724(0.416; 1.078)
$\alpha_{12,Cr1}$	2.844(1.048; 9.449)	2.184(0.977; 6.083)	1.965(0.932; 6.312)
$\alpha_{01,Cr2}$	1.115(0.746; 1.717)	1.086(0.815; 1.608)	1.066(0.833; 1.536)
$\alpha_{02,Cr2}$	0.955(0.740; 1.285)	0.978(0.789; 1.214)	0.981(0.788; 1.199)
$\alpha_{12,Cr2}$	0.747(0.238; 2.822)	0.839(0.306; 1.651)	0.902(0.343; 1.500)
$\alpha_{01,Cr3}$	1.070(0.955; 1.226)	1.050(0.979; 1.151)	1.048(0.972; 1.153)
$\alpha_{02,Cr3}$	1.022(0.955; 1.099)	1.014(0.964; 1.081)	1.011(0.950; 1.077)
$\alpha_{12,Cr3}$	0.824(0.660; 0.971)	0.853(0.710; 0.993)	0.867(0.707; 0.999)

4.7 References

Andrinopoulou, E.-R. and Rizopoulos, D. (2016). Bayesian shrinkage approach for a joint model of longitudinal and survival outcomes assuming different association structures. *Statistics in Medicine*, 35(26):4813–4823.

Andrinopoulou, E.-R., Rizopoulos, D., Takkenberg, J. J. M., and Lesaffre, E. (2014). Joint modeling of two longitudinal outcomes and competing risk data. *Statistics in Medicine*, 33(18):3167–3178.

Antonides, C. F., Schoenrath, F., de By, T. M., Muslem, R., Veen, K., Yalcin, Y. C., Netuka, I., Gummert, J., Potapov, E. V., Meyns, B., Özbaran, M., Schibilsky, D., Caliskan, K., and the EUROMACS investigators (2020). Outcomes of patients after successful left ventricular assist device explantation: a euromacs study. *ESC Heart Failure*, 7(3):1085–1094.

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

Ferrer, L., Rondeau, V., Dignam, J., Pickles, T., Jacqmin-Gadda, H., and Proust-Lima, C. (2016). Joint modelling of longitudinal and multi-state processes: application to clinical progressions in prostate cancer. *Statistics in Medicine*, 35(22):3933–3948.

Mauff, K., Steyerberg, E. W., Nijpels, G., van der Heijden, A. A., and Rizopoulos, D. (2017). Extension of the association structure in joint models to include weighted cumulative effects. *Statistics in Medicine*, 36(23):3746–3759.

Muslem, R., Caliskan, K., Akin, S., Sharma, K., Gilotra, N. A., Constantinescu, A. A., Houston, B., Whitman, G., Tedford, R. J., Hesselink, D. A., Bogers, A. J., Russell, S. D., and Manintveld, O. C. (2018). Acute kidney injury and 1-year mortality after left ventricular assist device implantation. *The Journal of Heart and Lung Transplantation*, 37(1):116 – 123.

Papageorgiou, G., Mokhles, M. M., Takkenberg, J. J. M., and Rizopoulos, D. (2019). Individualized dynamic prediction of survival with the presence of intermediate events. *Statistics in Medicine*, 38(30):5623–5640.

Proust-Lima, C. and Taylor, J. M. G. (2009). Development and validation of a dynamic prognostic tool for prostate cancer recurrence using repeated measures of posttreatment PSA: a joint modeling approach. *Biostatistics*, 10(3):535–549.

Rizopoulos, D. and Ghosh, P. (2011). A bayesian semiparametric multivariate joint model for multiple longitudinal outcomes and a time-to-event. *Statistics in Medicine*, 30(12):1366–1380.

Sène, M., Taylor, J. M., Dignam, J. J., Jacqmin-Gadda, H., and Proust-Lima, C. (2016). Individualized dynamic prediction of prostate cancer recurrence with and without the initiation of a second treatment: Development and validation. *Statistical Methods in Medical Research*, 25(6):2972–2991. PMID: 24847900.

Wulfsohn, M. S. and Tsiatis, A. A. (1997). A joint model for survival and longitudinal data measured with error. *Biometrics*, 53(1):330–339.

Yalcin, Y. C., Muslem, R., Veen, K. M., Soliman, O. I., Hesselink, D. A., Constantinescu, A. A., Brugts, J. J., Manintveld, O. C., Fudim, M., Russell, S. D., Tomashitis, B., Houston, B. A., Hsu, S., Tedford, R. J., Bogers, A. J., and Caliskan, K. (2020). Impact of continuous flow left ventricular assist device therapy on chronic kidney disease: A longitudinal multicenter study. *Journal of Cardiac Failure*, 26(4):333 – 341.

Chapter 5

The clinical impact of tricuspid regurgitation in patients with a biatrial orthotopic heart transplant

This chapter is based on:

Veen, K.M.*, Papageorgiou, G*, Zijderhand, C., Mokhles, M.M., Brugts, J., Manintveld, O.C., Constantinescu, A.A., Bekkers, J., Takkenberg, J.J.M., Bogers A.J.J.C. and Kaliskan, C. (2020) The clinical impact of tricuspid regurgitation in patients with a biatrial orthotopic heart transplant *Submitted to The Journal of Thoracic and Cardiovascular Surgery*

* both authors contributed equally

Abstract

Introduction
Tricuspid regurgitation (TR) is common in patients with after biatrial orthotopic heart transplant (OHT). Nevertheless, the clinical impact and long-term sequel of TR remains unclear. In this study, we aim to elucidate the clinical impact and long-term course of TR, taking into account its dynamic nature.
Methods
All consecutive adult patients undergoing biatrial OHT (1984-2017) and with an available follow-up echocardiogram were included in this study. Mixed-models were used to model the evolution of TR. Thereafter, the estimated trajectory was inserted into a Cox model, under the joint-model framework, in order to address the association of the dynamic TR with mortality.
Results
In total, 572 patients were included (median age: 50 years, males:74.9%). Approximately 32% of patients had moderate-to-severe TR immediately after surgery. However, this declined to approximately 11% on 5 years and 9% on 10 years after of surgery, adjusted for survival bias. Pre-implant mechanical support was associated with less TR during follow-up, whereas concurrent LV dysfunction was significantly associated with more TR during follow-up. Survival at 1, 5, 10, 20 years was 97±1%, 88±1%, 66±2% and 23±2%, respectively. The presence of moderate-to-severe TR during follow-up was associated with higher mortality (HR:1.07,95%CI[1.02-1.12],p=0.006). The course of TR was positively correlated with the course of creatinine (R=0.45).
Conclusion
TR during follow-up is significantly associated with higher mortality and worse renal function. Nevertheless, probability of TR is the highest immediately after OHT and decreases thereafter. Therefore, it may be reasonable to refrain from surgical intervention for TR during earlier phase after OHT.

5.1 Introduction

Tricuspid regurgitation (TR) is common in patients post biatrial orthotropic heart transplant (OHT) (Wong et al., 2008). Risk factors for TR after OHT include endomyocardial biopsies, allograft rejection, mismatch between the donor heart size and pericardial cavity dimensions (De Simone et al., 1995; Aziz et al., 1999; Caliskan et al., 2018). Additionally, several studies identified a biatrial anastomoses technique (vs a bicaval anastomoses) as independent risk factor for TR after OHT (Aziz et al., 1999; Wartig et al., 2014; Sun et al., 2007). Nevertheless, the clinical impact of TR remains unclear, partly because post-OHT TR is a dynamic disease that changes over time in individual patients. Due to these complex characteristics the clinical impact of post-OHT TR cannot be approached using traditional statistical tools. In this study, we aim to elucidate the clinical impact and long-term course of TR, taking into account its dynamic nature, by using novel statistical models to link the course of post-OHT TR to survival and renal function.

5.2 Methods

Patients

Consecutive patients that underwent biatrial OHT from 1984 to 2016 in Erasmus MC were included in this retrospective cohort study (n=687). Patients whom echocardiograms results were not retrievable or had recorded echocardiograms without TR measurements were excluded (n=115) resulting in 572 patients eligible for analyses. Of note, most patients that died within 30 days did not have a TR measurement on echocardiogram and, therefore, were excluded. (Supplementary Figure 5.6). Approval from the local Medical Ethical Committee was obtained to conduct this study (MEC-2017-421).

Data collection

Baseline characteristics were extracted from our institutional OHT database. Additionally, all echocardiographic measurements and creatinine measurements were collected longitudinally via automated extraction from the electronic patient records. Furthermore, echocardiographic measurements were supplemented with data acquired from paper patient records. The Dutch municipal civil registry was checked for the survival status.

Study outcome

The main outcome of this study is mortality in relation to the changing TR severity over time. Secondary outcomes include: the evolution of post-OHT TR grade and the evolution of post-OHT creatinine in relation to the changing TR severity.

Operation

All patients were operated with the biatrial anastomoses technique described in LOWER and SHUMWAY (1960). This technique entails an incision in the right atrium from the inferior vena cava toward the right atrial appendage to avoid sino-atrial node injury.

Statistics

Continuous data are presented as mean ± standard deviation (Gaussian distribution) or median [interquartile range (IQR)] (nonGaussian distribution). Categorical data are presented as frequencies (percentage).

Logistic mixed-effect models were used to assess probability of TR over time and investigate determinants of the longitudinal evolution over time. These models included random intercept and slope effects to capture the correlation of the repeated measurements in each patient. Natural splines with 2 knots placed at the 1st and 3rd quartiles were used to allow for flexibility of the subject-specific trajectories over time. Splines allow for non-linear trajectories over time. This is achieved by allowing a different spline-coefficient for each time interval defined by the knots (e.g. two knots define 3 such intervals). Survival probabilities were estimated and visualized by the Kaplan-Meier method. A joint model was developed to investigate determinants of mortality. More specifically, the mixed-effects model of TR and a relative risk model for the hazard of death (e.g. Cox model) were jointly modelled using shared-random effects. The subject-specific estimated longitudinal profiles were included in the relative risk model as predictors. Joint modelling has several benefits, such as the appropriate inclusion of endogenous covariates in relative risk models (TR), reduced bias and increased efficiency, while it can be used to derive dynamic predictions (Tsiatis and Davidian, 2004). At time point t one can investigate the effect of the current value of TR, the effect of the slope of TR (at which speed probability of TR is changing at time point t) and the cumulative effect of TR.

Global-local shrinkage priors were used for the regression coefficients of the relative risk sub-model for the selection of the current value of TR as predictor and this is presented in the article (Supplementary Tables: 5.4, 5.5, 5.6). The longitudinal

evolution of TR probability was correlated to the longitudinal evolution of creatinine by multivariate (multiple outcomes) mixed modelling. Correlation tests were done on the random effects D matrix. A p-value < 0.05 was considered statistically significant. Statistical analyses were done in R (R Core Team, 2020) with the use of statistical packages **GLMMadaptive**, **splines**, **JointAI**, **survival** and **JMbayes**.

5.3 Results

In total, 572 patients were included in this study. Baseline characteristics are presented in Table 5.1. Given the dynamic character of TR over time, baseline characteristics are not stratified on post-OHT TR grade. On average patients were 50 years old and 74.9% was male. Most frequently cyclosporine / prednisone (26.4%) and tacrolimus / prednisone (21.3%) are prescribed as immunosuppressive maintenance therapy. Median follow-up was 10.4 years (IQR: 6.4-15.3). Two patients were lost in follow-up, resulting in a completeness of 99.5% (C).

Tricuspid regurgitation evolution

In total, 8826 echocardiograms were collected (range: 1-50, mean: 15.4) and all echocardiograms are used in the analyses. The model predicting the evolution of TR over time is presented in Table 5.2. Probability of TR changed over time, as indicated by the significant times estimates (Table 5.2). The evolution of the probability of moderate-to-severe TR over time, as estimated by the mixed-model, is presented in Figure 5.1. On average, approximately 32% of patients have moderate-to-severe TR immediately after surgery. However, this declines to approximately 11% after 5 years and 9% after 10 years of surgery. Pre-implant mechanical support was significantly associated with lower probability of moderate-to-severe TR during follow-up (Table 5.2). Additionally, a worse LV function at the time of the TR measurement was significantly associated with a higher probability of moderate-to-severe TR (Table 5.2). Strikingly, the number of rejections in the first year was not associated with a higher probability of moderate-to-severe TR.

Mortality

During follow-up 357 patients died of which 5 (0.9%) within 30 days. Survival at 1, 5, 10, 20 years was 97±1%, 88±1%, 66±2% and 23±2%, respectively (Figure 5.2). The presence of moderate-to-severe TR during follow-up was associated with higher mortality (Table 5.3). Table 5.3 presents the estimates of the joint model. A higher age, the presence of pre-OHT diabetes, recipient female sex and dialysis

Table 5.1: Pre-, peri-, postoperative characteristics

Characteristics	**N = 572**
Recipient Age (median, IQR)	50.23 [41.87, 56.32]
Donor Age (median, IQR)	33.00 [22.00, 44.00]
Receiver female sex (n, %)	143 (25.1)
Donor female sex (n, %)	277 (49.0)
Primary diagnosis (n, %)	
Non ischemic	275 (49.4)
Ischemic	254 (45.6)
Other	28 (5.0)
Creatinine (median, IQR)	114.00 [93.00, 136.00]
Immunosuppression (n, %)	
Cyclosporine + azathioprine + prednisone	70 (12.3)
Cyclo + MMF + prednisone	32 (5.6)
Cyclosporine + prednisone	150 (26.4)
Tacrolimus + prednisone	121 (21.3)
Tacrolimus + MMF	21 (3.7)
Other	175 (30.8)
Number of prior cardiac operations (n, %)	
0	415 (72.8)
1	125 (21.9)
2	26 (4.6)
3	4 (0.7)
Urgency (n, %)	
0	370 (65.1)
1	99 (17.4)
2	93 (16.4)
3	6 (1.1)
Pre HTx diabetes (n, %)	35 (6.4)
Pre HTx mechical assistance (n, %)	
None	517 (92.8)
LVAD	33 (5.9)
ECMO	2 (0.4)
IABP	5 (0.9)
Ischemia time (median, IQR)	170.00 [143.00, 203.00]
Rethoractomy (n, %)	77 (13.5)
Dialysis* (n, %)	94 (16.8)
Pacemaker* (n, %)	70 (12.5)
Number rejection first year 1(median, IQR)	1.00 [1.00, 2.00]

Number of patients that received a pacemaker or dialysis during follow up

were significantly associated with mortality during follow-up. Moderate-to-severe TR remained significant a sensitivity analyses in which left ventricular dysfunction

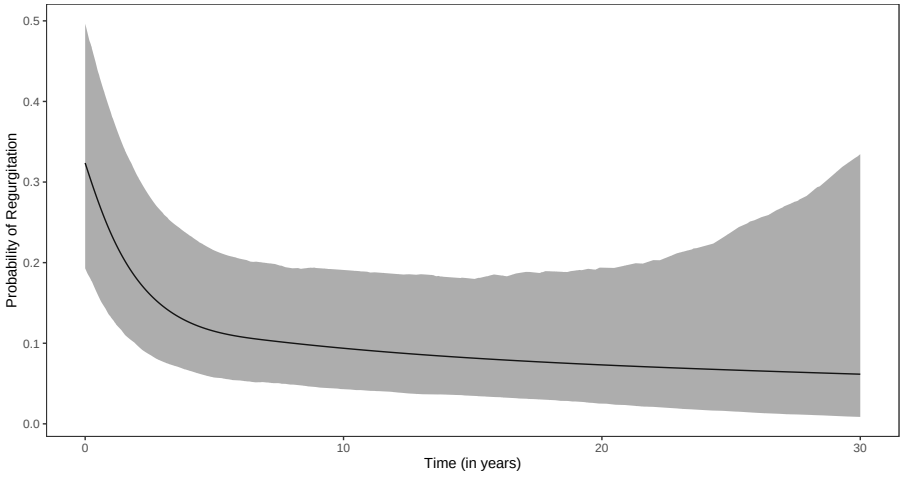

Figure 5.1: The marginal probability of moderate-to-severe TR during follow-up for an average patient.

was incorporated in the Cox model as time-varying covariate (Supplementary Table 5.7). Figure 5.3 and Figure 5.4 presents a dynamic survival probability plot for two patients. The first patient did not develop moderate-to-severe TR after approximately 3 years. At this moment, the survival probability of 10 years later is estimated to be 81% (Figure 5.3). The second patient did develop moderate-to-severe TR at 3 years, and the survival probability of 10 years later for this patients is estimated to be 77% (Figure 5.4).

Kidney function

Creatinine was collected at 4426 times simultaneously with an echocardiogram. The longitudinal evolution of creatinine is presented in Figure 5.5 as estimated by a mixed-model containing only the variable time with a spline function. The random slope of moderate-to-severe TR was highly positively correlated to the slope of creatinine (R = 0.45), meaning that if the probability moderate-to-severe TR increases, creatinine levels also increase in an individual patient. The intercept (starting point) of TR was not highly correlated the intercept (starting point) of creatinine (R = 0.04). The correlation matrix is shown in Supplementary Table 5.8. The current value of post-OHT moderate-to-severe TR was found to be predictive for dialysis dependence (HR 1.21 95% CI [1.04 to 1.44], P = 0.012) as estimated by a simple joint-model adjusting for baseline creatinine, sex and age (Supplementary Table 5.9).

Table 5.2: Odds Ratios (OR), 95% Credible Intervals (95% CI) and tail probabilities (P) from the logistic mixed-effects part of the joint model used to predict moderate-to-severe TR over time

Parameter	**OR**	**95% CI**	**P**
Intercept	0.10	(0.01; 1.35)	0.090
Time Spline_1 term[1]	0.23	(0.09; 0.56)	<0.001
Time Spline_2 term[1]	0.03	(0.01; 0.13)	<0.001
Time Spline_3 term[1]	0.18	(0.02; 1.22)	0.080
Receiver age	0.98	(0.95; 1.01)	0.138
Donor age	1.02	(0.99; 1.05)	0.114
Receiver sex = female	0.77	(0.36; 1.59)	0.466
Donor sex = female	1.66	(0.86; 3.18)	0.130
Ischemia time	0.96	(0.89; 1.04)	0.344
Cardiac reoperation	0.99	(0.56; 1.76)	0.952
Urgency 1 vs. 0	1.40	(0.56; 3.55)	0.456
Urgency 2/3 vs. 0	0.63	(0.27; 1.46)	0.322
No mechanical assistance prior HTx	6.29	(1.47; 27.31)	0.014
Pre HTx diabetes	0.60	(0.19; 1.94)	0.394
Number rejection first year	1.02	(0.88; 1.19)	0.742
Mildly impaired LV function vs normal[2]	1.73	(1.32; 2.31)	<0.001
Moderately impaired LV function vs normal[2]	4.03	(2.29; 7.18)	<0.001
Severely impaired LV function vs normal[2]	9.54	(2.82; 38.46)	<0.001
Pacemaker[2]	1.12	(0.59; 2.07)	0.718
Creatinine	1.00	(1; 1.01)	0.374

CI: Credible Interval, OR: Odds Ratio

1: Time was modelle using natural cubic splines with to knots placed at the 25th and 75th percentiles of observed time-points.

2: At the time of TR measurement (used as a timevarying covariate)

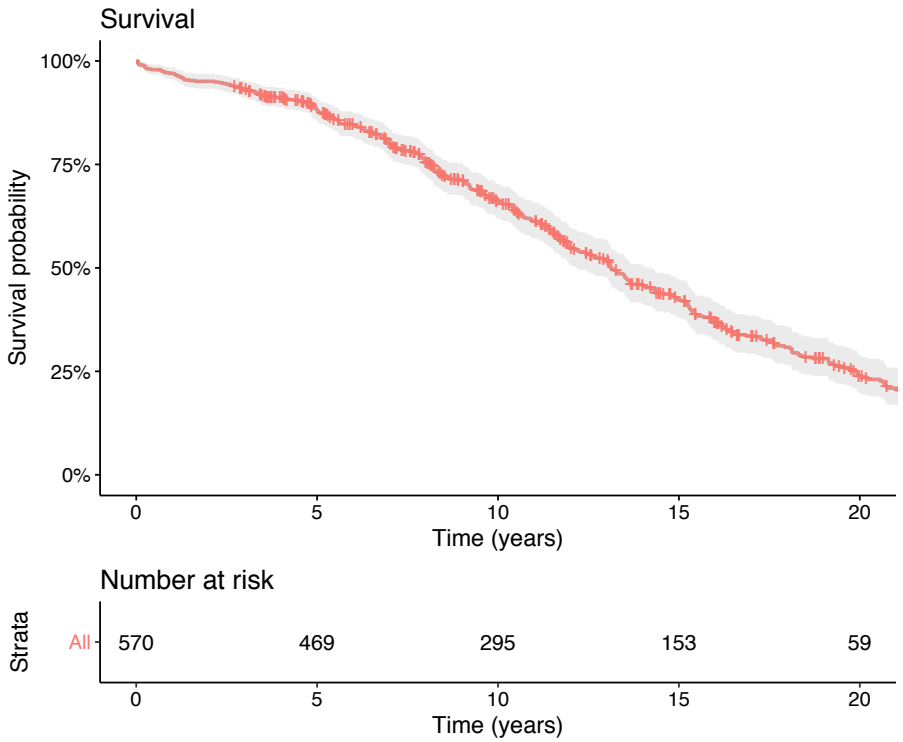

Figure 5.2: A Kaplan-Meier curve of overall survival.

5.4 Discussion

This study investigated the long-term course of moderate-to-severe TR and its impact on mortality and renal function. We found that moderate-to-severe TR during follow-up was associated with higher mortality and progressive decline of renal function. Specifically, moderate-to-severe TR was found to be a risk factor for dialysis. To the authors knowledge, this is the first study that accounts for the dynamic nature of TR during follow-up.

TR evolution

The etiology of TR after OHT is multifactorial in nature (Aziz et al., 1999). In older studies higher pulmonary pressures after OHT and endomyocardial biopsies were mainly found to be associated with TR (Aziz et al., 1999; Lewen et al., 1987; Bhatia et al., 1987; Hausen et al., 1995). Furthermore, the biatrial surgical technique

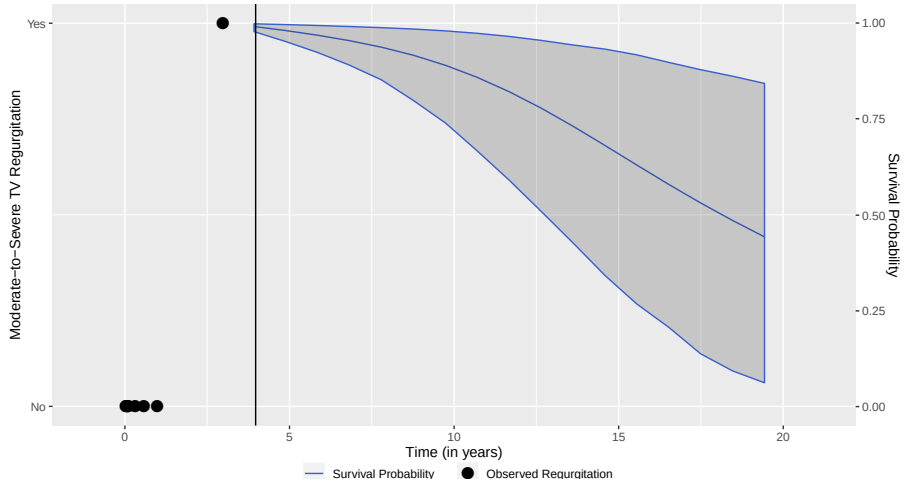

Figure 5.3: A dynamic prediction plot of Patient A who does not develop TR after 3 years.

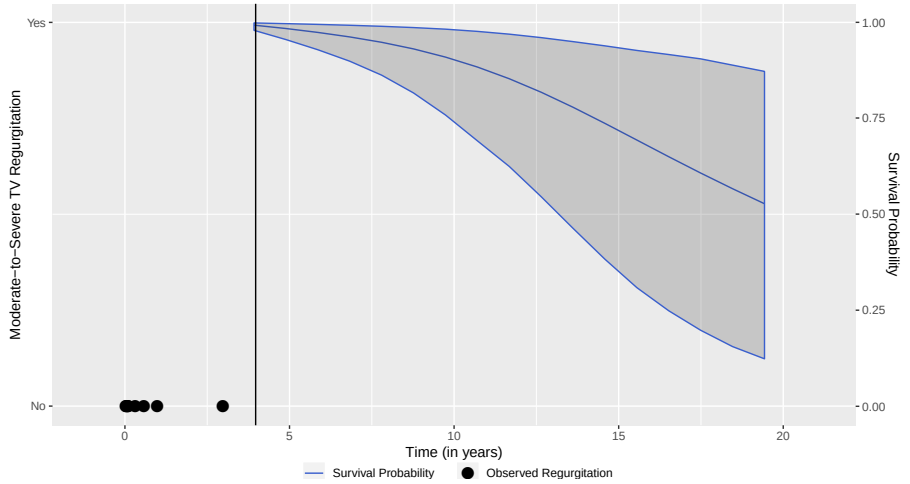

Figure 5.4: A dynamic prediction plot of Patient B who does develop TR after 3 years.

is found to be associated with more TR in multiple studies (Zijderhand et al., 2020).

In our study, left ventricular dysfunction at the time of TR measurement was significantly associated with the higher probabilities of moderate-to-severe TR, probably because worse LV function causes higher pulmonary pressures, subsequently leading

Table 5.3: Hazard Ratios (HR), 95% Credible Intervals (95% CI) and tail probabilities (P) from the relative risk submodel part of the joint model used to predict moderate-to-severe TR over time

Parameter	**HR**	**95% CI**	**P**
Receiver age	1.04	(1.03; 1.06)	<0.001
Receiver sex = female	1.44	(1.02; 2.03)	0.046
Donor sex = female	1.05	(0.78; 1.44)	0.768
Ischemia time	1.00	(1; 1.01)	0.064
Cardiac reoperation	1.22	(0.91; 1.6)	0.174
Urgency 1 vs 0	0.90	(0.61; 1.31)	0.590
Urgency 2 vs 0	0.78	(0.48; 1.22)	0.270
Urgency 3 vs 0	12.26	(2.32; 55.12)	0.012
No mechanical assistance prior HTx	0.75	(0.28; 2.19)	0.580
Donor age	0.99	(0.98; 1.01)	0.372
Non-ischemic CMP vs ischemic	1.17	(0.82; 1.66)	0.364
Other diagnosis vs ischemic	1.15	(0.55; 2.15)	0.668
Pre HTx diabetes	2.30	(1.3; 3.8)	<0.001
Ceatinine	1.00	(1.00; 1)	0.096
Pacemaker[1]	1.00	(0.64; 1.52)	0.972
Dialysis[1]	1.81	(1.27; 2.5)	0.002
Mod-sev TR	1.07	(1.02; 1.13)	0.006

CI: Credible Interval, HR: Hazard Ratio
1: Time-varying covariate

to RV dysfunction and dilatation, leading to functional TR. Moreover, we noted that patients who have mechanical assistance (LVAD, ECMO, IABP) prior OHT have a lower probability of post-OHT moderate-to-severe TR. It has been observed that left ventricular assist devices effectively unload the left ventricle and reduce pulmonary pressures (Atluri et al., 2013). Hence, patients with pre-OHT mechanical assistance will probably have lower pulmonary pressures, resulting in less right ventricle dysfunction, annulus dilation and, secondary TR immediately after OHT.

Other studies noted initially a decrease in TR severity after OHT, but a relative

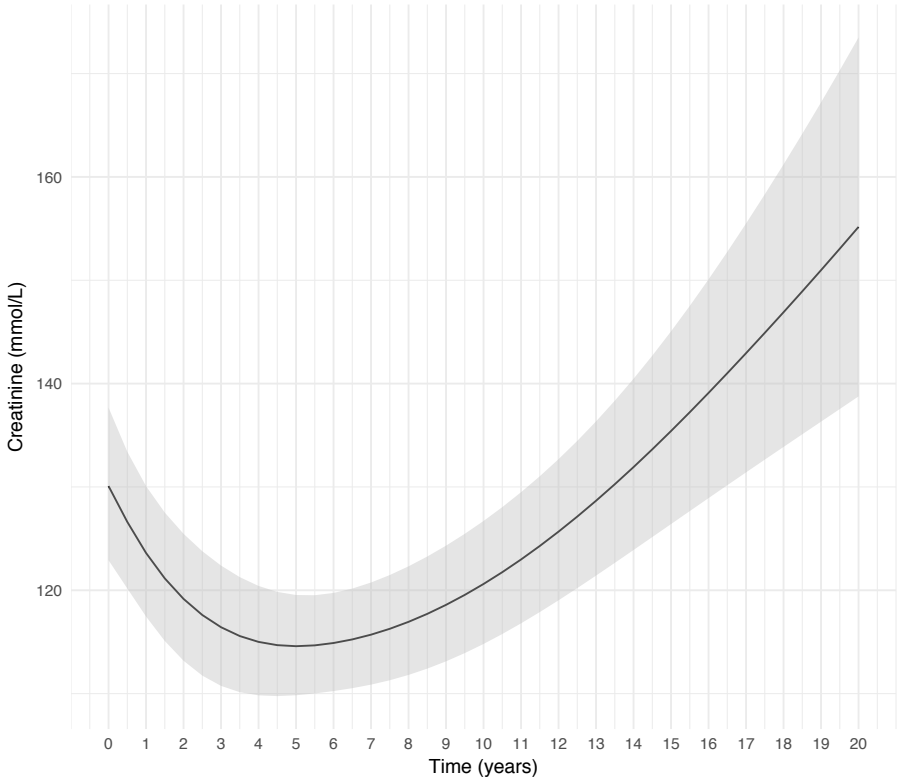

Figure 5.5: The predicted evolution of creatinine after HTx

increase later in follow up, or even a gradual increase in TR over time (Wartig et al., 2014; Hausen et al., 1995; Berger et al., 2012). This study did not replicate these results. Nevertheless, change over time was not significantly decreasing over time later in follow-up. The results of prior studies can partly be explained by the used methodology, which does not take into account the correlations within patients vs between patients nor does take into account the dropout of patients (either due to death or censoring), whereas the joint modeling framework does take these phenomena into account.

Mortality & Morbidity

In this cohort we only included patients with a follow-up echocardiogram, as the focus was on evolution of TR. Previously we reported the outcomes of the entire

cohort (Zijlstra et al., 2015). In patient who die early an echocardiogram may not be performed or TR in this echocardiogram is not recorded, explaining the low 30-day mortality (0.9%) in this subset of the entire cohort (Supplementary Figure 5.6).

Previous studies noted that TR at discharge was associated with impaired late morality (Wartig et al., 2014; Anderson et al., 2004). Two other studies examined late TR and noted contradicting results in regard to the association with mortality (Berger et al., 2012; Aziz et al., 2002). This study models the dynamic nature of TR over time and the association with mortality. During follow-up developing TR is associated with higher probability of mortality. The dynamic predictions estimated that developing TR at 3 years after OHT is paired with a 4% reduction in survival 10 years later compared to a patient who does not develop TR, given that all the other variables are similar.

The observed association of TR with mortality does not inherently imply a causal association. An important factor in in this interplay is right ventricular dysfunction. We did not analyze right ventricular function in this study, because of high rate of missing data (it was only reported in 1216 of the 8826 echocardiograms) without possibilities for review of older echocardiograms, and it had a tendency to be reported more if significantly impaired (missing mechanism: not at random). Moreover, it is complicated to make causal inference in regard to right ventricular dysfunction and TR due to their circular relationship; TR leads to right ventricular dysfunction, which leads to dilatation, in turn leading to more TR. One needs to backtrack which phenomena starts first and starts the negative spiral, which is difficult to do in retrospective studies. Nevertheless, previous studies claim that it is the TR that may lead to right ventricular dysfunction (Wong et al., 2008; Berger et al., 2012; Jeevanandam et al., 2006).

Moreover, we could also link the longitudinal evolution of TR probabilities to the longitudinal evolution of creatinine. Previous studies also found an association between renal function and TR (Aziz et al., 2002). It is still debatable whether it is the TR or right ventricular dysfunction causing the renal dysfunction, however TR may contribute to renal dysfunction by increasing venous congestion (Maeder et al., 2008) and the combination of TR and right ventricular dysfunction is found to predictive of impaired renal function (Agricola et al., 2017). Furthermore, in a recent study Karam et al. noted stabilization of renal function and improvement of liver function in patients undergoing transcatheter tricuspid valve repair, suggesting a beneficial effect of eliminating TR (Karam et al., 2019).

Clinical Implications

In most cases TR is managed with diuretics, but in refractory cases a surgical intervention becomes necessary (1). Literature regarding surgery for TR after OHT is scarce. Nevertheless, it has been shown that surgery in these patients can be performed safely (Alharethi et al., 2006; Filsoufi et al., 2006). The authors who linked discharge TR to impaired survival suggest to surgically intervene if TR is not resolved by discharge (Wartig et al., 2014). However, our data shows that it may be reasonable to wait longer, as probability of TR continues to decrease after discharge, and TR usually remains asymptomatic for years (Caliskan et al., 2018). Notwithstanding, our data shows that after approximately five years post-OHT the decrease probability of moderate-to-severe TR negates. In patients with persistent TR at five years post-OHT surgical intervention may be most beneficial, assuming the association of TR and mortality / renal function is causal in nature. A small randomized clinical trial (n=60) in which patients received either prophylactic tricuspid annuloplasty vs. no annuloplasty concomitant to OHT noted a better cardiac survival in the annuloplasty groups, if they combined early and late deaths (Jeevanandam et al., 2006). No overall survival difference was noted. Furthermore, opportunities arise with emerging transcatheter devices to treat TR. However, these devices still need to be validated in this complex subgroup of patients. Other authors advocate the use of bicaval anastomosis to prevent TR in the first place (Sun et al., 2007; Solomon et al., 2004).

Strengths and limitations

The major advantage of this study is that we were able to collect 8826 echocardiograms, enabling us to use advanced statistical methods to model the dynamic nature of TR and making less biased inference of the impact of TR on mortality. Several limitations apply to this study common in retrospective analyses. We were not able to analyze right ventricular function due to limited and unreliable data. We did not consider cardiac allograft vasculopathy (CAV) explicitly in this study since CAV is diagnosed by coronary angiography and, therefore, there is a delay between development and diagnosis of CAV. Nevertheless, CAV manifests as LV dysfunction, which we were able to analyze, hence CAV is implicitly considered. Patients who died without an TR measurement on echocardiogram were excluded, which can introduce selection bias. Assessing TR remains challenging, but we dichotomized this variable to create a more robust measurement. Moreover, we it was not possible to determine the cause of TR (e.g. biopsy related vs functional). Lastly, LV function, pacemaker and dialyses were incorporated in the models as a time-varying exogenous variable, while in fact these variables are more likely to be endogenous.

5.5 Conclusions

TR during follow-up is significantly associated with higher mortality and progressive decline of renal function / end-stage renal failure. Nevertheless, probability of TR is the highest immediately after OHT and decreases thereafter. Therefore, it may be reasonable to refrain from surgical intervention during early phase after OHT with bi-atrial anastomoses.

Supplementary Material

5.A Supplementary Material

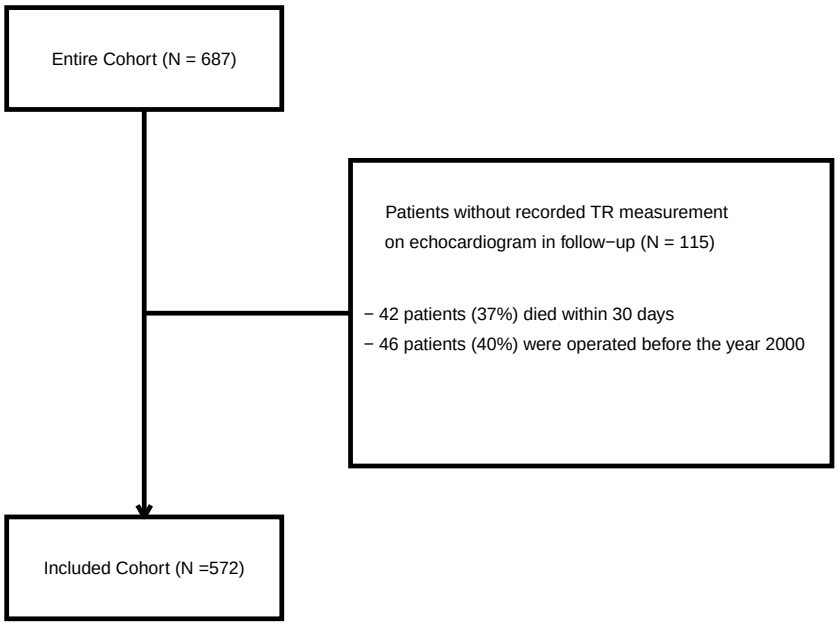

Figure 5.6: Flowchart of included patients

Table 5.4: Hazard Ratios (HR), 95% Credible Intervals (95% CI) and tail probabilities (P) from the relative risk submodel part of the joint model used to predict moderate-to-severe TR over time when using the horseshoe global-local shrinkage prior and value, slope and area under the curve association structures.

Parameter	**HR**	**95% CI**	**P**
Receiver age	1.01	(1.00; 1.03)	<0.050
Receiver sex = female	1.25	(0.95; 1.77)	0.186
Donor sex = female	1.03	(0.86; 1.26)	0.820
Ischemia time	1.00	(1; 1.01)	0.036
Cardiac reoperation	1.17	(0.95; 1.53)	0.198
Urgency 1 vs 0	0.98	(0.73; 1.22)	0.870
Urgency 2 vs 0	0.89	(0.56; 1.18)	0.544
Urgency 3 vs 0	6.98	(1.02; 26.60)	0.042
No mechanical assistance prior HTx	0.83	(0.32; 1.23)	0.608
Donor age	0.99	(0.98; 1.01)	0.404
Non-ischemic CMP vs ischemic	1.05	(0.86; 1.38)	0.656
Other diagnosis vs ischemic	1.09	(0.81; 1.75)	0.710
Pre HTx diabetes	2.17	(1.16; 3.79)	0.010
Ceatinine	1.00	(1.00; 1)	0.026
Pacemaker[1]	1.02	(0.80; 1.36)	0.888
Dialysis[1]	1.64	(1.09; 2.30)	0.020
Mod-sev TR (value)	1.07	(1.01; 1.15)	0.010
Mod-sev TR (slope)	0.97	(0.35; 1.95)	0.978
Mod-sev TR (area)	1.00	(1.00; 1.00)	0.046

CI: Credible Interval, HR: Hazard Ratio
1: Time-varying covariate

Table 5.5: Hazard Ratios (HR), 95% Credible Intervals (95% CI) and tail probabilities (P) from the relative risk submodel part of the joint model used to predict moderate-to-severe TR over time when using the ridge global-local shrinkage prior and value, slope and area under the curve association structures.

Parameter	**HR**	**95% CI**	**P**
Receiver age	1.01	(1.00; 1.03)	0.060
Receiver sex = female	1.41	(0.97; 1.97)	0.066
Donor sex = female	1.04	(0.76; 1.42)	0.802
Ischemia time	1.00	(1; 1.01)	0.064
Cardiac reoperation	1.22	(0.90; 1.62)	0.188
Urgency 1 vs 0	0.93	(0.63; 1.38)	0.722
Urgency 2 vs 0	0.76	(0.47; 1.21)	0.232
Urgency 3 vs 0	7.57	(1.42; 32.30)	0.022
No mechanical assistance prior HTx	0.63	(0.24; 1.68)	0.308
Donor age	0.99	(0.98; 1.01)	0.302
Non-ischemic CMP vs ischemic	1.14	(0.84; 1.61)	0.402
Other diagnosis vs ischemic	1.15	(0.59; 2.07)	0.618
Pre HTx diabetes	2.45	(1.53; 3.96)	0.002
Ceatinine	1.00	(1.00; 1)	0.040
Pacemaker[1]	1.03	(0.66; 1.56)	0.902
Dialysis[1]	1.78	(1.21; 2.49)	0.006
Mod-sev TR (value)	1.09	(1.02; 1.15)	0.006
Mod-sev TR (slope)	0.95	(0.53; 1.42)	0.770
Mod-sev TR (area)	1.00	(1.00; 1.00)	0.050

CI: Credible Interval, HR: Hazard Ratio
1: Time-varying covariate

Table 5.6: Hazard Ratios (HR), 95% Credible Intervals (95% CI) and tail probabilities (P) from the relative risk submodel part of the joint model used to predict moderate-to-severe TR over time when using the value, slope and area under the curve association structures without shrinkage priors.

Parameter	**HR**	**95% CI**	**P**
Receiver age	1.01	(1.00; 1.03)	0.158
Receiver sex = female	1.44	(0.96; 2.12)	0.074
Donor sex = female	1.03	(0.76; 1.43)	0.904
Ischemia time	1.00	(1; 1.01)	0.112
Cardiac reoperation	1.24	(0.91; 1.70)	0.218
Urgency 1 vs 0	0.92	(0.62; 1.37)	0.720
Urgency 2 vs 0	0.77	(0.44; 1.22)	0.308
Urgency 3 vs 0	8.74	(1.39; 48.33)	0.028
No mechanical assistance prior HTx	0.67	(0.23; 2.15)	0.474
Donor age	0.99	(0.98; 1.01)	0.348
Non-ischemic CMP vs ischemic	1.15	(0.81; 1.67)	0.442
Other diagnosis vs ischemic	1.12	(0.56; 2.15)	0.752
Pre HTx diabetes	2.52	(1.46; 4.22)	<0.001
Ceatinine	1.00	(0.99; 1)	0.070
Pacemaker[1]	1.01	(0.65; 1.50)	0.918
Dialysis[1]	1.83	(1.26; 2.64)	0.002
Mod-sev TR (value)	1.12	(1.03; 1.22)	0.010
Mod-sev TR (slope)	0.53	(0.07; 3.27)	0.546
Mod-sev TR (area)	1.00	(1.00; 1.00)	0.046

CI: Credible Interval, HR: Hazard Ratio
1: Time-varying covariate

Table 5.7: Hazard Ratios (HR), 95% Credible Intervals (95% CI) and tail probabilities (P) from the relative risk submodel part of the joint model used to predict moderate-to-severe TR over time when using the value association, without shrinkage priors but with left vertical function included as a time-varying covariate.

Parameter	**HR**	**95% CI**	**P**
Receiver age	1.04	(1.03; 1.06)	<0.001
Receiver sex = female	1.16	(0.84; 1.64)	0.344
Donor sex = female	1.37	(1.03; 1.82)	0.04
Ischemia time	1.00	(1.00; 1.00)	0.534
Cardiac reoperation	1.49	(1.08; 2.03)	0.012
Urgency 1 vs 0	1.16	(0.78; 1.78)	0.436
Urgency 2 vs 0	0.80	(0.49; 1.30)	0.356
Urgency 3 vs 0	11.69	(2.08; 45.83)	0.018
No mechanical assistance prior HTx	1.66	(0.63; 4.58)	0.296
Donor age	0.98	(0.97; 0.99)	<0.001
Non-ischemic CMP vs ischemic	1.19	(0.85; 1.62)	0.272
Other diagnosis vs ischemic	0.92	(0.48; 1.75)	0.764
Pre HTx diabetes	1.43	(0.85; 2.27)	0.136
Ceatinine	1.00	(1.00; 1.00)	0.658
Pacemaker[1]	0.99	(0.66; 1.45)	0.932
Dialysis[1]	1.37	(0.94; 2.03)	0.102
Moderate/Severe LV function vs. normal[1]	110.23	(53.51; 252.98)	<0.001
Mod-sev TR (value)	1.07	(1.02; 1.12)	0.008

CI: Credible Interval, HR: Hazard Ratio
1: Time-varying covariate

Table 5.8: Random-effects correlation matrix of the multivariate longitudinal model with creatinine and tricuspid regurgitation. Random effect were: intercept for patients and slope over time in both models. No splines were added in the random effects for time in order to enhance interpretability.

	Intercept TR	**Slope TR**	**Intercept Creatinine**	**Slope Creatinine**
Intercept TR	1	X	X	X
Slope TR	0.1791	1	X	X
Intercept Creatinine	0.0464	0.0775	1	X
Slope Creatinine	-0.6112	0.4540	-0.5634	1

Table 5.9: Hazard Ratios (HR), 95% Credible Intervals (95% CI) and tail probabilities (P) from the relative risk submodel part of the joint model used to predict

Variable	**HR**	**95% CI**	**P**
Receiver Age	0.99	(0.97; 1.02)	0.546
Baseline Creatinine	1.00	(0.99; 1.01)	0.784
Receiver sex = female	0.58	(0.26; 1.21)	0.150
Moderate-to-severe TR	1.21	(1.04; 1.44)	0.012

5.2 References

Agricola, E., Marini, C., Stella, S., Monello, A., Fisicaro, A., Tufaro, V., Slavich, M., Oppizzi, M., Castiglioni, A., Cappelletti, A., and Margonato, A. (2017). Effects of functional tricuspid regurgitation on renal function and long-term prognosis in patients with heart failure. *Journal of Cardiovascular Medicine*, 18(2):60 – 68.

Alharethi, R., Bader, F., Kfoury, A. G., Hammond, M. E., Karwande, S. V., Gilbert, E. M., Doty, D. B., Hagan, M. E., Thomas, H., and Renlund, D. G. (2006). Tricuspid valve replacement after cardiac transplantation. *The Journal of Heart and Lung Transplantation*, 25(1):48–52.

Anderson, C. A., Shernan, S. K., Leacche, M., Rawn, J. D., Paul, S., Mihaljevic, T., Jarcho, J. A., Stevenson, L. W., Fang, J. C.-T., Lewis, E. F., Couper, G. S., Mudge, G. H., and Byrne, J. G. (2004). Severity of intraoperative tricuspid regurgitation predicts poor late survival following cardiac transplantation. *The Annals of Thoracic Surgery*, 78(5):1635–1642.

Atluri, P., Fairman, A. S., MacArthur, J. W., Goldstone, A. B., Cohen, J. E., Howard, J. L., Zalewski, C. M., Shudo, Y., and Woo, Y. J. (2013). Continuous flow left ventricular assist device implant significantly improves pulmonary hypertension, right ventricular contractility, and tricuspid valve competence. *Journal of Cardiac Surgery*, 28(6):770–775.

Aziz, T. M., Burgess, M. I., Rahman, A. N., Campbell, C. S., Deiraniya, A. K., and Yonan, N. A. (1999). Risk factors for tricuspid valve regurgitation after orthotopic heart transplantation. *The Annals of Thoracic Surgery*, 68(4):1247 – 1251.

Aziz, T. M., Saad, R. A., Burgess, M. I., Campbell, C. S., and Yonan, N. A. (2002). Clinical significance of tricuspid valve dysfunction after orthotopic heart transplantation. *The Journal of Heart and Lung Transplantation*, 21(10):1101 – 1108.

Berger, Y., Har Zahav, Y., Kassif, Y., Kogan, A., Kuperstein, R., Freimark, D., and Lavee, J. (2012). Tricuspid valve regurgitation after orthotopic heart transplantation: Prevalence and etiology. *Journal of Transplantation*.

Bhatia, S. J., Kirshenbaum, J. M., Shemin, R. J., Cohn, L. H., Collins, J. J., Sesa, V. J. D., Young, P. J., Mudge, G. H., and Sutton, M. G. (1987). Time course of resolution of pulmonary hypertension and right ventricular remodeling after orthotopic cardiac transplantation. *Circulation*, 76(4):819–826.

Caliskan, K., Strachinaru, M., and Soliman, O. I. (2018). *Tricuspid Regurgitation in Patients with Heart Transplant*, pages 49–58. Springer International Publishing, Cham.

De Simone, R., Lange, R., Sack, F.-U., Mehmanesh, H., and Hagl, S. (1995). Atrioventricular valve insufficiency and atrial geometry after orthotopic heart transplantation. *The Annals of Thoracic Surgery*, 60(6):1686 – 1693. Thirty-second Annual Meeting.

Filsoufi, F., Salzberg, S. P., Anderson, C. A., Couper, G. S., Cohn, L. H., and Adams, D. H. (2006). Optimal surgical management of severe tricuspid regurgitation in cardiac transplant patients. *The Journal of Heart and Lung Transplantation*, 25(3):289–293.

Hausen, B., Albes, J. M., Rohde, R., Demertzis, S., Mügge, A., and Schäfers, H.-J. (1995). Tricuspid valve regurgitation attributable to endomyocardial biopsies and rejection in heart transplantation. *The Annals of Thoracic Surgery*, 59(5):1134 – 1140.

Jeevanandam, V., Russell, H., Mather, P., Furukawa, S., Anderson, A., and Raman, J. (2006). Donor tricuspid annuloplasty during orthotopic heart transplantation: Long-term results of a prospective controlled study. *The Annals of Thoracic Surgery*, 82(6):2089 – 2095.

Karam, N., Braun, D., Mehr, M., Orban, M., Stocker, T. J., Deseive, S., Orban, M., Hagl, C., Näbauer, M., Massberg, S., and Hausleiter, J. (2019). Impact of transcatheter tricuspid valve repair for severe tricuspid regurgitation on kidney and liver function. *JACC: Cardiovascular Interventions*, 12(15):1413–1420.

Lewen, M. K., Bryg, R. J., Miller, L. W., Williams, G. A., and Labovitz, A. J. (1987). Tricuspid regurgitation by doppler echocardiography after orthotopic cardiac transplantation. *The American Journal of Cardiology*, 59(15):1371 – 1374.

LOWER, R. and SHUMWAY, N. (1960). Studies on orthotopic homotransplantation of the canine heart. *Surgical forum*, 11:18—19.

Maeder, M. T., Holst, D. P., and Kaye, D. M. (2008). Tricuspid regurgitation contributes to renal dysfunction in patients with heart failure. *Journal of Cardiac Failure*, 14(10):824 – 830.

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Solomon, N. A. G., McGiven, J., Chen, X.-Z., Alison, P. M., Graham, K. J., and Gibbs, H. (2004). Biatrial or bicaval technique for orthotopic heart transplantation: which is better? *Heart Lung Circulation*, 13(4):389—394.

Sun, J. P., Niu, J., Banbury, M. K., Zhou, L., Taylor, D. O., Starling, R. C., Garcia, M. J., Stewart, W. J., and Thomas, J. D. (2007). Influence of different implantation techniques on long-term survival after orthotopic heart transplantation: An echocardiographic study. *The Journal of Heart and Lung Transplantation*, 26(12):1243 – 1248.

Tsiatis, A. A. and Davidian, M. (2004). Joint modeling of longitudinal and time-to-event data: An overview. *Statistica Sinica*, 14(3):809–834.

Wartig, M., Tesan, S., Gäbel, J., Jeppsson, A., Selimovic, N., Holmberg, E., and Dellgren, G. (2014). Tricuspid regurgitation influences outcome after heart transplantation. *The Journal of Heart and Lung Transplantation*, 33(8):829 – 835.

Wong, R. C.-C., Abrahams, Z., Hanna, M., Pangrace, J., Gonzalez-Stawinski, G., Starling, R., and Taylor, D. (2008). Tricuspid regurgitation after cardiac transplantation: An old problem revisited. *The Journal of Heart and Lung Transplantation*, 27(3):247 – 252.

Zijderhand, C. F., Veen, K. M., Caliskan, K., Schoonen, T., Mokhles, M. M., Bekkers, J. A., Manintveld, O. C., Constantinescu, A. A., Brugts, J. J., Bogers, A. J., and Takkenberg, J. J. (2020). Biatrial versus bicaval orthotopic heart transplantation: A systematic review and meta-analysis. *The Annals of Thoracic Surgery*, 110(2):684 – 691.

Zijlstra, L. E., Constantinescu, A. A., Manintveld, O., Birim, O., Hesselink, D. A., van Thiel, R., van Domburg, R., Balk, A. H. M., and Caliskan, K. (2015). Improved long-term survival in dutch heart transplant patients despite increasing donor age: the rotterdam experience. *Transplant International*, 28(8):962–971.

Part III

Missing Data

Chapter 6

An alternative characterization of MAR in shared parameter models for incomplete longitudinal data and its utilization for sensitivity analysis

This chapter is based on:

Papageorgiou, G. and Rizopoulos, D. (2020), An alternative characterization of MAR in shared parameter models for incomplete longitudinal data and its utilization for sensitivity analysis *Statistical Modelling*, Online First. doi: https://doi.org/10.1177/1471082X20927114

Abstract

Dropout is a common complication in longitudinal studies, especially since the distinction between random (MAR) and non-random (MNAR) dropout is intractable. Consequently, one starts with an analysis that is valid under MAR and then performs a sensitivity analysis by considering MNAR departures from it. To this end, specific classes of joint models, such as pattern-mixture models (PMMs) and selection models (SeMs), have been proposed. On the contrary, shared-parameter models (SPMs) have received less attention, possibly because they do not embody a characterisation of MAR. A few approaches to achieve MAR in SPMs exist, but are difficult to implement in existing software. In this paper, we focus on SPMs for incomplete longitudinal and time-to-dropout data and propose an alternative characterisation of MAR by exploiting the conditional independence assumption, under which outcome and missingness are independent given a set of random effects. By doing so, the censoring distribution can be utilised to cover a wide range of assumptions for the missing data mechanism on the subject-specific level. This approach offers substantial advantages over its counterparts and can be easily implemented in existing software. More specifically, it offers flexibility over the assumption for the missing data generating mechanism that governs dropout by allowing subject-specific perturbations of the censoring distribution, whereas in PMMs and SeMs dropout is considered MNAR strictly.

6.1 Introduction

Follow-up studies that include human subjects are known to be highly susceptible to dropout. The extent to which dropout will complicate the subsequent analysis depends on how the missingness is associated with the observed and unobserved longitudinal outcomes. According to the taxonomy introduced by Rubin (1976) and Little and Rubin (2002) this may happen according to three distinct missing data mechanisms: missing completely at random (MCAR), missing at random (MAR) and missing not at random (MNAR). Under MCAR, the missingness depends on neither observed nor unobserved outcomes. If the missingness is independent of the unobserved outcomes after conditioning on the observed outcomes, then the missing data mechanism is considered to be MAR. Conversely, under MNAR, conditioning on the observed outcomes cannot disentangle the dependence between the missingness process and the unobserved outcomes.

What separates MNAR from its counterparts in applied practice, though, is the concept of ignorability. That is under the (Bayesian) likelihood framework, one does not need to consider a model for the missingness process and hence 'ignore' it. Ignorability holds under the assumption that the process that generates the missing data is MAR and that the missingness and measurement processes depend on independent sets of parameters. While the innate convenience of ignorability popularized MAR models, one needs to consider if such an assumption holds carefully. That is especially the case since the distinction between MAR and MNAR is intractable as for every MNAR model, there exists a MAR counterpart with the same fit to the observed data (Molenberghs et al., 2008). Hence, one should start with an analysis that is valid under MAR and then perform a sensitivity analysis considering plausible MNAR departures from it.

To showcase the importance of such sensitivity analysis, consider the illustrative example in Figure 6.1 which is based on simulated data. The data were simulated assuming a joint model for longitudinal and survival data. To mimic MAR missingness, censoring was based on the value of previous repeated measurements and therefore the missing longitudinal trajectories (depicted with dotted gray) consist of the subjects that fulfilled the MAR criterion or experienced the event. We then fitted a model to the complete data whereas for the incomplete data we fitted an MAR and an MNAR model. All three models resulted in different fits to the data, which consequently lead to different subject-specific predictions as shown in the lower panel of Figure 6.1. Therefore, sensitivity analysis is necessary to be able to investigate the robustness of the model and the plausibility of the assumptions concerning the mechanism that generated the missing data.

To this purpose, three well established general frameworks exist, distinguished by their approach in factoring the joint distribution of the outcome and missingness processes to conditionally independent components. In a pattern mixture model (PMM), the joint distribution is expressed as the product of the marginal distribution of the missingness process and the conditional distribution of the outcome given the missingness process. Contrariwise, in a selection model (SeM), the joint distribution is expressed as the product of the marginal distribution of the outcome and the conditional distribution of the missingness process given the outcome. Alternatively, in a shared-parameter model (SPM), the outcome and the missingness processes are assumed to be driven by a set of unobserved variables, the so-called random effects. The use of latent variables gives rise to the conditional independence assumption under which the outcome and the missingness processes are independent, given the set of random effects.

The SeM framework naturally encompasses MAR while Molenberghs et al. (1998) provided a characterization of MAR under the PMM setting, specifically for the case of longitudinal data with dropout. As a result, sensitivity analysis is well-developed in the SeM and the PMM frameworks (Molenberghs and Vereke, 2005; Molenberghs and Kenward, 2007), whereas the SPM has received less attention. More specifically, Creemers et al. (2011) provided a characterization of MAR for the SPM case, by introducing the general SPM (GSPM) and then proposed a sensitivity analysis approach utilizing this framework (Creemers et al., 2010). Later, Njagi et al. (2014) proposed a characterization of MAR for the case of shared-parameter models for longitudinal and time-to-event data under the GSPM.

In this paper, we introduce an alternative characterization of MAR in SPMs for longitudinal and time-to-event data by exploiting the conditional independence assumption and utilizing the censoring distribution. By doing so, we cover a wide range of assumptions for the missing data mechanism on the subject-specific level. Previous work on the topic focused on the GSPM, which includes potentially a broad set of random effects. Here we focus on the standard SPM (Wu and Carroll, 1988; Wu and Bailey, 1988,9; Hogan and Laird, 1997,9), a particular case of the GSPM, that is most commonly encountered in practice. The benefit of doing so is that our approach is intuitively appealing and can be easily implemented in existing software. Furthermore, it allows for subject-specific perturbations of the censoring distribution, concerning the missing data generating mechanism, whereas in PMMs and SeMs dropout is considered MNAR strictly.

The rest of the paper is structured as follows. In Section 2, we propose an alternative characterization of MAR for the SPM model. In Section 3, we discuss its estimation under the Bayesian framework. In Section 4, we illustrate with an application how

the SPM can be used for sensitivity analysis with existing software. In Section 5, we present a simulation study to assess the behavior of the proposed sensitivity analysis under different scenarios with respect to the amount of MAR and MNAR missingness. Discussion follows in Section 6.

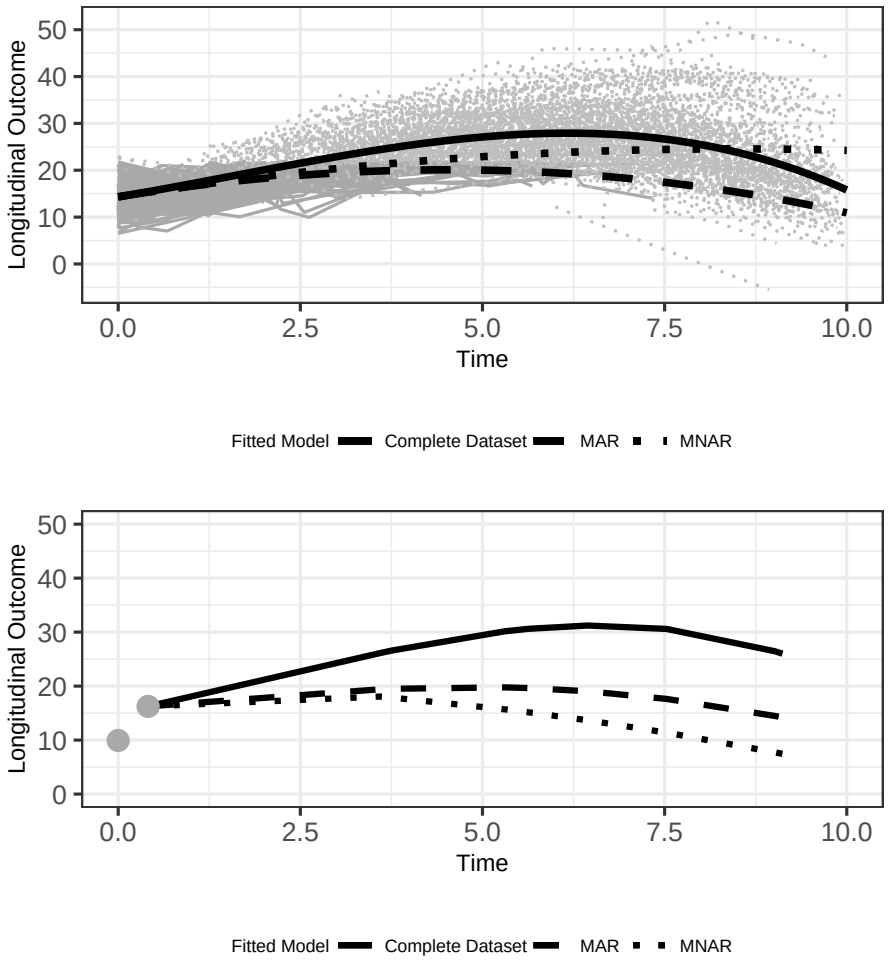

Figure 6.1: Illustrative Example: (Upper panel) observed longitudinal trajectories (black lines), missing longitudinal trajectories (gray) and fitted models using complete data and using incomplete data under MAR and MNAR. (Lower Panel) observed longitudinal measurements of a new subject and subject-specific predictions based on the three different models.

6.2 A Characterization of MAR for the Shared Parameter Model

For each subject i out of a sample of size N from the target population, let $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{in_i})^\top$ denote the vector of size $n_i \times 1$ of planned repeated measurements, with y_{ij} being the value of the longitudinal outcome at time point t_{ij}, $j = 1, \ldots, n_i$, out of a set of $\mathcal{J}$ planned measurement occasions. The vector of planned repeated measurements may further be partitioned into two vectors: $\boldsymbol{y}_i = (\boldsymbol{y}_i^o, \boldsymbol{y}_i^m)$, of observed and missing repeated measurements, respectively. Next, let T_i^* be the true dropout time, C_i the censoring time and $T_i = \min(T_i^*, C_i)$ the observed dropout or censoring time. Furthermore, let $\boldsymbol{\theta}$ and $\psi = \left\{\psi^{T^*}, \psi^C\right\}$ be the vectors of parameters for the measurement, dropout and censoring processes respectively. Under these settings, the vectors of observed and missing repeated measurements can be expressed as $\boldsymbol{y}_i^o = \{y_{ij}; t_{ij} < T_i, \forall j \in \mathcal{J}\}$ and $\boldsymbol{y}_i^m = \{y_{ij}; t_{ij} > T_i, \forall j \in \mathcal{J}\}$. In other words, the vector of observed measurements includes all the observations which were recorded before neither dropout nor censoring occurred. On the contrary, the vector of missing measurements includes the observations which were not recorded due to either dropout or censoring. Note that we regard dropout and censoring as different causes of missingness, a distinction that we will exploit to achieve MAR characterization of SPMs.

What differentiates dropout from censoring in such a setting, pertains to the causes of leaving the study. As an example, consider the case of a study, with death as the main outcome and where a longitudinal outcome is also planned to be recorded for a time period. The planned longitudinal outcome measurements will be missing for both the subjects that died during the study but also for subjects that dropped out from the study due to any other reason (e.g. they moved to another country). Depending on the information available and the research setting, one may define death as the only cause of dropout and assume all the other reasons are non-informative and therefore record them as censoring. In the case where no information is available for the cause of dropout, then one would consider only those who completed the study as censored.

Under this framework and with covariate vectors suppressed from the notation, the shared parameter model can be specified as:

$$p\left(\boldsymbol{y}_i^o, \boldsymbol{y}_i^m, T_i^*, C_i; \boldsymbol{\theta}, \psi\right) = \int p\left(\boldsymbol{y}_i^o, \boldsymbol{y}_i^m, T_i^*, C_i, \boldsymbol{b}_i; \boldsymbol{\theta}, \psi, \boldsymbol{D}\right) d\boldsymbol{b}_i, \tag{6.1}$$

where $\boldsymbol{b}_i \sim \mathcal{N}(0, \boldsymbol{D})$ is a vector of shared random-effects assumed to induce the dependence between the processes involved in the joint density of the SPM. That is by

conditioning on the random effects the joint density of the SPM can be decomposed to the following product of conditional densities

$$\int p\left(T_i^*, \mid \boldsymbol{b}_i; \boldsymbol{\psi}^{T^*}\right) p\left(C_i \mid \boldsymbol{y}_i^o, \boldsymbol{y}_i^m, \boldsymbol{b}_i; \boldsymbol{\psi}^C\right) p\left(\boldsymbol{y}_i^o, \boldsymbol{y}_i^m \mid \boldsymbol{b}_i; \boldsymbol{\theta}\right) p\left(\boldsymbol{b}_i; \boldsymbol{D}\right) d\boldsymbol{b}_i, \tag{6.2}$$

which brings to light the core assumption of conditional independence in SPMs, under which the measurement and the dropout processes are independent conditional on the random effects. Note that the conditional independence assumption commonly does not imply independence of the censoring and the measurement processes. More specifically, in the general case censoring is allowed to depend on the measurement process. Note that the conditional independence assumption may be extended to allow for independence of the censoring and the measurement processes as well. Finally, the assumption of non-informative censoring, which we discuss later, may be used to relax both these assumptions and allow for independence between the censoring and measurement processes.

Model (6.2) is a conventional SPM for longitudinal and time-to-event data. The term conventional here is in reference to the GSPM framework in the sense that a single common underlying random-effects structure is used, instead of multiple random-effects structures. Another important note for model (6.1) is that by definition, it allows for dependence between the dropout process T_i^* and the censoring process C_i. When there is no information from the data on the joint distribution of (T_i^*, C_i), the assumption of non-informative censoring may be utilized (Tsiatis, 1975) to further simplify the model. Under both the assumptions of non-informative censoring and conditional independence, (6.2) further simplifies to

$$\int p\left(T_i^* \mid \boldsymbol{b}_i; \boldsymbol{\psi}^{T^*}\right) p\left(C_i \mid \boldsymbol{y}_i^o; \boldsymbol{\psi}^C\right) p\left(\boldsymbol{y}_i^o, \boldsymbol{y}_i^m \mid \boldsymbol{b}_i; \boldsymbol{\theta}\right) p\left(\boldsymbol{b}_i; \boldsymbol{D}\right) d\boldsymbol{b}_i, \tag{6.3}$$

which, in other words, means that the censoring process no longer depends on the random-effects and the missing observations. It should be noted that the assumption of non-informative censoring while commonly applicable, may not always be a reasonable choice depending on the setting under study. Nevertheless, it is a necessary condition for achieving an MAR characterization in SPMs without introducing more latent variables as in Creemers et al. (2011).

Note that by definition a subject may either be classified as a dropout or as censored and never both. This means that the decomposition of the joint density as described in (6.3) is, on a subject-specific level, further decomposed to

$$\begin{cases} \int p\left(T_i^* \mid \boldsymbol{b}_i; \boldsymbol{\psi}^{T^*}\right) p\left(\boldsymbol{y}_i^o, \boldsymbol{y}_i^m \mid \boldsymbol{b}_i; \boldsymbol{\theta}\right) p\left(\boldsymbol{b}_i; \boldsymbol{D}\right) d\boldsymbol{b}_i, & i : \text{dropout} \\ \int p\left(C_i \mid \boldsymbol{y}_i^o; \boldsymbol{\psi}^C\right) p\left(\boldsymbol{y}_i^o, \boldsymbol{y}_i^m \mid \boldsymbol{b}_i; \boldsymbol{\theta}\right) p\left(\boldsymbol{b}_i; \boldsymbol{D}\right) d\boldsymbol{b}_i, & i : \text{censored.} \end{cases} \tag{6.4}$$

The decomposition in (6.4) reveals that SPMs encompasses MAR on the subject-specific level. That is for a subject who dropped out from the study, the model is MNAR because both the dropout, $p\left(T_i^* \mid \boldsymbol{b}_i; \boldsymbol{\psi}^{T^*}\right)$, and measurement, $p\left(\boldsymbol{y}_i^o, \boldsymbol{y}_i^m \mid \boldsymbol{b}_i; \boldsymbol{\theta}\right)$ processes depend on the missing observations through the random-effects $\boldsymbol{b}_i$. For a censored subject, though, the model is MAR since the random-effects and the missing observations appear only in factor $p\left(\boldsymbol{y}_i^o, \boldsymbol{y}_i^m \mid \boldsymbol{b}_i; \boldsymbol{\theta}\right)$. Moreover, due to ignorability, the joint density reduces to the density of the observed measurement process $\int p\left(\boldsymbol{y}_i^o \mid \boldsymbol{b}_i; \boldsymbol{\theta}\right) p\left(\boldsymbol{b}_i\right) d\boldsymbol{b}_i$. Ignorability is a key consequence of MAR in SPMs. It entails that there are not any information to be gained from either the dropout or the censoring processes with respect to the measurement generating process and therefore MAR can be explored within the SPM framework. This is achieved without broadening the definition of the conventional SPM by introducing additional random-effects. Furthermore, the fact that MAR in SPMs is achieved on the subject-specific level allows for the potential of a subject- or cause-specific MAR sensitivity analysis. This is in contrast to the SeM and PMM frameworks which are MNAR or MAR for all the individuals under study. Finally, as we are going to illustrate in the upcoming Section, this decomposition allows exploiting existing software for its estimation.

6.3 Estimation of SPM under MAR

Estimation of the SPM under MAR follows from the conventional joint model for longitudinal and time-to-event data, which exploits the decomposition of the full joint-likelihood function to conditional independent components given the random effects. More specifically the likelihood contribution of the i^{th} subject is given by:

$$\int p\left(T_i^* \mid \boldsymbol{y}_i, \boldsymbol{b}_i; \boldsymbol{\psi}^{T^*}\right) p\left(C_i \mid \boldsymbol{y}_i; \boldsymbol{\psi}^C\right) p\left(\boldsymbol{y}_i \mid \boldsymbol{b}_i; \boldsymbol{\theta}\right) p\left(\boldsymbol{b}_i; \boldsymbol{D}\right) d\boldsymbol{b}_i, \tag{6.5}$$

The form of the conditional densities depends then on the model specification of each component. Let the risk of dropout for subject i depend on a function of the true underlying subject-specific trajectory $\boldsymbol{\eta}_i\left(t\right)$. Then the instantaneous risk of dropout can be expressed as

$$h_i\left(t \mid \mathcal{H}_i\left(t\right), \boldsymbol{b}_i, \boldsymbol{w}_i\right) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \Pr\left\{t \leq T_i^* < t + \Delta t \mid T_i^* \geq t, \mathcal{H}_i\left(t\right), \boldsymbol{w}_i\right\}$$
$$= h_0\left(t\right) \exp\left[\boldsymbol{\gamma}^\top \boldsymbol{w}_i + f\left\{\eta_i\left(t\right), \boldsymbol{b}_i, \boldsymbol{\alpha}\right\}\right], \quad t > 0, \tag{6.6}$$

where $\mathcal{H}_i\left(t\right) = \left\{\eta_i\left(s\right), 0 \leq s \leq t\right\}$ denotes the history of the true underlying longitudinal process up to time t (and therefore depends on the random-effects as well), $h_0\left(\cdot\right)$ is the baseline hazard function and $\boldsymbol{w}_i$ is a vector of baseline covariates with corresponding regression coefficients $\boldsymbol{\gamma}$. The association between the features of the longitudinal measurement process, $p\left(\boldsymbol{y}_i^o, \boldsymbol{y}_i^m \mid \boldsymbol{b}_i; \boldsymbol{\theta}\right)$, and the instantaneous risk for dropout, $h_i\left(t \mid \mathcal{H}_i\left(t\right), \boldsymbol{b}_i, \boldsymbol{w}_i\right)$, is quantified by parameter vector $\boldsymbol{\alpha}$. Function $f\left(\cdot\right)$ may take various forms but for the remainder of the paper we will assume that $f\left(\eta_i\left(t\right), \boldsymbol{b}_i, \alpha\right) = \alpha\eta_i\left(t\right)$, the so-called current value association, which postulates that the risk of dropout at time t depends only on the current value of the true underlying subject-specific trajectory at the same time $\boldsymbol{\eta}_i\left(t\right)$. Let $\boldsymbol{u}_i$ be a vector of baseline covariates with corresponding regression coefficients $\boldsymbol{\gamma}_C$ and $g\left(\boldsymbol{y}_i^o\right)$ a function of the observed measurements. Then, analogously, the instantaneous risk to be censored can be expressed as

$$\lambda_i\left(t \mid \boldsymbol{y}_i^o, \boldsymbol{u}_i\right) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \Pr\left\{t \leq C_i < t + \Delta t \mid C_i \geq t, \boldsymbol{y}_i^o, \boldsymbol{u}_i\right\}$$
$$= h_0\left(t\right) \exp\left[\boldsymbol{\gamma}_C^\top \boldsymbol{u}_i + g\left(\boldsymbol{y}_i^o\right)\right], \quad t > 0, \tag{6.7}$$

since based on the assumption of non-informative censoring, it cannot depend on the missing observations and the random-effects.

The probability of not dropping out from the study can then be obtained by:

$$S_i\left(t \mid \mathcal{H}_i\left(t\right), \boldsymbol{b}_i, \boldsymbol{w}_i\right) = \Pr\left(T_i^* > t \mid \mathcal{H}_i\left(t\right), \boldsymbol{w}_i\right)$$
$$= \exp\left(-\int_0^t h_0\left(s\right) \exp\left\{\boldsymbol{\gamma}^\top \boldsymbol{w}_i + \alpha\eta_i\left(s\right)\right\} ds\right), \tag{6.8}$$

which means that unlike the instantaneous risk of dropout, the probability of remaining in the study depends on the whole true underlying subject-specific trajectory up to time t. Let $\mathcal{Y}_i^o\left(t\right) = \left\{\boldsymbol{y}_i^o\left(l\right), 0 \leq l \leq t\right\}$ denote the history of the observed repeated measurements.Then, analogously, the probability of not being censored can be expressed as

$$V_i\left(t \mid \mathcal{Y}_i^o(t), \boldsymbol{u}_i\right) = \Pr\left(C_i > t \mid \mathcal{Y}_i^o(t), \boldsymbol{u}_i\right) = \exp\left(-\int_0^t \lambda_0(s) \exp\left\{\boldsymbol{\gamma}_C^\top \boldsymbol{u}_i + g\left(\boldsymbol{y}_i^o\right)\right\} ds\right), \tag{6.9}$$

omitting the conditioning on missing values and the random-effects due to the non-informative censoring assumption. Finally, to distinguish between censored subjects and dropouts, let δ_i be an indicator variable which takes the value 1 if a subject dropped out from the study, that is $T_i^* < C_i$ and 0 otherwise.

Under these assumptions and suppressing covariates $\boldsymbol{w}_i$ and $\boldsymbol{u}_i$ from the notation, the conditional density of the dropout process can be expressed as

$$p\left(T_i^* \mid \boldsymbol{b}_i; \boldsymbol{\psi}^{T^*}\right) = \begin{cases} h_i\left(t \mid \mathcal{H}_i(t), \boldsymbol{b}_i\right) S_i\left(t \mid \mathcal{H}_i(t), \boldsymbol{b}_i\right), & \delta_i = 1 \\ S_i\left(t \mid \mathcal{H}_i(t), \boldsymbol{b}_i\right), & \delta_i = 0, \end{cases} \tag{6.10}$$

while the conditional density of the censoring process is given by

$$p\left(C_i \mid \boldsymbol{y}_i^o; \boldsymbol{\psi}^C\right) = \begin{cases} \lambda_i(t) V_i(t), & \delta_i = 0 \\ V_i(t), & \delta_i = 1, \end{cases} \tag{6.11}$$

Depending on the dropout and censoring status of the i^{th} individual, the joint-likelihood in (6.5) can be written as:

$$\begin{cases} V_i(t) \int h_i\left(t \mid \mathcal{H}_i(t), \boldsymbol{b}_i\right) S_i\left(t \mid \mathcal{H}_i(t), \boldsymbol{b}_i\right) p\left(\boldsymbol{y}_i \mid \boldsymbol{b}_i; \boldsymbol{\theta}\right) p\left(\boldsymbol{b}_i; \boldsymbol{D}\right) d\boldsymbol{b}_i, & \delta_i = 1 \\ V_i(t) \lambda_i(t) \int S_i\left(t \mid \mathcal{H}_i(t), \boldsymbol{b}_i\right) p\left(\boldsymbol{y}_i \mid \boldsymbol{b}_i; \boldsymbol{\theta}\right) p\left(\boldsymbol{b}_i; \boldsymbol{D}\right) d\boldsymbol{b}_i, & \delta_i = 0. \end{cases} \tag{6.12}$$

Under (6.12) the model is MNAR for both uncensored and censored subjects. The latter is because even for censored subjects the term $S_i\left(t \mid \mathcal{H}_i(t), \boldsymbol{b}_i\right)$ still appears in the likelihood decomposition. However, for the case that a subjects is censored, this term equals to 1 reducing (6.12) to

$$V_i(t) \lambda_i(t) \int p\left(\boldsymbol{y}_i \mid \boldsymbol{b}_i; \boldsymbol{\theta}\right) p\left(\boldsymbol{b}_i; \boldsymbol{D}\right) d\boldsymbol{b}_i, \tag{6.13}$$

which due to ignorability further reduces to

$$\int p\left(\boldsymbol{y}_i \mid \boldsymbol{b}_i; \boldsymbol{\theta}\right) p\left(\boldsymbol{b}_i; \boldsymbol{D}\right) d\boldsymbol{b}_i, \tag{6.14}$$

an MAR model for the observed longitudinal measurements. The transition from equation (6.13) to (6.14) holds due to the fact that the instantaneous risk to be censored $\lambda_i(t)$ and the probability of not being censored $V_i(t)$ are ignorable, since they strictly depend on the history of the observed repeated measurements and not the missing ones (see equations (6.7) and (6.9)). Then when a subject is censored, instead of dropout, he/she is considered MAR allowing for straightforward sensitivity analysis since it can be achieved with existing software just by doing a data manipulation step and changing all the dropout indicators to zero.

As far as the true underlying longitudinal trajectory is concerned, let $\boldsymbol{y}_i(t)$ be the contaminated, with measurement error $\boldsymbol{\epsilon}_i$, observed values of the true longitudinal trajectory, $\boldsymbol{\eta}_i$. Then the longitudinal process can be expressed using a mixed-effects model defined as

$$\begin{cases} \boldsymbol{y}_i(t) = \boldsymbol{\eta}_i + \boldsymbol{\epsilon}_i(t), \\ \boldsymbol{\eta}_i(t) = E\left[\boldsymbol{y}_i(t) \mid \boldsymbol{b}_i\right] = \boldsymbol{x}_i^{\top}(t)\boldsymbol{\beta} + \boldsymbol{z}_i^{\top}(t)\boldsymbol{b}_i, \\ \boldsymbol{\epsilon}_i \sim \mathcal{N}\left(\boldsymbol{0}, \boldsymbol{\Sigma}_i\right), \\ \boldsymbol{b}_i \sim \mathcal{N}\left(\boldsymbol{0}, \boldsymbol{D}\right), \\ \boldsymbol{b}_i \perp\!\!\!\perp \boldsymbol{\epsilon}_i, \end{cases} \tag{6.15}$$

where the true longitudinal trajectory $\boldsymbol{\eta}_i$ is assumed to be a function of some general time-dependent design vectors $\boldsymbol{x}_i^{\top}(t)$ and $\boldsymbol{z}_i^{\top}(t)$ which are associated with a set of fixed-effects $\boldsymbol{\beta}$ and a set of subject-specific random effects $\boldsymbol{b}_i$ respectively. The errors are typically assumed to be normally distributed with mean zero and a general variance-covariance matrix $\boldsymbol{\Sigma}_i$ which is commonly specified as $\boldsymbol{I} \times \sigma^2$. The random effects $\boldsymbol{b}_i$ are also assumed to be normally distributed with mean zero and a general variance-covariance matrix $\boldsymbol{D}$, and they are also assumed to be independent from the errors.

The conditional density for the longitudinal responses is then given by

$$p\left(\boldsymbol{y}_i \mid \boldsymbol{b}_i; \boldsymbol{\theta}\right) = \left(2\pi\sigma^2\right)^{n_i/2} \exp\left\{-||\boldsymbol{y}_i - \boldsymbol{x}_i\boldsymbol{\beta} - \boldsymbol{z}_i\boldsymbol{b}_i||^2/2\sigma^2\right\},$$

with $||x|| = \sum_i \left\{x_i^2\right\}^{1/2}$ denoting the Euclidean vector norm. The density of the random-effects is, then, given by

$$p\left(\boldsymbol{b}_i; \boldsymbol{D}\right) = (2\pi)^{-q/2} \det\left(\boldsymbol{D}\right)^{-1/2} \exp\left(-\boldsymbol{b}_i^{\top}\boldsymbol{D}^{-1}\boldsymbol{b}_i/2\right). \tag{6.16}$$

Under the Bayesian framework, estimation of the parameters of the SPM is achieved using Markov chain Monte Carlo (MCMC) algorithms. Let $\boldsymbol{\Theta} = \left\{\boldsymbol{\beta}, \boldsymbol{\Sigma}_i, \boldsymbol{\psi}^{T^*}, \boldsymbol{\psi}^C, \boldsymbol{D}\right\}$ be a vector of unknown parameters, then the posterior distribution is proportional to:

$$p\left(\boldsymbol{\Theta}, \boldsymbol{b}\right) \propto \prod_i p\left(\boldsymbol{y}_i \mid \boldsymbol{b}_i; \boldsymbol{\theta}\right) p\left(T_i^* \mid \boldsymbol{b}_i; \boldsymbol{\psi}^{T^*}\right) p\left(C_i \mid \boldsymbol{y}_i^o; \boldsymbol{\psi}^C\right) p\left(\boldsymbol{b}_i; \boldsymbol{D}\right) p\left(\boldsymbol{\Theta}\right). \tag{6.17}$$

The computation of the integrals involved in $p\left(T_i^* \mid \boldsymbol{b}_i; \boldsymbol{\psi}^{T^*}\right)$ and $p\left(C_i \mid \boldsymbol{y}_i^o; \boldsymbol{\psi}^C\right)$ is achieved using standard Gauss-Kronrod and Gauss-Legendre quadrature rules. For the parameters in $\boldsymbol{\Theta}$ standard prior distributions are used. More specifically, for the vector of regression coefficients $\boldsymbol{\beta}$ of the longitudinal submodel, for the vector of regression coefficients of the dropout submodel $\boldsymbol{\gamma}$ we use independent univariate diffuse normal priors. For the covariance matrix of the random-effects and inverse Wishart prior is used while for the variance of the error terms σ^2 an inverse-Gamma prior is used. For the baseline hazard we use a B-spline approximation expressed as:

$$h_0\left(t\right) = \exp\left(\gamma_{h_0,0} + \sum_{q=1}^{Q} \gamma_{h_0,q} B_q\left(t, k\right)\right), \tag{6.18}$$

with $B_q\left(t, k\right)$ denoting the q^{th} basis function of a B-spline with knots $k_1, \ldots, k_q$ and $\boldsymbol{\gamma}_{h_0}$ the vector of coefficient corresponding to the spline terms. To achieve the B-spline approximation to the baseline hazard under the Bayesian framework the following prior is specified:

$$p\left(\boldsymbol{\gamma}_{h_0} \mid \tau_h\right) \propto \tau_h^{\rho(K)/2} \exp\left(-\frac{\tau_h}{2} \boldsymbol{\gamma}_{h_0}^\top \boldsymbol{K} \boldsymbol{\gamma}_{h_0}\right). \tag{6.19}$$

In the above specification, τ_h is a smoothing parameter with a Gamma $(1, 0.005)$ hyper-prior, $\boldsymbol{K}$ is a penalty matrix and $\rho(\boldsymbol{K})$ is its rank. For more details the reader is referred to Rizopoulos (2016).

6.4 The HIV CD4 Data: A Sensitivity Analysis

As an illustrative case study, we will use data from a randomized clinical trial designed to compare the efficacy and safety of Didanosine (ddI) versus Zalcitabine (ddC) in HIV Patients (Abrams et al., 1994; Goldman et al., 1996). In total, 467 advanced HIV patients were included in the study and were randomized to ddI (230; 49.2%) or ddC (237; 50.8%). The primary goal of the trial was to compare survival between

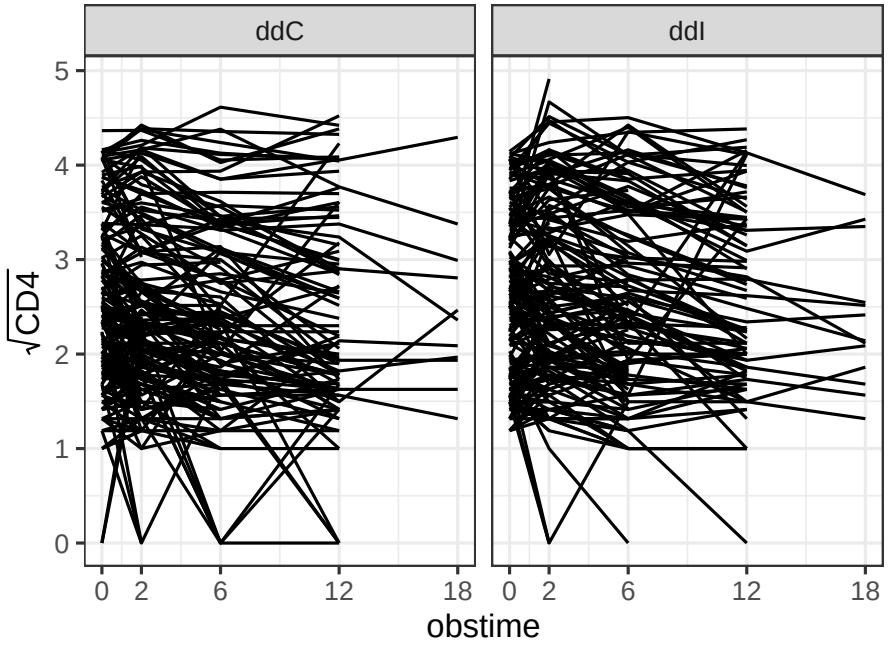

Figure 6.2: Observed longitudinal trajectories of $\sqrt{CD4}$ by treatment arm.

the two treatment arms while a secondary goal of the study was to investigate the association between CD4 cell count (a marker for the strength of the immune system) and the risk of death. For the latter goal, measurements of the CD4 cell count were recorded at scheduled visits at baseline, 2, 6, 12 and 18 months after the start of the study. Figure 6.2 shows the observed trajectories of the square root CD4 cell count. Until the end of the study, 184 (39.4%) died while only 1405 (60.1%) out of the 2335 planned visits were recorded, leading to 930 (39.9%) missing observations. In total, only 24(5.1%) patients were present for all five planned visits, 382(81.8%) dropped out from the study due to death or other reasons, and 61(13.0%) missed at least one planned visit without dropping out though from the remainder of the study. Table 6.1 shows the number of available measurements per dropout pattern per treatment arm, excluding the 61 cases of intermittent missingness.

We observe that dropout rates are similar between the treatment groups, the closer we look at the beginning of the study, but they slightly start to deviate towards the end of the study. Nevertheless, it is not possible to distinguish if the missing data are missing completely at random, missing at random or missing not at random. This

Table 6.1: Number of observed CD4 measurements per dropout pattern and per treatment arm. 'OXXXX' denotes a dropout pattern where only the first planned measurement was obtained (denoted with 'O') while the rest were not (denoted with 'X') whereas 'OOOOO' denotes a pattern for a completer (all planned measurements were obtained).

	ddC		ddI	
Dropout Pattern	N	%	N	%
OXXXX	29	14.4%	32	15.6%
OOXXX	35	17.4%	37	18.0%
OOOXX	41	20.4%	47	22.9%
OOOOX	85	42.3%	76	37.1%
OOOOO	11	5.5%	13	6.3%
Total	201	100%	205	100%

highlights the importance of being able to explore different scenarios with respect to missing data in the SPM framework via sensitivity analysis.

To model the longitudinal measurements of CD4 cell count we used a linear mixed-effects model with an interaction between time and ddI in the fixed-effects structure. We also used random-intercepts and random-slopes to allow for subject-specific deviations. The square root of the CD4 cell count was used in order to meet the normality and homoscedasticity assumptions of the model, which is defined as:

$$\sqrt{y_{ij}^o} = (\beta_0 + b_{i0}) + (\beta_1 + b_{i1})\, t_{ij} + \beta_2 \text{ddI} + \beta_3 \left(t_{ij} \times \text{ddI}\right) + \epsilon_{ij}, \tag{6.20}$$

where $\boldsymbol{b}_i \sim \mathcal{N}(0, \boldsymbol{D})$, independent of $\epsilon_{ij} \sim \mathcal{N}\left(0, \sigma^2\right)$.

For the time-to-dropout T_i we assumed a relative risk model of the form:

$$h_i\left(t \mid \mathcal{H}_i(t), \boldsymbol{b}_i\right) = h_0(t) \exp\left[\gamma_1 \text{ddI} + \boldsymbol{\alpha} \eta_i(t)\right], \tag{6.21}$$

where the baseline hazard $h_0(t)$ was approximated using penalized B-splines with 3 internal knots placed at quantiles of the observed event times (Eilers and Marx, 1996). The number of knots was kept small due to the low number of time-points per subject.

These two sub-models define the MNAR joint-model for the CD4 cell count and the dropout. We then assumed two different scenarios with respect to the cause

of dropout. Under Scenario 1, we assume no information concerning the cause of dropout, which means we treat dropout due to death or any other reason the same. Contrariwise in Scenario 2, we consider as dropouts only the subjects who died during the study. Finally, to achieve MAR, we assume that all the subjects are censored instead of dropping out. As shown in Section 3, this reduces the model to the mixed-effects sub-model, which is considered MAR. All models were fitted using **R** version **3.6.1** and package **JMbayes** (Rizopoulos, 2016). Table 6.2 shows the parameter estimates after fitting all three different models.

Table 6.2: Posterior means (standard errors) for an MAR and MNAR analyses.

		Scenario 1	Scenario 2	
Effect	Parameter	MNAR	MNAR	MAR
Intercept	β_0	2.4671 (0.0640)	2.4680 (0.0629)	2.4423 (0.0647)
Time	β_1	-0.0655 (0.1701)	-0.0674 (0.1715)	-0.0399 (0.0050)
ddI	β_2	0.1038 (0.0908)	0.1067 (0.0879)	0.1188 (0.0946)
ddI $\times$ Time	β_3	0.0219 (0.2332)	0.0290 (0.2376)	0.0089 (0.0070)

While the results of all the models are close, there are quantitative differences which indicate that the findings might not be stable. This is especially the case when comparing the MAR model with both MNAR models. In Figures 6.3 and 6.4 the posterior means of the random-effects estimated from the MAR model against the respective estimated random-effects from each MNAR model are shown. Similarly, in Figure 6.5 the posterior means of the random-effects estimated from the MNAR model under the 1st Scenario are plotted against the respective estimated random-effects from the MNAR model under the 2nd Scenario. These plots give insight into the differences between the three models. We see that especially for the slope components, the differences in the random-effects' estimates are more intense between the MAR model and the MNAR models under both Scenarios. These differences become weaker when we compare the MNAR model under the 1st Scenario and the MANR model under the 2nd Scenario, which utilizes the information and distinguishes between dropout due to an event and dropout due to any other reason. This highlights the importance of sensitivity analysis in this context since the differences in the random-effects estimates would translate to differences in the subject-specific predictions based on each of the models. The results are stable under different MNAR scenarios but substantial differences are observed when using an MAR model.

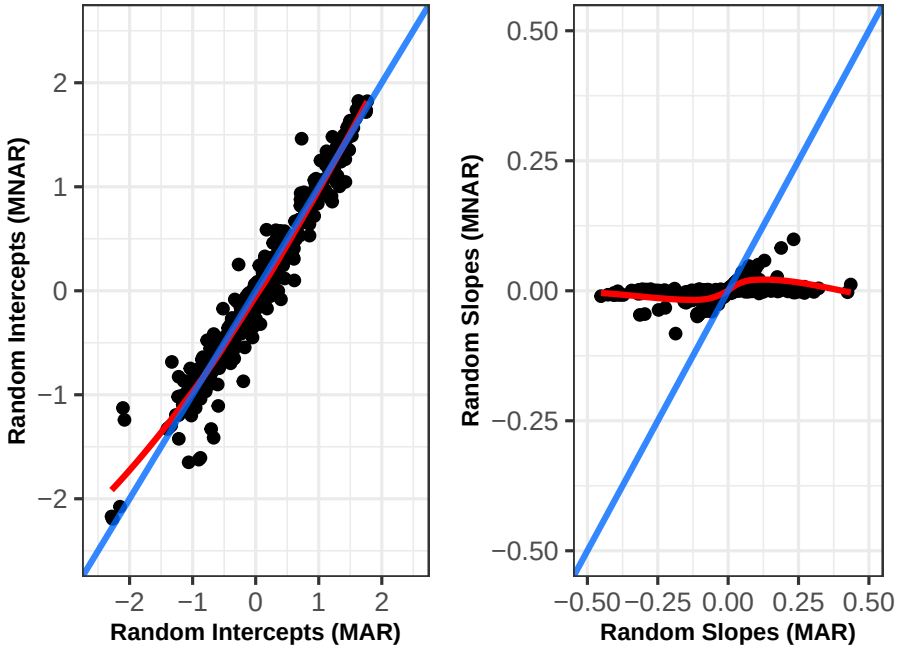

Figure 6.3: Scenario 1: Posterior means of the random-effects under MNAR and MAR.

6.5 Simulation Study

We conducted a simulation study to evaluate the performance of the following three models (used in the analysis of the HIV CD4 data):

- Model 1) all dropout cases considered MNAR (as the MNAR model used under Scenario 1 MNAR in the analysis of the HIV dataset),
- Model 2) all dropout cases considered MAR,
- Model 3) dropout cases considered MNAR if dropout due to the event or MAR otherwise (as the MNAR model used under Scenario 1 MNAR in the analysis of the HIV dataset),

under different scenarios for the amount of MAR and MNAR dropouts. More specifically, assuming a total dropout rate of 50%, we considered three different scenarios concerning the amount of dropout type: 1) 5% MAR dropouts versus 45%

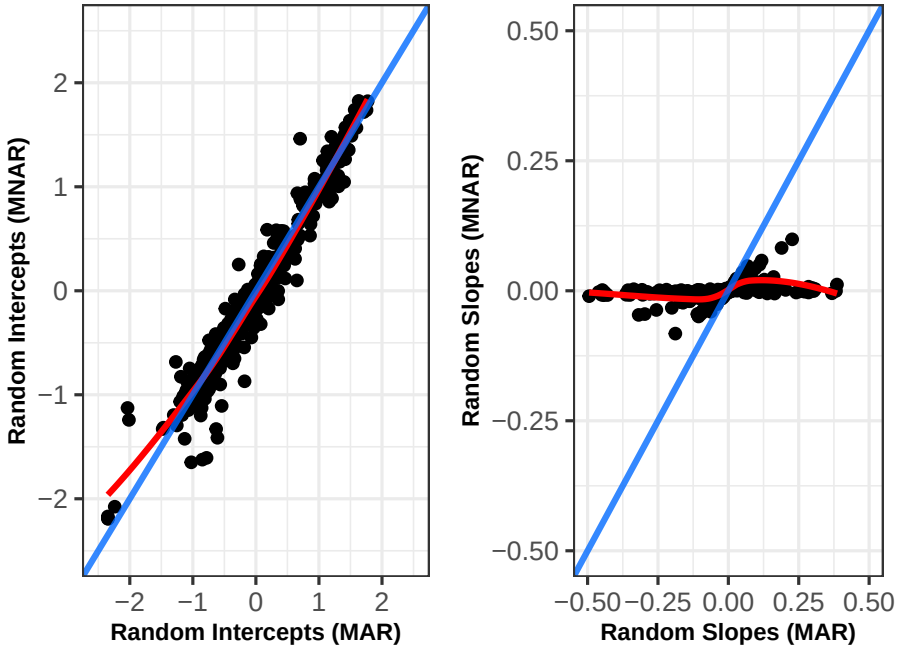

Figure 6.4: Scenario 2: Posterior means of the random-effects under MNAR and MAR.

MNAR dropouts, 2) 25% MAR dropouts versus 25% MNAR dropouts and 3) 45% MAR dropouts versus 5% MNAR dropouts. Our goal is to investigate the behavior between these models under each scenario.

We assumed 600 subjects and then randomly selected follow-up visits, t_{ij} from a uniform distribution between 0 and 10. We set the total number of measurements per subject to 10, but the final number of measurements may vary depending on whether a subject was a dropout or not. The type of dropout was then determined as follows: if the value of the longitudinal trajectory of a subject exceeded a pre-specified value, then the next time point was selected as the candidate MAR dropout time, which was then compared to the candidate MNAR dropout time. We simulated the MNAR dropout time using a joint model. Depending on which of these time points occurred sooner (if occurred), the subject was classified as MAR dropout, MNAR dropout, or completer. The threshold value for MAR varied between the three scenarios in order to attain the target percentages of dropout per type.

For the continuous longitudinal outcome, the data were simulated from a linear

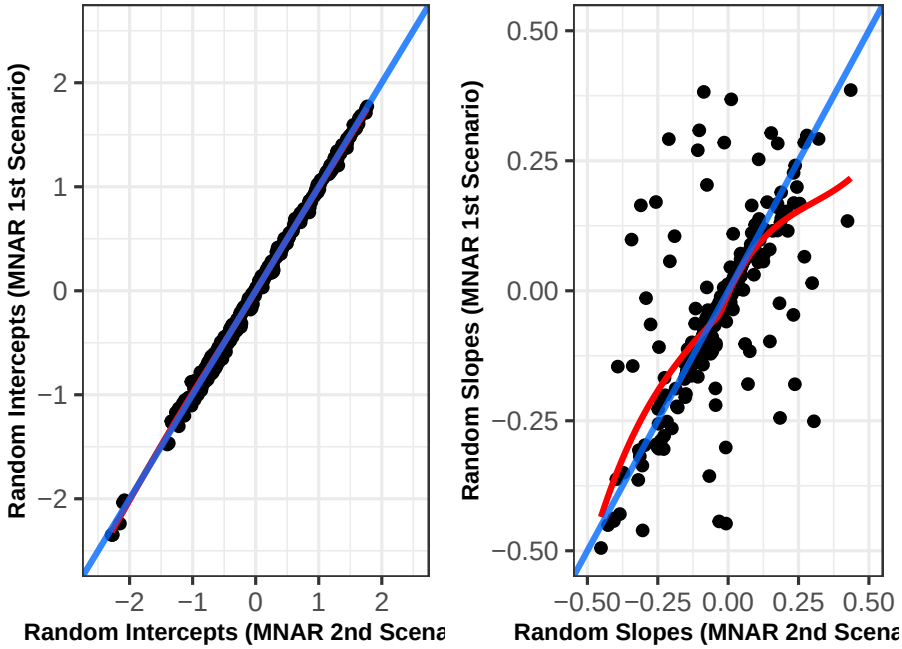

Figure 6.5: Scenario 2: Posterior means of the random-effects under MNAR (1st Scenario) and MNAR (2nd Scenario).

mixed-effects model as follows:

$$y_{ij} = (\beta_0 + b_{i0}) + (\beta_1 + b_{i1})\, t_{ij} + \beta_2 \text{Xcov} + \beta_3\, (\text{Xcov} \times t_{ij}) + \epsilon_{ij}, \tag{6.22}$$

where Xcov is a categorical variable with two groups, $\epsilon_i(t) \sim \mathcal{N}\left(0, \sigma^2 I_{n_i}\right)$ and $\boldsymbol{b} \sim \mathcal{N}(0, \boldsymbol{D})$. The parameter values used are: $\beta_0 = 2.47$, $\beta_1 = -0.067$, $\beta_2 = 0.107$, $\beta_3 = 0.029$, $\sigma = 1.431$ and for the variance-covariance matrix of the random effects:

$$\boldsymbol{D} = \begin{bmatrix} d_{11} = 0.765 & d_{12} = -0.038 \\ d_{21} = -0.038 & d_{22} = 4.910 \end{bmatrix}.$$

The values of the parameters were based on the analysis of the HIV CD4 data. For time-to-dropout, we assumed a relative risk model of the form:

$$h_i(t \mid \mathcal{H}_i(t), \boldsymbol{b}_i) = h_0(t) \exp\left[\gamma_1 \text{Xcov} + \boldsymbol{\alpha} \eta_i(t)\right], \tag{6.23}$$

where the baseline hazard was simulated from a Weibull distribution $h0\,(t) = \xi t^{\xi-1}$ with $\xi = 3.765$.

We simulated 500 datasets per scenario, and we fitted each of the three models on each dataset. To assess the behavior of the three models, we calculated the Bias and Root Mean Squared Error (RMSE) for both the regression coefficients and the variance parameter estimates. Table 6.3 summarizes the results for the regression coefficients while Table 6.4 summarizes the results for the variance parameters. The results suggest that when the major cause of missingness is MAR, Models 2 and 3 perform similarly and slightly better than Model 1, as far as the regression coefficients are concerned. As the amount of MNAR dropout increases, the three models seem to have similar performance in the estimation of the regression coefficients with only slight differences.

On the other hand, when looking at the Bias and RMSE for the estimates of variance parameters for the random-effects, there are more distinct differences between the models. More specifically, the results suggest that Models 1 and 3 perform better than Model 2. Moreover, for the variances of the random-effects, under Model 2, the Bias seems to increase as the balance between dropout types moves from MAR dominance to MNAR dominance, whereas the case is the opposite for Model 3. The differences in the variance parameters' estimates of the random-effects between the models also imply that there are considerable differences in the estimated random-effects. The difference in the random-effects means that the subject-specific predictions derived from these models will differ.

6.6 Discussion

In this paper, we have proposed an alternative characterization of MAR for the conventional SPM and proposed its application as a sensitivity analysis tool towards MNAR deviations from the MAR assumption. In doing so, we did not broaden the definition of the SPM by adding additional random-effects structures and hence retaining the computational feasibility of the model.

Furthermore, we argued how the subject-specific nature of the MAR characterization in SPM comes with the advantage of more flexible comparisons regarding the causes of missingness. This is an important advantage over other MNAR frameworks such as the PMM and SeM since it allows distinguishing groups of subjects as MAR or MNAR depending on the information available for the causes of missingness. This was illustrated in the application to the HIV data, where we were able to consider different MNAR scenarios with respect to the information available on the causes of

missingness.

Finally, it should be mentioned that this characterization of MAR in SPMs allows for potential extensions to account for multiple causes of missingness by using for example a multi-state model and/or generalized mixed-effects models for categorical longitudinal outcomes. Dropout is present in most longitudinal studies, and our approach showed how one could perform a sensitivity analysis using standard joint modeling software to explore different causes of missingness. By doing so we do hope that such sensitivity analyses will become standard procedure and will be routinely reported in longitudinal studies with dropout.

Acknowledgements The authors would like to acknowledge support by the Netherlands Organization for Scientific Research VIDI (grant number: 016.146.301). .

Table 6.3: Mean, Bias and RMSE for the regression coefficients from the three models under different scenarios for the amount of dropout caused by each missing data generating mechanism.

	Scenario 1: 5% MNAR - 45% MAR			Scenario 2: 25% MNAR - 25% MAR			Scenario 3: 45% MNAR - 5% MAR		
Parameter	Mean	Bias	RMSE	Mean	Bias	RMSE	Mean	Bias	RMSE
Model 1									
β_0	2.4532	-0.0147	0.0753	2.4667	-0.0012	0.0715	2.4705	0.0025	0.0711
β_1	0.0166	0.0840	0.1477	-0.0088	0.0585	0.1339	-0.0129	0.0544	0.1322
β_2	-0.0091	-0.1158	0.1819	-0.0109	-0.1176	0.1827	-0.0114	-0.1181	0.1821
β_3	0.1056	0.0766	0.1045	0.1045	0.0755	0.1002	0.1048	0.0758	0.0993
Model 2									
β_0	2.4688	0.0009	0.0738	2.4707	0.0028	0.0716	2.4723	0.0043	0.0712
β_1	-0.0136	0.0538	0.1322	-0.0127	0.0547	0.1322	-0.0139	0.0535	0.1319
β_2	-0.0106	-0.1174	0.1808	-0.0107	-0.1174	0.1825	-0.0116	-0.1183	0.1823
β_3	0.1060	0.0770	0.1043	0.1051	0.0761	0.1002	0.1046	0.0756	0.0994
Model 3									
β_0	2.4680	0.0001	0.0738	2.4687	0.0007	0.0717	2.4694	0.0014	0.0713
β_1	-0.0129	0.0545	0.1325	-0.0115	0.0558	0.1328	-0.0121	0.0553	0.1323
β_2	-0.0107	-0.1174	0.1810	-0.0105	-0.1172	0.1827	-0.0114	-0.1181	0.1820
β_3	0.1060	0.0770	0.1042	0.1051	0.0761	0.1005	0.1047	0.0757	0.0995

Table 6.4: Mean, Bias and RMSE for the variance parameters from the three models under different scenarios for the amount of dropout caused by each missing data generating mechanism.

	Scenario 1: 5% MNAR - 45% MAR			Scenario 2: 25% MNAR - 25% MAR			Scenario 3: 45% MNAR - 5% MAR		
Parameter	Mean	Bias	RMSE	Mean	Bias	RMSE	Mean	Bias	RMSE
Model 1									
d_{11}	0.7577	-0.0074	0.0910	0.7397	-0.0254	0.0848	0.7394	-0.0257	0.0824
d_{12}	-0.0514	-0.0134	0.1206	-0.0419	-0.0040	0.1132	-0.0265	0.0115	0.1107
d_{22}	4.7665	-0.1439	0.3248	4.6891	-0.2214	0.3543	4.6768	-0.2337	0.3600
σ	1.4318	0.0003	0.0182	1.4314	0.0000	0.0160	1.4293	-0.0022	0.0157
Model 2									
d_{11}	0.1406	-0.6245	0.6258	0.1381	-0.6270	0.6279	0.1381	-0.6270	0.6278
d_{12}	-0.0222	0.0158	0.0742	-0.0208	0.0171	0.0708	-0.0202	0.0178	0.0694
d_{22}	5.4293	0.5188	0.8668	5.4355	0.5251	0.8401	5.4484	0.5380	0.8477
σ	1.4296	-0.0018	0.0180	1.4302	-0.0013	0.0159	1.4301	-0.0014	0.0155
Model 3									
d_{11}	0.7469	-0.0182	0.0908	0.7427	-0.0224	0.0841	0.7392	-0.0259	0.0829
d_{12}	-0.0271	0.0108	0.1184	-0.0321	0.0059	0.1133	-0.0343	0.0037	0.1098
d_{22}	4.6621	-0.2484	0.3817	4.6715	-0.2389	0.3643	4.6798	-0.2306	0.3576
σ	1.4299	-0.0015	0.0181	1.4300	-0.0014	0.0160	1.4299	-0.0016	0.0156

6.7 References

Abrams, D. I., Goldman, A. I., Launer, C., Korvick, J. A., Neaton, J. D., Crane, L. R., Grodesky, M., Wakefield, S., Muth, K., Kornegay, S., Cohn, D. L., Harris, A., Luskin-Hawk, R., Markowitz, N., Sampson, J. H., Thompson, M., and Deyton, L. (1994). A comparative trial of didanosine or zalcitabine after treatment with zidovudine in patients with human immunodeficiency virus infection. *New England Journal of Medicine*, 330(10):657–662.

Creemers, A., Hens, N., Aerts, M., Molenberghs, G., Verbeke, G., and Kenward, M. G. (2010). A sensitivity analysis for shared-parameter models for incomplete longitudinal outcomes. *Biometrical Journal*, 52(1):111–125.

Creemers, A., Hens, N., Aerts, M., Molenberghs, G., Verbeke, G., and Kenward, M. G. (2011). Generalized shared-parameter models and missingness at random. *Statistical Modelling*, 11(4):279–310.

Eilers, P. and Marx, B. (1996). Flexible smoothing with b-splines and penalties. *Statistical Science*, 11:89–121.

Goldman, A., Carlin, B., Crane, L., Launer, C., Korvick, J., Deyton, L., and D., A. (1996). Response of cd4+ and clinical consequences to treatment using ddi or ddc in patients with advanced hiv infection. *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology*, 11:161–169.

Hogan, J. W. and Laird, N. M. (1997). Mixture models for the joint distribution of repeated measures and event times. *Statistics in Medicine*, 16(3):239–257.

Hogan, J. W. and Laird, N. M. (1998). Increasing efficiency from censored survival data by using random effects to model longitudinal covariates. *Statistical Methods in Medical Research*, 7(1):28–48.

Little, R. J. A. and Rubin, D. B. (2002). *Statistical Analysis with Missing Data*. Wiley, New York.

Molenberghs, G., Beunckens, C., Sotto, C., and Kenward, M. G. (2008). Every missingness not at random model has a missingness at random counterpart with equal fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(2):371–388.

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

Molenberghs, G., Michiels, B., Kenward, M. G., and Diggle, P. J. (1998). Monotone missing data and pattern-mixture models. *Statistica Neerlandica*, 52(2):153–161.

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

Njagi, E. N., Molenberghs, G., Kenward, M. G., Verbeke, G., and Rizopoulos, D. (2014). A characterization of missingness at random in a generalized shared-parameter joint modeling framework for longitudinal and time-to-event data, and sensitivity analysis. *Biometrical Journal*, 56(6):1001–1015.

Rizopoulos, D. (2016). The r package jmbayes for fitting joint models for longitudinal and time-to-event data using mcmc. *Journal of Statistical Software, Articles*, 72(7):1–46.

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63(3):581–592.

Tsiatis, A. A. (1975). A characterization of missingness at random in a generalized shared-parameter joint modeling framework for longitudinal and time-to-event data, and sensitivity analysis. *Proceedings of the National Academy of Sciences*, 72:20–22.

Wu, M. C. and Bailey, K. (1988). Analysing changes in the presence of informative right censoring caused by death and withdrawal. *Statistics in Medicine*, 7(1⊠2):337–346.

Wu, M. C. and Bailey, K. R. (1989). Estimation and comparison of changes in the presence of informative right censoring: Conditional linear model. *Biometrics*, 45(3):939–955.

Wu, M. C. and Carroll, R. J. (1988). Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. *Biometrics*, 44(1):175–188.

Chapter 7

Statistical Primer: How to deal with missing data in scientific research

This chapter is based on:

Papageorgiou, G, Grant, S. W., Takkenberg, J.J.M., and Mokhles, M.M. (2018), Statistical Primer: How to deal with missing data in scientific research *Interactive Cardio-Vascular and Thoracic Surgery*, 27(2): 153–158. doi: https://doi-org.eur.idm.oclc.org/10.1093/icvts/ivy102

Abstract

Missing data are a common challenge encountered in research which can compromise the results of statistical inference when not handled appropriately. This paper aims to introduce basic concepts of missing data to a non-statistical audience, list and compare some of the most popular approaches for handling missing data in practice and provide guidelines and recommendations for dealing with and reporting missing data in scientific research. Complete case analysis and single imputation are simple approaches for handling missing data and are popular in practice, however, in most cases they are not guaranteed to provide valid inferences. Multiple imputation is a robust and general alternative which is appropriate for data missing at random, surpassing the disadvantages of the simpler approaches, but should always be conducted with care. The aforementioned approaches are illustrated and compared in an example application using Cox regression.

7.1 Introduction

Missing data are a common challenge encountered by researchers while undertaking clinical research. It can occur across all types of studies including randomized controlled trials, cohort studies, case–control studies and clinical registries. The optimum approach to missing data is to ensure that strategies are devised to ensure that the amount of missing data in a study is as small as possible. Such strategies are commonly utilized in prospectively designed clinical trials as if statistical assumptions due to missing data are required, then the protection of randomization will be broken down and unbiased estimates of treatment effect will be lost. Strategies to minimize missing data in large multicentre cohort or registry studies may be employed however, data desired for research purposes may often be missing due to the retrospective nature of the study or because the data fall outside the primary purpose of the registry (Bell et al., 2013; Little et al., 2012).

Dealing with missing data may be low on the list of priorities for a researcher when undertaking a study but it is a vital step in data analysis as inappropriate handing of missing data can lead to a variety of problems. These included a loss of statistical power, loss of representation of key subgroups of the cohort, biased or inaccurate estimates of treatment effects and increased complexity of the statistical analysis.

To ensure that missing data are handled appropriately, there are a number of steps to follow: first, taking any necessary steps to complete or reduce the amount of missing data wherever possible; second, understanding the mechanism behind the remaining missing data; third, handling the missing data using appropriate methodology and finally, performing sensitivity analyses where appropriate. Focusing primarily on the framework of missing covariate data in non-randomized studies, this article introduces the concept behind different types of missing data and compares some of the most popular approaches for handling missing data in practice. Guidelines and recommendations for dealing with and reporting missing data in scientific research are also presented along with a simulated exercise on handling missing data.

7.2 Methodology

7.2.1 Missing data mechanisms

Before discussing methods for handling missing data, it is important to review the types of missingness. Commonly, these are classified as missing completely at random (MCAR), missing at random (MAR) and missing not at random (MNAR) (Little and Rubin, 2002). An analysis of missing data patterns across contributing participants or

centres, over time, or between key treatment groups should be performed to establish the mechanisms behind the missing data (Bell et al., 2013).

Missing completely at random

Observations of all subjects are equally likely to be missing. That is, there are no systematic differences between subjects with observed and unobserved values meaning that the observed values can be treated as a random sample of the population. For example, echocardiographic measurements might be missing due to sporadic ultrasound malfunction.

Missing at random

The likelihood of a value to be missing depends on other, observed variables. Hence, any systematic difference between missing and observed values can be attributed to observed data. That is, the relationships observed in the data at hand can be utilized to 'recover' the missing data. For example, missing echocardiographic measurements might be more normal than the observed ones because younger patients are more likely to miss an appointment.

Missing not at random

The likelihood to be missing depends on the (unobserved) value itself, and thus, systematic differences between the missing and the observed values remain, even after accounting for all other available information. In other words, there is extra information associated with the missing data that cannot be recovered by utilizing the relationships observed in the data. For example, missing echocardiographic measurements might be worse than the observed ones because patients with severe valve disease are more likely to miss a clinic visit because they are unable to visit the hospital.

Although there are a few methods proposed to test whether the data are MCAR or MAR, their practical value is dubious (Enders, 2010). On the other hand, test distinguishing between MAR and MNAR always depends on data that are not observed meaning it is not possible to make this distinction based only on the observed data. Therefore, a researcher should always evaluate the plausibility of each missing data mechanism with respect to the method used to analyse the data and importantly on how the data were collected. The missing data mechanism should be regarded as an assumption that either supports an analysis or not rather than as an inherent and identifiable feature of a dataset. If that assumption is false, results may be biased. While under MCAR, most standard statistical tools will lead to valid results, that is

not the case for MAR and MNAR, for which appropriate methods need to be employed. Table 7.1 summarizes the basic differences between the 3 missing data types and lists which of the methods discussed in the following section can be used to draw valid inference with respect to each missing data type.

Table 7.1: Summary of missing data mechanisms.

Missing Data Mechanism	**Related to**	**Not Related to**	**Probability to be missing**	**Valid Analysis**
MCAR		Observed or missing data	Equal for every data point	Complete case analysis, single and multiple imputation
MNAR	Observed data	Missing data	Equal for data points within groups	Multiple imputation
MAR	Missing data		Unequal and unknown	Sensitivity Analysis

7.2.2 Methods for handling missing data

There are various approaches for an incomplete data analysis. Two common approaches encountered in practice are complete case analysis and single imputation. Although these approaches are easily implemented, they may not be statistically valid and can result in bias when the data are not missing completely at random (J. and M.J., 2007; Vach and Blettner, 1991). On the other hand, multiple imputation is a more general approach that overcomes the main disadvantages of the aforementioned approaches when data are missing (completely) at random (van Buuren, 2012; Carpenter and M.J., 2013; Schafer, 1997).

Complete case analysis

The easiest way to deal with missing data is to drop all cases that have one or more values missing in any of the variables required for analysis. Although under MCAR this does not lead to bias of the results, it may result in significant loss of data and associated loss of power (e.g. wider confidence intervals) because the sample size is reduced. The extent of this loss of power is associated with the amount of missing data. If the data are MAR, this approach will lead to biased results. Complete case analysis may be appropriate for missing data related to the primary outcome of the study.

Single imputation

Alternatively, missing values in any variable could be replaced with a single value that is thought to best represent the mechanism of the missing data. This could be the mean of a normally distributed continuous variable, the median/mode of a categorical variable, the predicted value from a regression equation (that is, utilizing the complete observations to predict the values of the missing observations) or the best/worst observation carried forward. There may be cases where the missing risk factor data are believed to be highly likely due to the absence of a risk factor, and in this situation, it may be reasonable to impute the absence of the risk factor.

Although this approach allows the researcher to include all subjects in the analysis, it may lead to biased results. Moreover, the uncertainty of parameter estimates of the imputed variables will not necessarily improve when compared with the complete case analysis because the imputation is not conditional on the values of the outcome variable. How large the induced bias is depends on the variability of the imputed variable and on the proportion of missing values. Single imputation is also invalid under MAR since it does not account for the inter-relationships between the variables of interest. Single imputation may, however, be used to perform sensitivity analyses for missing covariate information or primary outcome data to ensure that the reported results are valid under the worst or best-case scenario.

Multiple imputation

Multiple imputation offers an alternative to overcome the disadvantages of the complete case analysis or single imputation approach. It allows the uncertainty, which is due to missing data, to be appropriately considered and can be thought of in three distinct steps: imputation, analysis and pooling of the results.

At the first step of imputation, multiple copies of the original incomplete dataset are generated. In each dataset, the missing values are replaced by values which are randomly sampled from the predictive distribution of the observed data, conditional on all other variables. The process of sampling induces variation in the imputed values which reflects the uncertainty of those imputed values.

In the analysis step, the model of interest is fitted to each imputed dataset. The results derived from each analysis will differ slightly due to the variability of the imputed values. In the third step, the results are pooled by taking the average of the estimates from the separate analyses to derive the pooled estimate and by applying Rubin's rules, which incorporate the within and between imputation uncertainty, to derive the associated standard errors. More details on Rubin's rules and the formulas that are used to obtain the pooled estimates can be found in Supplementary Material 7.A.

In contrast to the complete case analysis, multiple imputation provides valid results when the data are MAR while avoiding the loss of power due to sample size reduction. However, loss of power may still occur when using multiple imputation if there is high uncertainty in the distributions which are used to impute the missing data. Unlike single imputation, multiple imputation provides valid results if the data are MAR. This is because systematic differences between the missing and the observed values are due to information already present in the data at hand, and this is considered in the predictive distributions utilized in the first step of the multiple imputation procedure. In addition, single imputation does not account for the between-imputation variance that leads to overestimation of accuracy (small standard errors). Table 7.2 briefly summarizes the advantages and disadvantages of each method.

7.2.3 Multiple imputation: considerations and limitations

Multiple imputation is a general approach with numerous applications, and it is easily accessible through standard statistical software packages such as R (R Core Team, 2020), SPSS®, SAS® and STATA®. However, it should be highlighted that it is not a panacea for every incomplete data setting (Erler et al., 2016; White et al., 2011). Although multiple imputation is often considered as an out of the box method that can be easily applied in any missing data problem, this is not true. Its application requires the user to carefully consider the plausibility of each of the possible causes of missingness, thoroughly select an appropriate imputation model and choose appropriate variables to include with respect to both clinical relevance and the missing at random assumption.

Some common points and special cases to consider when performing multiple impu-

Table 7.2: Comparison of incomplete data analysis methods.

Methods	Pros	Cons
Complete Case Analysis	Simple to implement	Loss of power and efficiency and invalid under MAR
Single Imputation	Simple to implement and avoids loss of power	Does not appropriately account for uncertainty in results and invalid under MAR
Multiple Imputation	Avoids loss of power, retains efficiency and is valid under MAR	Time consuming and requires more statistical knowledge

MAR: Missing at Random

tation are as follows:

- Missing outcome information: It should be noted that up to this point, this article has focused primarily on missing covariate information. That is because when there are missing outcome data, it has been argued that the complete case analysis is more appropriate as imputed outcome data can lead to misleading results (Von Hippel, 2007; Little, 1992). Single imputation of the worst or best-case scenario for missing outcome data may be used as sensitivity analysis to ensure the validity of trial results. Multiple imputation of missing outcome data may also be performed if there are auxiliary variables that are highly correlated with the outcome and the probability that the outcome is missing. However, this can only help in reducing the loss in accuracy of the estimates due to missing data and only if the data are at most MAR. Nevertheless, the complete case analysis should be regarded as the principle analysis in the case of missing outcome data.

- The number of imputed datasets: Although 5 imputed datasets are considered adequate, it is always advised to increase the number to improve the efficiency and the reproducibility of the results (White et al., 2011).

- The number of iterations: Since multiple imputation is based on an iterative algorithm, the convergence criteria should always be assessed and if necessary, the number of iterations increased (van Buuren, 2012; van Buuren and Groothuis-Oudshoorn, 2011).

- Inclusion of the outcome in the imputation model: The outcome should be included in the first step of the multiple imputation procedure to take into account the association between outcome and incomplete covariates (Moons et al., 2006).

- Longitudinal studies: Common software packages usually require the transformation of long datasets (a row per measurement) to their wide (a row per subject) counterparts to perform multiple imputation. This implies that current implementation of multiple imputation in longitudinal settings works best in balanced studies (e.g. subjects are measured at the same time points).

- Survival analysis: Because of the complex nature of the outcome variable in such cases (pairing of a binary event indicator variable with a time-to-event variable), several approaches have been proposed on how to include it in the imputation model (Barzi and Woodward, 2004; van Buuren et al., 1999; Clark and Altman, 2003). The most recent research findings, however, propose to use the Nelson–Aalen estimator along with the event indicator in the imputation model rather than the event indicator along the time-to-event variable (White and Royston, 2009).

- Acceptable amount of missingness: There is no standard rule of how much missing data is too much. Theoretically, multiple imputation can handle large amounts of missingness. Nevertheless, the quality of the results is related to the complexity of the imputation model used, whether there are few or many variables with a large amount of missingness, the total sample size and the variability of the variables which are subject to missingness. For example, 50% missingness may be acceptable if the remaining 50% of the data allow accurate estimation of the predictive distribution used to draw imputed values. In settings with a small sample size, large variability and/or a heterogeneous study population, this may not be the case.

Given the potential complexities, it is clear that multiple imputation should be conducted carefully with respect to the challenges of each analysis. Advice from statistical experts is, therefore, highly recommended when considering multiple imputation to address missing data.

7.3 Reporting

Because performing analysis on incomplete data requires a lot of considerations, decisions and assumptions, it is recommended that authors provide a thorough description of the imputation procedure to ensure the transparency and reproducibility of the analysis. Often, such a description can be moved to the Supplementary Material accompanying a manuscript. Table 7.3 provides an extensive list of points that should be included when conducting incomplete data analysis.

7.4 Data Example

To illustrate the above points with a data example, we consider a simple scenario for survival analysis. The data come from a follow-up study of patients with congenital heart disease who received a human tissue allograft in the aortic position. The aim is to investigate the association between postoperative aortic gradient (mmHg) and risk of death while accounting for baseline factors such as age at operation (years), gender, donor age (years) and allograft diameter (mm). An overview of the data for the 'all cases' scenario (before excluding any case to artificially generate missing data scenarios) is provided in the Supplementary Material 7.B, Table S4.

To briefly illustrate a few of the points presented throughout this article for dealing with missing data, we artificially generated 40% missingness on the postoperative aortic gradient under the 3 missingness scenarios: MCAR, MAR and MNAR. Under MCAR, randomly chosen values were deleted. Under MAR, the aortic gradient measurements of younger patients (age less than the mean age in the dataset) were deleted. Finally, for MNAR, the missing values were selected to be patients with a high postoperative aortic gradient (higher than the 65th percentile of the postoperative gradient in the dataset) assuming that they are more likely to be unable to go to the hospital. We then applied Cox regression using complete cases, single mean imputation and multiple imputation under each scenario for the mechanism that generated the missing data and compared the corresponding results with those obtained when all the cases were used (no missing data). The analysis was conducted in R (R Core Team, 2020) using the packages **mice** (van Buuren and Groothuis-Oudshoorn, 2011) and **mitools** (Lumley, 2019). A sample R code for conducting multiple imputation in R is given in Supplementary Material 7.C.

The results are summarized in Fig. 1, where the red dot and black lines represent the estimated hazard ratios and their corresponding 95% confidence intervals, respectively. As shown in this figure, under MCAR and MAR, multiple imputation provided results that were slightly closer to those of the complete data (before val-

ues were removed; 'all cases') than the results from the simpler approaches for this specific example. Nevertheless, the differences are small, and both the complete case analysis and single mean imputation are theoretically valid under MCAR. The loss in efficiency due to the reduced sample size when using only the complete cases is evident from the wider confidence intervals. Under MNAR, all approaches provided biased estimates. In this situation, further sensitivity analyses or explicit accounting of the missing data mechanism would be required (Carpenter and M.J., 2013).

7.5 Discussion

Missing data are common in clinical research and should be minimized wherever possible through good study design and data collection protocols. However, in most cases, it is not possible to reduce the amount of missing data to zero. As demonstrated in the example presented in this article, inappropriate handling of missing data can potentially lead to biased results or significant loss of power. Although simpler approaches in handling missing data such as the complete case analysis or single imputation may be appropriate if the amount of missing data is small and the mechanisms behind the missing data are clearly understood, in most cases multiple imputation is accepted as the preferred strategy for handling missing data. Although multiple imputation deals with a number of pitfalls related to complete case analysis or single imputation, it does significantly increase the complexity of the analysis and can potentially lead to bias if the data are not missing at random.

It is important to approach the handling of missing data in a systematic manner and clearly report the steps that have been undertaken in the handling of missing data as outlined in the guidelines in Table 7.3. Although this article is intended to give an overview for clinicians on how to handle missing data, it is strongly recommended that complex approaches to handle missing data should be performed under the guidance of a statistician.

Acknowledgements We would like to thank Nicole S. Erler for the helpful discussions and valuable comments.

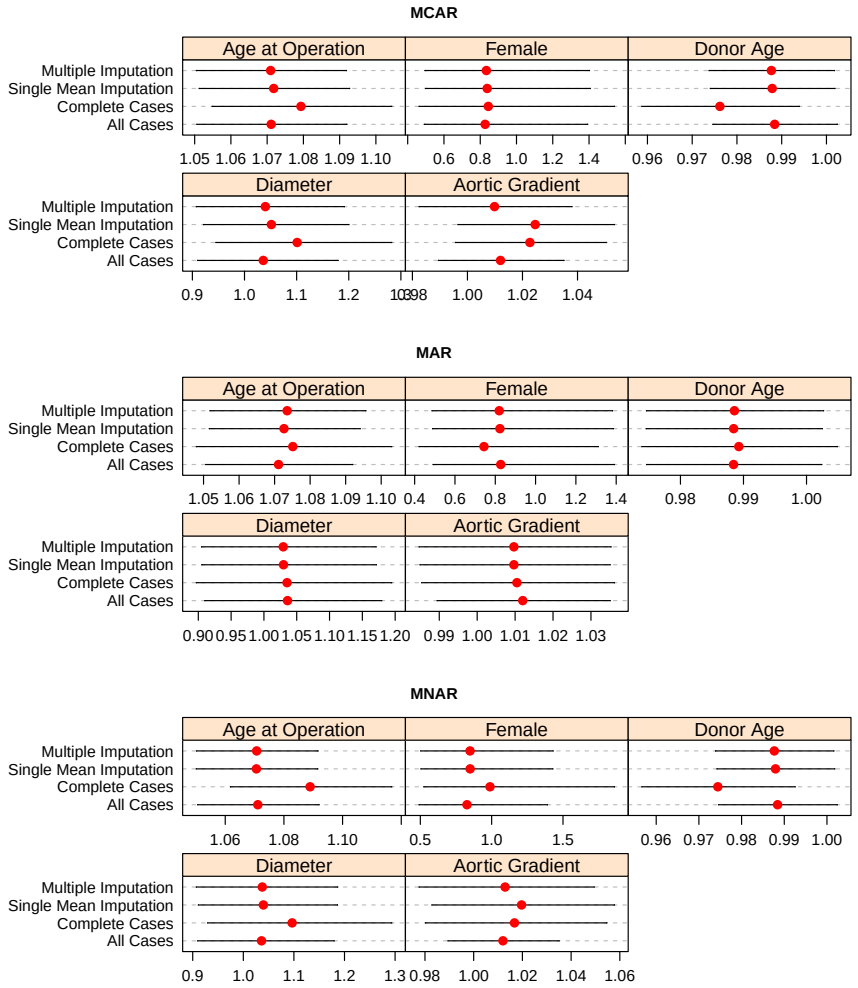

Figure 7.1: Hazard ratios and 95% confidence intervals of 'all cases', complete cases, single mean imputation and multiple imputation analyses under 3 missing data mechanisms. MAR: missing at random; MCAR: missing completely at random; MNAR: missing not at random.

Table 7.3: Guidelines for reporting incomplete data analysis in scientific manuscripts

Report the number/proportion of missing values per variable of interest:

- Alternatively/complementary report number of complete cases
- If possible, discuss potential causes for missing data

Provide comparison of complete and incomplete cases:

- Table or figure comparing the distributions of variables of interest

State the methodology used for incomplete data analysis:

- Which one was used: the complete case analysis, multiple imputation, etc.
- Report the assumptions that were made regarding the cause of missingness: MCAR, MAR or MNAR

Indicate the software (including version number) that was used in handling missing data:

- Optionally add any changes to the default settings/features of the software or/and functions which were used

Report the number of imputed datasets and number of iterations

List the variables that were included in the imputation model

Mention any higher order functions of imputed variables:

- Were higher order terms such as interactions, polynomials or spline transformations of original variables used in the analysis?
- If yes, were these higher order terms included in the imputation model as well?

Assess the robustness of MAR assumption by conducting sensitivity analysis for various MNAR scenarios

MAR: missing at random; MCAR: missing completely at random; MNAR: missing not at random.

Supplementary Material

7.A A

If we denote with $\hat{\theta}_i$ the estimate of the quantity of interest which we obtained from the i^{th} imputed dataset, then the combined estimate $\hat{\theta}$ is simply the average over the imputation-specific estimates:

$$\hat{\theta} = \frac{1}{M} \sum_{i=1}^{M} \hat{\theta}_i, i = 1, \ldots, M$$

Moreover, if we let V_i denote the estimated variance of $\hat{\theta}_i$ then the within imputation variance V is calculated as an average over all the individual variances:

$$V = \frac{1}{M} \sum_{i=1}^{M} V_i, i = 1, \ldots, M$$

while the between-imputation variance Q is calculated based on the squared differences between the individual estimates and the combined estimate:

$$Q = \frac{1}{M-1} \sum_{i=1}^{M} \left(\hat{\theta}_i - \hat{\theta}\right)^2, i = 1, \ldots, M$$

The total variance of the combined estimate $\hat{theta}$ is then formed by accounting for both the within- and between-imputation variances as follows:

$$\text{var}\left(\hat{\theta}\right) = V + \left(1 + \frac{1}{M}\right) Q$$

7.B B

Table 7.4: Baseline characteristics for “all cases” dataset used in the data example.

Total number of patients (n)	**374**
Total number of events (n)	**72**
Follow-Up (median; IQR)	**10.95**; (7.67, 15.35)
Age at Operation (median; IQR)	**47.15**; (34.40, 56.43)
Female Sex (n)	**105**
Donor Age (median; IQR)	**58.79**; (41.97, 69.53)
Diameter (median; IQR)	**23**; (21, 24)
Post-op Aortic Gradient (median; IQR)	**26.94**; (17.78, 33.80)

MAR: Missing at Random

7.C C

```
#### multiple imputation
# Load 'mice' package
library(mice)

# Load 'mitools' package
library(mitools)

# Generate 40 Imputed datasets
mice.data <- mice(data, m = 40)

# Convert to imputationList (not necessary)
# Note that the code used inside function
# imputationList():lapply(1:40, complete, x = mice.data)
# just converts the mice object to a list of data frames
# where each data frame corresponds
to one imputed dataset.
mice.data.ImpList <- imputationList(lapply(1:40, complete, x = mice. data))

# Apply Cox regression to each imputed dataset

models.mice <- with(mice.data.ImpList, coxph(Surv(LastFUP, death) ~ AgeatOP + Sex +
DonorAge + diameter + AoGradient))

# Extract pooled estimates using MIcombine() to apply Rubin's Rules
res.mice <- summary(MIcombine(models.mice))
```

7.4 References

Barzi, F. and Woodward, M. (2004). Imputations of Missing Values in Practice: Results from Imputations of Serum Cholesterol in 28 Cohort Studies. *American Journal of Epidemiology*, 160(1):34–45.

Bell, M. L., Kenward, M. G., Fairclough, D. L., and Horton, N. J. (2013). Differential dropout and bias in randomised controlled trials: when it matters and when it may not. *BMJ*, 346:e8668.

Carpenter, J. and M.J., K. (2013). *Multiple Imputation and Its Application*. John Wiley & Sons, Chichester, West Sussex.

Clark, T. G. and Altman, D. G. (2003). Developing a prognostic model in the presence of missing data: an ovarian cancer case study. *Journal of Clinical Epidemiology*, 56(1):28 – 37.

Enders, C. (2010). *Applied Missing Data Analysis*. Guilford Press, New York.

Erler, N. S., Rizopoulos, D., Rosmalen, J. v., Jaddoe, V. W. V., Franco, O. H., and Lesaffre, E. M. E. H. (2016). Dealing with missing covariates in epidemiologic studies: a comparison between multiple imputation and a full bayesian approach. *Statistics in Medicine*, 35(17):2955–2974.

J., C. and M.J., K. (2007). *A critique of common approaches to missing data. In: Missing data in randomised controlled trials-—a practical guide.* National Institute for Health Research, Birmingham, AL.

Little, J. and Rubin, D. (2002). *Statistical Analysis with Missing Data*. Wiley, Hoboken, NJ, 2nd edition.

Little, R. J., D'Agostino, R., Cohen, M. L., Dickersin, K., Emerson, S. S., Farrar, J. T., Frangakis, C., Hogan, J. W., Molenberghs, G., Murphy, S. A., Neaton, J. D., Rotnitzky, A., Scharfstein, D., Shih, W. J., Siegel, J. P., and Stern, H. (2012). The prevention and treatment of missing data in clinical trials. *New England Journal of Medicine*, 367(14):1355–1360.

Little, R. J. A. (1992). Regression with missing x's: A review. *Journal of the American Statistical Association*, 87(420):1227–1237.

Lumley, T. (2019). *mitools: Tools for Multiple Imputation of Missing Data*. R package version 2.4.

Moons, K. G., Donders, R. A., Stijnen, T., and Harrell, F. E. (2006). Using the outcome for imputation of missing predictor values was preferred. *Journal of Clinical Epidemiology*, 59(10):1092 – 1101.

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Schafer, J. (1997). *Analysis of Incomplete Multivariate Data*. Chapman and Hall, London.

Vach, W. and Blettner, M. (1991). Biased Estimation of the Odds Ratio in Case-Control Studies due to the Use of Ad Hoc Methods of Correcting for Missing Values for Confounding Variables. *American Journal of Epidemiology*, 134(8):895–907.

van Buuren, S. (2012). *Flexible Imputation of Missing Data*. Taylor & Francis, Boca Raton, FL.

van Buuren, S., Boshuizen, H. C., and Knook, D. L. (1999). Multiple imputation of missing blood pressure covariates in survival analysis. *Statistics in Medicine*, 18(6):681–694.

van Buuren, S. and Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in r. *Journal of Statistical Software, Articles*, 45(3):1–67.

Von Hippel, P. T. (2007). Regression with missing ys: An improved strategy for analyzing multiply imputed data. *Sociological Methodology*, 37(1):83–117.

White, I. R. and Royston, P. (2009). Imputing missing covariate values for the cox model. *Statistics in Medicine*, 28(15):1982–1998.

White, I. R., Royston, P., and Wood, A. M. (2011). Multiple imputation using chained equations: Issues and guidance for practice. *Statistics in Medicine*, 30(4):377–399.

Part IV

Discussion

Chapter 8

Discussion

In this thesis, we developed extensions for the joint modeling framework for longitudinal and time-to-event data, motivated by various clinical research questions in heart disease. This chapter will discuss these extensions, summarize our findings from their implementation, and discuss possible directions for future work.

In **Part I** we extended the joint modeling framework for cases where intermediate events may occur during follow-up that may alter the course of the disease for a patient. By doing so, we were able to develop subject-specific prediction tools that are dynamic (e.g., can be updated with new information) and adaptive to different scenarios concerning the occurrence of intermediate events in the future. This is not only important because it adds a new dimension to dynamic predictions - adapting to different scenarios with respect to intermediate events, that is - but also because the accuracy of predictions in such settings is substantially influenced by the correct specification of the subject-specific longitudinal profiles (Rizopoulos et al., 2017; Ferrer et al., 2019; Suresh et al., 2017), which is in accordance to our results. Apart from the pure methodological point of view, this can directly improve clinical decision-making and patient prognosis in practice. As we have shown in our simulation study, ignoring the occurrence of the intermediate events leads to worse predictive performance. This highlights the importance of appropriately incorporating information on such intermediate events when developing prediction tools. There are numerous ways in how these intermediate events can influence the subject-specific trajectories. We suggested the use of a general function:

$$g\{R_i(t), t_{i+}\} = \tilde{\boldsymbol{x}}_i^\top(t)\,\tilde{\boldsymbol{\beta}} + \tilde{\boldsymbol{z}}_i^\top(t)\,\tilde{\boldsymbol{b}}_i$$

, which is a function of time-varying variables that describe the occurrence of the intermediate event. We then illustrated how one could specify this function to accommodate different situations in two applications. First, for the study on the pulmonary gradient, we encountered a particular and obvious shape for the longitudinal trajectories of the patients who experienced re-intervention (the intermediate event in this case). This allowed us to specify function $g\{R_i(t), t_{i+}\}$ to accommodate this shape of trajectories explicitly. On the other hand, for the SPRINT trial and the occurrence of serious adverse events, this was not the case, and thus, we opted for a flexible specification based on spline functions of the time-varying variables. In practice, cases, where the impact of the intermediate event on the shape of the longitudinal trajectory is so obvious as for the case of re-intervention and the pulmonary gradient might be rare, and therefore the flexible specification based on splines is to be preferred. This was the case for investigating the influence of pregnancy on the long-term durability

of allografts in the right ventricular outflow tract. In that case, the impact of the onset of pregnancy on the shape of pulmonary gradient and pulmonary regurgitation trajectories could not be easily identified. Therefore we opted for a flexible specification based on spline functions of the time-varying variables. By doing so, we were able to quantify the impact of the onset of pregnancy on the risk of valve replacement.

In **Part II** we investigated the use of Bayesian global-local shrinkage priors as a tool to automate feature selection of biomarkers in the context of multivariate longitudinal and time-to-event data. The selection of an appropriate association structure between the longitudinal and the time-to-event outcomes is a complicated task with a substantial impact on the derived predictions' quality. However, in practice, it is often the case that in the absence of previous studies on the matter, the form of this association structure is not straightforward. In such cases, part of the objective of these studies is to identify an appropriate association structure between the outcomes. As we discussed in Chapter 4 this becomes increasingly complicated in settings where the interest lies in the association between multiple longitudinal markers and multi-state data as the number of possible association parameters increases significantly along with the potential issue of collinearity between the parameters. To this end, we developed an extension of the standard joint model that can accommodate multiple longitudinal outcomes and multi-state processes, and we then used global-local shrinkage priors to aid in selecting the appropriate association structure. The two candidates we proposed, the ridge and horseshoe priors, did not show considerable differences between them in the amount of shrinkage applied to the association parameters. That was the case for both the results of our simulation study and the results in the application to the LVAD dataset. Apart from the use in high-dimensional settings where multiple longitudinal outcomes and multi-state processes are considered, the application of the global shrinkage priors can be used in less complex settings. This was the case for the study on the association between tricuspid regurgitation and survival. By applying the global shrinkage priors, we were able to identify the current-value association structure as the best option to capture the relationship between the log-odds of moderate to severe tricuspid regurgitation and survival.

Finally, in **Part III** we worked on the topic of missing data. Missing data is an unavoidable reality in clinical research that can impact patient prognosis, prediction, and clinical decision-making when not accounted for and handled appropriately. Joint models for longitudinal and time-to-event data belong in the broader family of SPMs and are commonly used as MNAR models for dropout in longitudinal studies. If in a longitudinal study there is dropout due to both random censoring and due to an event associated with the longitudinal outcome of interest, simple approaches such as mixed-effects models or time-dependent Cox models will give biased results. When

ignoring the additional information that part of the dropout is due to an informative for the outcome cause, the mechanism that generates the missing data is MNAR, but the mixed-effects model and the time-dependent Cox model treat it as MAR and MCAR, respectively. Joint models solve this by using shared-random effects to explain the association between the two. However, in practice, information about the actual cause of dropout is rarely available. In addition to that, the distinction between MNAR and MAR is intractable. To this end, the standard approach is for one to start with a model valid under MAR and then explore deviations from these assumptions for different MNAR scenarios. An evergrowing significant body of literature with essential contributions during the last two decades created an arsenal of tools for the latter. However, this mainly focused on the use of PMM and SeMs. Our contribution was to add an intuitive and readily available in standard software approach to explore MAR and MNAR scenarios using the joint modeling framework for longitudinal and survival data. An advantage of our approach is that this can be achieved on the subject-specific level in contrast to other approaches, which prospectively allows for exploring subject-specific MAR and MNAR assumptions. Another important result that arose from our simulation study and application to the HIV data was that misspecification of the missing data generating mechanism does not necessarily lead to biased fixed-effect estimates but is absorbed by the random effects in SPMs. This means that while inference on the population parameters remains stable across different assumptions for the data generating mechanism, this is not the case for subject-specific inference and in extent prediction. This is important for practicing researchers as it reflects the importance of exploring MNAR and MAR with sensitivity analysis. Apart from the methodological contribution, there must be a common language between statistical methodologists and clinical researchers in the medical field. Missing data is an example of frequent misinterpretation as it is an active area of research, it is challenging to understand for clinicians, and there is a wide range of methodology available in different software with varying levels of accessibility for clinical researchers. Thus, apart from developing new methods, it is also essential to ensure that clear and practical guidelines are well documented and communicated to aid researchers in handling missing data. In our work, we tried to do so by creating a small overview of the simple approaches to handle missing data with software examples as we believe that good communication of existing methodology is as important in practice as the development of methodology itself.

Future Research

In this thesis we focused on extending the joint modeling framework for longitudinal and time-to-event outcomes in several directions. Specifically, we developed dynamic prediction tools that can be updated as new data are collected and can be adaptive to different scenarios for the future occurrence of intermediate events; we proposed a joint model that can handle multivariate longitudinal data and multi-state processes under the Bayesian framework and explored the use of global-local shrinkage priors in feature selection, and finally, we sketched a framework for sensitivity analysis under different assumptions for missing data generating mechanisms under the joint modeling framework. These three themes were motivated by existing research questions and were applied in heart disease data.

Of course, the role of research is not only to answer questions but to generate new ones to guide future research. Before discussing potential future research directions based on these three main themes, this is a good point to highlight the importance of software in exercising research. The methods presented in this thesis can be implemented using package **JMbayes** (Rizopoulos, 2016) in R (R Core Team, 2020). Package **JMbayes** was the building block upon which we were able to develop, test, and finally apply and make available to practicing scientists the extensions discussed in this thesis. During our research work, we programmed extensions such as dynamic predictions for multivariate longitudinal outcomes and a single survival outcome, joint models for multiple longitudinal and multi-state processes, dynamic predictions with intermediate events, shrinkage priors for selecting association parameters, and sensitivity analysis. This means that these methods are now readily available in standard software and can be easily be implemented in applications by practicing researchers. Most importantly, however, the knowledge and experience we gained during extending the existing **JMbayes** package lead to the birth of its evolution: **JMbayes2**. A new package that can be used to implement joint models to even more types of longitudinal data (e.g., ordinal, beta, etc.), more types of time-to-event outcomes (competing risks, multi-state, recurrent), has improved convergence behavior and diagnostics, is more user friendly and considerably faster. **JMbayes2** not only combines all our existing knowledge and experience both in theoretical and programming terms, but we also do hope that it will serve as the building block for future research as its architecture was built with future extensions in mind.

For the remainder of this section, we will discuss future research concerning each of the main themes covered in this thesis.

In **PART I** we discussed how joint models could be used to develop dynamic prediction tools that are adaptive to the future occurrence (or not) of intermediate events.

While this is a very useful tool that can aid practicing physicians in their decision-making, it can be further improved. More specifically, in our applications, we were only able to show if there was an improvement in the predicted clinical outcome of a patient under a hypothetical future scenario regarding the occurrence of a re-intervention, serious adverse event, and pregnancy. What would be even more interesting, though, especially for re-intervention where the physician drives the decision, is how we can utilize these predictions to inform for the optimal timing of such interventions. To do so, we need in collaboration with the physicians to come up with a benefit-loss function that can aid us in quantifying the impact of the timing of such re-interventions on the clinical outcome of a patient. This is similar to the work that has been done in the field of prostate cancer and personalized schedules (Tomer et al., 2019) where a benefit-loss function was used to detect the optimal timing for future biopsy visits. We believe that utilizing the adaptive dynamic predictions to quantify benefit and risk will allow us to use it as an informative tool for optimal timing of treatments in applications where timing is essential such as the application of pulmonary grafts and re-intervention.

In **PART II** we proposed and explored the use of Bayesian global-local shrinkage priors to aid in the selection of the association structure between longitudinal and multi-state processes. While this showed to work satisfactorily in the applications considered in this thesis, a natural extension would be to assess if the application of such shrinkage priors leads to improved prediction performance. To achieve so, a comprehensive simulation study should be performed comparing different feature selection methods in joint models such as local shrinkage, bayesian global-local shrinkage, bayesian averaging in terms of measures of predictive accuracy such as the AUC the PE. We are already working towards this direction by implementing in **JMbayes2** the capability to derive predictions from joint models for longitudinal and multi-state processes. This will enable us to perform such simulations and further investigate the benefit (or not) of shrinkage selection in the joint modeling framework towards prediction modeling. Such models can translate the complex relationships between multiple outcomes into tools that can aid physicians in everyday practice. While automation of future selection sounds like an elegant tool that can replace biological reasoning and clinical thinking and make model specification an easy task for everyone, this is merely the case. To successfully translate the reality into models, statisticians and clinicians need to communicate and work together to translate years of clinical experience and knowledge into tools that can be used in practice and do not solely rely on automated and data-driven procedures. Automated and data-driven approaches can aid this communication but should not replace critical thinking.

Finally, in **PART III** we discussed how the classic joint modeling framework could

be used to conduct sensitivity analysis under different assumptions with respect to the missing data generating mechanisms. The handling of missing data has become a major topic during the last two decades, and now it is considered necessary (as it should) to conduct sensitivity analysis to explore deviations from standard assumptions about missing data. We endorse this, and we believe that future work on the topic should be focused on making these approaches easy to implement in practice as this will improve the quality of clinical research.

General Conclusion

In this thesis, we developed and applied extensions of joint models for longitudinal and time-to-event data to address different research questions, mainly in the field of heart disease. Our research covered three main themes: the occurrence of intermediate events, feature selection for joint models with multiple longitudinal and time-to-event outcomes, and missing data. The methods extend and improve the existing joint modeling framework by tackling the unique features that were posed by each research question. Dynamic predictions that are adaptive to the occurrence of intermediate events can quantify the effect of the future occurrence of such intermediate events. Bayesian global-local shrinkage priors can be used to aid in feature selection in multicollinear settings with multiple biomarkers and states of interest. Finally, joint models can be used to explore the sensitivity of assumptions with respect to missing data. While these methods might seem complicated at first, their user-friendly implementation in standard software makes its implementation easy. As more and more data are collected continuously on the individual level, the necessity and benefit of such models and prediction tools that can be dynamically updated and provide subject-specific inference will become more and more apparent. Our contribution to the joint modeling framework extends their range of use. It makes their implementation easier, providing practicing researchers with a powerful tool that can be used to derive subject-specific predictions, to understand the inter-relationships between different outcomes, and to investigate the impact of different assumptions regarding missing data. This will eventually lead to prediction models that will allow us to optimize treatment decisions and hopefully translate to improved patient outcomes.

8.1 References

Ferrer, L., Putter, H., and Proust-Lima, C. (2019). Individual dynamic predictions using landmarking and joint modelling: Validation of estimators and robustness assessment. *Statistical Methods in Medical Research*, 28(12):3649–3666. PMID: 30463497.

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rizopoulos, D. (2016). The R package JMbayes for fitting joint models for longitudinal and time-to-event data using mcmc. *Journal of Statistical Software*, 72(7):1–45.

Rizopoulos, D., Molenberghs, G., and Lesaffre, E. M. (2017). Dynamic predictions with time-dependent covariates in survival analysis using joint modeling and landmarking. *Biometrical Journal*, 59(6):1261–1276.

Suresh, K., Taylor, J. M., Spratt, D. E., Daignault, S., and Tsodikov, A. (2017). Comparison of joint modeling and landmarking for dynamic prediction under an illness-death model. *Biometrical Journal*, 59(6):1277–1300.

Tomer, A., Nieboer, D., Roobol, M. J., Steyerberg, E. W., and Rizopoulos, D. (2019). Personalized schedules for surveillance of low-risk prostate cancer patients. *Biometrics*, 75(1):153–162.

APPENDIX

Summary

In this thesis, we developed extensions for the joint modeling framework for longitudinal and time-to-event data, motivated by various clinical research questions in cardiothoracic surgery. Heart diseased subjects are commonly followed up over long periods, and multiple outcomes, both longitudinal and clinical, are monitored. The association of these outcomes is of primary interest as its understanding and quantification can lead to prediction models that can help optimize the treatment and the decision-making for these patients. However, such prediction models are difficult to develop as there are several factors that can influence our understanding of the relationship between these outcomes. Intermediate events such as additional treatments and serious adverse events may occur during follow-up and can alter the course of the disease. Specifying the association structure between multiple correlated outcomes is challenging as its form may be unknown. Missing data and dropout can substantially reduce available information and pose additional challenges to our inference depending on the unknown cause of missingness. Motivated by these challenges, we extend the joint modeling framework to try and overcome them.

In **Part I**, Chapter 2, we developed dynamic prediction tools that are adaptive to the future occurrence (or not) of intermediate events. Motivated by two applications in heart disease, we estimated joint models that helped us quantify the impact of re-intervention in CHD patients and SAEs in subjects with increased cardiovascular risk. We then built prediction tools that are adaptive to different scenarios for such intermediate events in the future. We compared the performance of these tools with existing simpler approaches and showed their benefit in predictive performance. In Chapter 3 we used the joint modeling framework to quantify the impact of the onset of pregnancy on biomarkers that describe the function of the pulmonary valve and clinical outcomes for these patients. We then built prediction models that were adaptive to different choices for the timing of the onset of pregnancy in the future and were able to show that there was no expected difference in the clinical outcome of women depending on the onset of pregnancy in the future.

In **Part II**, Chapter 4, we extended the joint modeling framework for the case of multiple longitudinal and multi-state processes. We discussed the challenges in choosing the association structure between these outcomes due to the increased dimensionality of the parameter space and the correlation between these outcomes. We proposed using Bayesian global-local shrinkage priors to aid in selecting relevant features that quantify the association between longitudinal and multi-state outcomes. We showcased the approach using a dataset of patients who received an LVAD after heart failure and were able to identify the association of the current value as the best

candidate for both TB and Cr and the transitions between HF, complications, and death. In Chapter 5, we applied the global-local shrinkage priors in a simpler setting with one longitudinal outcome and one time-to-event outcome to help us identify the stronger association structure. By doing so, we were able to identify the current value association as the stronger one between TR and survival.

In **Part III**, Chapter 6, we proposed an alternative parametrization of MAR for SPMs. We illustrated how this parametrization could be used in practice to perform sensitivity analysis between MAR and MNAR using the joint modeling framework. We showed how MAR and MNAR models may lead to similar effect estimates with an extensive simulation study but can substantially differ in terms of subject-specific prediction. Finally, we provided basic practical guidelines for handling missing data in clinical research to make existing methods more accessible to practicing clinicians.

General Conclusion

In this thesis, we developed and applied extensions of joint models for longitudinal and time-to-event data to address different research questions, mainly in the field of heart disease. Our research covered three main themes: the occurrence of intermediate events, feature selection for joint models with multiple longitudinal and time-to-event outcomes, and missing data. The methods extend and improve the existing joint modeling framework by tackling the unique features that were posed by each research question. Dynamic predictions that are adaptive to the occurrence of intermediate events can quantify the effect of the future occurrence of such intermediate events. Bayesian global-local shrinkage priors can be used to aid in feature selection in multicollinear settings with multiple biomarkers and states of interest. Finally, joint models can be used to explore the sensitivity of assumptions with respect to missing data. While these methods might seem complicated at first, their user-friendly implementation in standard software makes its implementation easy. As more and more data are collected continuously on the individual level, the necessity and benefit of such models and prediction tools that can be dynamically updated and provide subject-specific inference will become more and more apparent. Our contribution to the joint modeling framework extends their range of use. It makes their implementation easier, providing practicing researchers with a powerful tool that can be used to derive subject-specific predictions, to understand the inter-relationships between different outcomes, and to investigate the impact of different assumptions regarding missing data. This will eventually lead to prediction models that will allow us to optimize treatment decisions and hopefully translate to improved patient outcomes.

Nederlandse Samenvatting

In dit proefschrift hebben we uitbreidingen ontwikkeld voor het joint model framework voor longitudinale en overlevingsdata, gemotiveerd door verschillende klinische onderzoeksvragen in de cardiothoracale chirurgie. Patiënten met hartaandoeningen worden vaak gedurende lange perioden gevolgd, hetgeen resulteert in meerdere metingen, zowel longitudinaal als klinisch. De relatie tussen de uitkomsten van deze metingen is van primair belang, omdat het begrip en de kwantificering ervan kan leiden tot voorspelmodellen die kunnen helpen bij het optimaliseren van de behandeling en de besluitvorming bij deze patiënten. Dergelijke voorspelmodellen zijn echter moeilijk te ontwikkelen omdat er diverse factoren zijn die ons begrip van de relatie tussen deze uitkomsten kunnen beïnvloeden. Tussentijdse gebeurtenissen, zoals aanvullende behandelingen en ernstige bijwerkingen, kunnen optreden tijdens de follow-up en mogelijk het verloop van de ziekte veranderen. Het specificeren van de associatiestructuur tussen meerdere gecorreleerde uitkomsten is een uitdaging omdat de vorm ervan onbekend kan zijn. Ook kan er een substantiële afname in beschikbare informatie ontstaan door ontbrekende gegevens of uitval van patiënten. Dit vormt, afhankelijk van de oorzaak van de afwezigheid van data, een extra uitdaging voor onze gevolgtrekking. In ons proefschrift breiden we het joint model framework uit met als doel voor deze uitdagingen een oplossing te vinden.

In **Deel I**, Hoofdstuk 2, hebben we dynamische voorspellingstools ontwikkeld die aangepast zijn aan het toekomstige optreden (of niet) van tussentijdse gebeurtenissen. Gemotiveerd door twee toepassingen bij hartaandoeningen schatten we joint models die ons hielpen de impact te kwantificeren van herinterventie bij CHD-patiënten en SAE's bij personen met een verhoogd cardiovasculair risico. Vervolgens hebben we voorspellingstools gebouwd die zich aanpassen aan verschillende scenario's voor dergelijke tussentijdse gebeurtenissen in de toekomst. We vergeleken de prestaties van deze tools met bestaande eenvoudigere benaderingen en konden hun voorspellende gave aantonen. In Hoofdstuk 3 hebben we het joint modeling framework gebruikt om de impact te kwantificeren van het begin van de zwangerschap op biomarkers die de functie van de longklep en klinische uitkomsten voor deze patiënten beschrijven. Vervolgens hebben we voorspelmodellen gebouwd die zich aanpasten aan verschillende keuzes voor de timing van het begin van de zwangerschap in de toekomst. Tevens konden we aantonen dat er geen verwacht verschil was in de klinische uitkomst van vrouwen afhankelijk van het begin van de zwangerschap in de toekomst.

In **Deel II**, Hoofdstuk 4, hebben we het joint model framework uitgebreid voor het inzetten van meerdere longitudinale en multi-state processen. We zetten de uitdagingen uiteen bij het kiezen van de relatiestructuur tussen deze uitkomsten vanwege de

toegenomen dimensionaliteit van de parameterruimte en de correlatie tussen deze uitkomsten. We stelden voor om Bayesiaanse global-lokale krimpprioriteiten te gebruiken om te helpen de relevante kenmerken te selecteren die de associatie tussen longitudinale en multi-state uitkomsten kwantificeren. We presenteerden de aanpak met behulp van een dataset van patiënten die een LVAD kregen na hartfalen en waren in staat om de associatie van de huidige waarde als de beste kandidaat voor zowel TB als Cr en de overgangen tussen HF, complicaties en overlijden te identificeren. In Hoofdstuk 5 hebben we de global-local shrinkage priors toegepast in een eenvoudigere setting met één longitudinale uitkomst en één uitkomst uit overlevingsdata om ons te helpen de sterkere associatiestructuur te identificeren. Hierdoor konden we de huidige waarde-associatie identificeren als de sterkere van TR en overleving.

In **Deel III**, Hoofdstuk 6, hebben we een alternatieve parametrisering van MAR voor SPM's voorgesteld. We hebben geïllustreerd hoe deze parametrisering in de praktijk kan worden gebruikt om gevoeligheidsanalyses uit te voeren tussen MAR en MNAR met behulp van het joint modeling framework. We hebben laten zien hoe MAR- en MNAR-modellen kunnen leiden tot vergelijkbare effectschattingen met een uitgebreide simulatiestudie, maar aanzienlijk kunnen verschillen in termen van onderwerpspecifieke voorspelling. Ten slotte hebben we praktische basisrichtlijnen gegeven voor het omgaan met ontbrekende gegevens in klinisch onderzoek om bestaande methoden toegankelijker te maken voor praktiserende clinici.

Algemene conclusie

In dit proefschrift hebben we uitbreidingen van gewrichtsmodellen voor longitudinale en overlevingsdata ontwikkeld en toegepast om verschillende onderzoeksvragen te beantwoorden, voornamelijk op het gebied van hartziekten. Ons onderzoek omvatte drie hoofdthema's: het optreden van tussentijdse gebeurtenissen, functieselectie voor joint models met meerdere longitudinale en time-to-event uitkomsten, en ontbrekende gegevens. De methoden verbeteren het bestaande kader voor joint modeling, en breiden deze uit door de unieke kenmerken van elke onderzoeksvraag te ondervangen. Dynamische voorspellingen die zich aanpassen aan het optreden van tussentijdse gebeurtenissen kunnen het effect van het toekomstige optreden van dergelijke tussentijdse gebeurtenissen kwantificeren. Bayesiaanse global-lokale krimpprioriteiten kunnen worden gebruikt om te helpen bij het selecteren van kenmerken in multicollineaire instellingen met meerdere biomarkers en een verschillende staat van belang. Ten slotte kunnen gezamenlijke modellen worden gebruikt om de gevoeligheid van aannames met betrekking tot ontbrekende gegevens te onderzoeken. Hoewel deze methoden op het eerste gezicht misschien ingewikkeld lijken, maakt hun gebruiksvriendelijke implementatie in standaardsoftware de implementatie ervan een-

voudig. Naarmate doorlopend meer en meer gegevens op individueel niveau worden verzameld, zullen de noodzaak en het voordeel van dergelijke modellen en voorspellingstools die dynamisch kunnen worden bijgewerkt en onderwerpspecifieke gevolgtrekkingen bieden, steeds duidelijker worden. Onze bijdrage aan het joint model framework breidt hun gebruiksbereik uit. Het maakt de implementatie ervan eenvoudiger en biedt praktiserende onderzoekers een krachtig hulpmiddel dat kan worden gebruikt om onderwerpspecifieke voorspellingen af te leiden, de onderlinge relaties tussen verschillende uitkomsten te begrijpen en de impact van verschillende aannames met betrekking tot ontbrekende gegevens te onderzoeken. Dit zal uiteindelijk leiden tot voorspellingsmodellen waarmee we behandelbeslissingen kunnen optimaliseren en deze hopelijk kunnen vertalen naar verbeterde patiëntresultaten.

Publications

1. **Papageorgiou, G.**, Mokhles, M.M., Muslem, R., Takkenberg, J.J.M., and Rizopoulos, D. (2021), Feature Selection of Longitudinal Biomarkers in Multivariate Joint Models for Longitudinal and Multi-State Processes. *In Preparation*

2. **Papageorgiou, G.**, and Rizopoulos, D. (2021), An alternative characterization of MAR in shared parameter models for incomplete longitudinal data and its utilization for sensitivity analysis. *Statistical Modelling*, 21(1-2), 95–1114.

3. Romeo, J.L.R., **Papageorgiou, G.**, da Costa, F.F.D., Sievers, H.H., Bogers, A.J.J.C., el-Hamansy, I., Skillington, P.D., Wynne, R., Mastrobuoni, S., El Khoury, G., Takkenberg, J.J.M., and Mokhles, M.M. (2021), Long-term Clinical and Echocardiographic Outcomes in Young and Middle-aged Adults Undergoing the Ross Procedure *JAMA Cardiology*, Published Online First

4. Gharbharan A., Jordans C.C.E., GeurtsvanKessel C., den Hollander J.G., Karim F., Mollema F.P.N., Stalenhoef – Schukken J.E., Dofferhoff A., Ludwig I., Koster A., Hassing R-J., Bos J.C., van Pottelberge G.R., Vlasveld I.N., Ammerlaan H.S.M., van Leeuwen – Segarceanu E.M., Miedema J., van der Eerden M., Schrama T.J., **Papageorgiou, G.**, te Boekhorst P., Swaneveld F.H., Mueller Y.M., Schreurs M.W.J., van Kampen J.J.A., Rockx B., Okba N.M.A., Katsikis P.D., Koopmans M.P.G., Haagmans B.L., Rokx C., Rijnders B.J.A. (2021), Effects of potent neutralizing antibodies from convalescent plasma in patients hospitalized for severe SARS-CoV-2 infection *Nature Communications*, 12, 3189

5. Schauwvlieghe A., Dunbar A., Storme E., Vlak A., Aerts R., Maertens J., Sciot B., Van Der Wel T., **Papageorgiou, G.**, Moors I., Cornelissen J.J., Rijnders B.J.A., Mercier T., (2021), Stopping antibiotic therapy after 72 h in patients with febrile neutropenia following intensive chemotherapy for AML/MDS (safe study): A retrospective comparative cohort study, *EClinicalMedicine*, 35

6. Huizing E., Schreve M.A., Kum S., **Papageorgiou, G.**, de Vries J-P PM, de Borst G.J., Ünlü C. (2021), Development of a Prediction Model for the Occurrence of Stenosis or Occlusion after Percutaneous Deep Venous Arterialization, *Diagnostics*, 11(6) 1008

7. Romeo, J.L.R., **Papageorgiou, G.**, Takkenberg, J.J.M., Roos-Hesselink, J.W., van Leeuwen, W.J., Cornette, J.M.J., Rizopoulos, D., Bogers, A.J.J.C., and Mokhles, M.M. (2020), Influence of pregnancy on long-term durability of allografts in right ventricular outflow tract *Journal of Thoracic and Cardiovascular Surgery*, 159(4), 1508–1516.

8. Veen, K., **Papageorgiou, G.**, Zijderhand, C., Brugts, J., Mokhles, M.M., Manintveld, O., Constantinescu, A., Bekkers, J., Takkenberg, J.J.M., Bogers, A.J.J.D., Caliskan, K. (2020), The Clinical Impact and Long-Term Outcome of Tricuspid Regurgitation in Patients with Orthotopic Heart Transplantation(*The Journal of heart and lung transplantation : the official publication of the International Society for Heart Transplantation*, 39(4), S291–S292.

9. Abjigitova, D., Mokhles, M.M., **Papageorgiou, G.**, Bekkers, J.A., Bogers, A.J.J.C. (2020), Outcomes of different aortic arch replacement techniques *Journal of Cardiac Surgery*, 35(2), 367–374.

10. Wester, V., de Groot, S, Kanters, T, Wagner, L., Ardesch, J, Corro Ramos, I., Enders-Slegers, M.-J, de Ruiter, M, le Cessie, S, Los, J, **Papageorgiou, G.**, van Exel, J, Versteegh, M, EPISODE-team (2020), Evaluating the Effectiveness and Cost-Effectiveness of Seizure Dogs in Persons With Medically Refractory Epilepsy in the Netherlands: Study Protocol for a Stepped Wedge Randomized Controlled Trial (EPISODE) *Frontiers in Neurology*, 11, Article no. 3

11. van der Slot, W.M.A., Benner, J.L., Brunton, L., Engel, J.M., Gallien, P., Hilberink, S.R., Månum, G., Morgan, P., Opheim, A., Riquelme, I., Rodby-Bousquet, E., Şimşek, T.T., Thorpe, D.E., van den Berg-Emons, R.J.G., Vogtle, L.K., **Papageorgiou, G.** Roebroeck, M.E. (2020), Pain in adults with cerebral palsy: A systematic review and meta-analysis of individual participant data *Annals of Physical and Rehabilitation Medicine*, Article in press.

12. **Papageorgiou, G.**, Mokhles, M.M., Takkenberg, J.J.M., and Rizopoulos, D. (2019), Individualized dynamic prediction of survival with the presence of intermediate events *Statistics in Medicine*, 38(30), 5623–5640.

13. Knol, W.G., **Papageorgiou, G.**, and Mokhles, M.M. (2019), What to do with the equipoise? *European Heart Journal*, 40(22), 1815.

14. **Papageorgiou, G.**, Mauff, K., Tomer, A., and Rizopoulos, D. (2019), An overview of joint modeling of time-to-event and longitudinal outcomes *Annual Review of Statistics and Its Application*, 6, 223–240.

15. Colijn, J.M, Verzijden, T., Meester-Smoor, M.A., Klaver, C.C.W., Demirkan, A.b, Meester-Smoor, M.A., Ahmad, S., van Duijn, C.M., Klaver, C.C.W., den Hollander, A.I., Kersten, E., Hoyng, C.B., Klaver, C.C.W., Cougnard-Grégoire, A., Merle, B.M.J., Korobelnik, J.-F., Delcourt, C., **Papageorgiou, G.**, et al. (2019), Increased High-Density Lipoprotein Levels Associated with Age-Related Macular Degeneration: Evidence from the EYE-RISK and European Eye Epidemiology Consortia *Opthalmology*, 126(3), 393–406.

16. Etnel, J.R.G., Huygens, S.A., Grashuis, P., Pekbay, B, **Papageorgiou, G.**, Roos Hesselink, J.W., Bogers, A.J.J.C., Takkenberg, J.J.M. (2019), Bioprosthetic aortic valve replacement in nonelderly adults: A systematic review, meta-analysis, and microsimulation. *Circulation: Cardiovascular Quality and Outcomes*, 12(2), e005481.

17. Rigters, S.C., Van Der Schroeff, M.P., **Papageorgiou, G.**, Baatenburg De Jong, R.J., Goedegebure, A. (2019), Progression of hearing loss in the aging population: Repeated auditory measurements in the Rotterdam study. *Audiology and Neurotology*, 23(5), 290–297.

18. Romeo, J.L.R., **Papageorgiou, G.**, Takkenberg, J.J.M., Mokhles, M.M. (2018), The right time-dependent statistics: this is the moment. *European journal of cardiothoracic surgery*, 54(6), 1145.

19. Etnel, J.R.G., Grashuis, P., Huygens, S.A., Pekbay, B., **Papageorgiou, G.**, Helbing, W.A., Roos-Hesselink, J.W., Bogers, A.J.J.C., Mokhles, M.M., Takkenberg, J.J.M. (2018), The Ross Procedure: A Systematic Review, Meta-Analysis, and Microsimulation. *Circulation. Cardiovascular quality and outcomes*, 11(12), e004748.

20. Head, S.J., **Papageorgiou, G.**, Milojevic, M., Stone, G.W., Kappetein, A.P. (2018), Interpretation of results of pooled analysis of individual patient data – Authors' reply. *The Lancet*, 392(10150), 818.

21. **Papageorgiou, G.**, Grant, S.W., Takkenberg, J.J.M., Mokhles, M.M. (2018), Statistical primer: How to deal with missing data in scientific research? *Interactive Cardiovascular and Thoracic Surgery*, 27(2), 153–158.

22. Romeo, J.L.R., **Papageorgiou, G.**, Van De Woestijne, P.C., Takkenberg, J.J.M., Westenberg, L.E.H., Van Beynum, I., Bogers, A.J.J.C., Mokhles, M.M. (2018), Downsized cryopreserved and standard-sized allografts for right ventricular outflow tract reconstruction in children: Long-Term single-institutional experience. *Interactive Cardiovascular and Thoracic Surgery*, 27(2), 257–263.

23. Head, S.J., Milojevic, M., Daemen, J., Ahn, J.-M., Boersma, E., Christiansen, E.H., Domanski, M.J., Farkouh, M.E., Flather, M., Fuster, V., Hlatky, M.A., Holm, N.R., Hueb, W.A., Kamalesh, M., Kim, Y.-H., Mäkikallio, T., Mohr, F.W., **Papageorgiou, G.**, Park, S.-J., Rodriguez, A.E., Sabik, J.F., Stables, R.H., Stone, G.W., Serruys, P.W., Kappetein, A.P. (2018), Stroke Rates Following Surgical Versus Percutaneous Coronary Revascularization. *Journal of the American College of Cardiology*, 72(4), 386–398.

24. Head, S.J., Milojevic, M., Daemen, J., Ahn, J.-M., Boersma, E., Christiansen, E.H., Domanski, M.J., Farkouh, M.E., Flather, M., Fuster, V., Hlatky, M.A., Holm, N.R., Hueb, W.A., Kamalesh, M., Kim, Y.-H., Mäkikallio, T., Mohr, F.W., **Papageorgiou, G.**, Park, S.-J., Rodriguez, A.E., Sabik, J.F., Stables, R.H., Stone, G.W., Serruys, P.W., Kappetein, A.P. (2018), Mortality after coronary artery bypass grafting versus percutaneous coronary intervention with stenting for coronary artery disease: a pooled analysis of individual patient data. *The Lancet*, 391(10124), 939–948.

PhD Portfolio

PhD Candidate:	Papageorgiou Grigorios
Department:	Department of Biostatistics & Department of Cardio-Thoracic Surgery Erasmus Medical Center Rotterdam
PhD Period:	2015 – 2021
Promotors:	Prof. dr. Dimitris Rizopoulos Prof. dr. Johanna J. M. Takkenberg
co-Promotor:	dr. Mostafa M. Mokhles

Presentations

International Conferences	Year	ECTS
9^{th} EMR & Italian Region of IBS Conference, Thessaloniki, Greece	2017	1.0
38^{th} ISCB Conference, Vigo, Spain	2017	1.0
29^{th} IBS Conference, Barcelona, Spain	2018	1.0
40^{th} ISCB Conference, Leuven, Belgium	2019	1.0
41^{st} ISCB Conference, Krakow, Poland	2020	1.0

Other	Year	ECTS
Erasmus Statistics Day, Rotterdam, the Netherlands	2017	1.0
EMC & LUMC Joint Meeting, Leiden, the Netherlands	2020	1.0
CQM Symposium: Prediction to Personalize Medicine, Rotterdam, the Netherlands	2020	1.0

Teaching

Teaching Assistant	Year	ECTS
SPSS Practicals, MSc Medicine, Erasmus MC	2015 - 2020	1.5
Biostatistical Methods II (EP03), NIHES	2015 - 2020	1.5
Repeated Measurements (CE08), NIHES	2015 - 2020	1.5
Missing Data in Clinical Research (EP16), NIHES	2019	0.5

Courses

	Year	ECTS
Scientific Integrity, Erasmus MC	2017	0.5
Multivariate Dimension Reduction for Biological Data Integration, IBC	2018	0.5
Advanced methods for group sequential and adaptive trials, ISCB	2019	0.5
Regression Modeling Strategies, ISCB	2020	0.5

Seminars and Workshops

	Year	ECTS
Erasmus MC bi-weekly Biostatistics CQM Seminar	2015 - 2020	1
Erasmus MC monthly Thorax Journal Club	2017 - 2019	0.5
BMS-ANed Meeting, Leiden, the Netherlands	2017	0.2
BMS-ANed Meeting, Amsterdam, the Netherlands	2017	0.2
BMS-ANed Meeting, Rotterdam, the Netherlands	2018	0.2

Consulting

	Year	ECTS
Statistical Consultant for the Consultation centre for Patient-Oriented research (CPO)	2015 - 2020	20

Other

	Year	ECTS
Organization PhD Day BMS-ANed	2017	3

About the Author

Grigorios Papageorgiou was born in Athens, Greece. He graduated high school in 2005 and enrolled at the University of Piraeus to study towards the BSc in Statistics and Acturial Science in 2006. After successfully obtaining his bachelor degree, he applied and was accepted to the Masters program in Statistics at the Katholieke Universiteit Leuven in Belgium. During his studies he specialized in Biometrics and graduated with honours in 2015. Immediately after graduation he started his PhD project at the Biostatistics department at Erasmus MC, in Rotterdam under the supervision of prof. dr. Dimitris Rizopoulos, prof. dr. Johanna J. M. Takkenberg and dr. Mostafa Mokhles. Currently he works as a post doc at the department of Biostatistics, as a statistician for the department of Infectious Diseases and as a methodological assessor for CBG-MEB.

Acknowledgements

First and foremost, I would like to thank Dimitris Rizopoulos, Hanneke Takkenberg, and Mostafa Mokhles for giving me the opportunity to pursue this PhD. None of this would have been possible without your constant support, scientific guidance, and overall mentorship.

The story of a PhD has as much to do with this thesis's contents as it has to do with the people you meet, you collaborate with, and the relationships you develop with them. And I have been fortunate in that respect.

Dimitris, this is probably version 7.3.8 of this paragraph where I attempt to thank you, and like all the previous versions, it will not be up to the task. Thank you for all your generosity, the guidance and for all the opportunities you have given me. When I first arrived at the department, your first question was what I expected from you as a supervisor. I said I wanted to feel comfortable enough to ask questions frequently and even ask questions I was supposed to know. You replied: "Yes, sure. We can even discuss 1 + 1 = 2 if you want. Just knock on my door". Indeed, you were always there to answer every knock at your door. And the truth is I did quite some knocking at this door. However, guidance was not the only thing I found there. Because while, at first, a supervisor has been answering the door, with each of these knocks over the years, the projects, the programming, the Crunch challenges, the requests for professional and sometimes even personal advice, there is now a mentor, a colleague and a true friend who answers that door.

Mostafa and Hanneke, thank you for all your support and guidance through my PhD. The field of Thorax Surgery was entirely new for me when I started, and without your mentorship, I would have been lost. You were always there for me with advice and ideas on how to move forward with my PhD. Furthermore, you always motivated me to get involved with as many projects as possible in the department of Thorax Surgery, and I have to admit that while this felt like too much work sometimes, I now can look back at a rich scientific output that I can be proud of, and I am grateful to you for that.

Nicole, Kazem, Elrozy, and Katya thank you for welcoming me into the department when I first started. You did everything to make me feel comfortable back then. With the four of you and Sara who joined our group a bit later, it might have taken quite some time, a considerable amount of drinks, endless discussions about how to survive the PhD, conferences, and most surprisingly: sneaker shopping, escape rooms, aubergine spreads with too much garlic, spare ribs that became tofu that became spare ribs 2.0, smoking, quitting smoking, kayaking, western paviljoen, pictures of cats, conference dinner party dancing, but we became good friends, and I look forward to

more such moments to come.

Eline, thank you for everything. I couldn't do it without your help. Your willingness to help as much as you can with any sort of issue that might be of concern, work-related or not, is amazing. Thank you for making our office feel like home.

To all my current colleagues in the department of Biostatistics: Eline, Nicole, Elrozy, Sara, Joost, Sten, Floor, Lidia, Tobias, Hongchao, Pedro, Anja, Aglina, Teun, Zhenwei, and also to all past colleagues, thank you. With you around, working is much more fun.

I would also like to thank all my colleagues in the department of Thorax Surgery, who always made me feel welcomed and part of their group even if I was not there so often.

Jamie, Rahat, and Kevin, this thesis would not have been completed without your tireless work and help. Thank you for everything. It was an honor to collaborate with you.

Maroula, when we first met, I was already one year in my Bachelor's degree studies and had failed to pass any of the courses. But then, everything changed. Being around you is not only fun but also gives me purpose. I could not have made it that far without you. Thank you for these wonderful years we have spent together so far and all your support. Thank you for your patience and loyalty through all these years, all these different countries, homes, and changes of plans. I am so lucky to have found my partner in life at such an early age. I am looking forward to the rest of our lives together.

Last but by no means least, I would like to thank my family. Mom, there are no words to thank you. Thank you for all the love, support, motivation, and happiness that you so generously give me. Dad, thank you for all your love and support. Katerina, thank you for all your love and support and for being such a cool sister. Finally, I would like to thank and commemorate in this thesis my late grandfather Grigoris, my two late grandmothers Katerina and Tasia, and my childhood hero and grandfather Giannis.